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)	Cohort Profile: Early School Years follow up of the Asking Questions in Alcohol Longitudinal Study in Melbourne, Australia (AQUA at 6)
AUTHORS	Muggli, Evelyne; Halliday, Jane; Elliott, Elizabeth; Penington, Anthony; Thompson, Deanne; Spittle, Alicia; Forster, Della; Anderson, Peter Lewis, Sharon; Hearps, Stephen

VERSION 1 – REVIEW

REVIEWER	Suttie, Michael University of Oxford, Nuffield Department of Women's & Reproductive Health
REVIEW RETURNED	01-Sep-2021

GENERAL COMMENTS	Longitudinal observation for PAE is a vital for tracking changes throughout the lifespan, an in turn providing age-appropriate support where necessary. This is an extremely concise, detailed, and well written study profile with very strong collaborative links for further analysis of collected data. I only have a few minor points: No citation for the quoted ABCD study in 4th paragraph of introduction. Cohort Description Was there any intervention or support offered to those with alcohol dependence/addiction? Is there any validation for the telehealth assessments performed compared to the in-person assessments? Exposure Assessment A pictorial drinks guide is a noteworthy method for assessing intake. It may be worth including this as a resource as a supplementary figure.

REVIEWER	Aiton, Neil Brighton and Sussex University Hospital NHS Trust, Neonatology
REVIEW RETURNED	13-Oct-2021

GENERAL COMMENTS	I am happy for these comments to be shared with the authors and for open publication. This is an important study which will contribute significantly to our understanding of the impact of prenatal alcohol exposure on children as they grow up. The authors collectively have the breadth of experience for the study.

	In particular the attempts to carefully collect data at different timepoints (trimesters) during pregnancy and to collect information on important confounding variable is important. The methods of original recruitment/selection and cohort drop-out over time are clearly delineated. Basic detail about original source/method of recruitment should be provided in abstract Given the variability internationally in the definition of an alcohol unit (and the common use of a percentage to define concentration) it would be helpful to have the definition in volume as well as weight (p12 manuscript L42, p13 of pdf). -the original AQUA manuscript looked carefully at confounders, but there is no mention of smoking as an important confounder(is this included in the definition of maternal/paternal drug use or not). More rigorous attention to the hypotheses would be helpful: In this section perhaps given the attention to the careful collection of data in the original study – perhaps the authors could reconsider hypothesis 2 which is very loosely worded? (and) The outline approach which will be used for statistical analysis is mentioned but no detail is provided. The authors comment that this is a 'journal instruction' (see Strobe statement checklist), hence editorial. However, although detailed perhaps this slightly glibly glosses over the complexities of longitudinal comparisons in performance and the complexities of dealing with missing data The initial bullet-pointed section titled "strengths and limitations" (p3 on manuscript) is poorly constructed and unclear. There is also a separate "strengths and limitations section on p19 which is more analytical and objective (so is the first section even needed?) but in this section reference to external validated measure of neurocognitive performance used in assessment of FASD elsewhere would be helpful. Many of the references given relate to follow-up of children related to prematurity (references 56-59) and do demonstrate local expertise available. In some ways I feel that they 'undersell' themselves as far as the real strengths of this study are concerned with respect to the attention paid to the original collection of the data during pregnancy, and the ability to relate this to outcomes There multiple outcome measures based on a battery of psychological measures (supplementary table 1) eg WISC-V amongst others. Although they have attempted to provide a power calculation which does demonstrate a relative sensitivity to potential differences in overall IQ score, this does require some context: is it really the most sensitive way of differentiating children who may have been affected by prenatal alcohol exposure? How does that relate to commonly accepted diagnostic frameworks? Are there other behavioural assessments which might be more sensitive? (eg. measures of attention/executive function?) Likewise with MRI - what is the context of a reduction of brain volume of 64cc in primary school age children? What's the normal range we see in children, for example, and how does this relate to other measures of performance? Perhaps reference to a review article on this topic would be helpful? As they mention, power calculation for craniofacial analysis is also complex, but they have used statistical methods to demonstrate

