PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The associations between biological and sociodemographic risks for developmental vulnerability in twins at age five: A population data linkage study in Western Australia.
AUTHORS	Dhamrait, Gursimran; Christensen, Daniel; Pereira, G.F; Taylor, Catherine

VERSION 1 – REVIEW

REVIEWER	Stacy Tzoumakis Griffith University, Australia
REVIEW RETURNED	23-Apr-2020

GENERAL COMMENTS	This is an interesting study that uses a large sample to examine the factors associated to the developmental vulnerability of twins. Overall, it is well-written and conceived; I have a few comments and suggestions detailed below – particularly regarding the methods. 1. It is was not until the acknowledgments that it became clear that the record linkage was conducted by an extremal organisation. The authors should include a section in the methods on the linkage procedures that includes this information as well as how was the linkage conducted, what personal identifiers were used, and any information on the reliability of the linkage (e.g., false positive/negative rates). Also, was it possible to link all children who completed the AEDC to the Birth Registrations and Midwives Notification System? 2. The authors take a data driven approach and include a large number of variables simultaneously in the multivariate model. Many of these variables are likely strongly related – it would be helpful to include more information on the model fit and that the assumptions were met. Have the authors attempted to include fewer variables in the model and take a more theoretical or stepwise approach? It is possible that more than 5 variables out of 24 would be significant. 3. Relatedly, the large confidence intervals for some of the significant variables (e.g., maternal age of 20 years or younger at time of twins' birth (Table 1: OR 8.69, 95% CI 1.52 to 49.69; Occupational Status Scale at Time of Child's Birth 7.97 [3.08-20.66]) is concerning as this can be an indication that the ORs aren't reliable. It may also be due to sparseness or low cell sizes of the variables included in the multivariate model. This could be minimised by including fewer variables in the final model and/or recoding the variables so that the distribution is more even where appropriate. For instance, for the maternal age at child's birth, why not create a binary variable with a slightly higher cut-off (e.g., 22 or 25 years)? Some of the categories in

	Occupational status could also be collapsed. If this cannot be fixed with additional analyses then the large confidence intervals need to be emphasised in the limitations. 4. Since pregnancy and birth complications are relatively rare, research using these risk factors typically use an obstetrics scale. There are formal scales that exist, or a simple sum of complications could be computed (e.g., 0, 1, 2, 3+ complications). It is not clear why the authors are retaining so many individual variables in the multivariate model that were not significant in the unadjusted models – especially considering the low base rates of these factors. For example, in Table 2 only 14 (6%) of cases had APH and DV2; as a result, in the adjusted model the 95% CI is 0.95- 37.40, why retain this in the multivariate model? 5. The authors do not include a separate limitations section in the manuscript. The limitations provided in the Strobe checklist are brief. As described above, there are some important limitations that need further discussion. 6. More specific information regarding Associations with domain-specific developmental vulnerability is needed – how are these results broadly consistent? 7. The first paragraph of the discussion is confusing and somewhat misleading– it refers to the findings being in comparison to singletons, but Figure 1 indicates that the final cohort consisted of twins only. These comparisons between the twins and the general population were not the focus of this study. It is not clear why this is emphasised in the discussion when it is not based on findings from these analyses. The discussion also ends with this as a main point: “The higher prevalence rates of DV1 and DV2 in twins observed in this study are indicative of the fact that twins form an at-risk group in terms of developmental vulnerability at the time at which children commence full-time school.”. Did the authors have access to the entire population (i.e., nontwin births)? Are they able to formally compare the prevalence rates between twins and singletons? If not, then the discussion should be revised to focus on the analyses on the risk factors that were conducted in this study. 8. Line 289 “The Louisville Twin Study also reported sex differences, with females scoring higher than males at ages four and five years, however, scores tended to converge at six years of age.” This sentence needs clarification – females scores higher on vulnerability or on competency? 9. Lines 305- 309 – can the authors elaborate on potential explanations for the difference between the significant variables in the South Australian study and their findings? Prenatal smoking in particular is a variable that is typically strongly associated with developmental vulnerability and in the unadjusted models if significant across analyses. This should be further addressed. Again, I wonder if it would remain significant in the multivariate model is a more parsimonious approach was taken to the inclusion of the other variables. 10. Can the conclusion highlight how these findings are different/the same as what we already know? What is the contribution to the research in this area? 11. There are several errors in the manuscript: Copy and paste error in the abstract line 70-71 “at time of AEDC completion (at time of AEDC completion”; Typo line 102 “that that”; Line 297 typo: “the need for further research into assess the effects”.

