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: The Australian Genetics of Depression Study
AUTHORS	Byrne, Enda; Kirk, Katherine; Medland, Sarah; McGrath, John; Colodro-Conde, Lucia; Parker, Richard; Cross, Simone; Sullivan, Lenore; Statham, Dixie; Levinson, Douglas; Licinio, Julio; Wray, Naomi; Hickie, Ian; Martin, Nicholas

VERSION 1 – REVIEW

REVIEWER	Michael Silverman Icahn School of Medicine at Mount Sinai, New York, USA
REVIEW RETURNED	05-Aug-2019

GENERAL COMMENTS	The manuscript reports on the attempt to acquire a population for behavioral and genetic study of depression. Results are descriptive of the sample population cohort obtained. Unfortunately, the rate of participation is low, prone to considerable response bias and limited in resolution - none of which appears to be addressed by the authors. As a research-based manuscript, it offers little to increase our knowledge base. I'm unsure of the agenda of the publication other than to attempt to validate what currently seems to be an unnormalizable cohort. The availability of such a cohort is unusual and extremely valuable towards the understanding of depression. I would encourage the investigators to consider ways of improving response rates. There are a few minor grammatical errors throughout. Example below. Page 5; Participant and Patient Involvement; last sentence; "data from the cohort will be sent" I believe should read in the past tense. Page 8; "merican" should read "American." I sincerely hope the authors continue to pursue this effort.

REVIEWER	Hanna van Loo University of Groningen
REVIEW RETURNED	19-Aug-2019

GENERAL COMMENTS	Review Byrne et al. “Cohort Profile: The Australian Genetics of Depression Study” This paper describes the design of a new cohort study for research into the etiology and antidepressant treatment of depression. This

	large cohort is an important contribution to science in psychiatry. Depression's etiology is still largely unknown and antidepressant treatment is not effective in quite a large proportion of patients. Large general population studies as the Australian Genetics of Depression study are therefore highly needed. Byrne and colleagues have included n~20,000 Australian residents, of whom ~3,000 were recruited through their prescription history, and ~17,700 through public appeal. About 75% of the sample provided saliva for genotyping. This sample will be studied to identify genetic and non-genetic risk factors for depression, antidepressant response, and side-effects, and to contribute to the common effort of the Psychiatric Genomics Consortium. The data will be available to all interested researchers through collaboration, and thus this study will be an important source for future research. See below a few comments and questions: <u>Major comments:</u> - As the authors mention, the recruitment strategy has some limitations. It led to a sample that is relatively young and highly educated. This sample may also miss the more severe cases of depression as the study excluded institutionalized individuals. Why was that? Were there other exclusion criteria? - A very low proportion of subjects was included via the pharmaceutical prescription history. If I understood it correctly, only ~3,000 subjects responded to the invitation whereas 110,000 invitation letters were sent to subjects who had received antidepressant prescriptions in the past 4.5 years. How do the authors explain this low number of respondents? Does it affect the representativeness of the cohort? - What was the response rate for the additional questionnaires? How many respondents provided data? - Side-effects from antidepressants are often hard to disentangle from symptoms of depression. Libido loss, insomnia, fatigue, weight gain/loss are all depression symptoms but can also be side-effects from antidepressants. Is it possible with this cohort study to differentiate between depression symptoms and side-effects? - Not all participants may have remembered their antidepressant use correctly. Will their reports also compared/combined with the recorded antidepressants in the PBS records?

	- The controls of QSkin were not screened with validated structured questionnaires for the absence/presence of lifetime depression, but only through self-report “have you ever been diagnosed with depression?”. This means that some of the controls might in fact be MD cases if the diagnosis was missed, they forgot about their diagnosis, or they never went to see a doctor. Also, if the included controls are still very young, there is quite a high proportion that eventually will have depression (~20%). Do the authors see solutions for this? Minor comments: - What is the ancestry of the participants? - At first, an upper age restriction was used in recruiting participants (18-30 years), later this restriction was removed. Why was that? - P6: Genotyping is supposed to be finished mid 2019. Has genotyping already finished? An update would be useful. - A legend may be useful for the “Sample age distribution” plot, to clarify how the recruitment strategy led to this age distribution.

