PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A study of antiretroviral drug class and anaemia risk in the current treatment era among people living with HIV in the United States: A clinical cohort study
AUTHORS	Harding, Barbara; Whitney, Bridget; Nance, Robin; Crane, HM; Burkholder, Greer; Moore, Richard; Mathews, William; Eron, Joseph; Hunt, Peter; Volberding, Paul; Rodriguez, Benigno; Mayer, Kenneth; Saag, Michael; Kitahata, Mari M.; Heckbert, Susan; Delaney, Joseph

VERSION 1 – REVIEW

REVIEWER	David Boettiger University of California, San Francisco. USA I'm currently based at UCSF and know Paul Volberding. We have never worked closely together but have spoken on the odd occasion.
REVIEW RETURNED	09-Jul-2019

GENERAL COMMENTS	Abstract Pg4, line6 – remove “during” Introduction Pg6, line55 – clearer to say ART failure rather than treatment failure Pg7, line22 – Sentence beginning “Since worsening…” is redundant Pg7, line 44 – The authors reference studies evaluating anemia and PI, and anemia and INSTI use. However, phases such as “increased rate” and “some participants discontinued” should be supported by actual numbers. These studies also warrant further comparison with the authors work in the discussion. Methods Pg8, line51 – should it be “2010 or” rather than “2010 and”? Pg9, line19-31 – The authors should describe here that they censored pts switching from one of the core drug groups to a non-core group. This becomes clear later on in the manuscript. Pg10, line4 – I suggest something like “first post-baseline Hb measure below XX..” Pg10, line31 – how was HCV diagnosed/defined? Pg10, line31 – I think you need to include baseline haemoglobin and AZT use in your models Pg10, line52 – The limited ART adherence data is mentioned but there is no description of how the authors used this data. I was expecting the authors to describe how they imputed the missing adherence data and included this variable in their models, or did

	sensitivity analyses where they only included pts with adherence data. Pg11, line 44 – please define administrative censoring Results Baseline table and models should include information on AZT use and baseline Hb. Figure 2 doesn't make sense to me. The study population only included PLHIV using a core ART class but this figure indicates about 15% were either not using a core class or were not using ART (not clear which – presumably the number of pts not using ART in CNICS is much <15%?) Discussion Pg15, line47 – could the authors comment on the likelihood of dialysis, etc being unevenly distributed among different ART class users Pg16, line20 - remove comment on generalizability of findings to PLHIV not on ART. This is a study on anemia risk with different ART regimens – of course it doesn't apply to people not using ART

REVIEWER	Annegret Pelchen-Matthews University College London, UK
REVIEW RETURNED	30-Sep-2019

GENERAL COMMENTS	The manuscript presents an interesting study which identifies a possible higher risk of anaemia associated with use of INSTIs in the current treatment era. This is an important question as anaemia has health implications for many comorbidities and may warrant clinical attention, especially with the increasing use of INSTIs in recent years. The authors use a stringent definition of anaemia and severe anaemia likely to identify clinically relevant health risks. However I have some concerns, and clarification is required in a number of areas. Major comments: 1. It is a strength of the study that it is focused on use of ARV core agents. However, given that associations of ARVs with anaemia have been mostly associated with NRTIs, I was confused by the lack of clarity about NRTI use in the exposure definitions. cART regimens involving a backbone of two NRTIs plus a core agent are briefly mentioned in the introduction (p7), hence I would assume that the study is restricted to individuals taking two NRTIs, though this is not explicitly stated in the methods? Please clarify, in the methods section on p. 8 (overview and setting, specifying as an enrolment criterion being on an ART regimen containing a PI, NNRTI or INSTI) and on P.9 (definition of the exposures). This is particularly problematic for understanding the 'multiple core agent classes' exposure: If all individuals are on 2 NRTIs, then those on multiple core agent classes would be taking regimens with >3 ARVs, which might include individuals with treatment failure or resistance for whom it is more difficult to construct appropriate treatment regimens and signifying poorer health status – confounding by indication for anaemia. If there was no requirement for NRTI use, then some NRTI-sparing two-drug regimens might also have been included in the multiple core agent classes category (e.g. INSTI + NNRTI regimens such as RAL+ETR, DTG+RPV, or INSTI + PI 2-drug regimens - particularly in recent times); reasons for prescribing of 2DRs are different from those for prescribing >3 ARVs. If it is the case that NRTI use may