REVIEWER	Katrina Scurrah University of Melbourne, Australia
REVIEW RETURNED	14-May-2020

GENERAL COMMENTS	Main points Overall this is a well-written, comprehensive and worthwhile paper, describing interesting analyses and excellent use of population-based linked data which provides a large sample size. The chosen statistical approach is appropriate and most aspects of the analyses are described clearly. Some minor revisions would improve the readability of the paper. Prevalence is reported in the results, but not mentioned in the aims. The association analyses could also be motivated a bit more by including additional discussion in the aims of the importance and implications of developmental vulnerability and associations of specific measured risk factors with it. This was addressed more clearly in the discussion. Can the authors comment on the completeness and coverage of the databases and the linkages between them, which groups of children (if any) are most likely to be missing (or missing linkage), and potential biases arising from this? I recognise that the databases are population-based but coverage is unlikely to be exactly 100% and I would expect that some groups are less likely to be included. The authors report overall prevalence of DV1 and DV2, but the number of concordant affected and discordant twin pairs should also be reported. The implications and potential strategies to address DV in twins could differ depending on whether the affected individuals tend to be part of discordant pairs (which might require an individual level approach) or concordant affected pairs (which might be more suited to family level approaches). Similarly, it would be interesting to see the results of within-pair association analyses for the few risk factors such as sex that can differ between twins in a pair (if there are enough outcome-discordant pairs to allow these analyses), which would control for the effects of shared family level factors. More information and discussion of some statistical issues would reassure the reader that the results are robust: i) Some of the reported CIs are very wide, eg for young maternal age (CI for OR of 1.5-49 in the multivariable model for DV1), which is concerning. This should be discussed and potential explanations considered. Was correlation between risk factors assessed and could this have contributed to wide CIs? ii) It is unclear how risk factors were entered or retained in the multivariable models – eg whether all risk factors were included, or whether a stepwise or other approach was used. Did the estimates and CIs change much depending on which risk factors were included or excluded, and were any sensitivity analyses, model checking or model comparisons performed? iii) Were any interactions between risk factors (especially sex) assessed? Did the authors check for cohort effects (eg differences between those with AEDCs in 2009 vs 2012 vs 2015) or effects of

	gestational age or birthweight (actual rather than percent of expected)? iv) Reporting actual p-values is more appropriate than using stars to represent strength of evidence. The p-value for overall association with DV for categorical exposures with > 2 categories (eg maternal age and occupation) should also be reported (eg from a likelihood ratio test), and considered when describing associations. Minor points The word “notion” (used several times) is too informal and is inappropriate here. Possible alternatives include “hypothesis”, or “theory”. In Tables 1 and 2, models including one risk factor would be better described as simple or univariable (not bivariate). Similarly, models with > 1 risk factor are multivariable models (not multivariate, which describes models for multiple outcomes analysed together). It would be interesting to know approximately how many extra children a 3% increased prevalence in WA twins corresponds to. Can this be added? Similarly, was any information on address (eg rural/remote/metropolitan) available and is this likely to be relevant? Is much information available about the risks or effects of DV later on in children’s (especially twins) school careers, and are the authors aware of more recent papers than Lorenz et al. (year missing from reference but appears to be 2012)? If twins have an increased DV risk at school commencement that disappears later this could be important. Line 110: Twin studies that estimate heritability don’t usually assess associations between genes and the environment, but estimate the contribution of each of these to variation in the outcome. Line 258: “These results were broadly consistent with the findings for the aggregate measures” – do the authors mean the association results for the 5 domains? A bit more information about these results would be worthwhile, along with a few specific examples. Line 323 - “Furthermore, differences in the prevalence rates of particular language groups in WA is likely to be different to those that are prevalent in British Colombia and the difference in findings between the Canadian study and our results may be attributable to this fact.”. This sentence is unclear – do the authors mean that the language groups represented are likely to be different, the prevalences of each are likely to be different or both? Can the authors list some of the most frequently spoken languages other than English in WA? Line 327: “language emergency” – should be “language emergence”.