REVIEWER	Laurie Hannigan Lovisenberg Diaconal Hospital (Norway) & University of Bristol (UK)
REVIEW RETURNED	03-Sep-2019

GENERAL COMMENTS	**Review: bmjopen-2019-032580** This article presents a cohort profile of the new Australian Genetics of Depression Study, including details on study design, recruitment strategy and included measures, as well as information on characteristics of the sample and preliminary results for core measures. Much of the information is well-presented and there are some useful figures and to help readers understand the structure of the data. My main comments are focused on improving the clarity and informativeness of this paper for what I consider to be its two primary audience groups, namely: 1. Those interested in depression genetics & the potential future contributions of these data 2. Those interested in aspects of large scale cohort study research design Main comments:

	1. I think the structure, at present, is disjointed and makes the reader work too hard to find relevant information. Borrowing from the STROBE guidelines for reporting on cohort studies, as far as you can within journal requirements, I would suggest restructuring to something along the lines of (sub-sections indented): Introduction Objectives Design Recruitment strategy Cases Controls Enrolment procedure Record linkage (currently 'Access to Medicare and PBS records') Measures Development and structure (currently under 'Questionnaire') Core questionnaire(s) Saliva collection and DNA extraction Participant/patient involvement Results Sample characteristics Descriptive data Mental health history CIDI depression Family history Antidepressant usage Discussion 2. Introduction: I think the clarity of the intro overall would be improved if you can you clearly state the objectives/priorities for the study design – i.e., was precedence given to depression case ascertainment vs study of antidepressant use? Or were both of equal

	importance? What role (if any) did an awareness of the issues around heterogeneity in depression GWAS play in influencing the design of AGDS? What are the implications of recruiting a sample based (in part) on antidepressant use? 3. Design: Given that recruitment strategies eventually overlapped, and given the predominance of the media/self-enrolment strategy in ultimately leading to participation, I think the sections describing them could be collapsed and condensed considerably 10,000 cases are mentioned in the intro – where did this target come from? Did you have case control ratio target? How did this feed into study design; did you have a response rate expectation? Were the company able to provide information on 'conversions' – impressions to link clicks to registrations – for digital campaign? The questionnaire section is good – would it be possible to add a figure (or supplementary figure) with screenshots demonstrating what that participant actually sees, with respect to the core and satellite modules? I think the Figure 1 Schematic should be introduced later as a summary of the design once all aspects have been mentioned, and should include information on recruitment periods I know you are focused on core measures for the preliminary results, but I think the completeness of this article as an overview of the cohort would be greatly increased by providing full information on all the 'satellite' measures – even if only in a supplementary appendix (I appreciate that there will be a lot of information). Or, if there is online data dictionary or similar, make clear where this can be accessed? Is it correct that reason for prescription/benefits/side-effects were only assessed for 10 most common antidepressants? Can you make sure this is clear – and also perhaps outline the rationale for this? Table 4 – can you clarify in legend are proportions of all individuals or all reporting taking top 10 antidepressants? 4. Results ('Findings to date'): Could you add an equivalent column with info on controls into Table 1 for context? And – more broadly – it would be worth conveying more clearly whether the Qskin controls should be considered as 'part' of the Australian Genetics of Depression Study cohort; this would be of interest, for example, to researchers interested in collaborating/using the data Can you give an overview of response rates broken down by module (i.e., for the core – though this is presumably everyone who consented? – and each satellite module)?

	Figure 2 - where does the data for median time-to-complete come from – piloting or the real thing? For completeness, have you considered getting a small test sample to complete the 'anxieties and phobias' and 'General Physical and Mental Health' satellites and including their median time-to-complete, to account for the timer failure (and noting that this was done)? Figure 4 – age at onset: I think the right-hand axis label is misleading, as it could be lead to the plot being isinterpreted as % participants with depression in each age bin (i.e., prevalence by age). Consider relabeling as per Figs 5&6? Figures 5&6 – is there a reason these analyses are not broken down by sex – this would be both interesting and consistent with the other figures Figures 7&8 – both of these feel a little redundant to me (or at least inefficient in terms of the amount of information conveyed vs space taken). I completely understand that "more detailed analyses… will follow" – but replacing these two figures with one showing the results of any one of a number of possible stratified analyses (e.g., number of antidepressants taken by duration of worst episode, etc) would really help to emphasise the potential that is inherent in these data (such an addition is at your discretion – just a suggestion to increase the impact on the reader) 5. Discussion – opening sentence could be shortened for clarity and to avoid unintentional implication that willingness to provide a saliva sample was a pre-requisite for recruitment More engagement with the issues around selection/response rate/possible biases in future analyses and potential strategies for handling them in the discussion would be beneficial to the readers. Paragraph 3 of the discussion, in particular, could be re-written to this effect (the comparison between the recruitment methods is interesting from a study design point of view, but not as relevant to future work as selection effects in the sample as a whole, which came predominantly from self-enrolment). In general, more balance is needed in the discussion between the interpretation of preliminary findings (which dominates currently, and could probably be reduced to one or two paragraphs), and discussion of the characteristics of cohort and selection issues (mentioned above), reflections

