Dear Editor,

Thank you for offering us the opportunity to submit a revised version of this paper. We have responded to all of the comments made by the reviewers and editors, and have given our responses to each comment below.

We have found the reviewers comments very helpful and hope that we have clarified the issues raised in the reviews of our original submission.

Reviewer: 1

Recommendation:

Comments:
This article is very clear, detailed, and discussing very important subjects. The sample size is sufficient to answer to sub-questions that are discussed. The lenght of follow-up is reasonnable.

I have some concerns about the method and the answer to one question.

-the method : the question that emerges immediately is "if some subjects are carrying known risks of CV events, GP will prescribe them the most secure drug, that will appear as the most associated with events, because of these prescription biases". The answer provided could be bettered : only adjusting on some covariables will not provide the good correction about precription biases : it would be probably better to use propensity scores for prescription of each drug, in association with risk factors for CHD, stroke and sudden death.

We considered using propensity scores in our analyses, however as we have treated antidepressant exposure as a time-varying factor, to allow for the complex pattern of antidepressant treatment in a real life setting, including starting and stopping different types of treatment and switching between them, we did not consider that the propensity score method would be viable as it is predominantly used to account for factors associated with prescribing of specific drugs at a single time point. Propensity scoring relies on more detailed and specific knowledge about the factors likely to influence prescription of a particular antidepressant in specific subgroups of patients than we currently have. We acknowledge that GPs may have been selective in their prescribing in relation to cardiovascular risk but advice on this matter has not been consistent e.g. GPs were discouraged from prescribing venlafaxine to people with hypertension and depression from 2004 to 2007 by NICE but then this was rescinded. Furthermore when studies have compared results from a propensity score analysis with an analysis that adjusts for confounding variables directly the results have been very similar (Shah et al., 2005).

-the question : atrial fibrillation is rarely a cause of sudden death, and is due tu ectopic pace makers. Torsades de pointes (TdP) are very frequently inducing sudden death and are linked to an excessive QT length : so I am not sure wether the observation of new atrial fibrillation (which is very interesting per se) could predict TdP and sudden death. Regarding this outcome, the propensity score associated with prescription of each drug could include other factors linked to QT length : basal QT, hypokaliemia and hypokaliemiant drugs, history of familial sudden deaths.
If this is impossible, conclusion should be changed regarding the reproaches done to Citalopram which are not ruled out by this study as written.

Response:
Only a small number of participants in this study who received antidepressants would be likely to have a baseline QT interval recorded and even fewer would have any information recorded systematically on it in their GP records. Similarly, sudden deaths are extremely rare events, and a history of familial sudden deaths would not be routinely ascertained, so would be very infrequently recorded. Drugs that affect potassium levels such as diuretics were recorded and controlled for in the analysis. Alcohol intake which has been shown to be associated with QT length was also adjusted for. Other factors such as hepatitis C and HIV status which have been shown to be linked to QT length (Girardin et al., 2013) would be rarely tested for although they would be recorded when they infrequently occur. Overall if this approach were to be used a propensity sample would be drawn up based on a number of rare, infrequently, unreliable and/or not systematically recorded variables that would apply to only a small proportion of the sample. Results would be highly unlikely to be different whether such a propensity score was accounted for or not. However we acknowledge that our findings for arrhythmia and citalopram do not exclude an association with the much rarer Torsades de pointes and have made this point in the Discussion (page 16).

However, if those remarks could be taken into account by modifying analyses or conclusion as proposed, the study appears of great interest.

Additional Questions:
Please enter your name: Péquignot

Reviewer: 2

Recommendation:

Comments:
This paper presents a cohort study about the relationship of the use of antidepressants (AD) with aryhtmia, myocardial infarction, and stroke/TIA among 238 963 patients with a first depression. The study fine-tunes earlier publications on this topic, focussing on young adults only (20-64), individual drugs, and dose and duration of use. The study is based on a very large database and it is written well. Yet, the paper suffers from a lack of focus, the subsequent great number of results, and a number of methodological limitations. I explain in more detail below.

Major limitations
In the introduction the authors present a number of interesting research questions. Instead of choosing one and working it out well, the authors set out to test all. For instance, the authors might have hypothesized why there is a significant association between SSRIs and stroke in elderly but not in younger adults (Shin 2014). Is it not weird that SSRIs are generally known to induce bleeding disorders, but apparently increase the risk of hemorrhagic AND ischemic stroke? Was history of stroke accounted for in all included studies? Were the age groups similar in case-control and cohort studies? Likewise, the authors refer to a number of observational studies reporting discrepant findings about the association between AD and arrhythmias, but again they do not give a potential interesting and testable explanation for these discrepancies. The authors do not explain either why they might want to test the association between AD and myocardial infarction. As a result it remains unclear what the clinical and scientific relevance of this study is.

Response:
Our study was designed to investigate the associations between antidepressants and a number of adverse outcomes (as listed in our published protocol paper). Previous studies of antidepressant safety have tended to focus on only one or two adverse outcomes, but this does not allow for a

comprehensive comparison of the risks associated with different antidepressants across a range of outcomes which can then inform decision-making with individuals. We included outcomes where previous studies have shown some associations with risk (either reductions or increases), although these may be conflicting. We have previously published results in the BMJ for the suicide and self-harm outcomes (Coupland et al., 2015), and have a paper in press (Hill et al., In press) relating to epilepsy and seizures; in the current manuscript we have focussed on the three cardiovascular outcomes, and highlighted in the Introduction that there is uncertainty and conflicting evidence for all three of these including myocardial infarction.

In the Discussion we have now added the individual study references and some further text to indicate which studies had restricted age groups, or did not account for depression. We have offered some possible explanations for the differences by age for myocardial infarction (page 17) and these might also explain the differences for stroke.

The general design and analytic approach are customary for this type of pharmaco-epidemiological studies. However, the authors used the same confounding factors for all three outcomes, even though it is unlikely that they are the same. Moreover, their definition of a confounding factor (i.e. affects the risk of the exposure OR outcome) is not in line with that in the literature (i.e affects exposure AND outcome) (Hernan 2002). Consequently, they list a long list of potential confounders. Over-adjusting might lead to bias just like under-adjustment.