	vary between participants, then it should be reported among the baseline characteristics and should be considered as a confounding variable in all analyses (especially since NRTI use has been shown to be associated with anaemia) which would then require major revisions! 2. CIs are reported consistently for all outcome measures. However P-values should also be reported as probabilities for testing the null hypotheses (e.g. in Table 2 and supplementary Table 1). Since 'statistically significant' terminology is used in the manuscript, the significance level should be specified in the methods. 3. Table 1 listing the baseline characteristics of PLWH included is not very informative as the data shown are for overlapping groups, i.e. those who developed anaemia or severe anaemia are included in the overall group of 16,505 participants. It would be more helpful to present characteristics for the subgroups used in the analyses of anaemia incidence and HR, e.g. comparing the 11586 individuals without prevalent anaemia who did not develop anaemia during FU and the 1040 persons without prevalent anaemia who did develop anaemia during FU (a total of 12,626 individuals without prevalent anaemia) and similarly for development of severe anaemia (for N=14869, N=488 and N=15357 individuals, respectively). This would avoid overlap between the groups presented, and would allow the identification of factors more strongly associated with development of anaemia or severe anaemia together with chi2 p-values for the comparisons. Also for Table 1: - under the heading 'ART class' for category 'Protease inhibitor', for those who developed anaemia during FU 558/1040 = 54%, not 71% as stated in the table - If required (see point 1), use of NRTIs should also be reported. 4. What is the denominator for the data shown in Figure 2? This cannot be proportion of the study population (of 16,505) since it there are around 15% of individuals using no core classes. It would also be helpful to list the number of individuals included in each year to understand how this figure relates to the data cited for the cohort overall in Table 1. In the description of the Figure (results section p.13) the text implies that the multiple core agent class all include an INSTI – is this correct? 5. It would be helpful if the reporting of the incidence of anaemia could include numbers of events and PYFU for the different classes. The results might be easier to understand if presented as a Forest plot. Also for Table 2: - Unadjusted HR for INSTI, incident severe anaemia in the CI: Replace the 0 with) - Kidney function is as eGFR (missing R) 6. Please clarify or correct the description of the decrease in haemoglobin level over time: the decrease of -0.06 g/dl/year seems very small and hardly clinically relevant. But then Table 3 suggests that his is not a decrease in HB level, but a relative adjusted decrease compared to that seen with NNRTIs. If so, a relative adjusted decrease should not be given the units of g/dl since it is a ratio. Please clarify, especially in Table 3. It may be

	helpful to list the crude decrease in HB for each ARV class (in g/dl/year) as well as the adjusted measures relative to NNRTIs and without units (this also applies for the sensitivity analysis). 7. As this is a cohort study, please mention confounding by indication among the strengths and limitations. 8. Not excluding participants with conditions strongly associated with anaemia or HB level would be confounding the result. In this regard, although the analysis for chronic anaemia gave similar results, it is notable that the CIs are large for this outcome, and all incidence rates for chronic anaemia were non-significant. Minor points: Abstract – Is this a retrospective or prospective cohort – it cannot be both? I suggest listing the design as a clinical cohort study. Abstract Results first sentence – remove the word 'during' at the end of the sentence. 'Overtime' should be changed to 'over time' (2 words) as these mean different things. Introduction: The precise definition of anaemia and severe anaemia according to haemoglobin level should be in the methods section, not in the Introduction, since the prior studies described use different definitions of anaemia.

VERSION 1 – AUTHOR RESPONSE

Responses to Reviewer 1 comments:

1. Reviewer 1 #1: Pg4, line6 – remove "during"

   Response to Reviewer 1 #1: We have removed the word "during" (page 3, line 20).

2. Reviewer 1 #2: Pg6, line55 – clearer to say ART failure rather than treatment failure

   Response to Reviewer 1 #2: We have replaced "treatment failure" with "ART failure" as suggested (page 4, line 9).

3. Reviewer 1 #3: Pg7, line22 – Sentence beginning "Since worsening…" is redundant

   Response to Reviewer 1 #3: We have condensed these sentences to reduce redundancy (page 4, lines 15-16).

4. Reviewer 1 #4: Pg7, line 44 – The authors reference studies evaluating anemia and PI, and anemia and INSTI use. However, phases such as "increased rate" and "some participants discontinued" should be supported by actual numbers. These studies also warrant further comparison with the authors work in the discussion.

   Response to Reviewer 1 #4: We have added more detail to the manuscript about the findings of these two studies in the introduction (page 4, lines 22-25) and have included mention of these in the discussion (page 8, lines 21-25).