	on the successes and shortcomings of recruitment strategies and other aspects of design, and potential unique contributions using these data in areas such as investigating heterogeneity of depression, treatment response etc – i.e., things for which a cohort profile paper is a specific vehicle (whereas the substantive other issues you touch on superficially will be better addressed in full detail in the many empirical papers that will results from this study). Minor comments The use of the AGDS acronym vs the full cohort name throughout is inconsistent – it would be easier on the reader and beneficial for future reference if the acronym was used consistently (including in the title?) p.2-line13 This total includes recruitment by both strategies, not just 'traditional and social media' 2-20 Include % agreeing to genotyping? 2-22/25 Give % values (or example % values) in place of descriptive text such as "overwhelming majority", "Rates…were high", and "Two-thirds" 2-47 "An online study… led to the sample being mostly younger people" It is not clear to me that this is correct, looking at the data. It is probably not a primary limitation in any case – perhaps make a more general point about low response rate (PBS) and self-selection related biases 3-53 "Participants were invited…" this sentence is not needed in the introduction 4-38 "294 responses" over what time period? Any changes based on this for second wave? Low response rate for the PBS recruitment should be covered in discussion – what were the reasons, should this kind of recruitment strategy be avoided? 4-57 missing word "been"? 8-7 missing "A" in American 10-23 this sentence is unclear and could be rephrased 10-33 unclear what "had" refers to here – rephrase this sentence 11-15 missing word "they"? ('but [they] will also')

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

The manuscript reports on the attempt to acquire a population for behavioral and genetic study of depression. Results are descriptive of the sample population cohort obtained. Unfortunately, the rate of participation is low, prone to considerable response bias and limited in resolution - none of which appears to be addressed by the authors.

We thank the reviewer for the comment. Indeed the rate of participation through the pharmaceutical benefits scheme was low, but overall the study was highly successful in recruiting a large number of participants in a short time. We note that the rate of recruitment into UK Biobank, the largest genetically informed sample in the world was 5-6% of those approached. We do not believe this will bias gene mapping analyses. We further address the response bias in the Discussion.

As a research-based manuscript, it offers little to increase our knowledge base. I'm unsure of the agenda of the publication other than to attempt to validate what currently seems to be an unnormalizable cohort. The availability of such a cohort is unusual and extremely valuable towards the understanding of depression. I would encourage the investigators to consider ways of improving response rates.

We thank the reviewer for this comment and indeed we are exploring methods to increase response rates. The aim of publication was to provide an overview of how the cohort was recruited as we believe it will be of interest to readers seeking to establishing similar studies and also to give a feel for the data and how it may be useful in understanding depression and antidepressant use. Further manuscripts that conduct more detailed analyses will contribute to the knowledge base.

There are a few minor grammatical errors throughout. Example below.

Page 5; Participant and Patient Involvement; last sentence; "data from the cohort will be sent" I believe should read in the past tense.

Apologies for the confusion. What was meant here was that any papers deriving from the study will be sent to the participants when they are published. We have made a slight correction to this sentence.

Page 8; "merican" should read "American."

Thank you for pointing this out. This error has been corrected.

I sincerely hope the authors continue to pursue this effort.

We thank the reviewer for the comment. We are continuing to pursue avenues to increase the response rate.

Reviewer 2:

As the authors mention, the recruitment strategy has some limitations. It led to a

sample that is relatively young and highly educated. This sample may also miss the

more severe cases of depression as the study excluded institutionalized individuals.

Why was that? Were there other exclusion criteria?

The exclusion of institutionalized individuals was only in the recruitment from the Pharmaceutical Benefits Scheme because the Department of Human Services could not send letters to people in institutions. There was also likely to be difficulties in getting saliva kits back from participants in institutions. Any person who was institutionalized and was able to participate could enrol in the online study.

People with severe depression may be less likely to respond to a media appeal. This is likely to be a weakness of any study design, even those that recruit samples in clinics. We are now seeking to recruit more patients in clinical settings such as psychiatric hospitals which may allow us to recruit more patients with severe depression. There were no other exclusion criteria except that participants had to be at least 18 years of age. Depending on their responses to the questionnaire, participants may be excluded from analyses that seek to address specific research questions.

A very low proportion of subjects was included via the pharmaceutical prescription

history. If I understood it correctly, only ~3,000 subjects responded to the invitation

whereas 110,000 invitation letters were sent to subjects who had received

antidepressant prescriptions in the past 4.5 years. How do the authors explain this

low number of respondents? Does it affect the representativeness of the cohort?

There are a number of possible explanations. It is our opinion that appeals on television, radio or social media where someone from the study discusses how the project can help to better understand depression has more effect on people. The advertising included a number of people with a history of depression who told their stories. A letter may be seen as impersonal and participants may have quickly discarded them.

It may affect the representativeness of the cohort. However, it’s not clear how another study design would have been able to recruit such a large number of participants in such a short timeframe that better represents the population. Nearly all study designs rely on volunteer participation and hence may not represent the population.

What was the response rate for the additional questionnaires? How many

respondents provided data?

We now provide the response rates to each module in a Supplementary Table.

Side-effects from antidepressants are often hard to disentangle from symptoms of depression. Libido loss, insomnia, fatigue, weight gain/loss are all depression symptoms but can also be side-effects from antidepressants. Is it possible with this cohort study to differentiate between depression symptoms and side-effects?

