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)	Change in prevalence rates of physical and sexual intimate partner violence against women: Data from two cross-sectional studies in New Zealand, 2003 and 2019.
AUTHORS	Fanslow, Janet; Hashemi, Ladan; Malihi, Zarintaj; Gulliver, Pauline; McIntosh, Tracey

VERSION 1 – REVIEW

REVIEWER	Yukiko Washio RTI International
REVIEW RETURNED	14-Oct-2020

GENERAL COMMENTS	Change in prevalence rates of women's physical and sexual IPV in New Zealand Nation wide household survey was conducted in 2003 and 2019, and data on IPV were analyzed specific to women populations. The project is significant and indicative of future research directions. Authors might want to reconsider the analyses by comparing women who responded to both years so there will be one on one matching in how the prevalence has changed over the years. Introduction – Association between IPV and women's belief need to be stated which direction (positive or inverse) Analytic procedure – moderate/severe physical IPV and any IPV but what about sexual IPV? Used both categories? Changes in physical IPV prevalence rates – 30% in both years? Lifetime physical IPV for multivariate analyses, as lifetime sexual IPV multivariate analyses?

REVIEWER	Carmen Vives-Cases University of Alicante, Spain
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	Congratulation for this interesting study. In my opinion, it should be improved in the following relevant aspects: 1. abstract: be sure that the objective here is written as it is in the introduction 2. Strengths and limitations: I do not think that you can affirm that your sample is representative of New Zealand. The second survey has a quite small sample. You must explain why the second one is so small in comparison with the first one. 3. Introduction: you should refer to the Spanish Macro-Survey of Violence Against women due to it has a real representative sample of spanish women and there are different since the first one that it was done in 1998. Here is the last one from 2019: https://violenciagenero.igualdad.gob.es/en/violenciaEnCifras/macro

	encuesta2015/Macroencuesta2019/home.htm. Here you have information about it in English: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221049 Discusion: You must deeply explain the differences, trends and associations founds in your study. The first paragraph of this section is not necessary. You should provide a sum up of the main results instead of describing again the objectives of the study. You made interesting comparisons and some implications but you must reinforce the interpretations of your results. In its current form, the manuscript lack of this important part of the discusion section. In limitation, you should consider the affirmations related to the representativiness of your sample.

REVIEWER	Ryu, Euijung Mayo Clinic, USA
REVIEW RETURNED	19-Dec-2020

GENERAL COMMENTS	This study reports potential changes in intimate partner violence over 15 years using two cross-sectional studies in New Zealand. Major comments 1. One of the study outcomes is 12-month IPV. However, the conclusions were based on very small sample size, which is difficult to know whether the result is simply data fluctuation or real observation. For example, in Table 2, association direction for age is opposite between lifetime and past 12-month (not necessarily talking about statistical significance). Does this make sense? Additionally, for the majority of the results for 12-month IPV lacks statistical power (i.e., the majority of cells in Tables have counts less than 10). Therefore, it would be inappropriate to claim “no significant difference” when in fact it would just reflect the lack of power. 2. This study was based on national population-based study. However, sample size included in this study was very small given that this is a national survey study (~2000 for 2003 and ~900 for 2019 survey). If the participation rate was high as stated by the authors, why is this study based on only this number of subjects? It would be required to describe the original study design (including the survey sampling design and the number of subjects invited). 3. The authors stated that this study is representative to general population. However, there is no data to justify the claim. 4. One of advantages of this study is using diverse population. I am guessing people with different culture would have different results. However, this study did not include the type of population (e.g., Maori, etc) as one of factors, although it would be likely a strong confounders in this study. 5. This is a national survey-based study, which will be likely based on some sort of sampling frame to represent the population. For studies like this, researchers usually use survey-based methods for the analysis in order to incorporate survey weights. However, this study did not consider that. 6. In Strength and limitation section, the authors mentioned that this is based on a large population-based study. But the authors

	need to add the lack of statistical power in limitation as 12-month outcomes lacks statistical power. 7. The 2019 data seems a bit contradicting. Compared to 2003 data, subjects included in the 2019 data have higher education, but lower proportion of independent income and live in more deprived area. Usually, higher education is associated with less deprivation. Do authors have any idea why this is the case in this cohort? It seems like this might indicate the 2019 cohort might have some sort of strong selection bias. There is a section for representativeness in page 7, the authors can provide more details to claim how representative this data is to the general population for each wave. 8. The statement about missing data is concerning. Exactly what percentages of subjects were excluded due to missing data? Did the authors check how much the results were biased because of these excluded people? 9. In page 17, the authors mentioned about interaction between study year and subject characteristics. If there is any interaction, one should not report the results based on main effect only model (e.g., the current multivariable model presented in the manuscript). 10. Why Table 4 (over 1000) has more subjects in 2019 than previous tables (~900)? Minor comments 1. This study is based on two cross-sectional studies. However, in the manuscript, the authors also mentioned “repeated” survey (e.g., in Strengths and limitations section), which usually refers to studies where the same surveys were asked to the same set of people. Please clarify to avoid the confusion. 2. Did the authors check whether there are the same people answered the both surveys, or did the survey already designed to select new people in 2019 survey. Please clarify. 3. It is not clear whether the survey was initiated for women only, or women were selected after the survey was done for this study. I.e, unclear what “household members” are supposed to be. 4. Some of results about age association is rather obvious. For example, it is possible that women less than 25 do not have enough time to experience in IPV yet. 5. Current marital status can be a confounder. The outcomes are the information that occurred before the survey and the marital status is the status at the time of survey. Isn’t it possible that those who experienced IPV in the past would likely not stay in marriage? 6. Page 16 is hard to follow. In general, this manuscript will benefit from having clear subheading (e.g., one subheading for physical, and separate subheading for sexual IPV, for instance). This study reports potential changes in intimate partner violence over 15 years using two cross-sectional studies in New Zealand. Major comments 1. One of the study outcomes is 12-month IPV. However, the conclusions were based on very small sample size, which is

	difficult to know whether the result is simply data fluctuation or real observation. For example, in Table 2, association direction for age is opposite between lifetime and past 12-month (not necessarily talking about statistical significance). Does this make sense? Additionally, for the majority of the results for 12-month IPV lacks statistical power (i.e., the majority of cells in Tables have counts less than 10). Therefore, it would be inappropriate to claim "no significant difference" when in fact it would just reflect the lack of power. 2. This study was based on national population-based study. However, sample size included in this study was very small given that this is a national survey study (~2000 for 2003 and ~900 for 2019 survey). If the participation rate was high as stated by the authors, why is this study based on only this number of subjects? It would be required to describe the original study design (including the survey sampling design and the number of subjects invited). 3. The authors stated that this study is representative to general population. However, there is no data to justify the claim. 4. One of advantages of this study is using diverse population. I am guessing people with different culture would have different results. However, this study did not include the type of population (e.g., Maori, etc) as one of factors, although it would be likely a strong confounders in this study. 5. This is a national survey-based study, which will be likely based on some sort of sampling frame to represent the population. For studies like this, researchers usually use survey-based methods for the analysis in order to incorporate survey weights. However, this study did not consider that. 6. In Strength and limitation section, the authors mentioned that this is based on a large population-based study. But the authors need to add the lack of statistical power in limitation as 12-month outcomes lacks statistical power. 7. The 2019 data seems a bit contradicting. Compared to 2003 data, subjects included in the 2019 data have higher education, but lower proportion of independent income and live in more deprived area. Usually, higher education is associated with less deprivation. Do authors have any idea why this is the case in this cohort? It seems like this might indicate the 2019 cohort might have some sort of strong selection bias. There is a section for representativeness in page 7, the authors can provide more details to claim how representative this data is to the general population for each wave. 8. The statement about missing data is concerning. Exactly what percentages of subjects were excluded due to missing data? Did the authors check how much the results were biased because of these excluded people? 9. In page 17, the authors mentioned about interaction between study year and subject characteristics. If there is any interaction, one should not report the results based on main effect only model (e.g., the current multivariable model presented in the manuscript).

	10. Why Table 4 (over 1000) has more subjects in 2019 than previous tables (~900)? Minor comments 1. This study is based on two cross-sectional studies. However, in the manuscript, the authors also mentioned "repeated" survey (e.g., in Strengths and limitations section), which usually refers to studies where the same surveys were asked to the same set of people. Please clarify to avoid the confusion. 2. Did the authors check whether there are the same people answered the both surveys, or did the survey already designed to select new people in 2019 survey. Please clarify. 3. It is not clear whether the survey was initiated for women only, or women were selected after the survey was done for this study. I.e, unclear what "household members" are supposed to be. 4. Some of results about age association is rather obvious. For example, it is possible that women less than 25 do not have enough time to experience in IPV yet. 5. Current marital status can be a confounder. The outcomes are the information that occurred before the survey and the marital status is the status at the time of survey. Isn't it possible that those who experienced IPV in the past would likely not stay in marriage? 6. Page 16 is hard to follow. In general, this manuscript will benefit from having clear subheading (e.g., one subheading for physical, and separate subheading for sexual IPV, for instance).

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Yukiko Washio,
Institution and Country: RTI International USA

The project is significant and indicative of future research directions.

Thank you, we are pleased that you think the project is significant.

Authors might want to reconsider the analyses by comparing women who responded to both years so there will be one on one matching in how the prevalence has changed over the years.

Thank you for this suggestion. Unfortunately, one to one matching is not possible, as we did not interview the same women at two different time points. Data collection was carried out with two independent samples.

Introduction – Association between IPV and women's belief need to be stated which direction (positive or inverse).

Noted and agreed. We have added the direction to the association (introduction, page 4).

Analytic procedure – moderate/severe physical IPV and any IPV but what about sexual IPV? Used both categories?

We have treated any sexual IPV as severe. This is indicated in the Methods section (page 7).

Changes in physical IPV prevalence rates – 30% in both years?

Yes, roughly 30% of the sample in both study years (32.2% in 2003 and 29.1% in 2019) reported having ever experienced at least one episode of physical IPV (lifetime prevalence). We rounded the numbers to make the comparison easier.

Lifetime physical IPV for multivariate analyses, as lifetime sexual IPV multivariate analyses?
We are not sure what question is being asked. In multivariate analysis, AORs for change in lifetime physical IPV remained non-significant after adjusting for socio-demographic characteristics however, AORs indicating a reduction in lifetime sexual IPV remained unchanged and significant. This indicates that the observed change in prevalence rates of lifetime sexual IPV cannot be explained by socio-demographic differences between the two samples.

Reviewer: 2
Dr. C Vives-Cases, Alicante University

Congratulation for this interesting study.

Thank you! We are glad that you found this study interesting.

1. abstract: be sure that the objective here is written as it is in the introduction
   Agreed. We have added two objectives to the abstract in accordance with the goals mentioned in the introduction.

2. Strengths and limitations: I do not think that you can affirm that your sample is representative of New Zealand.
   As mentioned in the method section, Representativeness sub-section: in both surveys, the ethnicity, marital status, and deprivation level distribution of the samples were closely comparable to the general population, however both samples were under-represented for younger women (ages 20-29 in 2003, 16-29 in 2019). We've added two references where representativeness of both samples has been discussed in more detail.

   -The second survey has a quite small sample. You must explain why the second one is so small in comparison with the first one.
   As explained in the method section/ Eligibility sub-section: The 2019 survey recruited both women and men aged 16 years and over while the 2003 survey only recruited women aged 18-64 years. For the purpose of this analysis, we only included women aged 18-64 years from the 2019 survey to make it comparable with the 2003 survey, which inevitably reduced the sample size for 2019 survey. A flow diagram has been added (Figure 1) which demonstrates the number of people invited and those who were interviewed and included in the analyses for each survey year (Supplementary Figure 1).