5. Reviewer 1 #5: Pg8, line51 – should it be “2010 or” rather than “2010 and”?

   Response to Reviewer 1 #5: Yes, this should be “2010 or” and has been corrected accordingly (page 4, line 46).

6. Reviewer 1 #6: Pg9, line19-31 – The authors should describe here that they censored pts switching from one of the core drug groups to a non-core group. This becomes clear later on in the manuscript.

   Response to Reviewer 1 #6: The reviewer is correct: when a participant switches to a different ART regimen that is not one of the core classes of interest, they are censored. We moved mention of this censoring to where it was suggested by the reviewer (page 5, lines 12-13).

7. Reviewer 1 # 7: Pg10, line4 – I suggest something like “first post-baseline Hb measure below XX..”

   Response to Reviewer 1 #7: We have added “post-baseline” to the outcome ascertainment section as suggested for clarification (page 5, lines 23 and 26).

8. Reviewer 1 #8: Pg10, line31 – how was HCV diagnosed/defined?

   Response to Reviewer 1 #8: HCV was defined as: defined as a detectable HCV RNA level, HCV genotype or HCV antibody (the definition used in other work (PMID: 26990826). We have clarified this (page 5, lines 33-34).

9. Reviewer 1 #9: Pg10, line31 – I think you need to include baseline haemoglobin and AZT use in your models

   Response to Reviewer 1 #9: As suggested, we now include baseline Hb in our models estimating the risk of incident anaemia, severe anaemia, and chronic anaemia. Only 4% of this population used zidovudine during the study years (we have added this to the results section of the manuscript [page 6, line 45-46]). With such a low prevalence of use, zidovudine use was unlikely to impact findings and thus we did not include it in models.

10. Reviewer 1 #10: Pg10, line52 – The limited ART adherence data is mentioned but there is no description of how the authors used this data. I was expecting the authors to describe how they imputed the missing adherence data and included this variable in their models, or did sensitivity analyses where they only included pts with adherence data.

    Response to Reviewer 1 #10: In the 55% of the study population with assessment of adherence, the adherence was very high, mean 98% (IQR: 93, 100). We did not impute adherence in those with missing data because there was so little variability.

11. Reviewer 1 #11: Pg11, line 44 – please define administrative censoring

    Response to Reviewer 1 # 11: Administrative censoring is the date through which each site had complete data. The date differed by site based on data collection activities at that site and

how quickly data were made available to the data repository. We have now included information on administrative censoring in our methods section (page 4, lines 36-37).

12. Reviewer 1 #12: Baseline table and models should include information on AZT use and baseline Hb.

    Response to Reviewer 1 #12: See response to Reviewer 1 #9 above.

13. Reviewer 1 #13: Figure 2 doesn't make sense to me. The study population only included PLHIV using a core ART class but this figure indicates about 15% were either not using a core class or were not using ART (not clear which – presumably the number of pts not using ART in CNICS is much <15%?)

    Response to Reviewer 1 #13: Thank you for pointing out the lack of clarity in Figure 2. We have revised Figure 2 to include only the 16,505 participants who were using a core ART class of interest (and have clarified this in the figure legend for Figure 2). The revised Figure 2 no longer includes the 743 people with no record of ART use of any kind or the 2,012 who were not using one of the ART core classes of interest, which has also been made more clear in our Figure 1 flow chart.

14. Reviewer 1 #14: Pg15, line47 – could the authors comment on the likelihood of dialysis, etc being unevenly distributed among different ART class users

    Response to Reviewer 1 #14: We do not have specific information on dialysis outcomes in this population (this is a potential future area of work for us). It is possible that some of the associations we are detecting might be due to channeling of ARV drugs to sicker patients and this is always a threat in observational studies. However, none of these ART core medications have standard recommendations to be avoided in patients at risk of anaemia. Because we are focusing on patients who developed anaemia during followup, if there was an inbalance of incident dialysis then the anaemia could be an indirect effect of the drug (by way of dialysis or another mediating health outcome), which is still relevant to comparative safety.

15. Reviewer 1 #15: Pg16, line20 - remove comment on generalizability of findings to PLHIV not on ART. This is a study on anemia risk with different ART regimens – of course it doesn't apply to people not using ART

    Response to Reviewer 1 #15: We have removed mention of non-generalizability to untreated participants (page 8, line 20).