	Line 349: This sentence isn't clear, I suggest splitting it into 2 sentences.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1

1. It is was not until the acknowledgments that it became clear that the record linkage was conducted by an extremal organisation. The authors should include a section in the methods on the linkage procedures that includes this information as well as how was the linkage conducted, what personal identifiers were used, and any information on the reliability of the linkage (e.g., false positive/negative rates). Also, was it possible to link all children who completed the AEDC to the Birth Registrations and Midwives Notification System?
Response: Paragraph headed 'Data Sources' has been reworded to address this. We have also included a reference which details the WA data linkage system linkage methodology.

2. The authors take a data driven approach and include a large number of variables simultaneously in the multivariate model. Many of these variables are likely strongly related – it would be helpful to include more information on the model fit and that the assumptions were met. Have the authors attempted to include fewer variables in the model and take a more theoretical or stepwise approach? It is possible that more than 5 variables out of 24 would be significant.
Response: Given the sparse research in this area we deliberately included a wide range of factors that are associated with adverse developmental outcomes in twins. Per note #3 below, we have collapsed some variables together based on cell sizes.

3. Relatedly, the large confidence intervals for some of the significant variables (e.g., maternal age of 20 years or younger at time of twins' birth (Table 1: OR 8.69, 95% CI 1.52 to 49.69; Occupational Status Scale at Time of Child's Birth 7.97 [3.08-20.66]) is concerning as this can be an indication that the ORs aren't reliable. It may also be due to sparseness or low cell sizes of the variables included in the multivariate model. This could be minimised by including fewer variables in the final model and/or recoding the variables so that the distribution is more even where appropriate. For instance, for the maternal age at child's birth, why not create a binary variable with a slightly higher cut-off (e.g., 22 or 25 years)? Some of the categories in Occupational status could also be collapsed. If this cannot be fixed with additional analyses, then the large confidence intervals need to be emphasised in the limitations.
Response: In light of the comments provided by the reviewer we have collapsed the six maternal age categories into four maternal age categories (revised categories are <25 years, 2529 years, 30-34 years and ³35 years of age). This is in line with the maternal age distributions used in descriptive statistics published by 'Twin Research Australia.' (please see link below).
https://twins.org.au/images/PDFs/Twin-Pregnancy-and-Birth-Trends-in-Australia-7.2.18.pdf We agree with the reviewer in terms of the maternal occupation scale variable and have collapsed the variable categories from quintiles into two categories; 1) the most disadvantaged quintile (occupational status scale scores of 0-20, inclusive) and 2) greater than the lowest quintile. The relevant changes have also been made to the methods section (under the heading 'Maternal Variables') to reflect this change.
The combined effect of the reduction in categories for both the maternal occupation and age variables has reduced the associated CIs.

4. Since pregnancy and birth complications are relatively rare, research using these risk factors typically use an obstetrics scale. There are formal scales that exist, or a simple sum of complications could be computed (e.g., 0, 1, 2, 3+ complications). It is not clear why the authors are retaining so

many individual variables in the multivariate model that were not significant in the unadjusted models – especially considering the low base rates of these factors. For example, in Table 2 only 14 (6%) of cases had APH and DV2; as a result, in the adjusted model the 95% CI is 0.95- 37.40, why retain this in the multivariate model?
Response: We note the importance of cumulative risks, however combining risks into a scale assumes all risks have equal weight. Given the sparsity of research in this area, we wanted to estimate the association between individual risk variables, adjusted for the effects of other variables, to estimate the unique contribution of each risk factor to developmental vulnerability. We did derive a variable 'other pregnancy related complications' which combined the less common pregnancy complications in a binary variable (p. 9). We increased the exclusion criteria for variables from the multivariable analysis to a total N<50 for a given category of a given variable. APH is retained for two reasons firstly, because it meets this increased exclusion criteria, and secondly, removing APH from the multivariable models results in the widening of CIs. APH was originally retained in the models (prior to collapsing of the categories of maternal age and occupation) for these same reasons.

