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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Antiretroviral therapy for pregnant women living with HIV or hepatitis B: a systematic review and meta-analysis
AUTHORS	Siemieniuk, Reed; Foroutan, Farid; Mirza, Reza; Mah Ming, Jinell; Alexander, Paul; Agarwal, Arnav; Lesi, Olufunmilayo; Merglen, Arnaud; Chang, Yaping; Yuan, Zhang; Mir, Hassan; Hepworth, Elliot; Lee, Yung; Zeraatkar, Dena; Guyatt, Gordon

VERSION 1 - REVIEW

REVIEWER	Giuliano, Marina Istituto Superiore di Sanita, Department of Drug Research and Evaluation
REVIEW RETURNED	04-May-2017

GENERAL COMMENTS	This review is timely and important. Controversial results have been reported for tenofovir-associated effects on maternal and infant outcomes. The authors mention the systematic review by Nachega et al in the Introduction section but they fail to discuss and try to explain the different results in the discussion section. This would be very important since different studies are included in the 2 reviews and different findings are reported. The conclusions of the paper are questionable. Specific recommendations Abstract Eligibility criteria It is reported that for child outcomes also studies with a placebo group were included. However, in the Study Selection section of the manuscript (pag. 11) the authors report that only for outcomes specific to women also living with HBV they included studies with a control group. Background Pag. 5 Line 27. "access to other effective interventions". It is not clear what the authors refer to. Pag. 6. Lines 3-6. I would not mention the economic data for TDF because is it outside the scope of the review. Pag. 6. Line 38. Actually the review that is mentioned reported that TDF was associated to increased neonatal mortality. Pag. 7. Lines 18-21. The authors report: "the recommendation is based on observational evidence and clinical experience that suggested tenofovir is safe" and cite the reference: Mofenson et al. AIDS 2017. However, the cited paper does not report only "observational evidence and clinical experience" but is a sytematic

	data review on cohort studies or trials (among RCTs the PROMISE paper mentioned above is included). Pag. 7. Line 24. The authors report: “after publication of the full report” and cite reference n. 20 which is the Introduction to the BMJ Rapid Recommendation project. Do they refer here to the paper by Fowler et al ? Study Selection Page. 11. Line 3. “fetal outcomes” should be “child outcomes” Summary Measures Page 13. Line 3. The authors include “detectable viral load at delivery” among the maternal outcomes although this measure is not available for any study. So they should modify the sentence and say that they used 6-month post-ART viral load as a proxy. Page 13. Line 6. The authors include “development of HBV resistance” among the maternal outcomes. However, they do not mention it in the results. Page 13. Lines 15-26. Very low birth weight is not included among the outcomes although it is then analyzed at Pag. 27 and in Appendix 5j. Results Pag. 16. Lines 24-29. The authors report that the selected studies compared a TDF-based regimen to alternative NRTI-based regimens in pregnant women. However, this is not true for all the studies since both the studies on PrEP and the study on HBV-positive only women do not have an alternative NRTI-based regimen. Pag. 17. Lines 38-44. The patient representation in the panel of the studies is only mentioned for the PROMISE study. Since the role was not clearly defined I would omit this information. Maternal outcomes Pag. 21. Line 38. “Undetectable viral load”. I would not use this subtitle since this parameter was not actually evaluated. I would call it 6-month post-ART viral load. Child Outcomes Pag. 23. Line 6. “51 more for 1000” is “40 more for 1000” in Table 3. Pag. 23. Lines 41-54. The authors report the data of observational studies on stillbirths and early neonatal mortality. At pag. 24. Line 12, they refer for them to table 3. However in Table 3 these data are not reported. Pag. 24. Line 52. The authors mention the PROMISE trial and one study on PrEP however, they include only the reference for PROMISE. Also, at page 25. Line 3, the authors report that “women were enrolled at a median of 26 weeks gestation” which refers to PROMISE. Pag. 25. Lines 44-52. Actually the figures reported for the study by Mugo et al about abortion refer to “pregnancy loss” including stillbirths (although the pregnancy losses occurred before 20 wks in 91% of the cases). Pag. 31. Line 47. The authors refer the Table 3 but in Table 3 “hepatitis B flares” are not mentioned. Discussion Page 33. Line 27 “spontaneous abortion” is reported among the outcomes without evidence of a between group difference although in the sentence above the authors report a possible increase in risk of spontaneous abortion with TDF/FTC. Pag. 34. Lines 27-30. Actually about 50% of the deaths in the TDF arm were attributed to prematurity. Pag. 34. Lines 38-44. This sentence is not clear to me. Page 34. Line 56. Reference n. 86 cannot be found. Pag. 36. “HBV resistance” is only mentioned here while it was

	included among the outcomes to be evaluated The authors should discuss differences and similarity between their meta-analysis and those recently published. Pag. 36. Lines 38-42. The conclusions of the paper are very strong and in my opinion not appropriate. There are other recent reviews that found different results and a conclusion saying that more data are needed to draw definite conclusions would be more appropriate. Table 3. Also here I would not mention “detectable viral load at delivery” since there are no data on this. The second part of Table 3 should have a different title such as “..are different in different settings”.

REVIEWER	Mofenson, Lynne Elizabeth Glaser Pediatric AIDS Foundation
REVIEW RETURNED	10-May-2017

GENERAL COMMENTS	This is a meta-analysis of the use of antiretroviral drugs during pregnancy in pregnant HIV or hepatitis B-infected women that was generally well done. However, the main conclusion – that TDF/FTC increases stillbirth/early neonatal death compared to ZDV/3TC – is based on one study only, the PROMISE trial. This trial had some unique characteristics that make this conclusion problematic. 1. PROMISE trial was done during two periods. In the first period – about 2/3 of enrollment - only HBV-coinfected women got randomized to TDF/FTC ART vs AZT/3TC ART vs AZT alone (very small numbers – about 3% of overall population). It was only in the second period, enrolling 35% of the overall population, were all women regardless of HBV status randomized to one of the two ART regimens. Therefore, the only comparisons of TDF/FTC ART to AZT/3TC ART could be in the second period. The problem with this is that almost all the neonatal deaths in the AZT/3TC ART group occurred during period 1 (15 of 17 deaths, 88%), resulting in a very low rate of neonatal mortality in AZT/3TC ART group in period 2. In contrast, in the AZT alone comparison group, 39% of deaths (11 of 28 deaths) occurred during period 1 and 61% during period 2. Consequently, the AZT/3TC ART group appears to have an artificially low rate of infant mortality during period 2 (0.6%), the time it was compared to TDF/FTC ART; this is supported by the fact that neonatal mortality was not a significant difference between TDF/FTC ART compared to AZT alone (4.4% vs 3.2%, p=0.43). This raises concerns regarding the validity of the mortality comparison – something the paper itself brings up in the discussion. 2. The additional complication is that the ART in this trial was protease inhibitor-based, using LPV/r, and the dose of LPV/r was increased in the 2nd trimester through delivery. There are potential pharmacokinetics interactions between LPV/r and TDF that could have resulted in elevated TDF levels, which could have been exacerbated by the increased dose of LPV in 3rd trimester. Thus, if TDF/FTC was associated with preterm delivery, it may have been because it was combined with a PI as opposed to it being related to TDF itself. Zash et al reported observational data which compared TDF-based ART that was efavirenz-based did not see any differences in preterm or very preterm delivery between TDF/3TC/EFV and AZT/3TC/EFV.