<u>Responses to Reviewer 2 comments:</u>

1. Reviewer 2 #1: It is a strength of the study that it is focused on use of ARV core agents. However, given that associations of ARVs with anaemia have been mostly associated with NRTIs, I was confused by the lack of clarity about NRTI use in the exposure definitions. cART regimens involving a backbone of two NRTIs plus a core agent are briefly mentioned in the introduction (p7), hence I would assume that the study is restricted to individuals taking two NRTIs, though this is not explicitly stated in the methods? Please clarify, in the methods section on p. 8 (overview and setting, specifying as an enrolment criterion being on an ART regimen containing a PI, NNRTI or INSTI) and on P.9 (definition of the

exposures). This is particularly problematic for understanding the 'multiple core agent classes' exposure: If all individuals are on 2 NRTIs, then those on multiple core agent classes would be taking regimens with >3 ARVs, which might include individuals with treatment failure or resistance for whom it is more difficult to construct appropriate treatment regimens and signifying poorer health status – confounding by indication for anaemia. If there was no requirement for NRTI use, then some NRTI-sparing two-drug regimens might also have been included in the multiple core agent classes category (e.g. INSTI + NNRTI regimens such as RAL+ETR, DTG+RPV, or INSTI + PI 2-drug regimens - particularly in recent times); reasons for prescribing of 2DRs are different from those for prescribing >3 ARVs. If it is the case that NRTI use may vary between participants, then it should be reported among the baseline characteristics and should be considered as a confounding variable in all analyses (especially since NRTI use has been shown to be associated with anaemia) which would then require major revisions!

Response to Reviewer 2 #1: Thank you for pointing out the lack of clarity on this issue.

All participants included in this analysis were on regimens containing a backbone of 2

NRTIs plus a PI, NNRTI, or INSTI. We have revised the methods section to describe this (page 5, line 2 and 10). In addition, to address the final part of this comment, we conducted a sensitivity analysis that included baseline NRTI backbone adjustment for

ART Regimen	**Hazard Rati › of incident anaemia (n=12,626) Minimally adjusted[a] Unadjusted**	**With abacavir at baseline in model[b]**
NNRTI (REF)	1.00	1.00
PI	1.26 (1.05-1.52)	1.09 (0.90, 1.32) 1.09 (0.90, 1.32)
INSTI	1.39 (1.11-1.75)	1.26 (1.00, 1.58) 1.27 (1.01, 1.60)
Multiple core classes	2.02 (1.65-3.48)	1.39 (1.13, 1.70) 1.37 (1.11, 1.68)

abacavir (vs. tenofovir) to assess if the adjustment for abacavir influenced the point estimates we were detecting for INSTI and other core regimens and found results were essentially unchanged (table below summarizes this). We have also added to the manuscript that we conducted this sensitivity analyses (page 6, lines 21-23) and that results were similar to main findings (page 7, lines 33-36).

[a]Minimally adjusted model includes adjustment for: age, sex, race, site, hepatitis C coinfection, CD4 category, VL, eGFR category and haemoglobin at baseline. [b]Model includes adjustment variables in minimally adjusted model plus a term for baseline abacavir use

2. Reviewer 2 #2: CIs are reported consistently for all outcome measures. However P-values should also be reported as probabilities for testing the null hypotheses (e.g. in Table 2 and

supplementary Table 1). Since 'statistically significant' terminology is used in the manuscript, the significance level should be specified in the methods.

Response to Reviewer 2 #2: We prefer to focus on confidence intervals to convey inferences about study findings, rather than p-values, and have removed the word "significant" from our manuscript. The p-values for associations were reasonable; for instance, in the updated models for the adjusted risk of anaemia associated with PI use: HR 1.09 (95% CI: 0.90, 1.32), p=0.370; with INSTI use: 1.26 (1.00, 1.58), p=0.046; multiple core use: 1.39 (1.13, 1.70), p=0.002. We do not think the p-values are necessary for the manuscript. However, if the editor prefers that p-values be included, we are willing to add them.

3. Reviewer 2 #3: Table 1 listing the baseline characteristics of PLWH included is not very informative as the data shown are for overlapping groups, i.e. those who developed anaemia or severe anaemia are included in the overall group of 16,505 participants. It would be more helpful to present characteristics for the subgroups used in the analyses of anaemia incidence and HR, e.g. comparing the 11586 individuals without prevalent anaemia who did not develop anaemia during FU and the 1040 persons without prevalent anaemia who did develop anaemia during FU (a total of 12,626 individuals without prevalent anaemia)
and similarly for development of severe anaemia (for N=14869, N=488 and N=15357 individuals, respectively).

This would avoid overlap between the groups presented, and would allow the identification of factors more strongly associated with development of anaemia or severe anaemia together with chi2 p-values for the comparisons.