We plan to conduct multivariate analyses that will investigate the covariance of depressive symptoms and side-effects. For instance, if side-effects are purely depression symptoms then we would expect that polygenic risk scores for depression symptoms will be predictive for side-effects.

Not all participants may have remembered their antidepressant use correctly. Will their reports also compared/combined with the recorded antidepressants in the PBS records?

Yes we plan to compare the self-report data with those from the PBS records. These records will only provide information for the last 4.5 years, but should be sufficient to give an overall estimate of the accuracy of self-report. We have compared the frequency of reported antidepressant use to the summary data from the PBS records for the whole population of Australia and the relative frequencies match very accurately.

The controls of QSkin were not screened with validated structured questionnaires for the absence/presence of lifetime depression, but only through self-report “have you ever been diagnosed with depression?”. This means that some of the controls might in fact be MD cases if the diagnosis was missed, they forgot about their diagnosis, or they never went to see a doctor. Also, if the included controls are still very young, there is quite a high proportion that eventually will have depression (~20%). Do the authors see solutions for this?

The age of the QSKIN samples was between 40 and 69 when completing the questionnaire about depression. While some may go on to a diagnosis depression, the sample is past the peak age at onset for depression which makes it ideal as a control sample. Some of the participants may have forgotten or misreported a diagnosis of depression. This is likely to be a feature of any control sample for genetic studies of psychiatric disorders. It is not uncommon for genetic studies to use completely unscreened controls, where the larger sample size overcomes the problem of some of the controls having the disorder of interest. It would require another large effort and investment to clinically screen tens of thousands of controls so we believe that while maybe not completely accurate, self-report is more than adequate for the purposes of the study.

Reviewer 3:

This article presents a cohort profile of the new Australian Genetics of Depression Study, including details on study design, recruitment strategy and included measures, as well as information on characteristics of the sample and preliminary results for core measures. Much of the information is well-presented and there are some useful figures and to help readers understand the structure of the data.

My main comments are focused on improving the clarity and informativeness of this paper for what I consider to be its two primary audience groups, namely:

1. Those interested in depression genetics & the potential future contributions of these data

2. Those interested in aspects of large scale cohort study research design

Main comments:

1. I think the structure, at present, is disjointed and makes the reader work too hard to find relevant information. Borrowing from the STROBE guidelines for reporting on cohort studies, as far as you can within journal requirements, I would suggest restructuring to something along the lines of (sub-sections indented):

Introduction

Objectives

Design

Recruitment strategy

Cases

Controls

Enrolment procedure

Record linkage (currently ‘Access to Medicare and PBS records’)

Measures

Development and structure (currently under ‘Questionnaire’)

Core questionnaire(s)

Saliva collection and DNA extraction

Participant/patient involvement

Results

Sample characteristics

Descriptive data

Mental health history

CIDI depression

Family history

Antidepressant usage

Discussion

We thank the Reviewer for this helpful suggestion about restructuring the manuscript and have now reorganised the manuscript according to their suggestion. We believe this has substantially improved the clarity.

2. Introduction: I think the clarity of the intro overall would be improved if you can you clearly state the objectives/priorities for the study design – i.e., was precedence given to

depression case ascertainment vs study of antidepressant use? Or were both of equal

importance? What role (if any) did an awareness of the issues around heterogeneity in

depression GWAS play in influencing the design of AGDS? What are the implications of

recruiting a sample based (in part) on antidepressant use?

We thank the reviewer for the comment and have added a section on the specific objectives of the project. The primary goal was to recruit depression cases and once they had been recruited, to ask them about their antidepressant use. Precedence wasn't given to one over the other, hence why the antidepressant questions were included in the core module that needed to be completed by everyone.

The issue of heterogeneity in depression did not play any role in the recruitment strategy and mainly came into the design of the secondary modules of the questionnaire, where a wider range of variables were included.

3. Design: Given that recruitment strategies eventually overlapped, and given the

predominance of the media/self-enrolment strategy in ultimately leading to participation, I

think the sections describing them could be collapsed and condensed considerably

10,000 cases are mentioned in the intro – where did this target come from? Did you have

case control ratio target? How did this feed into study design; did you have a response rate

expectation?

We take the reviewers point about the predominance of the media strategy in terms of the recruitment strategy. However, given this is to our knowledge one of the first studies to utilise the Australian

pharmaceutical records to recruit participants, we feel it's important to give a reasonably detailed account of this process and to subdivide this section. The target of 10,000 cases was not based on a power calculation but was an informal target that the investigators thought would represent a substantial contribution to the global effort to identify genes for depression.

Were the company able to provide information on 'conversions' – impressions to link clicks

to registrations – for digital campaign?

We were unable to get any information from the media company about this.

The questionnaire section is good – would it be possible to add a figure (or supplementary

figure) with screenshots demonstrating what that participant actually sees, with respect to

the core and satellite modules?

We thank the reviewer for the suggestion and now include some screenshots of the questionnaire as Supplementary Figures.

I think the Figure 1 Schematic should be introduced later as a summary of the design once

all aspects have been mentioned, and should include information on recruitment periods

I know you are focused on core measures for the preliminary results, but I think the

completeness of this article as an overview of the cohort would be greatly increased by

providing full information on all the 'satellite' measures – even if only in a supplementary

appendix (I appreciate that there will be a lot of information). Or, if there is online data

dictionary or similar, make clear where this can be accessed?

This is a good suggestion. We will include the entire questionnaire as a supplement for those who are interested.

Is it correct that reason for prescription/benefits/side-effects were only assessed for 10

most common antidepressants? Can you make sure this is clear – and also perhaps outline

the rationale for this?

Yes this is correct. The rationale was because the recruitment strategy in the PBS focussed on prescriptions of the most commonly used antidepressants and so the questionnaire was designed to align with the recruitment strategy. We now mention this in the methods section.

Table 4 – can you clarify in legend are proportions of all individuals or

all reporting taking top 10 antidepressants?

Thank you for the suggestion. This is the proportion of those who have taken one of the top 10 most frequent antidepressants. We have altered the legend to clarify this.

4. Results ('Findings to date'): Could you add an equivalent column with info on controls into

Table 1 for context?

And – more broadly – it would be worth conveying more clearly

whether the Qskin controls should be considered as 'part' of the Australian Genetics of

Depression Study cohort; this would be of interest, for example, to researchers interested in

collaborating/using the data

The QSKIN study is not part of the AGDS cohort, but is a separate cohort that we will use as controls for analyses. For those wishing to use QSKIN, they would need to contact the principal investigators of that study. We have tried to make this more clear in the manuscript.

Can you give an overview of response rates broken down by module (i.e., for the core –

though this is presumably everyone who consented? – and each satellite module)?

Thank you for the suggestion. We have added a Supplementary Table with the response rates by module.

Figure 2 - where does the data for median time-to-complete come from – piloting or the

real thing? For completeness, have you considered getting a small test sample to complete

the 'anxieties and phobias' and 'General Physical and Mental Health' satellites and including

their median time-to-complete, to account for the timer failure (and noting that this was

done)?

This data comes from piloting of the questionnaire. We have altered the legend of the figure to make it clear that it comes from piloting. We have also now piloted the anxiety disorders and General Physical and Mental Health modules and added the median time to the Figure.

Figure 4 – age at onset: I think the right-hand axis label is misleading, as it could be lead to

the plot being is interpreted as % participants with depression in each age bin (i.e.,

prevalence by age). Consider relabeling as per Figs 5&6?

Thank you for the suggestion. We have relabelled to make it consistent with Figs 5&6.

Figures 5&6 – is there a reason these analyses are not broken down by sex – this would be both interesting and consistent with the other figures

We have now altered Figures 5 and 6 to give the break down by sex.

Figures 7&8 – both of these feel a little redundant to me (or at least inefficient in terms of the amount of information conveyed vs space taken). I completely understand that "more detailed analyses… will follow" – but replacing these two figures with one showing the results of any one of a number of possible stratified analyses (e.g., number of antidepressants taken by duration of worst episode, etc) would really help to emphasise the potential that is inherent in these data (such an addition is at your discretion – just a suggestion to increase the impact on the reader)

We thank the reviewer for this suggestion. We agree that stratified analyses will be more informative, but would prefer to leave them for a forthcoming comprehensive analysis of patterns of antidepressant use.

5. Discussion – opening sentence could be shortened for clarity and to avoid unintentional implication that willingness to provide a saliva sample was a pre-requisite for recruitment

More engagement with the issues around selection/response rate/possible biases in future analyses and potential strategies for handling them in the discussion would be beneficial to the readers. Paragraph 3 of the discussion, in particular, could be re-written to this effect (the comparison between the recruitment methods is interesting from a study design point of view, but not as relevant to future work as selection effects in the sample as a whole, which came predominantly from self-enrolment).

In general, more balance is needed in the discussion between the interpretation of preliminary findings (which dominates currently, and could probably be reduced to one or two paragraphs), and discussion of the characteristics of cohort and selection issues (mentioned above), reflections on the successes and shortcomings of recruitment strategies

and other aspects of design, and potential unique contributions using these data in areas

such as investigating heterogeneity of depression, treatment response etc – i.e., things for

which a cohort profile paper is a specific vehicle (whereas the substantive other issues you

touch on superficially will be better addressed in full detail in the many empirical papers

that will results from this study).

We thank the reviewer for this comment and have now rewritten large parts of the discussion. In particular, we have included more discussion of the recruitment strategy and the successes and weaknesses. Furthermore, we have shortened the discussion of the initial study findings so that there is more balance.