3. Introduction: you should refer to the Spanish Macro-Survey of Violence Against women due to it has a real representative sample of Spanish women and there are different since the first one that it was done in 1998. Here is the last one from 2019:

Thank you for the suggestion. We have added Spain as one of the high-income countries where the violence against women study have been carried out (introduction, page 3).

4. Discussion: You must deeply explain the differences, trends and associations founds in your study. The first paragraph of this section is not necessary. You should provide a sum up of the main results instead of describing again the objectives of the study. You made interesting comparisons and some implications but you must reinforce the interpretations of your results. In its current form, the manuscript lack of this important part of the discussion section.

Thank you, and we agree that further depth of discussion is important. In line with the reviewer's suggestion, we have replaced the first paragraph of the discussion with one sentence and proceeded directly to the summary of the main results. We have also added new paragraphs directly discussing the possible interpretations of the findings and articulating the evidence that we think points to the most likely conclusion. We have also made more specific recommendations for future research which could further elucidate the present findings.

In limitation, you should consider the affirmations related to the representativeness of your sample.
Please see above (point 2) where we explained the representativeness of the both samples.

Reviewer: 3
Dr. Euijung Ryu, Mayo Clinic

Comments to the Author:
This study reports potential changes in intimate partner violence over 15 years using two cross-sectional studies in New Zealand.
Major comments
1. One of the study outcomes is 12-month IPV. However, the conclusions were based on very small sample size, which is difficult to know whether the result is simply data fluctuation or real observation.
We agree with this point and have added this to the limitation section.

For example, in Table 2, association direction for age is opposite between lifetime and past 12-month (not necessarily talking about statistical significance). Does this make sense?
The association with age depends on the timeframe of IPV considered. The observed association from our study is consistent with literature (Heise, 2011), with lifetime IPV showing a positive association with age as older people have longer exposure time to experience violence. However, it is also expected that as women (and their partners) get older (particularly after middle age) their likelihood of experiencing 12-month IPV decreases. Explanations for this include that, over time women may leave their abusive partner, or, if women remain in the relationship, their partners are less likely to use physical violence as they age.
Heise, Lori L. (2011). What works to prevent partner violence: An evidence overview. London: STRIVE. ISBN: 978 0 902657 85 2

Additionally, for the majority of the results for 12-month IPV lacks statistical power (i.e., the majority of cells in Tables have counts less than 10). Therefore, it would be inappropriate to claim "no significant difference" when in fact it would just reflect the lack of power.
This is the same as the previous point on the smaller sample size for 2019 raised by Reviewer 2. Please see our response above to Reviewer 2, point 2.

2. This study was based on national population-based study. However, sample size included in this study was very small given that this is a national survey study (~2000 for 2003 and ~900 for 2019 survey). If the participation rate was high as stated by the authors, why is this study based on only this number of subjects? It would be required to describe the original study design (including the survey sampling design and the number of subjects invited).

We agree that both surveys are better described as “population-based surveys” than “national surveys” so we have removed the word “national” and the two surveys are now described as population-based surveys.
Response rates for both surveys were about 64%. As explained in the method section, Eligibility sub-section: Participants of the 2003 study were only women aged 18-64 years. In 2019, the eligible population was expanded to include women and men aged 16 years and older. For the purpose of this paper 1423 men (all ages) and 479 women aged (16-17 or 65 and over) were excluded to make two samples comparable.
To add more clarification on the recruitment processes, a flow diagram has been added (Figure 1) which demonstrates the number of people invited, number of eligible households and individuals, and those who were interviewed and included in the analyses for both surveys.
Additionally, a paper providing a description of 2019 survey has been recently published (Jan 2021) has been added to the references.

3. The authors stated that this study is representative to general population. However, there is no data to justify the claim.
We’ve added two references to the Method section where representativeness of both samples has been discussed in more detail. However, we have not expanded on this point in the main text, due to word count limitations.

4. One of advantages of this study is using diverse population. I am guessing people with different culture would have different results. However, this study did not include the type of population (e.g., Maori, etc) as one of factors, although it would be likely a strong confounders in this study.
While we agree that ethnicity would be an important cofounder in this study, unfortunately, the smaller sample size in the 2019 survey has put restrictions on our ability to run sub-group analysis based on ethnicity (particularly for those with Māori background).

5. This is a national survey-based study, which will be likely based on some sort of sampling frame to represent the population. For studies like this, researchers usually use survey-based methods for the analysis in order to incorporate survey weights. However, this study did not consider that.

We have now included further information on the procedures undertaken to account for the sampling design. This includes use of ‘STRATA’ to account for the location where participants were recruited from, cluster sampling, and weighting based on number of total eligible women per household. This explanation is provided in the methods section: page 7.

6. In Strength and limitation section, the authors mentioned that this is based on a large population-based study. But the authors need to add the lack of statistical power in limitation as 12-month outcomes lacks statistical power.

We agree. We have now added this point to the limitation section.

7. The 2019 data seems a bit contradicting. Compared to 2003 data, subjects included in the 2019 data have higher education, but lower proportion of independent income and live in more deprived area. Usually, higher education is associated with less deprivation. Do authors have any idea why this is the case in this cohort? It seems like this might indicate the 2019 cohort might have some sort of strong selection bias. There is a section for representativeness in page 7, the authors can provide more details to claim how representative this data is to the general population for each wave.

As previously noted in our responses to reviewers, we have now included a reference to the 2019 study, which documents the representativeness of the sample relative to the general population. This suggests that there is not a selection bias.

Using weighted data, area-deprivation level is no longer different between the two surveys. Differences in personal income between the two samples also decreased with use of the weighted data. As we have reported, and the reviewer has noted, the 2019 sample is older compare with the 2003 sample; this could also contribute to the noted differences in education, as older women will have had more time to engage with higher education.

8. The statement about missing data is concerning. Exactly what percentages of subjects were excluded due to missing data? Did the authors check how much the results were biased because of these excluded people?
Problems with missingness were minor. We have added more information on percentage of missing data in both datasets (Method section, page 7). As less than 4% of data were missing for any variable, it is very unlikely that findings were biased because of missing data.

9. In page 17, the authors mentioned about interaction between study year and subject characteristics. If there is any interaction, one should not report the results based on main effect only model (e.g., the current multivariable model presented in the manuscript).

We used interaction terms only to determine if there were any differences in changes in prevalence rates varied by population subgoups. We did not report on interactions, as none of them reached significance.

10. Why Table 4 (over 1000) has more subjects in 2019 than previous tables (~900)?
Thank you for spotting this typo. The sample sizes have been fixed in Table 4.

Minor comments
1. This study is based on two cross-sectional studies. However, in the manuscript, the authors also mentioned “repeated” survey (e.g., in Strengths and limitations section), which usually refers to studies where the same surveys were asked to the same set of people. Please clarify to avoid the confusion.
Noted and agreed. The two surveys were independent, with similar (and in most cases identical questions) asked of two different set of people. To avoid the confusion, we have replaced the words “repeated surveys” with the words “comparable surveys”.

2. Did the authors check whether there are the same people answered the both surveys, or did the survey already designed to select new people in 2019 survey. Please clarify.
Data collection was carried out with two independent samples. These surveys are cross-sectional and no attempt was made to interview the same people.

3. It is not clear whether the survey was initiated for women only, or women were selected after the survey was done for this study. I.e, unclear what “household members” are supposed to be.
In 2003, only women were recruited. In 2019, both men and women were recruited, but only women have been included in the present analyses. For both surveys, household selection processes were undertaken to identify if there was one or more eligible participants who resided in the household (in 2003, a female in the eligible age range; in 2019, a person in the eligible age range, with the sample later restricted to women aged 18-64, for comparability with the 2003 study).

4. Some of results about age association is rather obvious. For example, it is possible that women less than 25 do not have enough time to experience in IPV yet.
We agree on this point. We are not sure if there is a question here that we need to address.