Also for Table 1: under the heading 'ART class' for category 'Protease inhibitor', for those who developed anaemia during FU 558/1040 = 54%, not 71% as stated in the table

Response to Reviewer 2 #3: We have now updated Table 1 to show the four columns as suggested. For the additional comment about 54% versus 71%, thank you for catching this error. We have now corrected the table to show that 54% of those who developed anaemia were using a PI at baseline.

4. Reviewer 2 #4: What is the denominator for the data shown in Figure 2? This cannot be proportion of the study population (of 16,505) since it there are around 15% of individuals using no core classes. It would also be helpful to list the number of individuals included in each year to understand how this figure relates to the data cited for the cohort overall in Table 1. In the description of the Figure (results section p.13) the text implies that the multiple core agent class all include an INSTI – is this correct?

Response to Reviewer 2 #4: Thank you for pointing out the lack of clarity in Figure 2. We have revised Figure 2 to include only the 16,505 participants who were using a core ART class of interest (and have clarified this in the figure legend for Figure 2). The revised Figure 2 no longer includes the 743 people with no record of ART use of any kind or the 2,012 who were not using one of the ART core classes of interest, which has also been made more clear in our Figure 1 flow chart.

To the second part of this comment, those in the multiple core agent class had a record of use of any two or more of the core classes of interest. On page 5, line 15-17 we explain, "some PLWH in this cohort had prescriptions for multiple core classes simultaneously. Participants with regimens containing more than 1 core class were categorized separately in analyses as users of "multiple core classes." Regarding changes over time, on page 6, lines

41-45, we further explained that “and among those prescribed multiple cores, the proportion comprised of INSTI plus another core class increased as study years progressed.”

5. Reviewer 2 #5: it would be helpful if the reporting of the incidence of anaemia could include numbers of events and PYFU for the different classes. The results might be easier to understand if presented as a Forest plot.
Also for Table 2: Unadjusted HR for INSTI, incident severe anaemia in the CI: Replace the 0 with

) and kidney function is as eGFR (missing R)

Response to Reviewer 2 #5: We have now included incidence measures in Table 2 and referenced this in the results section (page 7 lines 2-3).

In Table 2, we have corrected the 0 to be a) and corrected the footnote to read eGFR: thank you for pointing this out.

6. Reviewer 2 #6: Please clarify or correct the description of the decrease in haemoglobin level over time: the decrease of -0.06 g/dl/year seems very small and hardly clinically relevant. But then Table 3 suggests that his is not a decrease in HB level, but a relative adjusted decrease compared to that seen with NNRTIs. If so, a relative adjusted decrease should not be given the units of g/dl since it is a ratio. Please clarify, especially in Table 3. It may be helpful to list the crude decrease in HB for each ARV class (in g/dl/year) as well as the adjusted measures relative to NNRTIs and without units (this also applies for the sensitivity analysis).

Response to Reviewer 2 #6: The values presented in this table (now Table 4) should be interpreted as the change in haemoglobin per year (in g/dL/year) as indicated in the footnote. This is the mean change per year for a given ART core class minus the mean change per year in the reference group adjusted for characteristics in the model, which provides the adjusted change per year. We agree that-0.06 g/dL/year change is a small change.

7. Reviewer 2 #7: As this is a cohort study, please mention confounding by indication among the strengths and limitations.

Response to Reviewer 2 #7: Since anaemia is not a recognized adverse effect of NNRTIs, PIs, or INSTIs, differential prescribing due to concern about development of anaemia was unlikely. However, differential prescribing on the basis of other patient characteristics associated with anaemia remains possible. We have added discussion of confounding by indication as a limitation in the discussion section (page 8, lines 6-8) in addition to the language we already included about confounding by indication (page 8, lines 33-37).

8. Reviewer 2 #8: Not excluding participants with conditions strongly associated with anaemia or HB level would be confounding the result. In this regard, although the analysis for chronic anaemia gave similar results, it is notable that the CIs are large for this outcome, and all incidence rates for chronic anaemia were non-significant.

Response to Reviewer 2 #8: We agree with this comment.

9. Reviewer 2 #9: Abstract – Is this a retrospective or prospective cohort – it cannot be both? I suggest listing the design as a clinical cohort study.

Response to Reviewer 2 #9 (minor point 1): As suggested, we have revised the title to include the description of this study as a "clinical cohort study". The CNICS study is a prospective cohort study.

10. Reviewer 2 #10: Results first sentence – remove the word 'during' at the end of the sentence.

Response to Reviewer 2 #10 (minor point 2): We have removed the word "during" (page 3, line 20).

