Reviewer's report

Title: Socio-Economic determinants of HIV disease progression in Kenya.

Version: 3 Date: 21 March 2015

Reviewer: Peter Rebeiro

Reviewer's report:

1. Is the question posed by the authors well defined?
Yes. This study aims to describe socioeconomic differences in HIV disease progression in order to inform the prioritization of populations during ART scale-up according to WHO criteria.

2. Are the methods appropriate and well described?
Methods are generally appropriate; description of deaths during the follow-up period, if any, would be appreciated. Analysis would benefit from description of how/why categories for major predictor/exposure of interest (income) were formulated. I would also reformulate CD4 count when included in regression models to be scaled (per 100 cells/µL) or else modeled with restricted cubic splines. Please justify why important confounders such as age and sex are not included in the regression models, even if they do not achieve statistical significance in bivariate comparisons (which is likely due to small sample size); important known confounders of disease progression (such as age and sex) should be accounted for; in particular, if categorization of age creates sparse data (due to small sample size within different categories), then age should be modeled as a continuous covariate (either scaled, transformed, or using splines).

3. Are the data sound?
They appear to be.

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
Yes.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?
Yes.

6. Are the discussion and conclusions well balanced and adequately supported by the data?
Yes, with concerns about the methodology noted above.

7. Are limitations of the work clearly stated?
No. The “Discussion” section would benefit from a clear statement of limitations and strengths of this study, the design, the methods used, generalizability to low- and middle-income settings, etc.

8. Do the authors clearly acknowledge any work upon which they are building,
both published and unpublished?
Yes.

9. Do the title and abstract accurately convey what has been found?
Yes.

10. Is the writing acceptable?
Yes, generally; needs serious revision for grammar, spelling, syntax, and consistency, though.

• Major Compulsory Revisions

Methodology: Please describe how and why the cut-offs for the categories of income were used. Also, be clear if US dollars or Kenyan dollars are used as the currency in the analyses. You also do not describe the Kaplan-Meier survival analyses you conduct (or the log-rank tests you apparently use to compare survival functions by baseline CD4 or income strata); please include the appropriate description of using these methods in the “Methods” section.

Throughout, use the correct units for CD4 count (this is cells/µL or cells/mm3, not cells/mL)

Grammar, syntax, spelling, and content edits as follows:

Abstract:
Page 2, Line 4: (...“irrespective of CD4 count.”)
Page 2, Lines 5-6: “Few studies to date have sought to identify socioeconomic determinants of disease progression that may provide insight into which populations…” (please maintain consistency in spelling “socioeconomic”)
Page 2, Line 8: “A cohort of 312 HIV positive individuals was retrospectively analyzed.” (follow-up doesn’t occur retrospectively, even if the study design is retrospective)
Page 2, Line 11: “Outcome variables were progression to World Health Organization (WHO) clinical stage 3 or 4 and/or decline of CD4 count to <350 cells/µL during two years of follow-up.” (proper units for CD4 count are cells/µL or cells/mm3, not cells/mL)
Page 2, Line 15: “…univariate analysis and to control for confounders.”

Background:
Page 3, Lines 2-3: “The progressive loss of CD4+ T lymphocytes during HIV infection eventually results in the loss of an ability to mount a desirable immune response…”
Page 3, Line 5: “…of the greatest contributors to mortality in Africa.”
Page 3, Lines 5-6: would eliminate sentence beginning “A serodiscordant couple is a pair of stable…” and simply include citation for the definition (“3”) at the end of the following sentence. Providing a standalone definition here is disruptive to the flow of the text and should be general knowledge for the readership. If the definition must be included, would include it briefly in the sentence beginning “A
A large proportion of new HIV infections…” (also, choose either “serodiscordant” or “discordant” and remain consistent in use of terminology)

Page 3, Line 10: “The World Health Organization (WHO)…” (expand all acronyms the first time they are used)

Page 3, Lines 15: “…may be hindered due to financial considerations.”

Page 3, Line 16: “…sociodemographic data would be very helpful…”

Page 3, Line 17: “This study therefore focuses on the socioeconomic…”

Methods:

Page 4, Lines 3-4: “Data were analysed from discordant couples participating in a randomized clinical trial of acyclovir treatment as a preventative measure for HIV transmission.” (do not include name of the clinic again here; this just disrupts the flow of the sentence. Include the location, Nairobi, in the first sentence).

Page 4, Lines 7-8: “Out of the enrolled 416 individuals, 86 were excluded due to CD4 count <350 cells/µL at enrollment, meaning they were eligible for HAART according to national guidelines in place in 2011.” (be consistent in using HAART or else ART throughout)

Page 4, Line 9: would point out that the 18 individuals excluded were only 5% (18/330) of the eligible cohort

Page 4, Line 10: capitalize “Figure 1”

Page 4, Line 11: “Data were analysed…” (“data” is a plural noun)

Page 4, Line 12: “…progression by WHO criteria …”

Page 4, Line 13: “Sociodemographic and economic predictors…”

Page 4, Lines 15-16: “…WHO clinical disease progression according to staging criteria (stages three and four) and CD4 count of <350 cells/µL.”