5. Current marital status can be a confounder. The outcomes are the information that occurred before the survey and the marital status is the status at the time of survey. Isn't it possible that those who experienced IPV in the past would likely not stay in marriage?

We agree and have used the current marital status as a potential confounder in this study. Moreover, we have included both lifetime and 12-month IPV prevalence rates in this study in the recognition of the fact that some women may have already left the abusive partner. 12-month prevalence rates are more likely to capture current IPV which is happing at the time of survey.

6. Page 16 is hard to follow. In general, this manuscript will benefit from having clear subheading (e.g., one subheading for physical, and separate subheading for sexual IPV, for instance).

Noted and agreed. We have already introduced two main subheadings (one subheading for physical IPV and separate subheading for sexual IPV). Thank you for the suggestion.

VERSION 2 – REVIEW

REVIEWER	Yukiko Washio RTI International
REVIEW RETURNED	10-Feb-2021

GENERAL COMMENTS	Comments addressed

REVIEWER	Carmen Vives-Cases University of Alicante, Spain
REVIEW RETURNED	12-Feb-2021

GENERAL COMMENTS	Thank you very much for this revised version of the manuscript. You properly answered all my queries, but you should review a few issues in the interpretation of your results. In page 21, you explained the reduction in the prevalence of current physical IPV due to actual changes in perpetrator behaviour over time and changes due to differences in methods. In the first possible explanation, you added your own results about the observed descreased in violence acceptation among women as an evidence of the increase awaraness about this issue in society. This is ok but you should consider that you are puting in relation with this sentence a relationship between the risk of victimization and the violence acceptability among your sample (where there are women exposed to IPV). It has been evidenced the relationship between victimization and violence acceptatibility, but you should support this with references. Another issue is related to the interpretation of the results about the no changes observed in women contacting formal source of help. You interpret this results with concern and recommend greater efforts to make more accesible these formal resources. This is ok but you should also consider that the likelihood of seeking formal help among affected women tend to increase as the severity of the IPV does. You should consider therefore a more positive interpretation too: perhaps the no changes in the use of formal help among women is related with the reduction of current

	physical, and lifetime sexual IPV between the studied years and a possible decrease of high severity cases.

VERSION 2 – AUTHOR RESPONSE

Reviewer Reports:

Reviewer: 1

Dr. Yukiko Washio, RTI International

Comments to the Author:

Comments addressed

Thank you, we are pleased that you consider that we addressed your helpful comments appropriately.

Reviewer: 2

Dr. C Vives-Cases, Alicante University

Comments to the Author:

Thank you very much for this revised version of the manuscript. You properly answered all my queries, but you should review a few issues in the interpretation of your results. In page 21, you explained the reduction in the prevalence of current physical IPV due to actual changes in perpetrator behaviour over time and changes due to differences in methods. In the first possible explanation, you added your own results about the observed descreased in violence acceptation among women as an evidence of the increase awaraness about this issue in society. This is ok but you should consider that you are puting in relation with this sentence a relationship between the risk of victimization and the violence acceptability among your sample (where there are women exposed to IPV). It has been evidenced the relationship between victimization and violence acceptatibility, but you should support this with references.

Thank you, we are happy that you consider the revised version of this manuscript has appropriately addressed your queries.

With the respect to the need to more fully support the point about the relationship between victimisation and attitudes about the acceptability of violence, we have now addressed this by addition of one sentence and two further references to support this point on page 21.

Another issue is related to the interpretation of the results about the no changes observed in women contacting formal source of help. You interpret this results with concern and recommend greater efforts to make more accesible these formal resources. This is ok but you should also consider that the likelihood of seeking formal help among affected women tend to increase as the severity of the IPV does. You should consider therefore a more positive interpretation too: perhaps the no changes in the use of formal help among women is related with the reduction of current physical, and lifetime sexual IPV between the studied years and a possible decrease of high severity cases.

Thank you. We have now added this alternate interpretation on page 21 of the revised manuscript.