11. Reviewer 2 #11: 'Overtime' should be changed to 'over time' (2 words) as these mean different things.

Response to Reviewer 2 #11 (minor point 3): We have corrected this to read "over time" (page 5, line 32).

12. Reviewer 2 #12: Introduction: The precise definition of anaemia and severe anaemia according to haemoglobin level should be in the methods section, not in the Introduction, since the prior studies described use different definitions of anaemia.

Response to Reviewer 2 #12 (minor point 4): The definitions of anaemia and severe anaemia have been moved to the suggested location (page 5, lines 23-24).

VERSION 2 – REVIEW

REVIEWER	David Boettiger Kirby Institute, UNSW Sydney, Australia I am supported by a National Health and Medical Research Council Early Career Fellowship (APP1140503) and have received research funding from Gilead Sciences
REVIEW RETURNED	01-Nov-2019

GENERAL COMMENTS	I am satisfied with the majority of changes made, however, I still have concerns about the authors' not adjusting for zidovudine use in their models. Even though they now report that the prevalence of ZDV use was low (4%), ZDV is so strongly associated with anemia that I don't believe it can be ignored. At a minimum, I suggest either excluding patients who used ZDV during follow up (as sensitivity or for the main analysis) or doing an analysis (again, either as main or sensitivity) similar to the sensitivity analysis that includes adjustment for NRTI backbone but categorizing the backbone as ZDV v non-ZDV. It also seems like the current NRTI backbone sensitivity analysis needs some work as the authors appear to have only adjusted for baseline NRTI rather than time-updated NRTI, they only show the results for incident anemia in the response to reviewer letter rather than presenting the results for incident and severe anemia in the main paper or supplementary material, and they describe having adjusted for

	abacavir (vs. tenofovir) which only makes sense if all patients in the analysis were using either abacavir or tenofovir, which probably isn't correct.

REVIEWER	Annegret Pelchen-Matthews University College London, UK
REVIEW RETURNED	17-Nov-2019

GENERAL COMMENTS	Thank you for the revisions to the manuscript. The new Table 1 makes baseline comparisons much more informative, and the new Table 2 clarifies the crude incidence rates for anaemia and severe anaemia; I find it helpful to see the numbers of events that the models in the now Table 3 are based on. Thanks also for clarifying the use of CIs vs P-values – this approach now avoids the binary significant/not significant interpretation of results. Thank you also for the sensitivity analysis for use of ABC vs TDF. These changes have strengthened the manuscript. I do still have some concerns with the baseline table, Table 1, which seems to have a number of errors: - The total number of individuals who do not develop severe anaemia (heading to column 3) should be 14,869, NOT 14,896 as presented (i.e. last two digits are reversed. NB 15,357-488 = 14869, and categories of variables add up to this total). - For Years in CNICS at cohort entry – the "a" should be superscript (footnote), also please check the entries for "Develop anaemia" where the confidence intervals do not agree with the values in the previous version of the manuscript, and "Develop severe anaemia" – listed as 5.56 years, other entries and the CIs are only given to 1 decimal place. - BMI: Was there a "BMI unknown" category (column totals are less than the totals stated at the head of the table, i.e. 11,106, 1008, 14,275 and 473, respectively, but percentages add to 100). Please add a footnote to explain. In addition, please correct the percentage for BMI 25.0 to <30.0 for "Do not develop severe anaemia", which is listed as 345%. - ART core class for NNRTI in the column "Develop anaemia" – 117: is this a typographical error? The total for the categories does not add up, and the previous version of the paper listed 177 individuals (which would give 17% as listed). - Self-reported adherence: The values and CI for "Develop anaemia" and "Develop severe anaemia" differ slightly from the first version of the manuscript. These are the errors that I spotted, I would advise very careful checking of this table to ensure all entries are correct.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1 comment #1: I still have concerns about the authors' not adjusting for zidovudine use in their models. Even though they now report that the prevalence of ZDV use was low (4%), ZDV is so strongly associated with anaemia that I don't believe it can be ignored. At a minimum, I suggest either excluding patients who used ZDV during follow up (as sensitivity or for the main analysis) or doing an analysis (again, either as main or sensitivity) similar to the sensitivity analysis that includes adjustment for NRTI backbone but categorizing the backbone as ZDV v non-ZDV.