We used a broad range of confounding factors which we identified as including the major risk factors for the three outcomes such as age, diabetes, hypertension and smoking. We did not adjust for some other factors which have been shown to be associated with QT length in a secondary care mental health in-patient study (Girardin et al., 2013), and so potentially are risk factors for arrhythmia, such as hepatitis C, drug abuse and HIV status, but these would be rare in a primary care population so their confounding effect would be small. We have also now reported results for restricted sets of key confounders entered in blocks for each outcome so that readers can compare these results with the results after full adjustment. The results are shown in supplementary tables 12s to 14s, and show that adjustment for age, sex, deprivation, ethnic group and year of diagnosis has a marked effect on hazard ratios, but additional adjustment for further blocks of variables has a relatively small effect

We agree that for a variable to have a confounding effect it should affect the exposure and risk of the outcome, however as we determined our confounding variables *a priori*, we could not be certain that the confounding variables would affect both, and there is little evidence in the literature on variables associated with both antidepressant prescribing and cardiovascular outcomes, so we based our selection of confounders on those in the literature identified as associated with either the risk of exposure or the outcome, whilst anticipating that they would have little confounding effect unless they were associated with both.

The exclusion of patients with a history of arrhythmia, myocardial infarction, and stroke/TIA seems an omission because the results of the study are not generalizable to these patients. Yet, physicians will be extra keen on information about cardiovascular risk of AD in these patients (much less in young patients without such a history).

We excluded these patients separately from the analysis of each outcome (e.g. the analysis of stroke/TIA excluded patients with a previous stroke/TIA but did include patients with a history of arrhythmia or myocardial infarction), because it can be difficult in primary care databases to distinguish new events from ongoing treatment and assessment following a previous diagnosis. The

question of further cardiovascular risks associated with antidepressant use in patients who have had these events is an important one, which would need to be addressed in additional studies.

The authors thoroughly discuss residual confounding and information bias. What is not mentioned is that many strokes and myocardial infarction (esp. in women) go unnoticed. Could silent stroke and MI prior to depression have confounded the results? Could stroke and MI have been missed during follow-up and decreased the power of the study? In addition, selection bias needs to be discussed. It is possible that the study outcomes have been missed in patients that died even though death certificates were used.

We don't think it likely that silent stroke or MI would have confounded the results, since if these events had gone unnoticed they would not be likely to influence the exposure (selection of an antidepressant), even if they were risk factors for a subsequent diagnosed event. Some outcomes may have been missed during follow-up; however we used linked ONS records to reduce this possibility. Furthermore deaths from these outcomes in people in this age range are rare events, which are likely to be investigated, and in addition given the small numbers there will be little selection bias due to any missed outcomes in patients who died; for example there were only 83 deaths recorded on ONS records for myocardial infarction, 71 for stroke and 10 for arrhythmia.

Minor limitations/ questions

The standard comparison group is not described. What is it when non-users were excluded?

As this study analysis treated antidepressant treatment as a time-varying exposure to account for starting and stopping treatment during follow up, and also switching between different antidepressant treatments, there is not a single group of patients who are 'not exposed', but rather patients contribute follow-up time and events to the "unexposed person-years of follow-up" category when they have periods of unexposed time during follow-up, even if they were treated at other periods of time. This unexposed category also included events and person-years from the group of patients who were non-users throughout follow-up. The comparison is then between rates of the outcomes in exposed and unexposed periods of time throughout follow-up.

Why were the monoamine oxidase inhibitors not included in the group 'other AD'?

The group of "other antidepressants" we used is the specific subsection in the BNF (4.3.4) of which most prescriptions (90.7%) were for mirtazapine and venlafaxine. Monoamine oxidase inhibitors are a distinct class of antidepressants (BNF 4.3.2), which are rarely prescribed in primary care, and are known to have severe adverse effects, and given the small numbers prescribed in our study we decided to exclude them.

Were the proportional hazard assumptions confirmed?

Inspection of the log-minus-log plots showed the assumption of proportional hazards appeared to be valid for arrhythmia, but for myocardial infarction and stroke/TIA the curves by antidepressant class and individual drug tended to converge over time, which was confirmed by tests for proportional hazards using Schoenfeld residuals.

Could selective preference explain the fluoxetine findings as well? Is it not prescribed to the very fit/ young specifically?

We didn't find any evidence that fluoxetine is specifically prescribed to younger or fitter patients. Our suicide/self-harm paper using the same cohort of patients(Coupland et al., 2015) included supplementary tables comparing characteristics of patients prescribed the 11 individual antidepressants (supplementary tables 1s to 3s), and for example the mean age of patients first prescribed fluoxetine was 38.8 years, whereas for paroxetine it was 38.3, the proportion of patients with hypertension when first prescribed fluoxetine was 6.7%, whereas for paroxetine it was 5.3%. We have added some text to the current manuscript to refer to this (pages 11 and 15).

I do not understand the suggestion to test the cardio-preventive effect of fluoxetine in a trial. A large population at risk of cardiovascular disease and with depression would need to be included. Fluoxetine is generally avoided in elderly patients due to its long half time.

We have changed this sentence to say "The potential cardio-protective effects of selective serotonin reuptake inhibitors, particularly fluoxetine, warrant further investigation".

The abstract and text seem too long, and the number of tables and figures seems too large for a BMJ article (not taking supplementary content into account).
There are some textual mistakes.

We have not reduced the length of the text since other reviewers have said that the article is very clear and well presented. We would be happy to move some tables from the main article to the supplementary file if the BMJ would prefer this.

References
Shin D, Oh YH, Eom CS, Park SM. Use of selective serotonin reuptake inhibitors and risk of stroke: a systematic review and meta-analysis. J Neurol. 2014 Apr;261(4):686-95.

Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002 Jan 15;155(2):176-84.