	Given these two important issues, it seems problematic that the major conclusion in the abstract is that “Tenofovir/emtricitabine may increase stillbirth/early neonatal death and early premature delivery…” without any qualifications. Other comments: Background: Page 5, line 12: The authors state that the risk of transmission to the infant is 30%, citing a reference from 1995. The transmission rate to the infant depends on whether the infant is breastfeeding. In a non-breastfeeding population, transmission is approximately 15%, whereas in a breastfeeding population, it can be as high as 45% (see for example de Cock et al. JAMA 2000). Thus, it is incorrect to state the rate of vertical transmission is 30%. Page 5, line 24: The authors state that transmission is below 5% in low and middle income countries “when cART is universally available and routine antenatal HIV screening is in place”, referencing UNAIDS/WHO. However, this is incorrect as it does not account for breastfeeding transmission and the fact that most low-income countries have not really evaluated overall transmission at the end of breastfeeding, only early transmission rates. The WHO document “Getting on the Fast Track” page 20 provides data on current estimates of transmission at the end of breastfeeding in the priority sub-Saharan African countries. While some countries such as Botswana have achieved <5%, many countries, including Malawi with universal maternal ART and antenatal HIV screening in place since 2011 have not, with an overall transmission rate of 8.7% in Malawi. Thus, it is simplistic to imply that universal ART and HIV screening will by itself reduce transmission to <5%. Page 6, line 32-38: The authors refer to the Nachega systematic review, stating that it “assumed equal credibility in randomized and observational studies”. However, this is incorrect. The authors stated in the methods that they assessed the quality of evidence for the primary outcomes using the GRADE approach and provided a grade table that included judgements regarding the difference between observational data and clinical trials. Page 7, line 35: The authors show clear biases when they state that their analysis provides “…unconflicted and trustworthy recommendations…”, implying that prior systematic reviews and the judgements of the WHO after review are therefore conflicted and untrustworthy. In a scientific manuscript, the authors need to delete comments that are pejorative and keep their comments to their data and findings without providing judgmental comments on other studies. This is particularly a problem in that the authors are basing their apparent conclusions that TDF may not be safe on one randomized trial with potential problems with interpretation as noted earlier. It is inappropriate based on that trial to conclude that other studies and recommendations are therefore faulty, conflicted and shouldn’t be trusted. Methods: Page 9, line 36: The decision to include studies in non-pregnant individuals to evaluate maternal outcomes seems erroneous. There are many physiologic changes in pregnant women – cardiovascular,

	gastrointestinal, renal, hepatic enzyme modifications – that can affect both drug pharmacokinetics as well as potential drug toxicity during pregnancy that are not seen in non-pregnant individuals. Additionally, including data from ART studies in adults that include primarily men (e.g., Gallant, Sax) is also problematic, as some drug reactions (such as hypersensitivity to NVP) are more common in women than in men. Thus, including studies in non-pregnant individuals, primarily men, to supplement an evaluation of outcomes in pregnant women seems problematic and needs better justification (I don't see justification for this). Results: Page 17, line 35: The authors state the PROMISE trial included 694 women and that "most women in the ART group continued their ART regimen after giving birth". This is incorrect. Frist, the PROMISE study enrolled 3088 women overall, with 1230 enrolled during period 2, the period during which TDF ART and AZT ART could be compared. During this time, there were 407 women randomized to TDF ART, 410 women randomized to AZT ART and 413 randomized to AZT alone. Second, the PROMISE trial included a second randomization postpartum, in which breastfeeding mother/baby pairs were rerandomized (regardless of study arm during the antepartum period) to either maternal ART during breastfeeding compared to infant nevirapine with no maternal ART during breastfeeding. Thus, "most women in the ART groups" did not continue ART regimen postpartum – only until 2 weeks postpartum when then underwent a second randomization and half received ART while half did not. This s both discussed in the methods and clearly shown in the supplement with a figure S-1A showing the study design. Page 22, stillbirth/early infant mortality: The PROMISE trial did not combine the outcomes of stillbirth and early infant mortality and it is inappropriate to combine these in this analysis. In this study, there was not any significant difference in stillbirths between all the arms. The supplemental table for comparisons during period the number of stillbirth/spontaneous abortions during period 2 was 2.3% in the AZT alone group, 0.9% in the AZT/3TC ART group and 1.8% in the TDF/FTC ART group (p=0.79 for the comparison of ZDV/3TC ART and TDF/FTC ART and p=0,34 for the comparison of AZT alone vs TDF/FTC ART). The issues with early infant mortality has already been discussed above. Discussion: Page 32, line 53: The authors again appear to be indirectly criticizing other studies by stating that they have concluded "that TDF-based cART regimens are safe for women and their infants." However, neither of the studies they cite unequivocally concluded TDF is safe – both discussed the results from PROMISE and the potential problems with interpretation of this data and both gave nuanced statements that noted the need for further data. The Nachega paper conclusion stated "TDF-based ART in pregnancy appears generally safe for women and their infants. However, data remain limited and further studies are needed, particularly to assess neonatal mortality and infant growth/bone effects." Further, the Nachega paper in their grade table clearly show that the quality of the evidence is low and hence further data are needed to draw definitive conclusions.

	The Mofenson/WHO paper discussion stated, “Although additional surveillance is important, given the available safety data, the benefits of PrEP use for prevention by pregnant/lactating women at high risk of HIV acquisition (and its accompanying increased risk of mother to child HIV transmission) appear to far outweigh the potential risks of fetal, infant and maternal TDF exposure.” Page 34, line 31 and page 36, line 35: The authors state “The biology of the associated increase in stillbirths, however, remains unexplained”. The PROMISE study did not find any significant increase in stillbirths with TDF/FTC ART vs AZT/3TC ART, as discussed (and data provided) above. Therefore, it is not accurate to imply there was a difference in stillbirths with TDF/FTC, and the author’s analysis should not have combined stillbirth/early infant mortality into one variable since the PROMISE study (the only study cited) did not do that. The authors cannot state “…the adverse effect on stillbirths and neonatal mortality…”, since there was not an adverse effect on stillbirth in the trial. The discussion does not include any discussion of the problems in the PROMISE analysis noted above in this review – the problem between period 1 and 2 of the study for the AZT/3TC ART group and the pharmacokinetic interaction between TDF and LPV/r. The authors appear to conclude that the pharmacokinetic interaction is irrelevant but they have not considered the increased LVP/r dose in the second and third trimester in the PROMISE study (the dose was increased from 400/100 to 600/150 in the third trimester), which could have a further drug-drug interaction which does increase TDF above physiologic levels. The studies they cited used standard dosing of LPV/r with TDF in non-pregnant individuals. Only a study of drug levels in the PROMISE study will address this question. The conclusion of the authors – that TDF is not safe to use in pregnancy and that fully informed women would not choose this drug – appears as erroneous as a conclusion that TDF is unqualified safe for use in pregnancy. Based on the problems noted with the randomized trial regarding, it seems that there remains uncertainty regarding the outcomes of concern. Choice of ARV in pregnant HIV+ women (and HBV+ women) remains difficult because there remain limited data to make judgements as to true risks and that further data are needed, particularly data on the ART regimen currently being used in pregnancy – TDF/FTC with efavirenz. Given that there will likely be no further large randomized clinical trials to address this question, and that for most HIV+ women in the world, non-TDF-based fixed dose combinations are not available, quality observational data will be needed to better delineate the risks related to prematurity and early infant mortality.

REVIEWER	Deeks, Jonathan University of Birmingham, Public Health, Epidemiology and Biostatistics
REVIEW RETURNED	06-Jun-2017

GENERAL COMMENTS	The manuscript reports a systematic review, meta-analysis and network meta-analysis looking at effects of antiretroviral therapy. The review is part of a BMJ Rapid Recommendation. There are concerns about some of the methods used and the reporting.

	1. First the review is not reported in line with PRISMA guidelines. For example, the abstract does not state what the outcomes are which are assessed and has no mention of the assessment of methodological quality. 2. There is a lack of clear detail over the eligibility criteria, particularly for outcomes. It is usual to state the PICO components as eligibility criteria, as per PRISMA. Outcomes are reported in detail later in the review, but it would be good to up front state what outcomes were being considered as eligible. 3. Is it really OK to include non-pregnant women? Are there not important differences in physiology and metabolism which will influence how the drug behaves? More justification is needed that this is a relevant group to include. 4. What is a **semi-independent** rapid recommendation panel? In what way was it not fully independent? 5. The search strategy used for two of the three parts of the systematic review is to update searches from existing systematic reviews. 6. There are a whole number of study designs which could be classified as observational comparative designs. Could more information be given on what designs were deemed eligible? And how does the Antiretroviral Pregnancy Register fit within this? Only the briefest of study characteristics are reported for the observational studies in Appendix 6. More details of these studies are required. I also cannot see the results for all of these studies presented in the report – for example what data does Zhang 2016 provide in the meta-analyses. 7. The NOS scale is a rather out of date method for assessing risk of bias in non-randomised studies. Why was this method chosen in preference to more developed alternatives such as ROBINS-I? Also the modifications to the Cochrane Risk of Bias tool for RCTs are not justified nor described – it is not normal to see blinding of data collectors and outcome assessors separated out, nor is it clear how that is operationalised. Also, whilst the methods describes grading each item in this scale into four categories, the results are only reported using 2 categories. Why is this? 8. Why was a network meta-analysis approach used for HBV whereas a comparison of subgroups according to treatment comparator used for HIV outcomes? No explanation is given why two different analytical approaches have been applied. What is the validity of using the NMA approach for analysis of very rare events? Has this been established? 9. The methods describe excluding trials with no events for the HBV analysis. There is no statement as to what was done for the HIV trials, but clearly a different approach was used as for the mortality result in PROMISE the authors have computed a confidence interval for a risk difference when there were no events in either group. Confidence intervals are not normally estimable for such data, so clearly so arbitrary constant has been added. Justification and very cautious interpretation is required.

	10. Again for the HBV trials, the paper states “We imputed 0.5 events to both arms” is an ambiguous phrase. Did you add 0.5 events, or did you replace whatever was observed with 0.5 events? Was the same approach used when there are differences in total sample size? 11. Methodological research suggests that the Peto OR method is the most appropriate method to use for meta-analysis when events are rare, and that risk difference methods are particularly poor. This is recommended in several meta-analysis textbooks and backed up by papers in Statistics in Medicine. Also, there rarely is adequate data to reliably estimate between study variance terms for random effects models. Thus the authors’ choice of statistical method for this analysis is questionable. 12. I have found the results section difficult to read and understand. For example, the first section about acceptability findings on page 18 does not state whether the conclusion is that the acceptability was equal, different, or that these were too few data to state. There is a lot of discussion about the grading of the finding without being clear what the finding is. 13. The language used throughout the results and discussion, although compliant with the GRADE vocabulary is very difficult for a normal reader to plough through. It is severely in need of a Plain English intervention. 14. There are concerns about some of the entries in Table 3. Why are results which are compatible with no difference being interpreted as “may increase” – see Spontaneous Abortion 20 weeks outcome. Other rows of the table classify this as “probably no difference” (at the risk of a Type II error). I also cannot see what the differences are between the multiple rows for low and middle versus high resource settings in the last section of this table. Page 6 line 2 - £10.5 seems rather low. It is unclear why the random effects meta-analysis graphs are labelled as M-H – this is a fixed effect model.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Recommendation:

Comments:
This review is timely and important. Controversial results have been reported for tenofovir-associated effects on maternal and infant outcomes.
The authors mention the systematic review by Nachega et al in the Introduction section but they fail to discuss and try to explain the different results in the discussion section. This would be very important since different studies are included in the 2 reviews and different findings are reported.

The conclusions of the paper are questionable.

Thank you. The recent systematic review by Nachega and colleages concluded “the available data suggest that use of a TDF-containing ART regimen appears to be safe for HIV-infected pregnant women and their infants, the data remain limited and few studies addressed maternal toxicity or infant growth and bone effects.” (Nachega, 2017)

The primary difference between our review and theirs is that they report meta-analytic results that combine RCTs and observational studies. The result is that, applying the GRADE approach, they end up with very low certainty evidence. In GRADE, observational studies start at low certainty and there are serious concerns with inconsistency between studies and because of imprecise results acknowledged by Nachega and colleagues.

We also reviewed observational evidence and judged that the studies were also at high risk of bias – primarily because the observational studies did not control for several key confounders. Their decision to focus on very low quality data from observational studies flies in the face of a consensus regarding the relative trustworthiness of RCTs and observational studies, particularly when the observational studies do not implement optimal adjusted analysis and provide inconsistent results.

We have made changes to the following paragraph in the discussion to make it clear that we are comparing our results to those of the recent meta-analysis:

“Based on similar evidence, our review comes to a different conclusion than another recent meta-analysis13. The reason for this is because Nachega and colleagues pooled RCTs and observational studies which, given the much higher certainty associated RCTs, we consider inadvisable and, indeed, inappropriate.13..

This is particularly the case here because the available observational studies, already beginning as low quality evidence using the GRADE framework84, were further limited by failure to adjust for important confounders. For instance, AZT/lamivudine is an older drug combination than tenofovir/FTC. Thus, clinical care for women who received AZT/lamivudine was more likely limited or outdated for other aspects of their pregnancy. Observational studies also showed inconsistent results and pooled estimates were imprecise, further decreasing certainty of evidence.”

Nachega JB, et al. Safety of Tenofovir Disoproxil Fumarate-Based Antiretroviral Therapy Regimens in Pregnancy for HIV-Infected Women and Their Infants: A Systematic Review and Meta-Analysis. J Acquir Immune Defic Syndr. 2017 Mar 10. doi: 10.1097/QAI.0000000000001359. [Epub ahead of print]

Specific recommendations
Abstract
Eligibility criteria

It is reported that for child outcomes also studies with a placebo group were included. However, in the Study Selection section of the manuscript (pag. 11) the authors report that only for outcomes specific to women also living with HBV they included studies with a control group.

We now clarify:

"For HBV outcomes, we also included studies that compared antivirals to placebo."

Background

Pag. 5 Line 27. "access to other effective interventions". It is not clear what the authors refer to.

We removed this comment as it was not necessary for understanding.

Pag. 6. Lines 3-6. I would not mention the economic data for TDF because is it outside the scope of the review.

We removed this sentence.

Pag. 6. Line 38. Actually the review that is mentioned reported that TDF was associated to increased neonatal mortality.

Although the Nachega review does mention the increase in neonatal mortality in the PROMISE trial, it (we think erroneously) concludes:

"TDF-based ART in pregnancy appears generally safe for women and their infants."

Pag. 7. Lines 18-21. The authors report: "the recommendation is based on observational evidence and clinical experience that suggested tenofovir is safe" and cite the reference: Mofenson et al. AIDS 2017. However, the cited paper does not report only "observational evidence and clinical experience" but is a sytematic data review on cohort studies or trials (among RCTs the PROMISE paper mentioned above is included).

Thank you for pointing this out. We had originally intended to highlight the fact that the authors and guideline working group put greater emphasis on the observational than the RCT data. We have deleted the sentence.

Pag. 7. Line 24. The authors report: "after publication of the full report" and cite reference n. 20

which is the Introduction to the BMJRapid Recommendation project. Do they refer here to the paper by Fowler et al ?

We have updated the reference.

Study Selection
Page. 11. Line 3. "fetal outcomes" should be "child outcomes"

Changed.

Summary Measures
Page 13. Line 3. The authors include "detectable viral load at delivery" among the maternal outcomes although this measure is not available for any study. So they should modify the sentence and say that they used 6-month post-ART viral load as a proxy.

It now reads "*detectable viral load at 6 after starting cART as a proxy for the timing of delivery"*

Page 13. Line 6. The authors include "development of HBV resistance" among the maternal outcomes. However, they do not mention it in the results.

Thank you. We now include a section called *"Other hepatitis B outcomes"* and the sentence: *"One study reported the development of HBV resistance: HBV lamivudine resistance occurred in 1 of 25 (4.0%) of women52."*

Page 13. Lines 15-26. Very low birth weight is not included among the outcomes although it is then analyzed at Pag. 27 and in Appendix 5j.

Now included.

Results
Pag. 16. Lines 24-29. The authors report that the selected studies compared a TDF-based regimen to alternative NRTI-based regimens in pregnant women. However, this is not true for all the studies since both the studies on PrEP and the study on HBV-positive only women do not have an alternative NRTI-based regimen.

We say: *"For child outcomes, we included studies of PrEP (tenofovir/FTC vs. placebo)."*

Pag. 17. Lines 38-44. The patient representation in the panel of the studies is only mentioned

for the PROMISE study. Since the role was not clearly defined I would omit this information.
Maternal outcomes

Removed.

Pag. 21. Line 38. "Undetectable viral load". I would not use this subtitle since this parameter was not actually evaluated. I would call it 6-month post-ART viral load.

Now named *"Undetectable viral load 6 months after starting cART"*

Child Outcomes
Pag. 23. Line 6. "51 more for 1000" is "40 more for 1000" in Table 3.

Thank you. The text was correct and we have updated the table.

Pag. 23. Lines 41-54. The authors report the data of observational studies on stillbirths and early neonatal mortality. At pag. 24. Line 12, they refer for them to table 3. However in Table 3 these data are not reported.

We have removed reference to Table 3

Pag. 24. Line 52. The authors mention the PROMISE trial and one study on PrEP however, they include only the reference for PROMISE. Also, at page 25. Line 3, the authors report that "women were enrolled at a median of 26 weeks gestation" which refers to PROMISE.

Thank you, we have clarified this point.

Pag. 25. Lines 44-52. Actually the figures reported for the study by Mugo et al about abortion refer to "pregnancy loss" including stillbirths (although the pregnancy losses occurred before 20 wks in 91% of the cases).

Clarified.

Pag. 31. Line 47. The authors refer the Table 3 but in Table 3 "hepatitis B flares" are not mentioned.

We have now added Table 4: a separate table that includes outcomes that differ by setting.

Discussion

Page 33. Line 27 "spontaneous abortion" is reported among the outcomes without evidence of a between group difference although in the sentence above the authors report a possible increase in risk of spontaneous abortion with TDF/FTC.

Thank you. Based on comments from reviewer 3, we have removed mention of spontaneous abortion in the discussion.

Pag. 34. Lines 27-30. Actually about 50% of the deaths in the TDF arm were attributed to prematurity.

The number we cite also includes deaths that are likely complications of prematurity (e.g. respiratory distress syndrome).

Pag. 34. Lines 38-44. This sentence is not clear to me.

We have clarified:

"Another interpretation issue is whether the culprit drug is tenofovir or FTC, and circumstances in which the culprit drug would lead to increases in stillbirths and neonatal deaths. The culprit could be tenofovir or FTC, or the combination of the two."

Page 34. Line 56. Reference n. 86 cannot be found.

We have confirmed that the link for the reference is still available online and is correct.

86. World Health Organization. Press Release: First PROMISE study results confirm WHO recommendations to treat pregnant women and reduce mother-to-child-transmission of HIV 2014 [Available from: http://www.who.int/hiv/mediacentre/news/promise-study-result/en/ accessed March 20, 2017.

Pag. 36. "HBV resistance" is only mentioned here while it was included among the outcomes to be evaluated

We have now noted that there were no studies that reported HBV resistance in the mother.

The authors should discuss differences and similarity between their meta-analysis and those

recently published.

As noted above, we have now extensively discussed differences.

Pag. 36. Lines 38-42. The conclusions of the paper are very strong and in my opinion not appropriate. There are other recent reviews that found different results and a conclusion saying that more data are needed to draw definite conclusions would be more appropriate.

We have modified the conclusions in the abstract as follows:

Tenofovir/emtricitabine is likely to increase stillbirth/early neonatal death and early premature delivery compared to zidovudine/lamivudine, but certainty is low when combined with antiretrovirals other than lopinavir/ritonavir. Other outcomes are likely similar.

Table 3. Also here I would not mention "detectable viral load at delivery" since there are no data on this.

The outcome is now *"Detectable viral load 6 months after starting ART"*

The second part of Table 3 should have a different title such as "..are different in different settings".

Thank you. We have now separated the tables and the second part is now Table 4, with the title:

"Table 4. GRADE evidence profile: Tenofovir and emtricitabine versus alternative NRTI regimens in pregnant women living with HIV for outcomes that differ across settings* Full interactive evidence profile available at https://www.magicapp.org/goto/guideline/VLpr5E."

Additional Questions:
Please enter your name: Marina Giuliano

Job Title: Senior Researcher

Institution: Istituto Superiore di Sanità, Rome, Italy

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here:

Reviewer: 2

Recommendation:

Comments:
This is a meta-analysis of the use of antiretroviral drugs during pregnancy in pregnant HIV or hepatitis B-infected women that was generally well done. However, the main conclusion – that TDF/FTC increases stillbirth/early neonatal death compared to ZDV/3TC – is based on one study only, the PROMISE trial. This trial had some unique characteristics that make this conclusion problematic.

We have modified the conclusion in the abstract that now reads as follows:

Tenofovir/emtricitabine is likely to increase stillbirth/early neonatal death and early premature delivery compared to zidovudine/lamivudine, but certainty is low when combined with antiretrovirals other than lopinavir/ritonavir. Other outcomes are likely similar.

1. PROMISE trial was done during two periods. In the first period – about 2/3 of enrollment - only HBV-coinfected women got randomized to TDF/FTC ART vs AZT/3TC ART vs AZT alone (very small numbers – about 3% of overall population). It was only in the second period, enrolling 35% of the overall population, were all women regardless of HBV status randomized to one of the two ART regimens. Therefore, the only comparisons of TDF/FTC ART to AZT/3TC ART could be in the second period. The problem with this is that almost all the neonatal deaths in the AZT/3TC ART group occurred during period 1 (15 of 17 deaths, 88%), resulting in a very low rate of neonatal mortality in AZT/3TC ART group in period 2. In contrast, in the AZT alone comparison group, 39% of deaths (11 of 28 deaths) occurred during period 1 and 61% during period 2. Consequently, the AZT/3TC ART group appears to have an artificially low rate of infant mortality during period 2 (0.6%), the time it was compared to TDF/FTC ART; this is supported by the fact that neonatal mortality was not a significant difference between TDF/FTC ART compared to AZT alone (4.4% vs 3.2%, p=0.43). This raises concerns regarding the validity of the mortality comparison – something the paper itself brings up in the discussion.

We have responded in detail to this query in response to *The BMJ* editors above. Briefly, there were almost certainly differences in participants and co-interventions between the two periods. The relative difference in events of AZT and AZT/3TC ART should, however, unless there is an effect modifier, be similar regardless of the population. We therefore explored the possibility that the difference in relative effects were due to chance.

For early neonatal death, the risk ratio for AZT-cART vs. AZT alone was 0.89 (95CI 0.45 to 1.78) in period 1 and 0.18 (95CI 0.04 to 0.82) in period 2; the p-value did not reach statistical significance (interaction P=0.06).

Regardless, even if the interaction p-value was much lower, it would not tell us anything about applicability. In other words, any differences between tenofovir/FTC and AZT/lamivudine detected in period 2 might have been just as likely to have occurred in period 1 (or in any other similar setting).

We do however agree that the AZT cART group might have had, by chance, fewer than expected events. We therefore agree with the reviewer that the increase in prematurity and neonatal death may be an overestimate of the true harm. In the discussion we say:

"…the adverse effect on stillbirths and neonatal mortality is likely an overestimate and the mechanism and circumstances under which the effect exists remain uncertain."

2. The additional complication is that the ART in this trial was protease inhibitor-based, using LPV/r, and the dose of LPV/r was increased in the 2nd trimester through delivery. There are potential pharmacokinetics interactions between LPV/r and TDF that could have resulted in elevated TDF levels, which could have been exacerbated by the increased dose of LPV in 3rd trimester. Thus, if TDF/FTC was associated with preterm delivery, it may have been because it was combined with a PI as opposed to it being related to TDF itself. Zash et al reported observational data which compared TDF-based ART that was efavirenz-based did not see any differences in preterm or very preterm delivery between TDF/3TC/EFV and AZT/3TC/EFV.

We have carefully considered the implications of any possible interactions between protease inhibitors and TDF. Here, the reviewer says that "LPV/r and TDF that could have resulted in elevated TDF levels". This is true, as we detail in the discussion, but only by a small amount and by much less than the normal variability between patients. For example, a steady-state pharmacokinetic study found that the serum concentration of tenofovir was only increased by 15% (90CI 7% to 22%) in the presence of lopinavir/ritonavir – which is a magnitude lower than the variability between patients when tenofovir is given alone (patients in the 97.5th centile had 98% higher concentrations than the 2.5th centile) (Kearney, 2006). Lopinavir/ritonavir does not prolong the tenofovir half-life (Kearney, 2006). It is therefore implausible that tenofovir/FTC causes harm in the presence of lopinavir/ritonavir but does not when it is given alone or with other medications.

Our systematic review includes the observational study by Zash and colleagues that the reviewer mentions. We note in this response, and in the paper (as documented in response to reviewer 1) that

this and the other observational studies provide only very low quality evidence because of the limitations of any observational study and in addition their inconsistent results (other studies showed no effect or harm) and failure to adjust for important confounders. The RCT provides more trustworthy evidence.

Kearney BP, et al. Pharmacokinetics and safety of tenofovir disoproxil fumarate on coadministration with lopinavir/ritonavir. J Acquir Immune Defic Syndr. 2006 Nov 1;43(3):278-83.

Zash R, Jacobson D, Diseko M, et al. Adverse birth outcomes differ by ART regimen from conception in Botswana. Conference on Retroviruses and Opportunistic Infections. Seattle, Washington, 2017:Abstract #25.

Given these two important issues, it seems problematic that the major conclusion in the abstract is that "Tenofovir/emtricitabine may increase stillbirth/early neonatal death and early premature delivery…" without any qualifications.

Based on the reviewer's concerns, we now say:

"*Tenofovir/emtricitabine is likely to increase stillbirth/early neonatal death and early premature delivery compared to zidovudine/lamivudine, but certainty is low when combined with antiretrovirals other than lopinavir/ritonavir.*"

Other comments:

Background:

Page 5, line 12: The authors state that the risk of transmission to the infant is 30%, citing a reference from 1995. The transmission rate to the infant depends on whether the infant is breastfeeding. In a non-breastfeeding population, transmission is approximately 15%, whereas in a breastfeeding population, it can be as high as 45% (see for example de Cock et al. JAMA 2000). Thus, it is incorrect to state the rate of vertical transmission is 30%.

Thank you for the correction. We now use the 15 to 45% range instead and reference the de Cock paper.

Page 5, line 24: The authors state that transmission is below 5% in low and middle income countries "when cART is universally available and routine antenatal HIV screening is in place", referencing UNAIDS/WHO. However, this is incorrect as it does not account for breastfeeding transmission and the fact that most low-income countries have not really evaluated overall transmission at the end of breastfeeding, only early transmission rates. The WHO document "Getting on the Fast Track" page 20 provides data on current estimates of transmission at the end of breastfeeding in the priority sub-Saharan African countries. While some countries such as Botswana have achieved <5%, many countries, including Malawi with universal maternal ART and antenatal HIV screening in place since 2011 have not, with an overall transmission

rate of 8.7% in Malawi. Thus, it is simplistic to imply that universal ART and HIV screening will by itself reduce transmission to <5%.

We have added a qualifier to make it clear that not every country with universal health coverage is below 5%:

"… in several low and middle income countries when cART is universally available."

Page 6, line 32-38: The authors refer to the Nachega systematic review, stating that it "assumed equal credibility in randomized and observational studies". However, this is incorrect. The authors stated in the methods that they assessed the quality of evidence for the primary outcomes using the GRADE approach and provided a grade table that included judgements regarding the difference between observational data and clinical trials.

We appreciate that the authors of the Nachega paper provide a GRADE summary of findings table. However, lower quality observational data was (we think inappropriately) combined with higher quality RCT data and the authors ended up with very low quality evidence for most outcomes. Their assessments are consistent with ours for the observational data. Nachega and colleagues made no attempt to provide a certainty rating for the RCT evidence. When both observational and RCT data are available, GRADE guidance specifies that RCT evidence should be considered separately because it is more trustworthy than observational evidence (see Balshem, 2011 and Guyatt, 2012). Ultimately, RCTs provide the most trustworthy evidence.

Balshem H, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011 Apr;64(4):401-6. doi: 10.1016/j.jclinepi.2010.07.015. Epub 2011 Jan 5.

Guyatt GH, et al. GRADE guidelines 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol*. [Epub 2012 May 18]

Page 7, line 35: The authors show clear biases when they state that their analysis provides "…unconflicted and trustworthy recommendations…", implying that prior systematic reviews and the judgements of the WHO after review are therefore conflicted and untrustworthy. In a scientific manuscript, the authors need to delete comments that are pejorative and keep their comments to their data and findings without providing judgmental comments on other studies. This is particularly a problem in that the authors are basing their apparent conclusions that TDF may not be safe on one randomized trial with potential problems with interpretation as noted earlier. It is inappropriate based on that trial to conclude that other studies and recommendations are therefore faulty, conflicted and shouldn't be trusted.