Response to reviewer 1 comment #1: We have updated, refined, and expanded all the zidovudine analyses. We now include results from sensitivity analyses restricting the population to participants without zidovudine use during study time and present findings for anaemia risk, severe anaemia risk and change in Hb for non-zidovudine users. These analyses showed little change in the risk estimates. We have mentioned these sensitivity analyses in the methods section (page 6, lines 25-29), in the results section (page 7, lines 39-41) and have included results in supplemental tables (Supplemental Table 4 and Supplemental Table 5).

Reviewer 1 comment #2: It also seems like the current NRTI backbone sensitivity analysis needs some work as the authors appear to have only adjusted for baseline NRTI rather than time-updated NRTI, they only show the results for incident anaemia in the response to reviewer letter rather than presenting the results for incident and severe anaemia in the main paper or supplementary material, and they describe having adjusted for abacavir (vs. tenofovir) which only makes sense if all patients in the analysis were using either abacavir or tenofovir, which probably isn't correct.

Response to reviewer 1 comment #2: Thank you for pointing out the lack of clarity in the description of this sensitivity analysis. Supplemental Table 3 now shows incidence of anaemia and severe aneaemia as suggested with time-updated adjustment for abacavir use. In this time period, virtually all of the backbones were abacavir vs tenofovir. While a small portion were on zidovudine at baseline, this rapidly dropped off over time. Furthermore, while the use of TAF is currently quickly increasing in CNICS clinics, because of the timing of the end date of this study, it was not yet a phenomenon. This leaves virtually all of the included patients on abacavir vs. tenofovir for almost all of the study period, hence our decision was to use an approach with adjustment for abacavir, and as you suggested, this should be time-updated abacavir (results shown in supplemental table 3 are very similar to findings in the primary analyses). Furthermore, in sensitivity analyses where we are more restrictive with NRTI inclusion criteria (excluding participants with zidovudine use), findings are essentially the same although the population becomes somewhat less generalizable. Even in this analysis, there is essentially no impact as the numbers receiving anything but tenofovir and abacavir are small. We have updated the description of this sensitivity analysis (page 6, lines 22-25). This information is now also included in supplemental material for readers (Supplemental Table 3).

Reviewer 2 comment #1: I do still have some concerns with the baseline table, Table 1, which seems to have a number of errors:
a) The total number of individuals who do not develop severe anaemia (heading to column 3) should be 14,869, NOT 14,896 as presented (i.e. last two digits are reversed. NB 15,357-488 = 14869, and categories of variables add up to this total).

b) For Years in CNICS at cohort entry – the "a" should be superscript (footnote), also please check the entries for "Develop anaemia" where the confidence intervals do not agree with the values in the previous version of the manuscript, and "Develop severe anaemia" – listed as 5.56 years, other entries and the CIs are only given to 1 decimal place.

c) BMI: Was there a “BMI unknown” category (column totals are less than the totals stated at the head of the table, i.e. 11,106, 1008, 14,275 and 473, respectively, but percentages add to 100). Please add a footnote to explain. In addition, please correct the percentage for BMI 25.0 to <30.0 for “Do not develop severe anaemia”, which is listed as 345%.

d) ART core class for NNRTI in the column “Develop anaemia” – 117: is this a typographical error? The total for the categories does not add up, and the previous version of the paper listed 177 individuals (which would give 17% as listed).

e) Self-reported adherence: The values and CI for “Develop anaemia” and “Develop severe anaemia” differ slightly from the first version of the manuscript.

Response to reviewer #2 comment #1: Thank you very much for bringing these errors to our attention. We have made the following corrections to Table 1.

a) We have corrected the total number of individuals who do not develop severe anaemia to read “14,869” correcting the switch of the last two digits.

b) We have corrected the superscript and changed the number of significant digits reported to be consistent across all entries of “years in CNICS at cohort entry.”

c) Thank you for bringing this to our attention, all those included in the analytic sample had complete covariate information. There were a portion of participants with a very large BMI who were erroneously not included in the table among the ≥30.0 category as they should have been. This has now been corrected in the table and participants in each category of BMI add up to the overall total in each column.

d) We have corrected the ART core class for NNRTI in the develop anaemia column to read 177.

e) We have checked these confidence intervals and they are correct as currently presented. Furthermore, we have checked all other parts of this table and believe it to be complete and accurate.