Minor comments

The use of the AGDS acronym vs the full cohort name throughout is inconsistent – it would be easier on the reader and beneficial for future reference if the acronym was used consistently (including in the title?)

p.2-line13 This total includes recruitment by both strategies, not just 'traditional and social media'

2-20 Include % agreeing to genotyping?

2-22/25 Give % values (or example % values) in place of descriptive text such as "overwhelming

majority", "Rates…were high", and "Two-thirds"

2-47 "An online study… led to the sample being mostly younger people" It is not clear to me that

this is correct, looking at the data. It is probably not a primary limitation in any case – perhaps make

a more general point about low response rate (PBS) and self-selection related biases

3-53 "Participants were invited…" this sentence is not needed in the introduction

4-38 "294 responses" over what time period? Any changes based on this for second wave? Low response rate for the PBS recruitment should be covered in discussion – what were the reasons, should this kind of recruitment strategy be avoided?

4-57 missing word "been"?

8-7 missing "A" in American

10-23 this sentence is unclear and could be rephrased

10-33 unclear what "had" refers to here – rephrase this sentence

11-15 missing word "they"? ('but [they] will also')

We thank the reviewer for have addressed all of the above minor comments

VERSION 2 – REVIEW

REVIEWER	Laurie Hannigan Lovisenberg Diaconal Hospital, Norway & University of Bristol, UK
REVIEW RETURNED	27-Jan-2020

GENERAL COMMENTS	This was not an easy revision to review. On the face of it, the authors have been responsive to concerns raised by myself and other reviewers, in particular about selection/recruitment issues and relative lack of attention paid to these in the manuscript. Indeed, the changes to the discussion addressing this imbalance are - alongside the addition of demographic information for the QSkin controls, and of more detail about the outcomes - the most valuable improvements that have been made to the manuscript. However, overall the changes to the manuscript are relatively minor - and some of the assertions in the response to reviewers about changes are not backed up by evidence of actions in text (for example, the authors claim to have revised the structure to improve clarity, but apart from the addition of the Objectives section, I cannot see any changes to the order or content of sub-headings whatsoever; there is also no change to the axis label of Figure 4 in the new version I have, despite the authors statement to the contrary - though this latter point is clearly a minor issue). Other requests for the addition of supplementary information have largely been responded to - thank you, I think this helps. Overall, the concerns I had about the structural clarity of the manuscript are mostly still present. However, at some point this becomes an issue for the editor/authors to resolve to their satisfaction rather than mine. I do think this paper should be published as it now provides a reasonably comprehensive overview of this important sample, the process by which it was recruited, and the issues arising from that process. Whether that is in its current form, or subject to further formal revisions, is an editorial matter. Congratulations on recruiting the AGDS sample, which is sure to make an important contribution to the global MDD genetics effort. Minor comments: With the demographic info from the controls now included, a couple of things stand out. The age difference between the case and control samples, and implications thereof, should probably be commented upon. Similarly, the apparent education differences (7.7% vs 31.5% completing senior high school!? Is this a cohort effect, or a coding issue?). Finally, the genotyping rate differs, which may reflect the difference in the structure of the QSkin study (I guess people could not participate without providing saliva samples?). The implications of this should be considered - I already think your statement that selection issues "...are unlikely to affect gene mapping efforts" probably needs more unpacking. We know that participation is influenced by genetics overlapping with many psychiatric traits (see genetics of participation work from UKB and other cohorts), and just because it is - as you say - an issue for any study allowing for volunteer participation, doesn't

	necessarily mean it is not a problem we should be concerned about. You might at least be able to compare characteristics of saliva sample providers vs non-providers within the AGDS sample, to better characterise the issue and potentially develop sampling weights for use down the line.

VERSION 2 – AUTHOR RESPONSE

This was not an easy revision to review. On the face of it, the authors have been responsive to concerns raised by myself and other reviewers, in particular about selection/recruitment issues and relative lack of attention paid to these in the manuscript. Indeed, the changes to the discussion addressing this imbalance are - alongside the addition of demographic information for the QSkin controls, and of more detail about the outcomes - the most valuable improvements that have been made to the manuscript. However, overall the changes to the manuscript are relatively minor - and some of the assertions in the response to reviewers about changes are not backed up by evidence of actions in text (for example, the authors claim to have revised the structure to improve clarity, but apart from the addition of the Objectives section, I cannot see any changes to the order or content of sub-headings whatsoever; there is also no change to the axis label of Figure 4 in the new version I have, despite the authors statement to the contrary - though this latter point is clearly a minor issue). Other requests for the addition of supplementary information have largely been responded to - thank you, I think this helps.

Overall, the concerns I had about the structural clarity of the manuscript are mostly still present. However, at some point this becomes an issue for the editor/authors to resolve to their satisfaction rather than mine. I do think this paper should be published as it now provides a reasonably comprehensive overview of this important sample, the process by which it was recruited, and the issues arising from that process. Whether that is in its current form, or subject to further formal revisions, is an editorial matter.