One purpose of the BMJ Rapid Recommendations process is to demonstrate that guidelines can be created without financial conflicts and with as few intellectual and professional conflicts as possible, which we and others believe increases trustworthiness of clinical practice guidelines. Conflicts of

interest remain common among guideline committees and we note that the WHO guidelines to which the reviewer refers include several financially and intellectually conflicted panel members. Nevertheless, in the manuscript we state only that our guideline was free of conflict and say nothing about other guidelines in this regard. We disagree with the reviewer that our statement that the BMJ Rapid Recommendations process provides trustworthy guidelines implies anything about any other individual guideline. To us, inferring that our statement that our guidelines are trustworthy represents a pejorative statement about other guidelines, as the reviewer suggests, is very misguided. The reviewer is a coauthor of the Nachega paper and the first author of the Matheson paper. Under these circumstances, it is perhaps natural that she is at risk of making inferences regarding implicit criticisms that other readers might not make.

Methods:

Page 9, line 36: The decision to include studies in non-pregnant individuals to evaluate maternal outcomes seems erroneous. There are many physiologic changes in pregnant women – cardiovascular, gastrointestinal, renal, hepatic enzyme modifications – that can affect both drug pharmacokinetics as well as potential drug toxicity during pregnancy that are not seen in non-pregnant individuals. Additionally, including data from ART studies in adults that include primarily men (e.g., Gallant, Sax) is also problematic, as some drug reactions (such as hypersensitivity to NVP) are more common in women than in men. Thus, including studies in non-pregnant individuals, primarily men, to supplement an evaluation of outcomes in pregnant women seems problematic and needs better justification (I don't see justification for this).

We agree that applying evidence from non-pregnant adults to pregnant adults represents indirect evidence. Therefore, we rate down our certainty in evidence for indirectness whenever evidence from non-pregnant adults is used.

We have expanded our explanation for why and when we use evidence from non-pregnant adults:

"Because, for several critical outcomes, we found no direct evidence or the available evidence was of very low quality, for outcomes specific to the mother we also included RCTs that compared tenofovir to alternative NRTIs in non-pregnant adults living with HIV. We considered evidence exclusively from pregnant women before including evidence from non-pregnant adults."

Results:

Page 17, line 35: The authors state the PROMISE trial included 694 women and that "most women in the ART group continued their ART regimen after giving birth". This is incorrect. Frist, the PROMISE study enrolled 3088 women overall, with 1230 enrolled during period 2, the period during which TDF ART and AZT ART could be compared. During this time, there were 407 women randomized to TDF ART, 410 women randomized to AZT ART and 413 randomized to AZT alone. Second, the PROMISE trial included a second randomization postpartum, in which breastfeeding mother/baby pairs were rerandomized (regardless of study arm during the antepartum period) to either maternal ART during breastfeeding compared to infant nevirapine with no maternal ART during breastfeeding. Thus, "most women in the ART groups" did not continue ART regimen postpartum – only until 2 weeks postpartum when then

underwent a second randomization and half received ART while half did not. This s both discussed in the methods and clearly shown in the supplement with a figure S-1A showing the study design.

Thank you, we agree that we can provide more detail about the PROMISE trial. We now say:

"The PROMISE trial randomized 823 women to the comparison of interest, mostly in Africa. Women in the cART groups continued their cART regimen for 2 weeks after giving birth, at which time half discontinued cART."

Page 22, stillbirth/early infant mortality: The PROMISE trial did not combine the outcomes of stillbirth and early infant mortality and it is inappropriate to combine these in this analysis. In this study, there was not any significant difference in stillbirths between all the arms. The supplemental table for comparisons during period the number of stillbirth/spontaneous abortions during period 2 was 2.3% in the AZT alone group, 0.9% in the AZT/3TC ART group and 1.8% in the TDF/FTC ART group (p=0.79 for the comparison of ZDV/3TC ART and TDF/FTC ART and p=0,34 for the comparison of AZT alone vs TDF/FTC ART). The issues with early infant mortality has already been discussed above.

The decision to combine stillbirth and early infant mortality was prespecified and was made based on feedback from the linked Rapid Recommendation panel that included obstetricians, infectious diseases experts, methodologists, and patients. We have added a more detailed explanation for why we combined the outcomes:

"We combined stillbirths with early neonatal mortality because of a similar pathophysiology (most early neonatal deaths are caused by pregnancy-related factors) and because we believe that most women would place a similar value on the two events."

Discussion:

Page 32, line 53: The authors again appear to be indirectly criticizing other studies by stating that they have concluded "that TDF-based cART regimens are safe for women and their infants." However, neither of the studies they cite unequivocally concluded TDF is safe – both discussed the results from PROMISE and the potential problems with interpretation of this data and both gave nuanced statements that noted the need for further data.

The Nachega paper conclusion stated "TDF-based ART in pregnancy appears generally safe for women and their infants. However, data remain limited and further studies are needed, particularly to assess neonatal mortality and infant growth/bone effects." Further, the Nachega paper in their grade table clearly show that the quality of the evidence is low and hence further data are needed to draw definitive conclusions.

The Mofenson/WHO paper discussion stated, "Although additional surveillance is important, given the available safety data, the benefits of PrEP use for prevention by pregnant/lactating women at high risk of HIV acquisition (and its accompanying increased risk of mother to child

HIV transmission) appear to far outweigh the potential risks of fetal, infant and maternal TDF exposure."

We have changed the wording of this statement based on the reviewer's suggestion to now say:

"Others, including a recent meta-analysis, have suggested that on the basis of this observational data, tenofovir-based cART regimens appear to be safe for women and their infants13 19. Within the GRADE framework, this conclusion is based on very low quality evidence and cannot therefore reassure pregnant women living with HIV."

As we have noted several times in this response, and several times in the revised manuscript, the basis of the disagreement and varying conclusions are judgements regarding the trustworthiness of the relevant observational studies versus the PROMISE RCT. Relevant audiences, including public health officials, and women living with HIV with guidance from their health care providers, will have to decide on which conclusion they find more compelling, and on the implications of that conclusion (as we note in the last paragraph of the paper).

Page 34, line 31 and page 36, line 35: The authors state "The biology of the associated increase in stillbirths, however, remains unexplained". The PROMISE study did not find any significant increase in stillbirths with TDF/FTC ART vs AZT/3TC ART, as discussed (and data provided) above. Therefore, it is not accurate to imply there was a difference in stillbirths with TDF/FTC, and the author's analysis should not have combined stillbirth/early infant mortality into one variable since the PROMISE study (the only study cited) did not do that. The authors cannot state "…the adverse effect on stillbirths and neonatal mortality…", since there was not an adverse effect on stillbirth in the trial.

We have removed the sentence to which the reviewer refers.

The discussion does not include any discussion of the problems in the PROMISE analysis noted above in this review – the problem between period 1 and 2 of the study for the AZT/3TC ART group and the pharmacokinetic interaction between TDF and LPV/r. The authors appear to conclude that the pharmacokinetic interaction is irrelevant but they have not considered the increased LVP/r dose in the second and third trimester in the PROMISE study (the dose was increased from 400/100 to 600/150 in the third trimester), which could have a further drug-drug interaction which does increase TDF above physiologic levels. The studies they cited used standard dosing of LPV/r with TDF in non-pregnant individuals. Only a study of drug levels in the PROMISE study will address this question.

Thank you, we now include a statement in the discussion about the concern that the reviewer raises about differences between periods 1 and 2.

"Because some have raised concerns that the event rates in the AZT/lamivudine arm are lower than might have been anticipated, we also performed a sensitivity analysis that includes participants in the

AZT/lamivudine arm that were randomised early in the PROMISE study, before the tenofovir/FTC arm was added, (Fowler, 2016)."

