Reviewer's report

Title: Incidence rates of tuberculosis in chronic hepatitis C infected patients with or without interferon based therapy: a population-based cohort study in Taiwan

Version: 3
Date: 4 November 2014

Reviewer: Andrew Edmonds

Reviewer's report:

Major Compulsory Revisions
(which the author must respond to before a decision on publication can be reached)

1. I still think that the paragraph staring with line 184 should be clarified (even though you have attempted to explain in response 4b). You have defined the non-treated cohort as 7587 patients who did not receive IPT. Then, you say you randomly selected 2460 control subjects. First, you need to say that you selected the control subjects FROM the non-treated cohort. Second, the selection was NOT random even though you say it was – in fact, you have selected control subjects by matching (by age, sex, and year/month of visit) to the IBT treated cohort. This is not random. You say that the control subjects were matched with the non-treated cohort, even though the control subjects were matched (non-randomly) to the TREATED cohort. You should say that 95.98% of the IBT treated subjects were matched at a ratio of 1:4 with the non-treated cohort – not that 95.98% of the control subjects were matched at a ratio of 1:4 with the non-treated cohort.

2. You have added AHRs in the abstract but you still refer to incidence. Why? Why not say hazard? Also, it might be advisable to stay away from ascribing “significance” to your results merely based on the confidence interval (what I assume you are doing). There is no notion that the AHR of 2.81 has any meaning... you just say that it is not significant. One could argue that this is an effect that communicates both magnitude and directionality, but one that suffers from less than optimal precision (due to low numbers of events, e.g.). Instead it is just dismissed as not significant, and it is concluded that IBT is not a risk factor. What might you expect if you had a greater sample size? Are you really sure that IBT is not a risk factor, or might your data be limited in their ability to isolate its effects? I would at least qualify your main conclusion.

3. In response 14 (and in the paper), you say that you excluded patients who were less than 20 years old. Why not clarify that you took those who were 20 or older from the sample of one million (sampled from the 24 million), so the reader doesn’t think (as I did) that the sample of one million was drawn from the subset of the 24 million that was over 20 years old?

4. You have given the information requested in response 15, but why not add this
to the paper? It would help to clarify what you did – the reader might very well have the same questions that I did.

5. Looking at your response 16, I don’t think you understand my comment. My point is that you define HCV infection (line 153) as “individuals who had at least two service claims of ambulatory or inpatient care for the treatment of HCV between 2000 and 2008.” Then below you say “A total of 12,547 subjects with an HCV infection were identified.” So, the reader logically would think that you have found 12,547 individuals who had at least two service claims of ambulatory or inpatient care for the treatment of HCV between 2000 and 2008. But no – you have actually found 12,547 individuals WITH HEPATITIS C VIRUS and only subset of those had two service claims. So, do not define INFECTION as having two service claims. Can you clarify this?

6. Re: comment 17: You didn’t really address my questions in the response or in the paper. I still think it would be good to clarify what happened when TB was diagnosed AT THE SAME TIME as HCV was treated – were these cases removed? I still don’t know. Also I asked whether you were talking about HCV treatment rather than diagnosis, and you didn’t change anything in the paper (it still says diagnosis). But you are not talking about diagnosis with HCV; you are talking about treatment for HCV – right? I have the same comment on response 18.

7. Re: response 18b: Thanks for the explanation, but I think it would be good to explain in the paper (if possible) why these 77 were excluded; other readers may have the same question that I did.

8. Re: responses 20 and 21: It is still not clear to me how you are calling it random sampling of controls if you are matching by age, sex, and visit timing. You say “We randomly selected 2460 control subjects.” If you randomly selected from the 7587, how did they match on those characteristics? I really think this needs to be clarified; readers will have the same question. In the paper, you should also justify the decision to match based on the groups not being comparable, as you have in the response.

9. Re: response 22: I still think it is confusing to say that you allowed for at least a 1-year follow-up period even though individuals did not necessarily have one year of follow-up (due to death, TB diagnosis, drop out, etc.)

10. Re: response 23: OK, but can you at least acknowledge in the discussion that you’re not really looking at TB incidence? You’re looking at TB treatment which may be a proxy (or best possible approximation based on available data) of TB incidence.

11. Re: response 23: Thanks for giving the references in the response, but I think you should give the references in the paper to substantiate your statement that these factors need to be accounted for.

12. Re: response 26: OK, so if you’re not talking about DETECTION, don’t say detection. Say “The end point of follow-up in the subjects developing was the
date of 1) having taken two anti-TB medications for more than 90 days, and 2) having a TB-specific ICD-9 code. Do you see why I am confused about you calling it DETECTION? That implies identification of the organism. Re: “coded by the date of their last visit” – this must be a language issue. I think maybe you should just say “those lost to follow-up were censored on the date of last visit” if this is the case.

13. Re: response 27: I still think it would be nice to say how covariates were CODED (e.g., were all binary?)

14. Re: response 35c: Why not add this explanation to the paper? Also, my main point is that you (arguably) used a stringent definition of TB (i.e., treatment with two drugs). I was hoping you could speculate in the discussion on how your results may have been impacted if you would have used a different definition if the data were available (i.e. identification of the organism, a clearer reflection of incidence). This is also my point in comment 42.

Minor Essential Revisions
(such as missing labels on figures or the wrong use of a term which the author can be trusted to correct)

1. Thank you for adding the sentence to the abstract. But it does not make sense to say “to estimate the hazard ratio OF RATE OF active TB.” Just say “Cox proportional hazards models were used to estimate hazard ratios (HR) for active TB, and associated confidence intervals (CIs), comparing IBT to no IBT.” Or write something similar to that.

2. Line 234: Say hazard not risk.

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

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests