Author's response to reviews

Title: Cancer burden in China: a Bayesian approach

Authors:

Wanqing Chen (chenwq@cicams.ac.cn)
Bruce K Armstrong (bruce.armstrong@sydney.edu.au)
Rongshou Zheng (zhengrongshou@yahoo.com.cn)
Siwei Zhang (zhangsw@cicams.ac.cn)
Xueqin Yu (xuequiny@nswcc.org.au)
Mark Clements (mark.clements@ki.se)

Version: 3 Date: 17 September 2013

Author's response to reviews: see over
13 September 2013

Dr. Paige Bracci
Editor
BMC Cancer

Dear Dr Bracci,

RE: MS: 8806350928807230

Thank you for providing the opportunity for us to revise the manuscript. We thank the reviewers for their constructive comments, which we believe have led to significant improvements in the manuscript.

We were advised that we should prepare a revised manuscript by incorporating the reviewers’ comments and re-submit it for consideration for publication. We have taken this advice and revised the manuscript accordingly. We have addressed each of the specific comments and incorporated them into the new version as we thought necessary. We include a document outlining our responses to the comments and the changes we have made, point-by-point.

Thank you for considering this article for publication. We hope that the revisions are to your satisfaction and look forward to hearing about your decision.

Yours sincerely

Wanqing Chen
Ramon Clèries’ report
Major compulsory revisions:
1) Although Figures 1 and 2 shows MI ratios and age standardized rates by age group, do authors include childhood cancers in the fitting of the models? If so, how were smoothed the MI for these age groups? Therefore, which age groups were included in the analysis?

Response: We included all age groups in our analysis. We have now inserted the following into the Methods: “Age was divided into 19 subgroups, including 0 and 1-4 years, five year age groups from 5-9 years to 80-84 years, and 85 years or older.” All age groups were modelled within the same model using the same smoothing technique. We have also moved most of the models details from the additional file into the main text to improve clarity.

2) Following 1), although the approach to smooth model parameters is based on splines and it differs to that of Baker & Bray 2005 and Cleries et al 2013, these authors note that the use of younger age groups affect parameter estimates when these are smoothed. In this line, the use of younger age groups in the spline modeling may affect the performance of the model? Did authors try to obtain estimates of cancer incidence excluding these age groups?

Response: We thank the reviewer for this thoughtful suggestion and performed a sensitivity analysis by excluding ages less than 20 years. Under the Methods, we inserted the following paragraph: “Fifth, we examined whether the estimates were sensitive to the exclusion of ages less than 20 years. Re-fitting the models with this restriction, we found that the estimated number of cases did not vary significantly between models that excluded ages less 20 years compared with estimates from models that included all ages. For all sites, the relative differences were less than plus or minus 3%.”

3) Authors provide overall MI ratios (Table 1). However, when inspecting age-specific MI ratios in Figure 1, they should note in the discussion that MI ratios are closer or higher than 1 for certain cancer sites in certain age-groups or depending of rural-urban area (see Table 1 and discussion in page 14). In the case of liver, for example, this is an indicator of high lethality or poor diagnosis (See Bosch et al 2004). Do authors have information about differences in these MI ratios between Chinese cancer registries?

Response: Yes, we have data for all individual registries. An additional table shows the variation in overall MI ratios in 32 registries. MI ratios were higher in rural areas than in urban areas and MI ratios were close to or higher than 1 in some cancers with poorer prognosis, such as liver cancer, pancreas cancer, especially in age groups older than 70 (Table 1). We have added some text to this effect in Page 17, paragraph 2. The additional table X2 is attached in the Appendix.

4a) Globocan estimates of cancer incidence in China for 2008 (last accessed April 20th, 2013) was N= 2817210. The number of cancer cases of cancer incidence provided by the authors are 2,96 million for 2005, whereas Ren et al provided 2,58 million. Authors should mention these differences.

Response: The two other estimates were described in the Introduction. We have now moved them to the Discussion and suggest possible reasons for the differences.
Responses to reviewer's comment

4b) Since incidence estimates are based on data from cancer registries covering 5% of Chinese population, and these data belong to regions located in eastern China which is more developed economically than the west, authors should provide evidence about differences between these eastern and other Chinese regions in terms of cancer burden. In this line, could authors show a comparison of cancer mortality between eastern and other regions (Supplementary table)? If they find huge differences in certain cancer sites, they might assess differences in cancer incidence between areas and how it affects global estimates for China.

Response: Mortality has been known to be highly variable across China since the publication of maps of cancer mortality in China based on the first nationwide cancer mortality survey conducted in 1973-75. This variation is complex and depends on cancer site. The main issue is whether the 5% sample of registries will lead to biased estimates of the MI ratios: relative levels of incidence or mortality by region will not affect the national estimates of burden. The number of registries from the west was small in number and often selected for surveillance of a particular cancer, so their interpretation may be difficult. MI ratios by cancer registry are now shown in additional Table X4 and we have inserted several sentences into the Discussion on page 15, paragraph 1.

5) Discussion: Authors should mention that when incidence and mortality data is available in certain regions and these data will be used to estimate incidence in other regions, Bayesian modeling of the difference between incidence and mortality rates might be better than MI when estimating cancer incidence for certain cancer sites with low incidence (Clèries et al 2012).

Response: Thank you for this excellent suggestion. We have now added a paragraph to the Discussion on other possible prediction models, including the model by Clèries and colleagues.

6) Discussion: It would be interesting to assess if recent cancer incidence is rising or decreasing for certain cancer sites based on aggregated data from the registries. This information might be incorporated in the discussion.