We thank the reviewer for their comments in assessing the revised version of the paper. We agree with the reviewer that the changes to the structure of the manuscript were minor relative to the suggestions of the reviewer and have made further changes to the ordering and structure to bring it closer in line with the suggested outline in the original review.

We apologise for the oversight regarding Figure 4. This was a mistake where the original figure was uploaded instead of the updated version. We have now uploaded the figure with the correct axis label.

Minor comments:

With the demographic info from the controls now included, a couple of things stand out. The age difference between the case and control samples, and implications thereof, should probably be commented upon. Similarly, the apparent education differences (7.7% vs 31.5% completing senior high school!? Is this a cohort effect, or a coding issue?).

We have now added a section to the discussion on the differences between the case and control cohorts as regards differences in age and education – which are related to each other. There is likely

to be a significant cohort effect because the rate at which people obtain undergraduate and postgraduate degrees in Australia has increased linearly over time. For example, Australian census data showed that the proportion of the Australian population that holds a post-school qualification increased from 46% to 56% from 2006 to 2016.

On the other hand, it's likely that this doesn't explain all of the differences in education level between the cohorts. We now expand on this in more detail in the discussion.

Finally, the genotyping rate differs, which may reflect the difference in the structure of the QSkin study (I guess people could not participate without providing saliva samples?). The implications of this should be considered - I already think your statement that selection issues "...are unlikely to affect gene mapping efforts" probably needs more unpacking.

The reviewer is right that the differences in the genotyping rate may reflect the difference in structure of the QSkin study. When advertising the AGDS, it was made clear to potential participants that the focus of the study was genetic risk factors and that if they agreed to participate they would be sent a saliva kit in the mail. The QSkin study was established as an epidemiological study of risk factors for melanoma and genetic data was not collected initially. Therefore, collection of genetic data was not discussed with participants at the time of initial enrolment. The genetic arm of the study was only initiated later. This likely contributes to the differences in response rate to the genetic data collection between the two studies.

We know that participation is influenced by genetics overlapping with many psychiatric traits (see genetics of participation work from UKB and other cohorts), and just because it is - as you say - an issue for any study allowing for volunteer participation, doesn't necessarily mean it is not a problem we should be concerned about.

You might at least be able to compare characteristics of saliva sample providers vs non-providers within the AGDS sample, to better characterise the issue and potentially develop sampling weights for use down the line.

We thank the reviewer for this comment and now address the issue of volunteer participation and potential ways of address it in the discussion.

VERSION 3 – REVIEW

REVIEWER	Laurie Hannigan Lovisenberg Diaconal Hospital, Norway and University of Bristol, UK
REVIEW RETURNED	26-Feb-2020

GENERAL COMMENTS	I think your additions have improved the paper; thanks for being responsive.

VERSION 3 – AUTHOR RESPONSE

In accordance with the Editorial requests, we have added the Collaboration section and also responded to the comments of Reviewer 2 in the text. Attached please find our response to Reviewer 2's comments. Their comments are shown in black, our original response is shown in red, and how we have now addressed that in the text is shown in green. All changes to the text since the previous revision are highlighted in red in the main manuscript document.

Reviewer 2:

As the authors mention, the recruitment strategy has some limitations. It led to a

sample that is relatively young and highly educated. This sample may also miss the

more severe cases of depression as the study excluded institutionalized individuals.

Why was that? Were there other exclusion criteria?

The exclusion of institutionalized individuals was only in the recruitment from the Pharmaceutical Benefits Scheme because the Department of Human Services could not send letters to people in institutions. There was also likely to be difficulties in getting saliva kits back from participants in institutions. Any person who was institutionalized and was able to participate could enrol in the online study.

The sentence regarding institutions now reads: "individuals with residential locations in the PBS database corresponding to hospitals, aged-care facilities and correctional facilities were excluded as obtaining a saliva sample would not be possible"

People with severe depression may be less likely to respond to a media appeal. This is likely to be a weakness of any study design, even those that recruit samples in clinics. We are now seeking to recruit more patients in clinical settings such as psychiatric hospitals which may allow us to recruit more patients with severe depression. There were no other exclusion criteria except that participants had to be at least 18 years of age. Depending on their responses to the questionnaire, participants may be excluded from analyses that seek to address specific research questions.

We believe that the reviewer's comments about the recruitment of less severe patients has been addressed in the following paragraph in the Discussion that was added in a previous revision – "Volunteer participation could also cause bias towards recruiting participants with less severe forms of depression. We will endeavour to investigate this response bias by comparing results from our analyses with those from smaller datasets recruited in clinical settings and to other datasets with a broad spectrum of severity of depression. It has been shown that those with more severe depression have higher mean polygenic risk scores for depression than those with less severe depression. By comparing the distribution of polygenic risk scores to other samples, we can assess the effect of response bias on the severity of depression in AGDS. Our initial analyses suggest that many of the participants have had severe depression as they report large numbers of episodes and nearly 50%

report having had symptoms in the past 4 weeks. Likewise, the reported rates of response to the first prescribed antidepressant are nearly identical to those from the STAR*D clinical trial (33%) [23].Based on the self-report data on number of episodes and other measures of severity, the AGDS sample has high rates of severe depression."