VERSION 3 – REVIEW

REVIEWER	David Boettiger The Kirby Institute, Australia I am supported by a National Health and Medical Research Council Early Career Fellowship (APP1140503) and have received research funding from Gilead Sciences
REVIEW RETURNED	13-Jan-2020

GENERAL COMMENTS	I am satisfied with the corrections. I appreciate the inclusion of a sensitivity analysis excluding users of ZDV. For the sensitivity analysis on NRTI backbone, I suggest being clearer on the

	categorization used. The authors indicate that they adjust for abacavir use but it should really be referred to as adjustment for NRTI background where NRTI was categorized as abacavir, x, y, etc. As currently presented, I don't know what the non-abacavir groups are. It could actually make sense to combine the NRTI backbone and ZDV sensitivity analyses by keeping ZDV users in the analysis but categorizing the NRTI backbone as abacavir, ZDV, and other. This may address the generalizability concerns associated with excluding ZDV users which the authors mention in their response to reviewer comments.

REVIEWER	Annegret Pelchen-Matthews University College London, UK
REVIEW RETURNED	20-Jan-2020

GENERAL COMMENTS	Thank you for the revisions to this manuscript – I have no remaining concerns. You might like to note a few very minor suggestions: Methods - Participant characteristics: P.6 line 41: Remove the first "only" in "…only analyses only)" - Statistical analysis - P.7 AZT and anaemia line 430-44 - could cite the reference again (as in the Introduction) Results P.9 line 1: confidence intervals overlapped one. I think it might be clearer to use the numeric symbol (1) as you are referring to the null value in the analysis.

VERSION 3 – AUTHOR RESPONSE

We thank reviewers for their additional comments (shown in bold) and have responded to these below.

Reviewer 1:
I am satisfied with the corrections. I appreciate the inclusion of a sensitivity analysis excluding users of ZDV. For the sensitivity analysis on NRTI backbone, I suggest being clearer on the categorization used. The authors indicate that they adjust for abacavir use but it should really be referred to as adjustment for NRTI background where NRTI was categorized as abacavir, x, y, etc. As currently presented, I don't know what the non-abacavir groups are. It could actually make sense to combine the NRTI backbone and ZDV sensitivity analyses by keeping ZDV users in the analysis but categorizing the NRTI backbone as abacavir, ZDV, and other. This may address the generalizability concerns associated with excluding ZDV users which the authors mention in their response to reviewer comments.

Response to Reviewer 1:
We have created a supplemental table summarizing the four most common backbone component drugs at baseline. This provides additional information on what the non-abacavir group consists of in the sensitivity analysis addressing concerns that different backbone medications could be influencing results. The majority of participants have a backbone of emtricitabine/lamivudine (3TC) + tenofovir or 3TC + abacavir as their backbone regimen (~80%). So, we largely are comparing abacavir and tenofovir (drug 2 in the table below) as drug 1 is pretty consistent so cannot explain differences in core agents.

We have added text to the manuscript (page 7 lines 2-3) to explain that, "In the analytic study sample, backbone regimens mainly consisted of emtricitabine/lamivudine (3TC) plus tenofovir

or 3TC plus abacavir (Supplemental Table 1)." We additionally added text to Supplemental Table 4 (which is now the supplemental table providing results from the NRTI backbone regimen sensitivity analyses) to explain that we are applying time-updated adjustment for abacavir and that the other backbone regimens in our sample largely consist of tenofovir/3TC or complex backbone combinations/NRTI-sparing regimen/salvage therapies.

As we previously described, there is a small portion of our population that use ZDV during study follow-up. Because of the small amounts of ZDV use, we think it is best to restrict the population to non-ZDV users in a separate sensitivity analysis.

Supplemental Table 1: Backbone component drugs at baseline for participants in analytic sample (N=15,505)

Backbone Regimen	Drug 1	Drug 2	Percentage
1	Emtricitabine/Lamivudine	Tenofovir	69
2	Emtricitabine/Lamivudine	Abacavir	10
3	Emtricitabine/Lamivudine	Tenofovir/Abacavir	4
4	Emtricitabine/Lamivudine	Zidovudine	4
5	Other[a]		13

[a]Other backbone combinations include complex backbone combinations, NRTI-sparing regimens and salvage therapies.

Reviewer: 2
Reviewer Name: Annegret Pelchen-Matthews
Institution and Country: University College London, UK
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Thank you for the revisions to this manuscript – I have no remaining concerns.

You might like to note a few very minor suggestions:
Methods
- Participant characteristics: P.6 line 41: Remove the first "only" in "…only analyses only)"
- Statistical analysis - P.7 AZT and anaemia line 430-44 - could cite the reference again (as in the Introduction)
Results P.9 line 1: confidence intervals overlapped one. I think it might be clearer to use the numeric symbol (1) as you are referring to the null value in the analysis.

Response to Reviewer 2:
We have made the suggested changes (page 5 line 38, page 6 lines 25-26, and page 7 line 17).