Several studies have evaluated LPV/r dosing in pregnancy. Most women who continue with standard dosing in the third trimester have LPV serum concentrations lower than target (Stek, 2006). The dosing regimen used in the PROMISE study provides similar drug levels to non-pregnant women (Best, 2010). Most experts, including those in the PROMISE study, use the increased dosing during the third trimester. However, to be conservative we have rated down the certainty of evidence from moderate to low for women considering cART regimens that do not include LPV/r.

We have modified the paragraph discussing a possible interaction between LPV/r and tenofovir:

"Another mechanism postulates a role for LPV/r in the increase in stillbirths and neonatal deaths86. This cannot be a direct effect: patients in both the tenofovir/FTC and AZT/lamivudine groups received LPV/r12. Thus, the only possibility for implicating LPV/r is that it modifies the effect of tenofovir/FTC but not AZT/lamivudine on stillbirths and neonatal mortality. Were this true, tenofovir/FTC would have an adverse effect relative to AZT/lamivudine only when co-administered with LPV/r or similar antiretrovirals. The mechanism of such an interaction is unlikely to be increased LPV drug levels in the presence of tenofovir: if anything, tenofovir decreases LPV drug levels87-90. Further, protease inhibitors including LPV/r only slightly increase serum tenofovir levels 87 90 and implicating this drug-drug interaction would nonetheless implicate tenofovir at serum concentrations within the typical therapeutic range (the increase in tenofovir from concurrent LPV/r is a magnitude lower than normal variability between patients). Of note, the increased LPV/r dosing used in the PROMISE study during the third trimester provides similar serum drug concentrations to non-pregnant women taking LPV/r (Best 2010). Thus, the hypothesis that the adverse effects on fetal outcomes with tenofovir/FTC occur only with concomitant administration of LPV/r has no obvious biological basis. Nevertheless, we conservatively rate down our certainty for in the evidence for indirectness from moderate to low for key outcomes when tenofovir/FTC is combined with antiretrovirals other than LPV/r."

Stek AM, et al. Reduced lopinavir exposure during pregnancy. AIDS. 2006 Oct 3;20(15):1931-9.

Best BM, et al. Lopinavir tablet pharmacokinetics with an increased dose during pregnancy. J Acquir Immune Defic Syndr. 2010 Aug;54(4):381-8.

The conclusion of the authors – that TDF is not safe to use in pregnancy and that fully informed women would not choose this drug – appears as erroneous as a conclusion that TDF is unqualified safe for use in pregnancy. Based on the problems noted with the randomized trial regarding, it seems that there remains uncertainty regarding the outcomes of concern. Choice of ARV in pregnant HIV+ women (and HBV+ women) remains difficult because there remain limited data to make judgements as to true risks and that further data are needed, particularly data on the ART regimen currently being used in pregnancy – TDF/FTC with efavirenz. Given that there will likely be no further large randomized clinical

trials to address this question, and that for most HIV+ women in the world, non-TDF-based fixed dose combinations are not available, quality observational data will be needed to better delineate the risks related to prematurity and early infant mortality.

A statement as unequivocal as “TDF is not safe” would require high quality evidence. We have only moderate quality evidence, and so do not make unequivocal statements such as this, but rather use wording consistent with moderate quality evidence (i.e. “is likely”). As to what fully informed women would choose, our guideline panel, including – and indeed in particular - the patient panel members, considered this question carefully. Their judgment was that women place an extremely high value on avoiding stillbirth and neonatal death, and that the convenience advantages of TDF (there are no other advantages evident) would not compensate for the likelihood of an increased risk of these devastating outcomes. Perhaps the reviewer has evidence that this is not the case, but we think that unlikely and would speculate that her authorship on the papers that came to different conclusions than ours is not irrelevant to her positions.

Additional Questions:
Please enter your name: Lynne Mofenson MD

Job Title: Senior HIV Technical Advisor

Institution: Elizabeth Glaser Pediatric AIDS Foundation

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?:

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: I was involved in the development, conduct and publication of the PROMISE trial; I am a coauthor of the Nachega JAIDS manuscript meta-analysis (that compared TDF ART to non-TDF ART in HIV+ women); and am first author of the Mofenson AIDS manuscript (that compared TDF to non-TDF regimens in HIV and HBV+ women).

Reviewer: 3

Recommendation:

Comments:
The manuscript reports a systematic review, meta-analysis and network meta-analysis looking at effects of antiretroviral therapy. The review is part of a BMJ Rapid Recommendation. There are concerns about some of the methods used and the reporting.

1. First the review is not reported in line with PRISMA guidelines. For example, the abstract does not state what the outcomes are which are assessed and has no mention of the assessment of methodological quality.

We have paid careful attention to including each of the PRISMA requirements in our paper. The page number(s) for each can be found on the attached PRISMA checklist.

In the abstract, we include the outcomes assessed in the results section (although due to space constraints we focus on the most important ones). We also include a GRADE assessment with each outcome in the abstract. To clarify our quality assessment we now expand to say:

"We…used the GRADE framework to assess quality separately for each outcome."

2. There is a lack of clear detail over the eligibility criteria, particularly for outcomes. It is usual to state the PICO components as eligibility criteria, as per PRISMA. Outcomes are reported in detail later in the review, but it would be good to up front state what outcomes were being considered as eligible.

All of the outcomes are included in the methods under the heading "summary measures".

3. Is it really OK to include non-pregnant women? Are there not important differences in physiology and metabolism which will influence how the drug behaves? More justification is needed that this is a relevant group to include.

We agree that evidence is most applicable when study participants are pregnant women. However, as with most conditions in pregnancy, there is limited direct evidence and the best evidence, albeit limited by indirectness, often comes from low risk of bias, consistent, precise results from non-pregnant adults. For each outcome, we carefully consider whether higher quality evidence comes from studies of pregnant patients or from indirect evidence from non-pregnant adults. We now expand on our rationale to include studies of non-pregnant adults:

"Because, for several critical outcomes we anticipated finding no direct evidence or the evidence would be very low certainty, for outcomes specific to the mother we also included RCTs that compared tenofovir to alternative NRTIs in non-pregnant adults living with HIV. We considered evidence from pregnant women alone before including evidence from non-pregnant adults."

4. What is a **semi-independent rapid recommendation panel? In what way was it not fully independent?**

As part of the Rapid Recommendations process, some people participate in both the systematic review team and on the guideline panel. There is always at least one person (the systematic review lead) and never more than 3 or 4 people. We do this to keep lines of communication open and transparent between the two teams throughout the process.

5. The search strategy used for two of the three parts of the systematic review is to update searches from existing systematic reviews.

Yes, we did that to avoid duplication of work and felt comfortable doing so because the existing systematic reviews were credible and comprehensive.

6. There are a whole number of study designs which could be classified as observational comparative designs. Could more information be given on what designs were deemed eligible?

We now provide some examples of comparative observational studies:

"Comparative observational studies included cohort, case-control, and any other observational study type that attempted a direct and coincident comparison between any two of the eligible interventions."

And how does the Antiretroviral Pregnancy Register fit within this?

We included the Antiretroviral Pregnancy Register in a post-hoc sensitivity analysis at the advice of the Rapid Recommendation panel. We now include an expanded description of the ARV Pregnancy Register:

The registry is a frequently updated non-comparative database that tracks the incidence of birth defects in mothers who have taken antiretrovirals.

Only the briefest of study characteristics are reported for the observational studies in Appendix 6. More details of these studies are required. I also cannot see the results for all of these studies presented in the report – for example what data does Zhang 2016 provide in the meta-analyses.

We now include 3 additional columns: maternal HBV inclusion criteria, gestational age at start, and infant HBV test timing. Zhang 2016 is an observational study that compares tenofovir to no antiviral treatment. It is included in the forest plot (figure 5) and in the study outcomes table (Appendix 7). All

studies are included in the meta-analyses in which they reported the relevant outcome (see Appendix 7).

Ayres, 2011 was not included in any of the meta-analyses because it did not report any of those outcomes (lamivudine resistance only).

7. The NOS scale is a rather out of date method for assessing risk of bias in non-randomised studies. Why was this method chosen in preference to more developed alternatives such as ROBINS-I?

The NOS instrument remains in common use, is easily applied, readers understand it and are familiar with it, and the revised version we use addresses the limitations in the original instrument. ROBINS-I is far more complex, is unfamiliar to all but the cogniscient, and has no empirically demonstrated advantages over the modified NOS. Moreover, in this case, we rate almost all observational studies as high risk of bias for lack of adjustment for confounding and dissimilar populations, which is also a domain included in the ROBINS-I. Therefore, the more detailed ROBINS-I would not add any additional value - and it is a foregone conclusion that we will end up at high risk of bias.

Also the modifications to the Cochrane Risk of Bias tool for RCTs are not justified nor described – it is not normal to see blinding of data collectors and outcome assessors separated out, nor is it clear how that is operationalised. Also, whilst the methods describes grading each item in this scale into four categories, the results are only reported using 2 categories. Why is this?

We have expanded on the explanation and rationale of the modifications made to the risk of bias tools:

"We used a modified Cochrane Collaboration tool to assess risk of bias for RCTs[26] which substitutes response options of 'probably low risk' or 'probably high risk' for unclear; empricial evaluation has shown that reviewers can make these judgments accurately (Akl, 2012). Ultimately, we collapsed the low and probably low, and high and probably high risk, for presentation."

The detailed rationale is cited and available here:

Akl EA, Sun X, Busse JW, Johnston BC, Briel M, Mulla S, You JJ, Bassler D, Lamontagne F, Vera C, Alshurafa M, Katsios CM, Heels-Ansdell D, Zhou Q, Mills E, Guyatt GH. Specific instructions for estimating unclearly reported blinding status in randomized trails were reliable and valid. *J Clin Epidemiol* 2012 Mar;65(3):262-7.

Busse JWG, G. Modification of Cochrane tool to assess risk of bias in randomized trials 2013 [Available from: https://distillercer.com/wp-content/uploads/2014/02/Tool-to-Assess-Risk-of-Bias-in-Randomized-Controlled-Trials.docx accessed July 11, 2016.

8. Why was a network meta-analysis approach used for HBV whereas a comparison of subgroups according to treatment comparator used for HIV outcomes? No explanation is given why two different analytical approaches have been applied.

The explanation is included:

"For the comparisons of antivirals for HBV infection, we anticipated that there would be few if any direct comparisons between antivirals and therefore performed a network meta-analysis within a frequentist framework using RRs."

What is the validity of using the NMA approach for analysis of very rare events? Has this been established?

HBV transmission events occurred in 17/932 (1.8%) patients in the lamivudine arm and 87/736 (11.8%) in the placebo arm for comparison 1, and 3/335 (0.9%) in the tenofovir arm and 50/556 (9.0%) in the placebo arm for comparison 2. Although events were sparse in the tenofovir arm, they were not sparse in the others. There is no theoretical reason that NMA should be any more problematic than conventional meta-analysis when events are rare, nor are we aware of any methods papers in NMA have raised this concern.

9. The methods describe excluding trials with no events for the HBV analysis. There is no statement as to what was done for the HIV trials, but clearly a different approach was used as for the mortality result in PROMISE the authors have computed a confidence interval for a risk difference when there were no events in either group. Confidence intervals are not normally estimable for such data, so clearly so arbitrary constant has been added. Justification and very cautious interpretation is required.

We included studies with no events for HIV because we conducted the analyses with the risk difference (absolute) method rather than a relative effect measure. In this circumstance, a confidence interval can be calculated and for both outcomes (maternal mortality, HIV transmission), the additionally included studies added important information. We conducted the network meta-analysis in risk ratios. Because it is a relative measure, the variance for the studies with zero events cannot be included. We now have the following in the methods section:

"When events were rare across all studies (<2%), we performed meta-analysis directly with the Peto method unless one or more studies had zero events in both arms, in which case we used risk differences (RD) directly."

And

"…excluded trials with zero events in both arms because confidence intervals could not be calculated."

10. Again for the HBV trials, the paper states "We imputed 0.5 events to both arms" is an ambiguous phrase. Did you add 0.5 events, or did you replace whatever was observed with 0.5 events? Was the same approach used when there are differences in total sample size?

This now says we *added* 0.5 events to both arms. There were not major differences in total sample size. This is the standard approach that Cochrane uses.

11. Methodological research suggests that the Peto OR method is the most appropriate method to use for meta-analysis when events are rare, and that risk difference methods are particularly poor. This is recommended in several meta-analysis textbooks and backed up by papers in Statistics in Medicine. Also, there rarely is adequate data to reliably estimate between study variance terms for random effects models. Thus the authors' choice of statistical method for this analysis is questionable.

Please see response to The BMJ editors on the same issue above.

12. I have found the results section difficult to read and understand. For example, the first section about acceptability findings on page 18 does not state whether the conclusion is that the acceptability was equal, different, or that these were too few data to state. There is a lot of discussion about the grading of the finding without being clear what the finding is.

We have revised the text for each outcome to explicitly state the findings in plain English. For example, in the acceptability paragraph we say:

"The highest certainty evidence addressing acceptability came from medication discontinuation rates in the PROMISE trial11, in which there was no important difference between groups…"

13. The language used throughout the results and discussion, although compliant with the GRADE vocabulary is very difficult for a normal reader to plough through. It is severely in need of a Plain English intervention.

Thank you. We agree and have extensively revised the text.

14. There are concerns about some of the entries in Table 3. Why are results which are compatible with no difference being interpreted as "may increase" – see Spontaneous Abortion 20 weeks outcome. Other rows of the table classify this as "probably no difference" (at the risk of a Type II error). I also cannot see what the differences are between the multiple rows for low and middle versus high resource settings in the last section of this table.

We and GRADE place a de-emphasis on hypothesis testing and p-values, and instead place a higher value on confidence intervals (imprecision). In the case of spontaneous abortion, TDF appears to increase the risk by 10%. Most people would accept that a 10% increase in the risk of spontaneous abortion is important to almost all women. For that reason, we believe it is inappropriate to conclude that “there is no difference” or “no significant difference”. Because the confidence interval includes no effect (3.6% fewer to 30.4% more), we rate down certainty for imprecision. In this case, we say “TDF/FTC may increase the risk of spontaneous abortion.”, which we believe is the appropriate interpretation.

The other outcomes where we say “there may be no difference” are written in such a way because the estimated effect size is small enough to not be important to most people.

Page 6 line 2 - £10.5 seems rather low.

We have removed mention of dollar values at the suggestion of the reviewer above.

It is unclear why the random effects meta-analysis graphs are labelled as M-H – this is a fixed effect model.

The Mantel-Haenszel method was used for weighting individual studies. The Mantel-Haenszel method uses an alternative weighting scheme, which may be more robust even when data is sparse (and similar to inverse variance when data is not sparse). The random effects method refers to the calculation of the pooled estimate, which additionally incorporates a between-study heterogeneity estimate in the calculation and is unrelated to weighing of the studies.

Additional Questions:
Please enter your name: Prof Jon Deeks

Job Title: Professor of Biostatistics

Institution: University of Birmingham

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No