Page 4, Lines 18-19: “Chi-squared tests were used to determine differences in proportions with advanced disease (by both WHO staging and CD4 count criteria) across demographic, socioeconomic, and clinical characteristics. The Mann Whitney U test was used to compare medians of continuous variables by advanced disease status.”

Results:

The section called “Cohort Selection” is repetitive, and is described in the “Methods” section. This paragraph may be deleted.

Page 5, Lines 10-11: “All 312 individuals analysed entered the study in early HIV disease stages, either in WHO stages one and two or with CD4 count >350 cells/µL.”

Page 5, Line 11: “Median age at baseline was 30 years (Interquartile Range [IQR]: 26, 35).”

Page 5, Line 13: “Median household daily income was $3…”

Page 5, Line 17: “…and all individuals were in WHO stages one or two.”

Page 5, Line 18: “Only 15 (4.8%) of individuals…”
Page 5, Line 20: “…power to study predictors of disease progression by WHO disease stage alone.”

Page 5, Lines 21-22: delete sentence beginning “Median CD4 at entry…” (this repeats the information immediately above)

Page 5, Lines 22-23: “There was a significant (p=0.03) difference in median CD4 count for those who had progressed to <350 cells/µL and those who had not.”

Page 6, Line 1: “This difference indicated that CD4 at entry was a confounder to be included in the logistic regression model model during multivariate analyses.”

Page 6, Lines 3-5: This sentence is worded poorly- please rephrase and be clear about the reference category (likely those with CD4<350 cell/µL). The Kaplan-Meier for time to disease progression by baseline CD4 stratum should be Figure 2, not Figure 1 (which was already used to describe the cohort analyzed). Further, the Kaplan-Meier figure shows that disease progression was faster for those with LOWER baseline CD4 (those in the 350-500 category) compared to those with higher baseline CD4 (those in the >500 category), not the opposite, as you describe.

Page 6, Line 6: “There was no significant association found in bivariate analyses...”

Page 6, Line 9: Half of the sample population was reported to have an income (at all)? According to Table 1, you have income data for 311 of the 312 people in the sample. Please resolve this discrepancy.

Page 6, Lines 10-11: “…found to be significantly associated (p=0.02) with disease progression by CD4 count as illustrated...”

Page 6, Lines 12-19: Use consistent notation. Everywhere you describe an Odds Ratio or other measure of association, use the full term the first time and the acronym only after naming it, like so: “(Adjusted Odds Ratio [aOR]=4.0; 95% Confidence Interval [CI]: 1.0, 13.8)”. Do not include both p-values and 95% CI levels in the same result; choose one and be consistent. Again, please describe in the “Methods” section how you chose these categories for income; also describe the use of either the log-rank test or the Wilcoxon test for equality of survival functions in the “Methods” section.

Page 6, Line 19: You now note the use of US $ as the currency. Note this in the “Methods” section; afterward, just use “$”.

Discussion:

Page 7, Line 4: Do not repeat sentence about median baseline CD4. The first few sentences belong in the Results section, and are repetitive here. Would begin the “Discussion” with sentence beginning “Only a small proportion (4.8%) of HIV-infected individuals...,” then include mention of the fact that in your study population, a much larger proportion progressed using CD4 criteria.

Page 7, Line 12: “…explained by the fact that studies (13) have shown a rapid decline in CD4 count occurs among those with baseline CD4 >500 cells/µL, with a slower decline thereafter.”
Page 7, Line 14-15: the final sentence in this paragraph may be deleted, as it is redundant.

Page 7, Line 19: add period at end of sentence

Page 7, Line 23: “may have been a contributor”? This is the likeliest reason- the number in that age group is so low that there is probably insufficient power to detect differences in any characteristic by that group.

Page 8, Line 6: “…with increased level of education reported to have contributed…”

Page 8, Line 15: “services for HIV-infected individuals were made available (6,9).”

Page 8, Line 24: you have not defined PLWHIV before; should expand here “…people living with HIV (PLWHIV)…”

Page 8, Line 25: “…mentioned studies contributing to immune…”

Page 9, Line 5: “The association reported…”

Page 9, Lines 5-9: This is a fairly strong conclusion based on an analysis with limited power in a highly select population (those participating in a clinical trial); perhaps tempering this statement a little would be warranted?

Table 1: Be consistent in capitalization and use of bold font. For the WHO Staging category, you have listed an “N” of 330, while the title of your table indicates that the total population included is 312. Your CD4 count categories also sum to 330. Be clear and make sure you are including only the study participants in this table!

Table 3: expand “COR” and “AOR” to be “Crude Odds Ratio” and “Adjusted Odds Ratio” or else give the proper expansion in a table footnote.

Figures 2 and 3: Use clear labels for the X and Y axes (e.g., “Cumulative Survival” for the Y-axis, and “Duration of Follow-up (Months)” for the X-axis). Also use clear labels in the figure legends (e.g., “CD4 Baseline Category”). If using your own titles for figures, crop out the “default” titles of “Survival Functions” or “Kaplan-Meier survival estimates”.

• Minor Essential Revisions

None.

• Discretionary Revisions

None.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:

I declare that I have no competing interests.