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: 2  
Date: 12 September 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. Line 19: According to the title, the paper is about TB incidence rates in HCV-infected patients with or without IBT. Shouldn’t the running head be reflect the fact that you’re not just looking at patients *ON* IBT?

2. Line 40-41: Do you mean in patients in general, or amongst HCV-infected patients? Clarify. Also, isn’t it debated whether IBT increases/decreases the rate of TB, not only whether it increases the rate?

3. Lines 45-51: The methodology is not clear. Did you take a random sample of the 1,000,000 and find 8,321 HCV-infected people in that sample? If so, how large was the random sample? Or, did you just find 8,321 in the sample of 1,000,000?

4. Line 49: Can you clarify whether it is the case that the 621 subjects received IBT at time of HCV diagnosis, and whether active TB was excluded in these 621? Also, how did you get 2460? Is it four matched subjects for each IBT subject? If so, why did you not match 2,484?

5. Line 57: Suggest using incidence rate ratios and 95% confidence intervals rather than just p-values. This would need to be done in the main paper as well, obviously.


7. Line 61: Don’t you mean “rate” rather than “risk”?

8. Line 62: You’re not taking about TB infection… you’re talking about TB disease, right? I think you don’t have to say reactivation.

9. Line 64: Why active TB? If the concern is that IBT causes TB incidence, why would you be concerned if someone already has active TB? Is it because IBT would exacerbate the TB disease already ongoing? If so, I would explain this more in the main paper.

10. Line 63-65: This conclusion seems contradictory to your finding (IBT doesn’t cause TB disease, but patients should be evaluated for latent TB before taking IBT) – why? Something else needs to be explained in the abstract if you keep this conclusion there, and if you make this recommendation in the main paper, it...
needs to be explained.

11. Line 133: OK, now I see what you did. Your sample of one million is from a larger database of 24 million. Please clarify this in the abstract.

12. Lines 133-139: It would be helpful to provide more details about how you selected the one million people for your dataset. Did each of the 24 million people who had at least one day in the system have an equal probability of being selected for the sample population? I suppose so. If this is the case, then people with a longer duration of enrollment would have a higher probability of having HCV, IBT, and TB...

13. Line 151: How can this be a case-control design when you selected “controls” not based on their disease status, but rather based on their exposure status?

14. Line 153: So, it wasn’t a random sample of the 24 million. It was a random sample of the subset of the 24 that was over 20? (21 and older?)

15. Lines 154-155: Why at least two service claims? And why for treatment only? Why did you not call people HCV-infected based on laboratory diagnostic criteria?

16. Line 159: I don’t think you can say that 12,547 subjects with HCV infection were identified if your definition is treated at least two times for HCV.

17. Lines 160-164: So it was a random sample. But then you threw out some cases based on certain criteria? Why 8,312 when the abstract says 8,321? Why did the history of TB diagnosis have to come before the first HCV diagnosis; what if it was at the same time? Also, you’re not talking about diagnosis; you’re talking about treatment for HCV, right?

18. Lines 178-180: You’re talking about HCV treatment, not HCV diagnosis. Please clarify. Also, why exclude these 77? Give a justification. Need to add “for” after “those patients who had received IBT,” and add a space before “(n=621).”

19. Line 182: I thought it was ambulatory or inpatient?

20. Lines 186-187: Did they have to match exactly by age? How close did they have to match? Also, what was the index ambulatory (inpatient?) care visit for a “control”? Same as for the HCV treatment group?

21. Lines 183-189: This is a confusing. So you took a random sample of one million from 24 million, identified HCV cases from the one million (even though “HCV diagnosed” really means treated at least twice for HCV), split these cases into IBT and no IBT. There were 621 in the IBT group, and 7587 in the no IBT group. But rather than just looking at TB incidence in the no IBT group and comparing this to the TB incidence in the IBT group, you matched those in the IBT group to those from the no IBT group based on age, sex, year/month (1:4 ratio) and then looked at incidence in the two groups? Is this correct? Why not just compare incidence in the IBT and no IBT groups, particularly if the groups don’t differ by age, gender, and calendar time, and forget the matching altogether? How can you call it “randomly selected” if you’re matching by those three characteristics? If you randomly select 2,460 from 7,587, how would they match on the characteristics?
22. Lines 193-196: There wasn’t at least one year of follow-up if TB was diagnosed in the first year, right? What about competing risks (death, etc.)? There were more reasons for withdrawal than loss to follow-up, assuredly.

23. Lines 199-204: Why does treatment with TB figure into your definition of TB incidence? This is the same problem I’m having with needing to be treated for HCV in order to be called diagnosed with HCV. If you are looking at TB incidence, why does it matter if the person was treated for TB or how long they were treated for? Isn’t the most important thing whether they were diagnosed clinically or microbiologically?

24. Lines 207-220: It would be helpful to say here what you do with these factors. For example, what’s the point of clarifying the etiology and severity of liver disease? How did you decide on what factors to include?

25. Lines 223-237: Now there’s a time-to-event analysis? Where was this in the abstract?

26. Line 228: I thought the definition of TB was not detection (by culture/microscopy) but rather an ICD-9 code plus two anti-TB medications for more than 90 days. What does “were coded by the date of their last visit” mean? Do you just mean that those lost to follow-up were censored on the date of last visit?

27. Line 234: Do you mean “in the proportional hazard model, we adjusted for…”? How were these covariates coded? This would be good to specify in the preceding paragraph. Wasn’t IBT also included in the model? Did you estimate hazard ratios and 95% confidence intervals? If so, specify this.

28. Lines 241-251: I would recommend just giving all of these numbers here, rather than also in the Methods section. It is hard to reconcile the numbers in so many different places… why does it matter that there were 8,312 after exclusions in the Methods section (is this the same as the 8,321 in the abstract?) when that number is not here? The reader, like I did, will try to match the number he/she read earlier, and not find it… so just give the important numbers here and in the figure, and leave the rest out. Why does it say 7,586 here but 7,587 in the Methods? Edit carefully! Also, why say that 699 were enrolled for analysis when the actual number after an additional exclusion was 621? This whole paragraph seems like a repeat of the Methods section. Why not talk a little bit about the how the two groups compared in Table 1? Also, I would not call them “matched controls” – “control” implies disease status, not exposure status.


30. Line 261: I don’t think you defined “long-term” follow-up in the Methods section, nor did you say how these p-values were calculated. What do they represent? Again, why not give effect measures and 95% confidence intervals? Now that I look at Table 1, these p-values seem just to compare proportions in the different groups who got TB? Are they indeed not a comparison of the incidence rates, even though you attribute the p-values to the incidences? This all needs to be clarified.
31. Lines 262-263: The figure is labeled “cumulative incidence” but this is not what it is... if you used 1 minus Kaplan Meier to approximate cumulative incidence, say this. Cumulative incidence implies that you took into account competing risks, which you did not.

32. Lines 264-266: Why do you keep giving the definition of the IBT treated cohort (more than two months IBT)? Once was enough. The number who developed TB should be moved up to the part with the proportions and the p-values (line 261)... as written, it seems like this is something new, whereas in actuality it is more of the same. Why talk about the incidence of TB in a group that you excluded (n=77) even if it was for an unjustified reason?

33. Lines 269-271: Why not report the adjusted hazard ratios and 95% confidence intervals for your primary exposure (IBT)? Again, what is “modified”?

34. Line 274: Now you’re not calling it a case-control study. I would make the necessary correction.

35. Line 274: Hard to say it’s not associated when you don’t state the hazard ratio here or in the Results. Looking at the table, there is an indication that IBT is associated with TB in the short-term model, although the measure is imprecise. One wonders what would happen if you didn’t use such a stringent definition for the outcome (TB)?

36. Line 278: Again, you’re not comparing incidence rates, you’re comparing the proportions of the different groups that developed TB.

37. Line 287: It’s not just reactivation; in addition to latent TB developing into active TB, it could be primary infection leading to active TB.

38. Lines 288-295: What does HDV have to do with your study? Consider not including this report, and if you do, define before using abbreviation. Do you mean one case report, or one case? Or, do you mean that there is one case per case report? This paragraph describing isolated case reports does not seem to be placed in the context of your study... we don’t even know the details of the TB cases in your study population. Are there no population-level data whatsoever on TB incidence post-IBT? I don’t find these case reports to be informative, and would their removal from the text (and also remove Table 4).

39. Line 303: If you’re going to talk about a hepatic encephalopathy finding, at least give the hazard ratio in the Results section (or here).

40. Line 305: Did you look at dose-response? I don’t see anything about this.

41. Lines 315-329: How does your study support any of these recommendations? For example... you looked in a HCV/TB endemic area and appeared to find nothing of consequence, so why would you say that close monitoring is required in Turkey, Eastern Europe, etc.?