	significant differences when comparing groups in previous publications, although the methodology is not clearly provided. As with the multiple outcome measures of psychological assessment mentioned above (where they have provided figures on one overall aspect: intelligence core) it should be possible to perhaps pick out individual measures as far as craniofacial features are concerned. This could include comparison between previous 'standard' features involved in traditional dysmorphology assessment (eg lips, philtrum, eye size) or even features 'uncovered' by the 3D assessment which have not been used before such as mid-facial recession. Power calculation assessment (p18) should therefore probably include a more mature assessment of the difficulties involved, and how the team might address those difficulties with respect to the outcome measures being assessed when comparing groups within the cohort, and potentially impact/confound or create difficulties with cohort analysis. Perhaps the authors could explain why they consider (lines 44-47 page 12 manuscript) alcohol abstinence throughout pregnancy differently form lifetime abstinence. One of the important questions is the ability of the study to be replicated. This is not possible as the questions used in the AQUA study – to my knowledge – have not been published. It would be helpful if the study team could consider providing these as supplementary data: or at the very least indicate an intent to do so in the future. In summary – these are all relatively minor points which I hope will be helpful, and I would support publication of this manuscript subject to consideration of the above points.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Dr. Michael Suttie, University of Oxford, University of Oxford
Longitudinal observation for PAE is a vital for tracking changes throughout the lifespan, an in turn providing
age-appropriate support where necessary. This is an extremely concise, detailed, and well written study
profile with very strong collaborative links for further analysis of collected data. I only have a few minor points:
1. No citation for the quoted ABCD study in 4th paragraph of introduction.
Thank you, this has been added.
2. Cohort Description: Was there any intervention or support offered to those with alcohol dependence/addiction?
This was a cohort of women with low-risk pregnancies attending general antenatal clinics and did not include
women with alcohol dependence.
2
3. Cohort Description: Is there any validation for the telehealth assessments performed compared to the
in-person assessments?
Telehealth assessments were introduced as a result of Covid-19 lockdown restrictions. We were not in a

position to validate this approach as telehealth was only provided in situations where families were not able
to attend in person appointments. However, the telehealth assessments only involved carefully selected
measures that we deemed suitable for telehealth. There is growing interest in validating the telehealth administration of neuropsychological measures, with initial results being promising. To address the equivalence issue between telehealth and in-person assessments, we will perform our analyses with and
without the telehealth participants.
4. Exposure Assessment: A pictorial drinks guide is a noteworthy method for assessing intake. It may be
worth including this as a resource as a supplementary figure.
We have added the alcohol questions and drinks guide to the supplementary files.
Reviewer: 2 Dr. Neil Aiton, Brighton and Sussex University Hospital NHS Trust
This is an important study which will contribute significantly to our understanding of the impact of prenatal
alcohol exposure on children as they grow up. The authors collectively have the breadth of experience for
the study. In particular the attempts to carefully collect data at different timepoints (trimesters) during pregnancy and to collect information on important confounding variable is important. The methods of original recruitment/selection and cohort drop-out over time are clearly delineated.
1. Basic detail about original source/method of recruitment should be provided in abstract
We have added the following sentence: “Women attending general antenatal clinics in public hospitals in
Melbourne, Australia, were recruited in their first trimester and followed up three times during pregnancy
and at 12 and 24 months post partum.” As a result, we made minor changes to the wording of the remainder
of the abstract to comply with the 300 word limit.
2. Given the variability internationally in the definition of an alcohol unit (and the common use of a percentage to define concentration) it would be helpful to have the definition in volume as well as weight (p12 manuscript L42, p13 of pdf).
Alcohol content by volume (%ABV) varies greatly between alcoholic beverages drinks definitions of what
constitutes one unit of alcohol in a standard drink differs internationally. Our presentation of ‘grams of absolute alcohol’ consumed is a standardised method of defining alcohol consumption that is independent
of the strength (% ABV) and amount (volume) in different drinks.
3. the original AQUA manuscript looked carefully at confounders, but there is no mention of smoking as
an important confounder (is this included in the definition of maternal/paternal drug use or not).
3
Table 6 lists the co-factors we are considering and tobacco use is listed under ‘Mother health and lifestyle’.
4. More rigorous attention to the hypotheses would be helpful: In this section perhaps given the attention to the careful collection of data in the original study – perhaps the authors could reconsider hypothesis 2 which is very loosely worded? (and)
Any association between exposure and outcome is expected to be influenced by other factors, such as
confounders. There has been much speculation around the contribution of residual cofounding when reporting findings on the effects of lower levels of prenatal alcohol exposure and various child outcomes. We
included hypothesis 2 to highlight the ongoing issue of residual confounding being a possible reason for
conflicting reports on the effects of lower levels of prenatal alcohol exposure. It would be difficult for this
hypothesis to be more specific given the large number of confounders.
5. The outline approach which will be used for statistical analysis is mentioned but no detail is provided.
The authors comment that this is a ‘journal instruction’ (see Strobe statement checklist), hence