Response: We agree that an analysis of trends is a critical piece of evidence for cancer control. However, this would be a very substantial, additional piece of work that, we believe, is outside the scope of the current paper.

Supplementary file:
S1) When assessing the burden of Cancer, it would be valuable to provide estimates (number of cases) for selected age-groups based on the shapes of the age-specific incidence rates (Figure 2): as example younger than 35, 35-64 and older than 65. This could be provided as a Supplementary table.

Response: We have now included such results (age-specific incident cases and rates) as a supplementary file (additional table X3) as suggested.

S2) Authors should provide the WinBUGS code used to carry out the estimation process.

Response: The R/WinBUGS code for the model specification has been added to the additional file.
Responses to reviewer's comment

Minor Essential Revisions:
Page 7 last paragraph:
“it got…” please check this paragraph, capital letters might be required or some part of the paragraph is missing.

Response: We have amended the paragraph. We have also made additional changes to the text to improve the language.

Ming Wu’s report
Minor Essential Revisions
1. The title “Cancer burden in China” is too broad, as incidence and mortality are only parts of the disease burden.

Response: A complete description of cancer burden would require incidence, survival, mortality, prevalence and disability life years. Many articles have used the term ‘cancer burden’ in their title when cancer incidence and mortality were the only components of the burden reported, “Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010” and “The burden of prostate cancer in Asian nations. J Carcinogenesis 2012”. In line with these articles, we believe our title is appropriate.

2. Page 7, the end of Statistical Analysis part, should be “ We got……” or “ It got……”

Response: We have made the necessary changes.

Discretionary Revisions
1. It would be useful for readers if the Bayesian version of Poisson regression model is presented with more details in the manuscript, currently the description on the model looked like traditional method.

Response: We have now moved the technical description of the model from the additional file into the main manuscript.

2. In supplement file, details on the priors choosing on model parameters and iterative calculation were provided. For example, for younger age group, authors chose uniform priors between -4 and 4 for fixed effect parameters. For MCMC sampling, 25,000 sampling for burn-in and 1,000 samples from 25,000 sampling for calculation of CI were employed. It should be necessary to provide the reason for the settings-up or refer to the references related.

Response: The sampling properties based on graphical plots indicated adequate mixing for most parameters after a burn-in of approximately 5,000 and sampling for another 5,000. The choice of 25,000 for the burn-in and 25,000 for sampling was conservative to ensure good mixing for all of the sites. We have inserted this explanation into the manuscript.

Mohamed Amin Pourhoseingholi’s report
Major points:
The source of errors for rural areas:
According to the information provided in the method of manuscript, it seems that the data of mortality and incidence in rural are not accurate as same as urban; for example in these areas the verbal autopsies were used to ascertain the cause of death. The results for rural also indicated upper MI ratio compare to Urbana’s and the larger uncertainty which would be due
Responses to reviewer's comment

to the quality of rural data. Did the authors consider these sources of errors in their Bayesian model to avoid the underestimation?

Response: Since the models were fitted separately for urban and rural areas, the modelled variation in MI ratios between cancer registries was allowed to differ in urban and rural areas. However, this would not eliminate any underestimation in the rural areas. As described in the next response, we have now reported data quality indicators for each registry.

Misclassification:
It is known that misclassification in death registry and also incidence registry is a source of errors for developing countries. These errors could be one reason of underestimation and variation of burden among the type of cancers. Did the authors consider misclassification in the step of data preparing or analysis?

Response: We have now inserted several sentences in the Methods describing the quality assurance, and attached a table including indicators of data quality in the Appendix (additional Table X4).

Comparing criteria:
The authors mentioned that their priority was to improve on previous estimates of cancer incidence in China using the more updated national mortality data and regional cancer registries data with some technical refinements in estimation methods (Bayesian model). Despite their sensitivity analysis, it seems that there is no certain criterion to understand that this technique would be accurately estimates the burden of cancers. Calculating the more cases rather than other methods or registry dose not means the accuracy of model. I believe that at least a simulation study should be conducted to compare Bayesian models to other approaches (example Jensen et al. (1990)’s model).

Response: We thank the reviewer for this insightful comment. We have inserted text in the Discussion about potential weaknesses of the Globocan and Ren et al sets of estimates (page 13, paragraph 2 and page 14, paragraph 1 in the discussion). We have also inserted into the additional file a brief statistical development and simulation comparing our model with the model proposed by Jensen et al. (additional file 1). Finally, we noted in the Discussion that “Following a reviewer’s suggestion, future work could use simulation to compare several methods for burden calculation. This would allow the simulation of cancer incidence and mortality under different missingness mechanisms, or for bias in the selection of cancer registries, and then assess the degree of bias from the different methods relative to the hypothesized “truth”.

Generalization of the results:
The methodology of this study could be beneficial for those countries where have not complete and accurate registry system. But for the results of this manuscript (in which the title is Cancer burden in China) I’m still in doubts that the data would represent the total of country. According to the authors, these 32 cancer registries on which the estimated national cancer incidence is based, are predominately in the more economically developed areas of east China. So the estimation seems more accurate for this part of country, not for all.

Response: This is related to question 4b from Ramon Clèries. The main issue is whether the 5% sample of registries will lead to biased estimates of the MI ratios. As more cancer registries
become available, then a better characterisation of how survival varies across the nation will be possible and we will be able to better estimate the cancer burden in China. Our main purpose here was to use the best available data with the best available methods to make reasonable estimates for cancer burden in China in 2005. These estimates can certainly be improved as better data become available.