5. The authors do not include a separate limitations section in the manuscript. The limitations provided in the Strobe checklist are brief. As described above, there are some important limitations that need further discussion.
Response: The journal does not require a separate 'limitations' section in the manuscript however, we have addressed the limitations of the study in more depth and in line with the changes to the above-mentioned variables (page 4 and 14).

6. More specific information regarding Associations with domain-specific developmental vulnerability is needed – how are these results broadly consistent?
Response: We have added a brief sentence into the results to cover this comment (page 11). We are limited by the article length and thus have not gone into too much more detail.

7. The first paragraph of the discussion is confusing and somewhat misleading– it refers to the findings being in comparison to singletons, but Figure 1 indicates that the final cohort consisted of twins only. These comparisons between the twins and the general population were not the focus of this study. It is not clear why this is emphasised in the discussion when it is not based on findings from these analyses. The discussion also ends with this as a main point: "The higher prevalence rates of DV1 and DV2 in twins observed in this study are indicative of the fact that twins form an at-risk group in terms of developmental vulnerability at the time at which children commence full-time school.". Did the authors have access to the entire population (i.e., non- twin births)? Are they able to formally compare the prevalence rates between twins and singletons? If not, then the discussion should be revised to focus on the analyses on the risk factors that were conducted in this study.
Response: As the AEDC is a national census we compared the results of DV1 and DV2 between our study, which is conducted in twins and the state averages (i.e. the general population including both singletons and multiples) for DV1 and DV2. As this is the first study to our knowledge to assess AEDC outcomes in an exclusively twin population it was important to establish if twins are in fact at an elevated rate of developmental vulnerability compared to the general population. However, in light of the reviewers comments we have made some changes to the first paragraph of the discussion.

8. Line 289 "The Louisville Twin Study also reported sex differences, with females scoring higher than males at ages four and five years, however, scores tended to converge at six years of age." This sentence needs clarification – females scores higher on vulnerability or on competency?
Response: Edits have been made to this sentence.

9. Lines 305- 309 – can the authors elaborate on potential explanations for the difference between the significant variables in the South Australian study and their findings? Prenatal smoking in particular is a variable that is typically strongly associated with developmental vulnerability and in the

unadjusted models if significant across analyses. This should be further addressed. Again, I wonder if it would remain significant in the multivariate model is a more parsimonious approach was taken to the inclusion of the other variables.
Response: We have re-run the analysis and prenatal smoking is statistically insignificant in the multivariable model. We have added in a further explanatory sentence to the discussion to address differences between our results and that of the South Australian study.

10. Can the conclusion highlight how these findings are different/the same as what we already know? What is the contribution to the research in this area? Response: Edits have been made to this section.

11. There are several errors in the manuscript: Copy and paste error in the abstract line 70-71 “at time of AEDC completion (at time of AEDC completion”; Typo line 102 “that that”; Line 297 typo: “the need for further research into assess the effects”. Response: These errors have been corrected.

REVIEWER 2
1. Prevalence is reported in the results, but not mentioned in the aims. The association analyses could also be motivated a bit more by including additional discussion in the aims of the importance and implications of developmental vulnerability and associations of specific measured risk factors with it. This was addressed more clearly in the discussion.
Response: We have edited the aims of the study appropriately to include prevalence as a part of the aims.

2. Can the authors comment on the completeness and coverage of the databases and the linkages between them, which groups of children (if any) are most likely to be missing (or missing linkage), and potential biases arising from this? I recognise that the databases are populationbased but coverage is unlikely to be exactly 100% and I would expect that some groups are less likely to be included.
Response: We have also included a reference under the section headed ‘Data Sources’ which details the WA data linkage system linkage methodology. The AEDC has captured 99.6%, 99.0% and 98.7% of the estimated number of eligible WA children for the 2009, 2012 and 2015 respectively (see: https://www.aedc.gov.au/resources/detail/2015-aedc-national-report).