Additional Questions:
Please enter your name: HJ Luijendijk

Reviewer: 3

Recommendation:

Comments:
This paper considers the cardiovascular risks of antidepressant medication in adults. It presents the findings of a cohort study of almost a quarter of a million UK adults aged 20-64 with a first diagnosis of depression, and examines the risks of MI, CVA and arrhythmia in relation to different classes of and individual antidepressants. Data are drawn from the well-respected QResearch database of primary care electronic records, with supplementary information on cause of death from the Office of National Statistics. The authors provide a clear and cogent description of their methodology, including inclusion/exclusion criteria, confounders and statistical methods of analysis. The key elements of analysis include comparison of risks of adverse cardiovascular events when taking vs. when not taking antidepressants. Findings are presented clearly. There is a thorough and well-balanced discussion of study strengths and limitations - the latter including possibilities of selection

bias for lofepramine, and relatively small numbers on high dose citalopram - and of the implications of the findings for clinical practice.

This is a strong and well-presented paper, which provides convincing evidence that in adults aged 20-64 there are low risks of adverse cardiovascular events associated with antidepressant medication in general, and with SSRIs in particular. The specific evidence in relation to higher dose citalopram is potentially significant in policy terms, and is likely to be of interest to US and European drug monitoring agencies in view of existing advice regarding risk of prolonging QT intervals: the authors are suitably cautious on the implications of their findings. The suggestive evidence that SSRIs, especially fluoxetine, may reduce the risk of adverse cardiovascular events in adults is intriguing, and worthy of further investigation.

Thank you for these comments.

Additional Questions:
Please enter your name: Christopher Dowrick

Job Title: Professor of Primary Medical Care

Institution: University of Liverpool

Reviewer: 4

Recommendation:

Comments:
This is a well-written and well reported article, as expected from the set of authors involved. The authors have clearly worked hard to address an important topic using the primary care database available. It is good to see that a protocol was published for this cohort study. I have reviewed this from a statistical perspective, and although standards are generally good as expected, I have some comments for improvement and areas for clarification to be addressed in any subsequent revision:

1) When reading the article, my initial impression was that there are a lot of analyses here, for example across different 3 outcomes, different classes, individual drugs,, and different follow-up times. For the latter it says in the methods that 'As sensitivity analyses we repeated the analyses firstly restricted to the first year of follow-up, then including the entire follow-up period' and also time since starting treatment is investigated as categories, e.g. first 28 days.

But in the protocol, though 5 years and 28 days are mentioned, I cannot see mention about the 1-year analyses. Can the authors clarify please why they focused on 1-year in the end, if not mentioned in the protocol (perhaps I am missing something)?

Our primary results are for 5 years follow-up as specified in the protocol paper (Coupland et al., 2013). The time since starting treatment categories (1-28 days, 29-84 days etc) were pre-specified as categories of the duration of antidepressant exposure variable, but were not specifically related to the length of follow-up in the analyses which included treated and untreated exposure time. We did specify in the protocol paper that we would estimate absolute risks of the adverse events at 1 year, and this involved calculating the hazard ratios over 1 year and we felt these were the most relevant absolute risks to present in this paper since the median duration of use of antidepressants was 221 days, and only 5.5% of patients had five or more years of antidepressant treatment. In addition there

was some indication of non-proportional hazards over 5 years as described above and also the baseline characteristics are less likely to change during one year, and fewer switches occurred between different antidepressant drugs, so the results from the one year analysis are less likely to be influenced by residual confounding but the numbers of events in this analysis is smaller so there is less precision for these estimates. We have added this information to the paper (page 9).

2) In relation to this point, most analyses over the 5 year period are not significant, but there are more significant results by 1 year. This suggests that the hazard ratio is not proportional over time, but this is not evaluated formally (statistically) and raises the question about the HRs from years 1 to 2, and 2 to 3 etc. Therefore I find the focus on 1-year an incomplete picture, and wonder whether the authors could comment in the results about whether the proportional hazards assumption was appropriate (the methods say it was examined, but we don't see the details). I would find it strange that the HRs at 1-year are significant but not at 5-years, if the proportional hazards assumption is actually ok.

There was some evidence that hazard ratios were not proportional for myocardial infarction and stroke, particularly for the SSRIs. We have added further information on the validity of the proportional hazards assumption to the paper (pages 12 and 13).

Many 1-year results are the main message in the abstract and conclusions, yet they are only given in the supplementary material in the actual paper. I think they should be brought into the main article tables, and this may link to a more detailed investigation of the proportional hazards assumption (if the HR is constant over time, or what the HR is within each year interval upto 5 years).

We have moved the 1 year results to a table in the main article (Table 4). We have also carried out an additional analysis to estimate the risks separately over years 1-3, and 3-5 of follow-up, and described the results in the text (pages 9, 12 and 13) and have added tables showing these results to the supplementary file (Tables 6s and 7s). We were not able to calculate hazard ratios for each single year interval due to small numbers of events especially in the later years when more patients had stopped taking antidepressants or had been censored. Also there were only sufficient numbers to run these analyses for the 5 most commonly prescribed drugs rather than 11.

Of fundamental interest: if the SRIs are associated with benefit for the first year but overall the 5 years there is no difference, does this mean that the SRIs are associated with harm in the latter years?

Now that we have carried out an additional analysis for years 1-3 and 3-5 of follow-up we can see that SSRIs were not significantly associated with risk of myocardial infarction in years 1-3 of follow-up (adjusted hazard ratio 0.88, 95% CI 0.68 to 1.15) or years 3-5 (1.14, 95% CI 0.83 to 1.57). We have added this to the text (page 13).

3) Further, the authors look at 3 outcomes in the paper 'arrhythmia, myocardial infarction and stroke or transient ischaemic attack'. Yet, in the protocol there were far more than 3 outcomes listed (see below), and none were mentioned as primary outcomes. Can the authors clarify why they looked at these three outcomes in this paper as a priority over other outcomes listed below:

- all-cause mortality
- suicide (including open verdicts)
- attempted suicide/self-harm
- sudden death

•overdose/poisoning with an antidepressant •myocardial infarction •stroke/transient ischaemic attack (TIA) •cardiac arrhythmia •epilepsy/seizures •upper gastrointestinal bleeding •falls •fractures •adverse drug reactions (including bullous eruption) •motor vehicle crash.

