Epidemiological profile of tuberculosis patients by HIV status in Brazil: a retrospective cohort study

This MS has used a large data base of surveillance information from Brazil, a high TB burden country. The topic of this MS is important addition to the existing literature, but the MS writing and some analysis issues are to be addressed before publication. Detailed specific issues are detailed below.

One major issue for credibility of results by multivariate analysis (hierarchical) is missing data for certain variables. I see table 1 & 2 where authors have shown samples (numbers) included in the analysis for each of the variables. All the data was not available for all variables, including some as less as 50% of cases included in the final analysis. Data on skin test, Culture, histopathological report was available for <50% of the cases included.

Authors only had HIV information for just over 50% of the all TB cases captured from SISAN/TB database (43% missing from figure) but authors wrongly, state prevalence of TB-HIV co-infection as 10.8% by not excluding all TB cases for whom HIV status was unknown. The actual proportion (not prevalence since this isn't a survey using representative sample) is 19.07% shown in the figure.

Indeed, all TB cases included in the analysis are not representative of TB cases in Brazil since all TB cases, and those with and without HIV co-infection may not be reported to SINAN/TB as both conditions have social stigma.

Considering the missing data and for variables analyzed in regression models, and non-availability of HIV status in a major proportion of cases is a serious limitation but not withholding publication of this MS based on a large sample from Brazil. Authors should attempt imputations or sensitivity analysis. Though authors admit to this in discussion as demographic characteristics were similar, have not shown any numbers in the MS.

Thought this MS generally follows STROBE guidelines, most important missing information is about how SINAN/TB is organised in Brazil, what information is gathered, flow of the information i.e. health management information system
(HMIS) and how authors collected this data from SINAN for this MS.

The design is not retrospective cohort, but is a secondary data analysis for operational research as admitted by authos under limitations.

Information about general health services from which surveillance information tricked into SINAN/TB and TB control program, HIV testing policy in Brazil need a mention in the MS for international readers to comprehend their findings.

Title suggests epidemiological profile but, article presents factors associated with TB-HIV co-infection relative to TB only.

Authors should consider presenting descriptive picture of TB-HIV situation in Brazil in graphically without venturing into logistic regression analysis for factors associated with TB-HIV relative to TB (HIV-ve). This shed light on non-reporting, non-testing of HIV among TB and vice versa. Social and demographic profile of patients ofTB, EPTB and TB+EPTB with or without HIV would be interesting considering serious data limitations for logistic regression analysis.

Authors present both epidemiological (social, demographic), clinical and also treatment outcomes. Hence title is not apt and authors should reconsider this.

Another problem with the analysis is extra-pulmonary TB is more common in HIV+ve, but authors both forms regression analysis for factors associated with TB-HIV relative to TB alone (HIV-ve).

Background does not provide a sound rationale about which gaps in research on factors associated with TB-HIV co-infections this study aims to fill. Two of the cited references are too outdated.

Overall, MS writing is very poor, riddled with many spellings and typos. It is hard for reviewer to highlight all those in the comments. Results are presented in great details with too many numbers for odds ratios and p-values. instead present significant ones succinctly. I have not commented on the discussion but will if I re-review a revised version.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

'I declare that I have no competing interests'