editorial. However, although detailed perhaps this slightly glibly glosses over the complexities of longitudinal comparisons in performance and the complexities of dealing with missing data
We can ensure the reviewer that the complexities of dealing with missing data and longitudinal comparisons
etc will be carefully considered in our analyses and reported on in the papers reporting on respective outcomes. However, we decided to only provide an overview of the data analysis to be consistent with the
journal's instruction that "detailed statistical plans should not be reported".
6. The initial bullet-pointed section titled "strengths and limitations" (p3 on manuscript) is poorly constructed and unclear. There is also a separate "strengths and limitations section on p19 which is more analytical and objective (so is the first section even needed?) but in this section reference to external validated measure of neurocognitive performance used in assessment of FASD elsewhere would be helpful. Many of the references given relate to follow-up of children related to prematurity (references 56-59) and do demonstrate local expertise available.
A snapshot of the study's strengths and limitations as a separate bulleted list up front is required by the
journal and the style was modelled on previously published Cohort Profiles in BMJ Open. The section on p19
is a more in-depth version of this. References to validated measures in the assessment of FASD are provided
in 'Outcome measures: Neuropsychological assessments' on page 16. We have amended the relevant
sentence to clarify this.
7. In some ways I feel that they 'undersell' themselves as far as the real strengths of this study are concerned with respect to the attention paid to the original collection of the data during pregnancy, and the ability to relate this to outcomes
4
Thank you for your comment however we believe that we have made this point appropriately. We explain
that we collected extensive data throughout the study in the co-factor and the strengths and limitations
sections.
8. There multiple outcome measures based on a battery of psychological measures (supplementary table
1) eg WISC-V amongst others. Although they have attempted to provide a power calculation which does demonstrate a relative sensitivity to potential differences in overall IQ score, this does require some context: is it really the most sensitive way of differentiating children who may have been affected by prenatal alcohol exposure? How does that relate to commonly accepted diagnostic frameworks? Are there other behavioural assessments which might be more sensitive? (eg. measures
of attention/executive function?)
Our neuropsychological assessments measured performance on numerous domains of cognition based on
previous research in the field. Given the size of the sample is determined, and pre-specifying the statistical
power (80%) and significance level (2-sided 0.05), the magnitude of the effect will be consistent across all
measures (Cohen's f=0.12). For our power calculation we used full-scale IQ as an example, and we accept
that other domains may be more sensitive to subtle effects of low to moderate PAE including attention and
executive function. We had amended this section to try and clarify this point.
9. Likewise with MRI – what is the context of a reduction of brain volume of 64cc in primary school age
children? What's the normal range we see in children, for example, and how does this relate to other measures of performance? Perhaps reference to a review article on this topic would be helpful?
In a sample of typically developing 7-year olds, we have previously reported a mean intracranial volume of
1414cc (SD=99) – ref 62. Thus, a volumetric reduction of 54cc represents an effect of 0.54SD or