3. The authors report overall prevalence of DV1 and DV2, but the number of concordant affected and discordant twin pairs should also be reported. The implications and potential strategies to address DV in twins could differ depending on whether the affected individuals tend to be part of discordant pairs (which might require an individual level approach) or concordant affected pairs (which might be more suited to family level approaches).
Response: We have included a short section in the Results, under the heading ‘Prevalence of developmental vulnerability in twin’ which includes the rates of discordant pairs classified as DV1 and DV2. Further analysis in regards to the assessment of discordance cannot be conducted accurately as linked data does not included zygosity data.

4. Similarly, it would be interesting to see the results of within-pair association analyses for the few risk factors such as sex that can differ between twins in a pair (if there are enough outcomediscordant pairs to allow these analyses), which would control for the effects of shared family level factors.
Response: We appreciate this comment from the reviewer however, this was not within the scope of the study aims. We have included a sentence in the discussion section which highlights that future studies should aim to investigate within-pair associations.
Statistical Points

5. Some of the reported CIs are very wide, eg for young maternal age (CI for OR of 1.5-49 in the multivariable model for DV1), which is concerning. This should be discussed and potential explanations considered. Was correlation between risk factors assessed and could this have contributed to wide CIs?
Response: We have collapsed the six maternal age categories into four maternal age categories (revised categories are <25 years, 25-29 years, 30-34 years and ³35 years of age). This has reduced the size of the CIs.

6. It is unclear how risk factors were entered or retained in the multivariable models – eg whether all risk factors were included, or whether a stepwise or other approach was used. Did the estimates and CIs change much depending on which risk factors were included or excluded, and were any sensitivity analyses, model checking or model comparisons performed?
Response: We added all variables to the model simultaneously and have we have updated the methods to reflect this (see section headed 'Statistical Modelling').

7. Were any interactions between risk factors (especially sex) assessed? Did the authors check for cohort effects (eg differences between those with AEDCs in 2009 vs 2012 vs 2015) or effects of gestational age or birthweight (actual rather than percent of expected)?
Response: No. We did not have any strong a priori reasons for testing interactions or these other effects.

8. Reporting actual p-values is more appropriate than using stars to represent strength of evidence. The p-value for overall association with DV for categorical exposures with > 2 categories (eg maternal age and occupation) should also be reported (eg from a likelihood ratio test), and considered when describing associations.
Response: We have included the p-values in all of the data tables.
Minor points
9. The word "notion" (used several times) is too informal and is inappropriate here. Possible alternatives include "hypothesis", or "theory".
Response: Edits have been made.

10. In Tables 1 and 2, models including one risk factor would be better described as simple or univariable (not bivariate). Similarly, models with > 1 risk factor are multivariable models (not multivariate, which describes models for multiple outcomes analysed together). Response: Edits have been made.

11. It would be interesting to know approximately how many extra children a 3% increased prevalence in WA twins corresponds to. Can this be added? Similarly, was any information on address (eg rural/remote/metropolitan) available and is this likely to be relevant?
Response: No, we did not assess remoteness. 3% of the 1656 twins equates to an extra 50 DV1 children.

12. Is much information available about the risks or effects of DV later on in children's (especially twins) school careers, and are the authors aware of more recent papers than Lorenz et al. (year missing from reference but appears to be 2012)? If twins have an increased DV risk at school commencement that disappears later this could be important.
Response: Edits have been made to the reference to include the year of publication. We have included some further information in the introduction in regards to the associations between developmental vulnerability at school starting age and later academic outcomes. There are no twin specific studies to our knowledge so we have included general population and singleton studies.

13. Line 110: Twin studies that estimate heritability don't usually assess associations between genes and the environment but estimate the contribution of each of these to variation in the outcome.
Response: Sentence has been edited.