42. Lines 330-334: So you didn’t have any HCV or TB lab results? Or did you just not have viral genotype, viral load, and CD4. No wonder you had to use those exposure/outcome definitions. Even so, they were strict – can you discuss your choice of definitions? I don’t understand “a substantial proportion of patients exposed to corticosteroids would be underestimated.”
43. Line 351: What informs this recommendation?

44. In general, I am concerned about the selection of the “control” population (why?), the way incidences were compared (and the analysis overall), the exposure and outcome definitions…

45. Figure 2: How can this be both for one-year and long-term follow-up (as the legend says)... there are only two curves (treated and untreated) in the figure; is this figure supposed to represent either one-year or long-term (rather than both)?

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. Line 46: I recommend writing out the number instead of depending on the reader to make the calculation of 106.
2. Line 46: Say “of” not “from.” And, say approximately 1,000,000 (as there were not exactly 1,000,000).
3. Line 55: Need space before “per” (typo).
4. Line 59: Say “a risk factor” not “the risk factor.”
5. Line 61: Why not use the abbreviations IBT and TB?
6. Line 63: Say “is” not “was.”
7. Line 64 and 65: Why not use the abbreviation TB? Don’t give the abbreviation if you’re only going to use it sometimes.
8. Lines 78-79: References are needed to substantiate these statements.
9. Line 81: Why “therefore”? It would make sense that effective anti-HCV therapies are needed because HCV can lead to cancer and cirrhosis, not because many people are chronically and newly infected. Change the order of your sentences.
10. Line 87: In general, I would recommend including references when you make biological / “factual” statements like these.
11. Lines 90-91: In general, always spell out words before using abbreviations (e.g., TB, BCG).
12. Line 94: And then, use the abbreviation (e.g. TB) rather than spelling out the word in full.
13. Line 98: Now you’re defining BCG? Why not do this the first time you mention it? Also, what is ecomplex? Or is this a typo?
14. Line 100: Is “Th1type” a word?
16. Line 107: Say “are considered.”
17. Line 109: I think you can give the TB abbreviation sooner and also use the HCV abbreviation here.
18. Line 111: Do you mean “a difficulty” rather than “in difficulty”? 
19. Line 114: Do you mean “in different geographic areas”?
20. Lines 119-122: I think you mean to say that this study used the Longitudinal Health Insurance Database 2000, rather than assembling the database for the purpose of this study. Also, say “a risk factor” rather than “the risk factor,” insert a comma after “population,” and make sure to put a period at the end of the sentence.
21. Line 130: NHI isn’t just a registry, correct? Isn’t it a program or system?
22. Line 178: Need space after “excluded.”
23. Line 184: Need space after “cohort.”
24. Line 189: Need space after “excluded.”
25. Line 193, 196, 204, 275, 293, 294, 311, 325, etc.: Why not use “TB” abbreviation? Edit carefully!
26. Line 204: Need space after “90.”
27. Line 209: Need space after “silicosis.”
28. Line 211: Need space after “ICD-9.”
29. Line 212: Need space after “malignancy” and before “because.”
30. Line 217: Need space after “cirrhosis.”
31. Line 224. Need space after “(USA).” Distribution should be plural.
32. Line 227: Now endpoint is two words? Be consistent.
33. Lines 258-259: Say “the cumulative incidences were” and add a space after 0.151.
34. Line 283: What does anti-TNF therapy have to do with this study?
35: Line 291: Why give the abbreviation for HAART?
36. Line 309: You already gave the abbreviation for HCV.
37. Lines 315-322: You can’t start a complete sentence with 1) or 2) or 3) or 4).
38. Line 317: “May be” should be two words, not one.
39: Line 350: Use IBT abbreviation and insert a space after “active.”
40. Table 1: Make sure spaces are present where they should be. Why is “n(%)” not next to the “Outcome” title, like is for the other variables?

Discretionary Revisions

(which are recommendations for improvement but which the author can choose to ignore)

While it is arguable that several of the comments in the “Minor Essential Revisions” section are actually “Discretionary Revisions,” I would argue that all of the suggestions be strongly considered in order to strengthen the paper. Therefore, there are no comments in “Discretionary Revisions.”
Level of interest: An article whose findings are important to those with closely related research interests

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