approximately 4%. We have updated this section with more appropriate values. Brain MRI is included this
study as the sensitivity to detect subtle effects of low to moderate PAE may be greater using these metrics
than neuropsychological measures that are strongly associated with socio-demographic characteristics.
Please note that we have not hypothesised that brain volumes, or other MRI metrics, would be associated
with neuropsychological outcomes.
10. As they mention, power calculation for craniofacial analysis is also complex, but they have used statistical methods to demonstrate significant differences when comparing groups in previous publications, although the methodology is not clearly provided.
We added some further explanation of the methodology and added 2 extra references.
11. As with the multiple outcome measures of psychological assessment mentioned above (where they
have provided figures on one overall aspect: intelligence core) it should be possible to perhaps pick out individual measures as far as craniofacial features are concerned. This could include comparison between previous 'standard' features involved in traditional dysmorphology assessment (eg lips,
5
philtrum, eye size) or even features 'uncovered' by the 3D assessment which have not been used before such as mid-facial recession.
Our analytical approach is based on no preconception about where the differences in facial shape may
occur. If we find specific, geographic changes in facial shape, we may be able to pinpoint an area suitable for
an individual measure, which can then be included as an outcome measure. However, if we were to use our
images to compare our findings with previously identified features and new features we uncovered, we
would essentially be using the between-group differences already present in the data to in turn define a new
measure to compare between the same groups. We believe that this would likely result in spurious positive
findings which would be difficult t interpret.
12. Power calculation assessment (p18) should therefore probably include a more mature assessment of
the difficulties involved, and how the team might address those difficulties with respect to the outcome measures being assessed when comparing groups within the cohort, and potentially impact/confound or
create difficulties with cohort analysis.
We acknowledge that we have only given examples in our power calculations. We have a very large number
of outcomes across the different neuropsychological, MRI and craniofacial domains and it is not feasible to
provide power calculations across all these different outcomes. We have tried to amend the manuscript to
reflect that we are only providing examples, yet the power calculations are relevant for all outcome domains.
13. Perhaps the authors could explain why they consider (lines 44-47 page 12 manuscript) alcohol abstinence throughout pregnancy differently form lifetime abstinence.
The AQUA study was designed to investigate alcohol use in pregnancy and any potential effects on the
unborn baby in a general population of women who would normally consume alcohol, but may choose to
abstain during pregnancy. Women who are lifetime alcohol abstainers largely do so because of a particular
socio-cultural or religious background and therefore are not part of the study's target group. The sample size
of the lifetime abstainer group was not large and combining women who abstain during pregnancy and

women who never drink alcohol would likely introduce noise when considering confounding factors in our
analyses. Therefore, we excluded lifetime abstainers.
14. One of the important questions is the ability of the study to be replicated. This is not possible as the
questions used in the AQUA study – to my knowledge – have not been published. It would be helpful if
the study team could consider providing these as supplementary data: or at the very least indicate an intent to do so in the future.
We have added the alcohol questions and drinks guide to the supplementary materials.

VERSION 2 – REVIEW

REVIEWER	Aiton, Neil Brighton and Sussex University Hospital NHS Trust, Neonatology
REVIEW RETURNED	07-Dec-2021

GENERAL COMMENTS	The AQUA study is an important study which will contribute to understanding of the effects of prenatal alcohol exposure on long term neurodevelopment as well as facial characteristics. The revised submission contains more comprehensive information which would allow greater under understanding about how the study was conducted as well as allowing the possibility of replication. There is an excellent description about how the study was impacted by the coronavirus pandemic and the measure which were taken. In the structured review checklist, I have highlighted two issues which can easily be addressed: - see below Ethics and consent: issues over consent have become more nuanced over time - Although appropriate ethical review and informed consent has been covered(p23), the potential issue of “ongoing consent” is not clear. [see: Gupta UC 2013 https://doi.org/10.4103/2229-3485.106373] This is potentially important in an on-going cohort study where participants are no longer contributing to to the study, whether actively or passively, or through inability to contact. Eg. In particular making sure any active withdrawal from on-going participation id not necessarily imply total withdrawal from study. These things are difficult, but there should be an attempt to address them. The link between neuropsychological functioning and facial development: This is referred to (p19 L30-32) and is part of one hypothesis, but there is no associated reference or background explanation (eg https://doi.org/10.1542/peds.2012-1371, https://doi.org/10.1371/journal.pone.0043067, https://doi.org/10.1111/acer.13820] Comparative statistics: has any allowance been made for multiple comparisons (Table 2) Few small specific points which were not clear:

	The inclusion of percentages - particularly when discussing subgroups can be confusing The phrase “exposure-representative” or PAE-representative is not clear I assume that refers to stratified sampling? When were the 3 questionnaires administered? (p6 L33) - given that whole purpose of study is about alcohol exposure and outcome, this should probably be be mentioned rather than left to a reference. What Does “ complete information” (p6 L37-38) refer to? What is an “AQUA-6 assessment”? P8 L46 “What does “lived interstate” mean? P8 L48 “Exposure” in the context of photography, presumably refers to prenatal alcohol exposure p16 L9

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2 Dr. Neil Aiton, Brighton and Sussex University Hospital NHS Trust

The AQUA study is an important study which will contribute to understanding of the effects of prenatal alcohol exposure on long term neurodevelopment as well as facial characteristics.

The revised submission contains more comprehensive information which would allow greater under understanding about how the study was conducted as well as allowing the possibility of replication.

There is an excellent description about how the study was impacted by the coronavirus pandemic and the measure which were taken.

In the structured review checklist, I have highlighted two issues which can easily be addressed: - see below

1. Ethics and consent: issues over consent have become more nuanced over time - Although appropriate ethical review and informed consent has been covered(p23), the potential issue of “ongoing consent” is not clear. [see: Gupta UC 2013 https://doi.org/10.4103/2229-3485.106373] This is potentially important in an on-going cohort study where participants are no longer contributing to to the study, whether actively or passively, or through inability to contact. Eg. In particular making sure any active withdrawal from on-going participation id not necessarily imply total withdrawal from study. These things are difficult, but there should be an attempt to address them.

This is correct, consent in longitudinal studies is complex and requires careful attention. We added further detail in the manuscript to clarify our approach as follows:

a) To show which participants totally withdrew from the AQUA study upon invitation to the current follow-up, Table 1 now contains a new line separating out participants who did not take part in AQUA at 6 into opting out of AQUA at 6 (e.g. sorry we are just too busy right now; n=271) and withdrawing from the study altogether (e.g. please do not contact me again about this study; n=37).

b) To show how we will treat existing AQUA study data going forward, we added a new paragraph about consent to previous waves of the study to the section ‘Additional information’:

"Ethics approval and consent to participate in previous waves of the AQUA study: Ethical oversight of the cohort's recruitment and pre-birth and neonatal follow-ups was provided by the Eastern Health Research and Ethics Committee (E54/1011) and the Human Research Ethics Committees of Mercy Health (R11/14), Monash Health (11071), the Royal Women's Hospital (11/20) and the Royal Children's Hospital (31055). The latter also included approval of all procedures pertaining to the 12 and 24-month postpartum follow-ups. Families who have not actively withdrawn their consent to participate are ongoing study participants and their data may be included in future analyses by the project team if they are deemed to be in line with information that was provided to participants at the time of consent."

c) Further, under 'Availability of data and materials' there is a sentence to explain that "AQUA at 6 study families have the option to consent for their data to be used in future related and ethically approved projects". 93% of participants consented to this.