As stated above in response to reviewer 1 the aim of our study was to give a comprehensive overview of a range of adverse effects potentially associated with different antidepressants. We felt that there were too many results to cover adequately in a single paper for all of these outcomes and so we have decided to report them in a number of articles. We have previously published results in the BMJ for the suicide and self-harm outcomes (Coupland et al., 2015), and have a paper in press relating to epilepsy and seizures(Hill et al., In press). We have added the references to these papers earlier in the current paper to highlight that the full study included non-cardiovascular outcomes (page 6).

4) The authors adjust for confounding using Cox regression, and it is good to see that many confounders are indeed adjusted for. That being said, I would also have liked to see whether conclusions are robust to the use of propensity score matching methods. Or can they justify in the Discussion why this wasn't considered beneficial over traditional regression adjustment? Perhaps, due to the time-varying nature of the use of antidepressants, this was problematic

We did consider using a propensity score approach, but decided that this approach would be problematic due to the time-varying nature of the use of antidepressants, with complex patterns of starting and stopping and switching between different drugs over time, and there is not a single point in time where a prescribing decision is made.

5) Can the authors clarify in the paper the use of the time-varying antidepressants covariate and its interpretation for an individual who stopped. If an individual stops antidepressants, then do they then (for subsequent follow-up periods) move to the non-treatment group? If so, then how does this handle the potential for events to be due to the earlier use of antidepressants? Could it be that the lack of any differences between groups is because some of those who were on anti-depressants or moving into the non-treatment group, and therefore any genuine difference is being attenuated?

In our main analyses of antidepressant class and type of antidepressant patients were counted as exposed to an antidepressant while they were prescribed it, during gaps of up to 90 days between the end of one prescription and the start of the next (to allow for any accumulation of tablets over time) and also for an additional 90 days after stopping the antidepressant. This was so that any outcomes occurring during withdrawal periods would be attributed to the antidepressant and not to non-treatment which as the reviewer says would attenuate differences. We have clarified this in the text (page 8). In the analysis where we subdivided antidepressant exposure into time since starting/stopping the results show more precisely how the risks change in the periods of time after stopping treatment where the stopping date was defined as the estimated date of stopping treatment (the prescription end date), without adding 90 days.

6) "Even for doses of citalopram ≥ 40 mg/day there was no significantly increased risk (adjusted hazard ratio=1.11, 95% CI 0.72 to 1.71)." – though this statement is correct, the confidence interval is 0.72 to 1.71 and is therefore wide: is there low power? Indeed, there are not many events in many analyses. This is worthy of discussion please in the strengths and limitations section.

We have added text to this sentence to state that the number of events was small. We already have a section in the Discussion which says the 95% confidence interval is wide for this estimate and that increases in risk of up to 71% cannot be excluded (page 15).

7) I am also concerned about missing data: the authors say 'We included all eligible patients in the database in our analyses to maximise power' – but there are no details about how missing data were handled. I notice that under a table it says '5.0% of prescriptions had missing information on dosage.', so there is some missing data – but how was it handled? It is also not mentioned in the protocol.

We have clarified our handling of missing data in the Methods section (pages 8 and 9). We included all eligible patients in the descriptive and unadjusted analyses. For the adjusted analyses we excluded patients with missing Townsend scores (3.4% of cohort), and for analyses of dose we also excluded periods of follow-up time when patients had a prescription for a treatment where the prescribed dose was not known (around 2% of total follow-up time in the 5 year analyses). For smoking and alcohol we included categories for not recorded.

8) The authors used 'robust standard errors to allow for clustering of patients within practices' – such methods are done when the model used is mis-specified (or the correct model is difficult to actually fit), and therefore the 'robust' standard errors used to inflate uncertainty accordingly. However, here I do not understand why the clustering within practices was not accounted for by using, for example, using a stratified Cox model or adding a frailty term (with a random effect on the baseline hazard to allow for separate one for each practice). Though this is a minor point, I would like the article to clarify if alternative approaches to accounting for clustering affected the conclusions.

We have run additional analyses for each of the three outcomes using a stratified Cox model for our main 5 year analyses of antidepressant class and the 11 individual drugs. The results for antidepressant exposure were similar to our original analyses using robust standard errors (see additional tables 1-3 below). We have tried to run Cox models with frailty terms but due to the size of the datasets (each has more than 1 million rows of data due to time-varying terms) these models have not run over a period of many hours. We have added text to the article on this alternative approach (pages 9-10 and page 13), but have not added these tables since there are already a large number of tables but they could be added if required.

9) In places, the authors infer a difference between individual drugs, which is often not justified. This is most apparent in the 'absolute risks' section, where they say 'Absolute risks of arrhythmia and myocardial infarction were highest for lofepramine' – this is not justified, as the CIs for the risks and NNH are very wide and overlap with the other drugs. This therefore needs to be re-written. Please check elsewhere for this issue too.

We have re-worded this section and have checked throughout.

In summary, this is an important piece of work, and I hope my comments help to improve the article further, especially in regard to the outcome investigated, the time-points considered and the use of time-varying covariate.

Thank you for your comments; we believe they have helped us to substantially improve the article.

Additional Questions:
Please enter your name: Richard Riley

Job Title: Professor of Biostatistics

Institution: Keele University

Reviewer: 5

Recommendation:

Comments:
This is a very well written and organized paper by Coupland et al. that examined the association between different antidepressant prescriptions and 3 cardiovascular outomes within a large UK primary care database. The principal finding is that there there is no overall significant association between different antidepressant classes and the onset of MI, stroke or TIA, or arrhythmia.

There are a number of strengths:
1. Clinically relevant question. The regulatory warnings had a clear effect on comfort level in prescribing antidepressants in general and at particular dosages, therefore the findings have clinical application.
2. Very large dataset
3. The authors have tried to address the main limitations to the data (e.g. validity of primary outcomes.
4. Excellent analytic plan.

Main issues:
1. The precision of the key aims. The mechanism by which antidepressants may raise the risk of arrhythmias is very different than the mechanism that may impact rates of stroke, TIAs and MIs. Furthermore, the time frame for the effect is vastly different. The impact on arrhythmia being much shorter than the others. Lumping them together without adequate consideration of these issues is problematic.

