Reviewer's report

**Title:** Direct economic burden of hepatitis B virus related diseases: Evidences from Shandong, China

**Version:** 5  **Date:** 14 December 2012

**Reviewer:** Fujie Xu

**Reviewer's report:**

I was curious about the Major Epidemic Network database (is it the National Notifiable Diseases Surveillance System) and the universe the case patients were sampled from.

When the patients were interviewed, during hospitalization or right after discharge? If during hospitalization, do the patients have to forecast the cost of their current hospital stay?

Some data are relatively unimportant. For Example, given the data about income we already have, why should we care about occupation? The authors provided the numbers: 19.91% were laborers, 31.77% were farmers7.94% were civil servants, 9.06% were professionals, 7.49% worked in the service industry, 10.85% were retired, 4.81% were unemployed, and 2.91% were students/children—probably can delete these.

“Household monthly income was $437.70 for the overall sample; $439.17 for the insured and $410.84 for the uninsured group. Individual monthly income was $189.72 for the overall sample; $189.66 for the insured and $190.81 for the uninsured.”—round costs and income to dollar (remove the two decimals).

Table 3.

Proportions of direct medical cost( %) are consistently high: The numbers in the first 3 columns ranged from 91%-97% and I don’t think the authors have to show the specific numbers. These columns can be deleted from the table and a sentence about the 91-97% in the Results section should be sufficient.

Table 4:

Why the Chi-square test? The cost is a continuous the differences between groups should be t-test when the sample size is small.

The standard errors are all large. This is not surprising but reflects the unreliable nature of the mean estimates presented in the table. When the relative standard error (defined as sd/mean in this table) is greater than 30%, in statistics we determine the estimates as unreliable. Typically for cost and income data, the preferred summary statistic is the median, not the mean and sd. The authors can present both mean and median if so desired.

The most important data in the table is to show where the costs come from, but the numbers are too many and the table is unnecessarily busy—please simplify. Options to simplify the table include: 1) remove the decimals—keep in mind all
the estimates are not precise estimates and you don’t need precise numbers to the penny level (decimal); 2) the differences between the insurance and uninsured groups are relatively small even when the p value is significant (statistically significant but no significance in real life): can delete uninsured or combine the two groups. A typical reader may have only 30 seconds to read your table, please help them to see the main picture!

Table 5. As Table 4, an unnecessarily busy table. Because over 90% of the total costs are medical, the non-med costs are not at all the cost drivers in the overall picture. The details about transportation and in the insured and uninsured groups are like splitting hair and it can be a distraction. You can reduce the table to a couple of sentences in the text. All the cost numbers are estimates—precise accuracy is not possible and not needed.

Table 7: I think the data in this Table are among the most important findings of the study. Unfortunately the paragraph (pages 10-11) right before Table 7 should be completely written to introduce this important analysis (instead of the discussion about remove the uninsured group.)

Table 8 and Table 9: The costs by disease category are quite difference from the hospital source and from patient self-report. Why? What are the differences telling about the true costs, are they over-estimates or over-reporting? The authors presented the cost numbers but failed to discuss how the numbers are related to one another. I am confused about what these two tables are adding to the study.

Table 10: may be combined with 4 or replace Table 4.

Level of interest: An article of importance in its field

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

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

No to all
I declare that I have no competing interests