2. The link between neuropsychological functioning and facial development:

This is referred to (p19 L30-32) and is part of one hypothesis, but there is no associated reference or background explanation (eg https://doi.org/10.1542/peds.2012-1371, https://doi.org/10.1371/journal.pone.0043067, https://doi.org/10.1111/acer.13820]

Thank you for suggesting the above references, We have added Suttie et al 2018 to the following sentence in the second paragraph of the introduction: "High levels of alcohol exposure to the fetal brain can cause a spectrum of structural brain abnormalities, facial dysmorphology, neurological problems and neurodevelopmental impairments, collectively termed Fetal Alcohol Spectrum Disorder (FASD)."

References to support this statement are now:

- Riley E, Infante MA, Warren K. Fetal Alcohol Spectrum Disorders: An Overview. Neuropsychol Rev. 2011; 21:73-80.
- Mattson SN, Crocker N, Nguyen TT. Fetal alcohol spectrum disorders: neuropsychological and behavioral features. Neuropsychol Rev. 2011;21(2):81-101.
- Suttie, M. et al. Combined face–brain morphology and associated neurocognitive correlates in fetal alcohol spectrum disorders. Alcoholism: Clinical and Experimental Research 42, 1769-1782 (2018).
- Bower C, Elliot EJ for the Australian FASD Collaboration. Report to the Australian Government Department of Health: Australian Guide to the diagnosis of Fetal Alcohol Spectrum Disorder (FASD). 2016;ISBN. 978-0-6481297-4-5.

3. Comparative statistics: has any allowance been made for multiple comparisons (Table 2)

For the descriptive comparison of participants to non-participants, our preference is to retain a statistical significance level of $p<0.05$, given this is examining demographic characteristics of the sample and allows for comparisons to be more sensitive to small differences.

Few small specific points which were not clear:

4. The inclusion of percentages - particularly when discussing subgroups can be confusing

We appreciate that adding percentages add more numbers into tables and text, but we believe that this is an easy way for the reader to put the raw numbers into context. For example, the percentages in Table 4 allow the reader to easily review proportions across exposure groups, which would be difficult if only frequencies are reported. However, in one sentence we have deleted percentages reported in the text which is also reported in Table 4 directly below.

5. The phrase “exposure-representative” or PAE-representative is not clear I assume that refers to stratified sampling?

We did not actually stratify our cohort by exposure group for our MRI and earlier craniofacial studies. Rather, we continued to invite families taking part in the study to also complete a brain MRI or 3D phot (in the earlier AQUA follow-up) until we achieved the target number in each of the exposure groups. This approach is referred to as “representative sampling” and is commonly used to reduce the chance of exposure bias where we were only able to offer testing in a sub-sample of the cohort.

6. When were the 3 questionnaires administered? (p6 L33) - given that whole purpose of study is about alcohol exposure and outcome, this should probably be be mentioned rather than left to a reference.

We expanded this sentence to include the timing of the questionnaires: “During pregnancy, women completed three questionnaires, 1) at recruitment (<18 weeks’ gestation); 2) at 25 weeks’ gestation; and 3) at 35 weeks’ gestation.”

7. What Does “ complete information” (p6 L37-38) refer to?

We only included mother and child pairs in our follow-up studies where complete information was available on frequency and level of alcohol consumption at all pregnancy timepoints, thereby enabling us to assign an exposure classification.

8. What is an “AQUA-6 assessment”? P8 L46

To address this, we have amended the sentence as follows: “Following consent, we obtained externally assessed scores from the family’s private psychologist for nine children, which in two instances we complemented with a partial assessment of the remaining tests.”

9. “What does “lived interstate” mean? P8 L48

We have amended this to read “lived in another state of Australia”.

10. “Exposure” in the context of photography, presumably refers to prenatal alcohol exposure p16 L9

Correct. The exposure of interest in the AQUA study is prenatal alcohol exposure. Accordingly, we changed the wording in this line to “children’s prenatal alcohol exposure”.

VERSION 3 – REVIEW

REVIEWER	Aiton, Neil Brighton and Sussex University Hospital NHS Trust, Neonatology
REVIEW RETURNED	14-Dec-2021

GENERAL COMMENTS	All previous comments have been addressed. recommend for publication.