We believe we have considered and examined the impact of differing time frames by carrying out sensitivity analyses with different lengths of follow-up and by including in the analyses an exposure variable which categorised duration of exposure (1-28 days, 29-84 days, 85+ days), so we could see whether the impact on arrhythmia was shorter than for the other two outcomes.

2. The paper focuses on differences between antidepressants, yet the primary analyses use comparisons with periods of no antidepressant treatment. This complicates the interpretation of results, since there are many additional factors that influence prescribing of any antidepressant and which could impact the primary outcomes. The authors are encouraged to clarify why they did not select on antidepressant or class as the reference, and then compare others to this reference. This would have decreased the potential bias.

Our primary analyses use comparisons with periods of no antidepressant treatment since this allows comparison with other studies which have used this non-exposed group as the comparator, and also addresses directly whether antidepressant treatment overall is associated with increased risk of these outcomes. We have also included results from direct tests between antidepressant classes or individual types in the manuscript. To aid with the interpretation of results between antidepressants we have now added a table to the supplementary material (supplementary table 2s) which uses SSRI treatment as the reference group in the analysis of antidepressant class, mid-dose SSRIs as reference category for analysis of antidepressant dose and citalopram (the most commonly prescribed antidepressant) as the reference group in the analysis of individual antidepressants.

3. The results are important and meaningful, but are ultimately more confirmatory than new.

Minor issues:
4. It is unclear why the authors selected to only focus on patients with depression. Couldn't the results have included all antidepressant prescriptions? A sensitivity analysis could then examine whether there are unique results for the depression group. The explanation on page 12 that "....depression itself is an established risk factor for cardiovascular outcomes..." is unconvincing, since this same statement could be made for bipolar disorder or schizophrenia.

A priori, we decided to restrict our cohort to patients with depression, since the different antidepressants are prescribed for a range of conditions which have varying associations if any with the cardiovascular outcomes. It is not always possible to ascertain in primary care databases precisely what the indication is for a prescription, and this would be a major source of indication bias if the indication could not be adjusted for. By selecting a cohort with a diagnosis of depression we have included the largest group of patients likely to receive antidepressant prescriptions, and have largely removed indication bias by restricting the cohort to patients with the same indication for prescriptions. In terms of bipolar disorder or schizophrenia the use of antidepressants alone has been discouraged by NICE Guidelines for these conditions. They would be prescribed with antipsychotic drugs in schizophrenia and many bipolar disorder patients and with lithium and anticonvulsants in bipolar disorder (RM chaired the NICE Guideline for Bipolar Disorder in 2014 and was also on the guideline development group for its earlier edition). It is true that in bipolar disorder, many patients are misdiagnosed as unipolar depression and treated with antidepressants but treatment with antidepressants alone has not been recommended for many years in those with bipolar disorder. The analysis could not have separated the effect of condition from co-prescription of antidepressants with other medications that might in themselves increase or decrease the risk of arrhythmia, stroke or MI even if these conditions had been included because of confounding.

5. Related to this, it is unclear why the authors selected cohort from patients with a first recorded diagnosis of depression.

We selected patients with a first recorded diagnosis of depression so that they would be treatment naïve at the start of follow-up, in which case selection of an antidepressant would not be affected by previous experiences and preferences which would be difficult to account for in the analyses.

6. A rationale for the duration of follow-up should be provided.

We selected five years of follow-up for our main analyses as this can encompass periods of long term treatment, as there is evidence that the duration of antidepressant treatment has increased substantially in recent years(Moore et al., 2009). This length of follow-up also allows for more events to accrue adding to the power of the study.

7. A rationale for the age restriction of 25-64 should be provided.

The age restriction was 20 to 64. We used this age range as we have previously studied antidepressant safety in older people (aged 65+), and wanted to investigate whether the associations we found were similar in younger people. We did not include people below the age of 20 as the incidence of these adverse outcomes is very low in this age group, and different guidelines apply to the selection of an antidepressant.

8. Page 6: the wording regarding ages 20-64 is unclear. Is this current age or age at time of first diagnosis of depression?

We have clarified that this refers to the age at the time of first diagnosis of depression (page 6).

Additional Questions:
Please enter your name: Ayal Schaffer

Job Title: Head, Mood and Anxiety Disorders Program

Institution: Sunnybrook Health Sciences Center, University of Toronto

Reviewer: 6

Recommendation:

Comments:
This study describes a cohort study from family practices across the United Kingdom which tracks over 230,000 patients aged 20 to 64 who were diagnosed with depression and exposed to antidepressants, in terms of their risk for subsequent myocardial infarction, stroke or transient ischaemic attack or arrhythmia based on antidepressant exposure, over five years.

This study represents a very large database and is an important contribution to the literature; there are, however, some limitations.

1. Is the ratio of MI to CVA/TIA expected? Generally, the prevalence of coronary heart disease is higher rather than lower than cerebrovascular disease. Does this ratio match with the general population? If not, comment is warranted about how the exclusion criteria may have impacted this ratio and what importance that may have.

It is difficult to find a direct comparison for the same outcomes and age range as this study. Age-standardised prevalence values for the UK do show slightly higher values for stroke (not including TIA) in men and women than for myocardial infarction however (Townsend N et al., 2014.)

2. The abstract lists 25-64yo age range, looks like it should be 20-64yo.

We have corrected the mistake in the conclusions section of the Abstract – it should be 20 to 64 years not 25 to 64.

3. Reasons for excluding angina or other non-MI diagnoses reflecting atherosclerotic heart disease should be provided. There are many more patients with atherosclerotic heart disease (e.g. angina, require bypass/stent) that don't have a frank MI.

We did not include these in our outcomes, as previous research has focussed on myocardial infarction associated with antidepressant use and we wanted to be able to compare our findings.

4. The justification for the study is weak. Although the study is of value, the principal reason stated for this study is that previous studies haven't explored a younger age group. However, the literature review include studies that refer to subjects with a range of ages but the issue of age and how it might impact on these particular findings is not discussed.

We have added further justification for the study, and added further information to the Discussion on the age ranges included in other published studies.

5. Recruitment ended in 2011 and follow-up ended in 2012. Therefore, there is the possibility that a not insignificant proportion of subjects could not have reached five years of follow-up. As new medications became available over time, how might this have affected the chance of finding the outcome (or no outcome) with newer antidepressants?

We used a maximum of 5 years of follow-up for our main analyses, but patients were censored if they left the practice, died or the study period ended before 5 years, so not all members of the cohort had 5 years of follow-up (although 52% did), 70% had more than 3 years of follow-up before censoring. We adjusted for year of diagnosis of depression in our analyses to reduce confounding from changes in prescribing patterns over time. Comparison with an earlier paper that we published in the BMJ shows that there were no changes to the top 11 most prescribed antidepressants in UK from 1996- 2011 so the effects of new antidepressants on these results can be ruled out.

6. The analysis explored 4 categories of antidepressants and a subsequent analysis chose to look at the 11 most frequently prescribed antidepressant drugs. The justification for selecting 11 (As opposed to 10 or 12 or 13) most prescribed antidepressants is not provided.

We included these 11 antidepressants as we had included them in our previous study of antidepressant safety in older people(Coupland et al., 2011), and we wanted to be able to compare our results. In addition in the current study each of these 11 antidepressant drugs accounted for at least 1% of total antidepressant prescriptions (the proportions were similar for the bottom two – lofepramine and trazodone, so it did not seem logical to exclude one of them), and all other antidepressants had much smaller numbers of prescriptions.

7. There is no explanation as to why "year of diagnosis" is considered a risk factor.

As above, we included this as a confounder to account for time trends in both the outcomes and patterns of prescribing which could otherwise bias the results.

8. Severity of initial diagnosis was determined using codes previously published and “some additional classification by a member of the study team”. It is possible that severity of initial diagnosis could be an important predictor. The process of this “additional classification” is not described and deserves more clarity.

We have added further detail on the additional classification of the severity of depression diagnosis to the text (page 8).We have also added a sentence to the limitations on this classification (page 15).

9. No justification is provided for the eligibility criteria that the diagnosis of depression must have occurred at least 12 months after registration.

We used this criterion so that we could be sure this was a new diagnosis of depression, rather than a retrospective recording of depression following a new registration at a practice. This type of criterion has been applied in many studies using primary care databases.

10. Previous depression is an exclusion criteria at entrance. The reason for this exclusion is not provided. The presence of a previous stroke, myocardial infarction or arrhythmia is not an exclusion criteria at baseline. This might create a bias in that people with a previous event such as myocardial infarction may be preferentially prescribe one kind of antidepressant when they eventually become depressed. There is an exclusion of people with one of these outcomes at baseline but it is described in multiple different ways in different sections of the manuscript. This makes it hard to interpret. At one point the manuscript states, “patients were excluded from the analysis of each outcome if they

had the outcome recorded a baseline". In another place it is suggested that one confounding variable was "comorbidities at baseline (coronary heart disease, stroke/transient ischaemic attack (except when stroke/transient ischaemic attack was the outcome)". and then on page 9 the manuscript states "these patients were excluded from analysis of each respective outcome." Regardless of the way it is expressed (and interpretation of each definition is subtly different), there needs to be clear justification as to why a prior outcome will be included for some people but not others.

We have clarified the description of the exclusion criteria in the Methods section of the manuscript (pages 7), and added the numbers in the analysis for each cohort to the Results section (pages 11-12). We have justified our reason for excluding people with previous depression in the response to reviewer 2 above.

11. In general, there are many exclusion and inclusion criteria are not justified. For example, why are subjects excluded if they received prescriptions for antidepressant more than 36 months before the first recorded diagnosis of depression (i.e. how was 36 months chosen)? What is the justification for the categorization for the number of days taking antidepressant and the number of days after stopping treatment - why are these ranges of days chosen? Why are the classifications or categorization of the intensity of drug dose used?

We have tended not to justify every decision we made in this paper, since this would make it rather long and complex to read. We used the 36 months criterion since patients were often prescribed antidepressants before the diagnosis was recorded, and we did not want to exclude these from the analyses since that would reduce power. With a gap of more than 36 months we thought the prescription might be for an indication other than the subsequent diagnosis of depression. In total 49,179 patients (21% of cohort) received an antidepressant prescription before the date of diagnosis of depression, in the majority of these (64%) it was within the 12 months before diagnosis. We specified the categorization for the number of days taking antidepressant treatment *a priori* based on previous studies (Coupland et al., 2011, Tata et al., 2005).

12. There is no description of how eligible subjects would be handled if they left the practice, before 5 years; in other words, how is this outcome coded?

Patients were censored if they left the practice before 5 years. We have clarified this in the text (pages 8-9).

13. There needs to be more description of how confounding variables were coded (as individual dichotomous variables or as composite variables)

We have added this to the section describing confounding variables (page 8).

14. "Deprivation" was derived from patient's postcodes, "in fifths". Does this mean "quintiles"? Why quintiles, and how reliable is this scale?

Quintiles specifically refer to the cut-offs used to split the cohort into fifths. The deprivation data was provided in this form to help preserve anonymity. This deprivation score has been shown to be associated with many adverse health outcomes, including stroke, for example (Hippisley-Cox et al., 2013), and is a measure of material deprivation, which unlike the Index of Multiple deprivation does not include a health domain as this can produce misleading results(Adams and White, 2006) in the analysis of adverse health outcomes.

15. Ethnicity was categorized using “white/not recorded”. Is there evidence that this is an appropriate classification? What proportion of the "not recorded” are actually white? Or perhaps, those that don’t record ethnicity should not be included in the analysis of ethnicity, or included as their own group.

We used this categorization, as in previous studies (e.g. (Hippisley-Cox et al., 2013, Hippisley-Cox et al., 2008), as this gives proportions more comparable with census data, suggesting the ‘not recorded’ group are predominantly white. We have added the proportion where ethnic group was not recorded to Table 1.

16. On page 7, lines 51 to 54 there appears to be a missing parenthesis and maybe missing a short description as to why this list is included and whether these variables will be considered yes/no or some composite.

We have checked and clarified this.