A very low proportion of subjects was included via the pharmaceutical prescription

history. If I understood it correctly, only ~3,000 subjects responded to the invitation

whereas 110,000 invitation letters were sent to subjects who had received

antidepressant prescriptions in the past 4.5 years. How do the authors explain this

low number of respondents? Does it affect the representativeness of the cohort?

There are a number of possible explanations. It is our opinion that appeals on television, radio or social media where someone from the study discusses how the project can help to better understand depression has more effect on people. The advertising included a number of people with a history of depression who told their stories. A letter may be seen as impersonal and participants may have quickly discarded them.

It may affect the representativeness of the cohort. However, it's not clear how another study design would have been able to recruit such a large number of participants in such a short timeframe that better represents the population. Nearly all study designs rely on volunteer participation and hence may not represent the population.

We addressed the reviewer's concerns about the low rate of recruitment through the pharmaceutical benefits scheme about the recruitment of less severe patients has been addressed in the following paragraph in the Discussion that was added in a previous revision - "The media campaign was the more successful of the two methods as more than 80% of the sample was recruited in this way. Approximately 2.5% of those sent letters by the Department of Human Services enrolled in the study. There may be several reasons for the low rate of participation from this method. Firstly, as antidepressants are prescribed for a range of conditions, many of those sent letters may not have had depression and hence decided not to participate. Secondly, letters may be easily discarded by recipients as unsolicited mail may not be well received. Lastly, the media campaign included interviews with both study investigators and individuals with lived experience of depression who encouraged others to participate. As more information can be conveyed about the importance of the research through a TV or radio interview, it likely had a bigger impact on potential participants."

What was the response rate for the additional questionnaires? How many

respondents provided data?

We now provide the response rates to each module in a Supplementary Table.

Side-effects from antidepressants are often hard to disentangle from symptoms of

depression. Libido loss, insomnia, fatigue, weight gain/loss are all depression

symptoms but can also be side-effects from antidepressants. Is it possible with this

cohort study to differentiate between depression symptoms and side-effects?

We plan to conduct multivariate analyses that will investigate the covariance of depressive symptoms and side-effects. For instance, if side-effects are purely depression symptoms then we would expect that polygenic risk scores for depression symptoms will be predictive for side-effects.

Not all participants may have remembered their antidepressant use correctly. Will

their reports also compared/combined with the recorded antidepressants in the PBS

records?

Yes we plan to compare the self-report data with those from the PBS records. These records will only provide information for the last 4.5 years, but should be sufficient to give an overall estimate of the accuracy of self-report. We have compared the frequency of reported antidepressant use to the summary data from the PBS records for the whole population of Australia and the relative frequencies match very accurately.

We have added the following sentence to the Discussion: “When PBS records become available, we will be able to investigate the concordance with self-report information on drug response over the past 4.5 years.”

The controls of QSkin were not screened with validated structured questionnaires

for the absence/presence of lifetime depression, but only through self-report “have

you ever been diagnosed with depression?”. This means that some of the controls

might in fact be MD cases if the diagnosis was missed, they forgot about their

diagnosis, or they never went to see a doctor. Also, if the included controls are still

very young, there is quite a high proportion that eventually will have depression

(~20%). Do the authors see solutions for this?

The age of the QSKIN samples was between 40 and 69 when completing the questionnaire about depression. While some may go on to a diagnosis depression, the sample is past the peak age at

onset for depression which makes it ideal as a control sample. Some of the participants may have forgotten or misreported a diagnosis of depression. This is likely to be a feature of any control sample for genetic studies of psychiatric disorders. It is not uncommon for genetic studies to use completely unscreened controls, where the larger sample size overcomes the problem of some of the controls having the disorder of interest. It would require another large effort and investment to clinically screen tens of thousands of controls so we believe that while maybe not completely accurate, self-report is more than adequate for the purposes of the study.

We have added the following paragraph to the Discussion: “The primary focus of the study was to recruit cases because of the availability of the QSkin sample for use as controls for genetic analyses. QSkin participants have already been genotyped on the same SNP chip. However, the Qskin participants were not administered the full questionnaire and a single question about a prior diagnosis of psychiatric disorders is used to define controls for inclusion. Some participants may have had depression but did not receive a diagnosis and will be incorrectly included as controls. The Qskin cohort is older than the AGDS cohort (mean age 60.8 years vs 42.8 years). This means that most participants are past the peak age at onset for depression and are unlikely to go on to be diagnosed with depression.”