Author's response to reviews

Title: Geographical spread of gastric cancer incidence in the Caspian Sea region of Iran: spatial analysis of cancer registry data

Authors:

Mohammadreza Mohebbi (mrmmohebbi@yahoo.com)
Mahmooh Mahmoodi (mahmoodim@tums.ac.ir)
Rory Wolfe (Rory.Wolfe@med.monash.edu.au)
Keramat Nourijelyani (nouri4@yahoo.com)
Kazem Mohammad (mohamadk@tums.ac.ir)
Hojjat Zeraati (zeraatih@tums.ac.ir)
Akbar Fotouhi (afotouhi@tums.ac.ir)

Version: 2 Date: 11 February 2008

Author's response to reviews: see over
Answers to Farin Kamangar's review

1. The authors have examined geographic distributions of all gastrointestinal tract cancers not just gastric (stomach) cancer. Therefore, they should change "gastric" to "gastrointestinal tract" in the title, abstract, and other parts of the manuscript.

The mistaken expression was corrected.

2. In the Discussion Section, under after the title "liver cancer", please change "Hepatitis" to "Viral hepatitis".

The sentence was changed.

3. There are a number (very few) of typographical errors in the text. For example, in page 3 (Background), the correct spelling is IARC, not IRAC. Likewise, in that same page, age-adjusted is correct, not aged adjusted. In the last page of results, "indicted" should be replaced with "indicated". I suggest that the authors examine the text once more for typographical errors.

The text was revised for typographical errors.

4. In the Discussion section, under "stomach", the authors have stated that "..which may be an indication of common environmental risk factors for these two (esophageal and gastric) cancers." Indeed, there several environmental risk factors shared between esophageal and gastric cancers. Examples include smoking, low socio-economic status, low fruit and vegetable intake, and gastric atrophy. The authors could use there references: PMID: 14996860, PMID: 17198458, PMID: 1458460, PMID: 8064893, and PMID: 9259392.

The discussion was revised according to the reviewer's guidelines.

Answers to Martin Kulldorff's review

1. When using Moran's I on aggregated data, there is a bias if each incidence rate is treated as an observation rather than the individual counts. This is because the variance of the incidence rate is depends on the population size, and since the population size is different in the different wards the rates are not identically distributed as required by Moran's I. This bias has been described and explained in an important paper by CC McLaughlin and FP Boscoe: Effects of randomization methods on statistical inference in disease cluster detection, Health and Place, 2007 Mar;13(1):152-63. When using Moran's I, it is important to avoid bias by doing it in the way proposed by McLaughlin and Boscoe. In this paper, it is unclear how the method was applied in this new appropriate way or whether the
old standard approach was taken. If the former, the McLaughlin-Boscoe paper should be cited. If the latter, the calculations needs to be redone.

We used modified indices.

The original Moran's I index is as follows:

\[
I = \left( \frac{1}{s^2} \right) \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} (y_i - \bar{y})(y_j - \bar{y}) \tag{1}
\]

where \( \omega \) is spatial proximity matrix, and

\[
s^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{y})^2.
\]

We adjusted Moran's I for regional counts by comparing the observed count in each region with its expectation under the constant risk hypothesis, rather than comparing the count to the overall mean count. [1] The revised index is as follows:

\[
I_{cr} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} \left( \frac{y_i - r_{ni}}{\sqrt{r_{ni}}} \frac{y_j - r_{nj}}{\sqrt{r_{nj}}} \right)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij}} \tag{2}
\]
where $n_i$ denotes the population size for region I, and $r$ denotes the overall disease incidence rate specified a priori or estimated by the total number of cases observed divided by the total number of people at risk. Note that the scaling factor $s^2$ is replaced with the product of the region-specific (Poisson) standard deviations, to emphasize variation around each regional expectation rather than around an overall mean count.

We have added the relevant notes and references in the 4th paragraph of the discussion.

2. I am not as familiar with Geary's C. Does this method have the same problem as the one that McLaughlin and Boscoe describe for Moran's I? If so, the authors need to deal with that as well.

Geary's C is defined as

$$C = \frac{N - 1}{2\sum_{i=1}^{N} (y_i - \bar{y})^2} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} (y_i - y_j)^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij}}$$

We also adjusted Geary's C for counts in a manner similar to the modification of Moran's I denoted $I_{cr}$ and defined in equation 2. The key is to replace each regional count $y_i$ by a standardized value (thereby adjusting values $y_i$ and $y_j$ to have comparable variance) and to remove the overall measure of variation around the mean regional count. [2]

We have added the relevant notes and references in the 4th paragraph of the discussion.
3. **P7L1-3:** This type of imputation is appropriate, as the amount of missing data is very small. The authors may want to point out though, that if there is a small bias, then it will be in the conservative direction towards accepting the null hypothesis.

A paragraph was added to the 3rd paragraph of the discussion to address the potential bias of the imputation method.

4. **P7L13:** cancer

The sentence was corrected.

5. **P8L4:** Cite original references to these two methods.

The original references were added.

6. **P9L16:** Why oval rather than the more compact and natural choice of circular? I would have used the latter. Was the elliptic SaTScan option used? Or, did you use one specific oval shape by transforming either the x or y axis?

Since the overall shape of the study region was similar to a crescent rather than a square, we thought oval windows might be more appropriate. However since circular windows were used in the majority of published studies we realize the relevance of this method in order to have comparable results. The calculations were redone by circular radii and the results were very similar to using oval windows. We report the new findings based on circular radii.

7. **P9L17-18:** . . . radii ranging from zero to a user defined . . .

The sentence was corrected.

8. **P9L19:** SaTScan

The sentence was corrected.

9. **P9L20:** the log likelihood ratio was calculated and the p-values obtained by a Monte Carlo simulation procedure.

The sentence was corrected.
10. P10: It would be interesting to be able to compare the incidence rate in the Mazandaran/Golestan region with other parts of the world. For people to be able to do that, using past or future research studies, it would be great if for each cancer site; the authors would provide directly standardized incidence rates using the 1970 and 2000 standard world population.

The suggested directly standardized incidence rates were calculated using the 1970 (Segi’s World population) [3, 4] and 2000 (WHO World Population) [5] standard world population. The methodology for this is now described in the 8th paragraph of the methods section, and the findings are presented in Table 2.

11. P11: The maps are the most interesting results, and it would be very nice if there was a more complete set provided in the paper. I would suggest seven figures, one for each cancer site, where each figure has six maps in three rows and two columns. The three rows would be for (i) women only, (ii) men only and (iii) women and men combined. The left column would show the incidence rates in the same colors as in the current maps, with some type of shading (e.g. //////) or boundaries for the statistically significant clusters detected by the spatial scan statistic. The right column would show the smoothed relative risks.

Figures changed according to the comments. We have elected to not present figures for the remaining cancer sites because of the small numbers of cases and consequent relatively large error bounds.

12. Table 2-3: Rather than stars specify the calculated p-values. That is more informative.

P-values were added to Table 2-3.

13. Table 2-3: May help the reader to add a column clearly specifying which cancer sites are clustered (esophagus and stomach), which are dispersed (colorectal), which are inconclusive or random (the remaining four), or something to that effect.

A column describing type of the spatial pattern was added to Table 2.

14. Figure 3: A set of smoothed maps could replace this figure.

The Figure was deleted.
15. Figures 4-9: Would be nice to add the location of the spatial scan statistics to these maps, through some form of shading or thicker boundaries.

Figures changed according to the comments.

References