17. The main analysis is based on the first five years of follow-up. It is not clear why the cut-off was made of five years when indeed the median follow-up was 5.2 years.

We have justified the 5 year cut-off above (reviewer 5, point 6), and have now clarified in the manuscript that the 5.2 years value refers to the overall length of follow-up, and have added that 51.5% of patients in the cohort had at least 5 years of follow-up (page 10).

18. Rather than exclude the monoamine oxidase inhibitors and the 1700+ subjects that were treated with these compounds, could they not be amalgamated into the “other antidepressants” class.

As in our previous response to reviewer 1 we did not think it would be appropriate to combine the monoamine oxidase inhibitors with the other antidepressants. In addition whilst there were 1,791 prescriptions for MAOIs, they were only prescribed to a total of 156 patients who were excluded from subsequent analyses which comprise a small percent of the total cohort (0.07%). We have added the number of patients prescribed MAOIs to the text for clarification (page 10).

19. On page 10 the manuscript states that there was an increased risk for lower doses of lofepramine but the hazard ratio and confidence interval are not reported.

We have added this to the text (page 11).

20. The manuscript states on page 12 line 29 that all eligible patients were included. The more important question is whether all patients exposed antidepressants were included. Clearly they were not, for various reasons, some justified, some not. The most problematic limitation of the study maybe not the inclusiveness of the eligible sample but the exclusion criteria that were applied and may limit the generalizability of the findings.

We have added further comment on our exclusion criteria in response to other reviewers’ comments. We have added further justification in the text, and also in response to this comment have added extra text on the generalizability of findings. There is a balance to be had between generalizability and reliability of the results, and our exclusion criteria were selected to reduce indication biases that could otherwise distort the study findings. We have added a sentence to say that our findings can only be generalised to people with a diagnosis of depression (page14).

21. The manuscript seems to suggest that including only patients with a diagnosis of depression makes it easier to separate the effects of antidepressant treatments from those of depression. But, in fact, if patients had also been included that did not have a diagnosis of major depression but received the same medications, (for example anxiety disorders), then it may have been possible to distinguish the effects of antidepressants from diagnosis. Furthermore, exploring other medications that people with depression may be prescribed, for example, anxiolytics, may also have helped to disentangle diagnosis and medication risks.

As stated above we decided that if we included a range of different indications, this would make it more difficult to disentangle the effects of the antidepressants from that of the condition for which they were prescribed, particularly since it is not always possible to identify the precise indication for a prescription. We included anxiolytics as a potential confounding variable, but anxiolytic drugs other than antidepressants are not recommended in the treatment of depression and often contraindicated for the treatment of depression. Most of the evidence suggests that with a few exceptions such as alprazolam which is rarely prescribed in the UK, they are ineffective for the treatment of depression. A control group that included patients taking a range of ineffective anxiolytic drugs for depression would be hard to interpret.

Overall the findings are intriguing and this manuscript would be a excellent addition to world literature. Nonetheless, important clarifications are important and need to be addressed to approve this manuscript for publication.

Thank you for your helpful and thoughtful comments which have helped us improve the paper.

Additional Questions:
Please enter your name: Anthony Levitt

References

Adams, J., White, M. 2006. Removing the health domain from the Index of Multiple Deprivation 2004—effect on measured inequalities in census measure of health. *Journal of Public Health,* 28**,** 379-383.

Coupland, C., Dhiman, P., Morriss, R., Arthur, A., Barton, G., Hippisley-Cox, J. 2011. Antidepressant use and risk of adverse outcomes in older people: population based cohort study. *BMJ,* 343**,** d4551

Coupland, C., Hill, T., Morriss, R., Arthur, A., Moore, M., Hippisley-Cox, J. 2015. Antidepressant use and risk of suicide and attempted suicide or self harm in people aged 20 to 64: cohort study using a primary care database. *BMJ,* 350**,** h517.

Coupland, C., Morriss, R., Arthur, A., Moore, M., Hill, T., Hippisley-Cox, J. 2013. Safety of antidepressants in adults aged under 65: protocol for a cohort study using a large primary care database. *BMC Psychiatry,* 13**,** 135.

Girardin, F. R., Gex-Fabry, M., Berney, P., Shah, D., Gaspoz, J.-M., Dayer, P. 2013. Drug-Induced Long QT in adult psychiatric inpatients: The 5-year cross-sectional ECG Screening Outcome in Psychiatry Study. *American Journal of Psychiatry,* 170**,** 1468-1476.

Hill, T. A., Coupland, C. A., Morriss, R., Arthur, A., Moore, M., Hippisley-Cox, J. In press. Antidepressant use and risk of epilepsy and seizures in people aged 20 to 64 years: cohort study using a primary care database. *BMC Psychiatry*

Hippisley-Cox, J., Coupland, C., Brindle, P. 2013. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ,* 346**,** f2573.

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., Minhas, R., Sheikh, A., et al. 2008. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ,* 336**,** 1475-1482.

Moore, M., Yuen, H. M., Dunn, N., Mullee, M. A., Maskell, J., Kendrick, T. 2009. Explaining the rise in antidepressant prescribing: a descriptive study using the general practice research database. *BMJ,* 339**,** b3999-.
Shah, B. R., Laupacis, A., Hux, J. E., Austin, P. C. 2005. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *Journal of Clinical Epidemiology,* 58**,** 550-559.
Tata, L. J., West, J., Smith, C., Farrington, P., Card, T., Smeeth, L., et al. 2005. General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction. *Heart,* 91**,** 465-71.
Townsend N, Williams J, Bhatnagar P, Wickramasinghe K, M, R. 2014. Cardiovascular disease statistics, 2014. *British Heart Foundation: London.*

Additional tables

Table 1 Adjusted hazard ratios for arrhythmia by antidepressant class, and individual drug over 5 years follow-up including analysis stratified by GP practice