14. Line 258: "These results were broadly consistent with the findings for the aggregate measures" – do the authors mean the association results for the 5 domains? A bit more information about these results would be worthwhile, along with a few specific examples.
15. Response: We have added a brief sentence into the results to cover this comment. We are limited by the article length and thus have not gone into too much further detail.

16. Line 323 - "Furthermore, differences in the prevalence rates of particular language groups in WA is likely to be different to those that are prevalent in British Colombia and the difference in findings between the Canadian study and our results may be attributable to this fact.". This sentence is unclear – do the authors mean that the language groups represented are likely to be different, the prevalences of each are likely to be different or both? Can the authors list some of the most frequently spoken languages other than English in WA?
Response: We have revised this sentence and identified the most common language groups in WA and British Colombia.

VERSION 2 – REVIEW

REVIEWER	Stacy Tzoumakis Griffith University, Australia
REVIEW RETURNED	06-Jul-2020

GENERAL COMMENTS	The authors have addressed all comments/issues raised. The confidence intervals are now all reasonable considering the nature of the sample. The revised version is much improved.

REVIEWER	Katrina Scurrah University of Melbourne, Australia
REVIEW RETURNED	21-Jul-2020

GENERAL COMMENTS	Most of my previous comments have been satisfactorily addressed. A few remaining issues are discussed below (with numbers referring to points addressed by the authors in their response). 2. A note on coverage and completeness of the databases should be included in the mansuscript. 6. It is not yet clear which sensitivity analyses and model checking/model comparisons have been performed – these checks should be done and results briefly summarised. 8. "The p-values for overall association with DV for categorical exposures with > 2 categories (eg maternal age) should also be reported (eg from a likelihood ratio test), and considered when describing associations. " I note that p-values for each category have now been reported but not the p-values for the overall association of the categorical exposure variable with the outcome. The overall associations should be discussed (not just the ORs that were significantly different from the reference category) . P-values should be reported to about 2 significant figures – 4 decimal places for all p-values is too much detail.

	There are a few minor errors such as “Results from a study of an Australia wide study of 261,147 children, singletons and multiples” (line 312)

VERSION 2 – AUTHOR RESPONSE

REVIEWER 2

Most of my previous comments have been satisfactorily addressed. A few remaining issues are discussed below (with numbers referring to points addressed by the authors in their response).

2. A note on coverage and completeness of the databases should be included in the manuscript.

Response: A sentence has been included under the ‘Data Sources’ section which describes the participation rates for the 2009, 2012 and 2015 WA AEDC collections (line139).

6. It is not yet clear which sensitivity analyses and model checking/model comparisons have been performed – these checks should be done and results briefly summarised.

Response: No formal sensitivity analyses were undertaken because we were interested in the independent effect of each variable after accounting for all others. The pattern of results is broadly consistent across different specifications of the outcome (DV1, DV2, and the AEDC sub-domains). Between Revision 1 and Revision 2 we changed the treatment of a number of predictors (i.e. we collapsed the categories for the variables of maternal age and maternal occupation) but, again the pattern of results was broadly consistent.

8. “The p-values for overall association with DV for categorical exposures with > 2 categories (eg maternal age) should also be reported (eg from a likelihood ratio test) and considered when describing associations.“ I note that p-values for each category have now been reported but not the p-values for the overall association of the categorical exposure variable with the outcome. The overall associations should be discussed (not just the ORs that were significantly different from the reference category). P-values should be reported to about 2 significant figures – 4 decimal places for all p-values is too much detail.

Response: Overall p-values for each categorical exposure variable that had more than two categories have been included. Edits have been made to Tables 1-2 and Supplementary Tables 1-5. In all the tables p-values have been reported to 3 decimal places in order to allow for p-values <0.001 to be reported. This is consistent with BMJ requirements. If the Editorial team advises us that the p-values need to be altered further, we are happy make further changes if necessary. Updates have also been made to the “Prevalence of developmental vulnerability in twins” section to include a statement for significant associations. We are limited by the article length and thus have not gone into too much further detail.

There are a few minor errors such as “Results from a study of an Australia wide study of 261,147 children, singletons and multiples” (line 312).

Response: Errors have been corrected.