	Full model[1]		Stratified analysis[2]	
	Adjusted hazard ratio	95% CI	Adjusted hazard ratio	95% CI
Antidepressant class				
No current use	1.00		1.00	
TCAs	1.09	(0.97 to 1.49)	1.09	(0.88 to 1.35)
SSRIs	0.84	(0.80 to 1.06)	0.84	(0.74 to 0.96)
Other antidepressants	1.21	(1.04 to 1.67)	1.22	(0.94 to 1.56)
Combined antidepressants	1.07	(0.61 to 2.36)	1.03	(0.55 to 1.94)
Antidepressant drug				
No current use	1.00		1.00	
TCAs:				
Amitriptyline	1.16	(0.99 to 1.75)	1.14	(0.86 to 1.51)
Dosulepin	0.93	(0.64 to 1.47)	0.91	(0.60 to 1.36)
Lofepramine	1.67	(1.17 to 3.15)	1.87	(1.12 to 3.11)
Trazodone	0.72	(0.30 to 2.16)	0.74	(0.27 to 2.00)
SSRIs:				
Citalopram	0.86	(0.77 to 1.12)	0.86	(0.72 to 1.03)
Escitalopram	1.06	(0.75 to 1.64)	1.01	(0.68 to 1.51)
Fluoxetine	0.74	(0.65 to 1.02)	0.75	(0.61 to 0.92)
Paroxetine	0.97	(0.66 to 1.43)	0.94	(0.64 to 1.38)
Sertraline	0.97	(0.67 to 1.40)	0.96	(0.67 to 1.37)
Others:				
Mirtazapine	1.20	(0.91 to 1.97)	1.23	(0.83 to 1.82)
Venlafaxine	1.27	(0.95 to 1.87)	1.27	(0.91 to 1.78)
All other antidepressants	0.73	(0.41 to 1.63)	0.71	(0.35 to 1.43)
Combined antidepressants	1.06	(0.61 to 2.36)	1.03	(0.55 to 1.94)

[1] Full model using robust standard errors to account for clustering by practice

[2] Stratified Cox model with stratification by practice

Table 2 Adjusted hazard ratios for myocardial infarction by antidepressant class, and individual drug over 5 years follow-up including analysis stratified by GP practice

	Full model[1]		Stratified analysis[2]	
	Adjusted hazard ratio	95% CI	Adjusted hazard ratio	95% CI
Antidepressant class				
No current use	1.00		1.00	
TCAs	1.20	(0.94 to 1.52)	1.15	(0.87 to 1.52)
SSRIs	0.85	(0.71 to 1.00)	0.84	(0.70 to 1.01)
Other antidepressants	1.00	(0.70 to 1.42)	1.02	(0.71 to 1.47)
Combined antidepressants	0.57	(0.18 to 1.75)	0.66	(0.21 to 2.07)
Antidepressant drug				
No current use	1.00		1.00	
TCAs:				
Amitriptyline	1.17	(0.82 to 1.66)	1.13	(0.77 to 1.65)
Dosulepin	1.17	(0.75 to 1.83)	1.11	(0.68 to 1.81)
Lofepramine	2.02	(1.14 to 3.59)	2.05	(1.09 to 3.84)
Trazodone	0.57	(0.14 to 2.30)	0.52	(0.13 to 2.14)
SSRIs:				
Citalopram	0.88	(0.69 to 1.12)	0.87	(0.67 to 1.12)
Escitalopram	0.77	(0.41 to 1.44)	0.82	(0.43 to 1.57)
Fluoxetine	0.73	(0.54 to 0.98)	0.72	(0.54 to 0.97)
Paroxetine	0.76	(0.44 to 1.31)	0.75	(0.42 to 1.32)
Sertraline	1.27	(0.84 to 1.94)	1.25	(0.81 to 1.92)
Others:				
Mirtazapine	1.31	(0.81 to 2.12)	1.36	(0.85 to 2.17)
Venlafaxine	0.89	(0.52 to 1.51)	0.90	(0.53 to 1.55)
All other antidepressants	0.52	(0.17 to 1.60)	0.48	(0.15 to 1.53)
Combined antidepressants	0.57	(0.18 to 1.75)	0.66	(0.21 to 2.07)

[1] Full Cox model using robust standard errors to account for clustering by practice

[2] Stratified model with stratification by practice

Table 3 Adjusted hazard ratios for stroke/TIA by antidepressant class, and individual drug over 5 years follow-up including analysis stratified by GP practice

	Full model[1]		Stratified analysis[2]	
	Adjusted hazard ratio	95% CI	Adjusted hazard ratio	95% CI
Antidepressant class				
No current use	1.00		1.00	
TCAs	1.24	(1.10 to 1.77)	1.25	(0.99 to 1.58)
SSRIs	1.09	(1.01 to 1.37)	1.06	(0.92 to 1.23)
Other antidepressants	1.20	(1.01 to 1.77)	1.16	(0.86 to 1.56)
Combined antidepressants	1.54	(1.02 to 3.31)	1.55	(0.84 to 2.86)
Antidepressant drug				
No current use	1.00		1.00	
TCAs:				
Amitriptyline	1.35	(1.16 to 2.11)	1.37	(1.02 to 1.85)
Dosulepin	1.16	(0.81 to 1.85)	1.16	(0.76 to 1.76)
Lofepramine	1.75	(1.07 to 3.53)	1.91	(1.06 to 3.43)
Trazodone	0.43	(0.13 to 2.00)	0.40	(0.10 to 1.62)
SSRIs:				
Citalopram	1.06	(0.93 to 1.39)	1.04	(0.85 to 1.28)
Escitalopram	0.97	(0.64 to 1.71)	0.97	(0.59 to 1.59)
Fluoxetine	1.13	(1.03 to 1.54)	1.09	(0.88 to 1.34)
Paroxetine	0.95	(0.61 to 1.50)	0.92	(0.59 to 1.43)
Sertraline	1.26	(0.86 to 1.83)	1.27	(0.87 to 1.84)
Others:				
Mirtazapine	1.35	(1.05 to 2.33)	1.28	(0.84 to 1.95)
Venlafaxine	1.05	(0.72 to 1.73)	1.01	(0.65 to 1.57)
All other antidepressants	1.00	(0.56 to 2.26)	0.99	(0.49 to 2.01)
Combined antidepressants	1.55	(1.02 to 3.31)	1.55	(0.84 to 2.86)

[1] Full model using robust standard errors to account for clustering by practice

[2] Stratified model with stratification by practice