Reviewer's report

Title: Analysis of the EGFR, HER2, and TOP2A gene status and chromosomal aneusomy in gastric adenocarcinoma from Chinese patients

Version: 1 Date: 2 June 2008

Reviewer: Wei-Ting Hwang

Reviewer's report:

This is a statistical review.

- Major Compulsory Revisions

1. It is not clear what a Spearman test is involved. Most of the comparisons performed in the report are between two binary measures (e.g., overexpression versus not, disomy versus aneusomy), which can be done by Pearson’s Chi-square test or Fisher’s exact test (when small cell count exists). Meanwhile, it is also useful to provide a measure of association besides the results from a significance test. No spearman correlations or other measure of associations were reported. Because all the measures are often binary, or only have a few categories (<5), other measure of associations such as odds ratio would be more appropriate than a spearman correlation.

2. There are a lot of statistical comparisons done in this report. The issue of multiple comparisons should be noted and adjusted.

3. It is not clear why they wish to establish cut point(s) for NI and why only 5%, 10%, and 25% were explored. Can a continuous NI be also informative? A better and common method to establish cut points for these data would be using Receiver Operating Characteristic (ROC) curve analysis. Table 5 needs a better format and less information.

4. Figure 4 can be better presented with a histogram, which is able to show many characteristics of a distribution (e.g., central location, spread, and range). Those descriptive statistics should be reported. The bin for the histogram should be properly selected to capture the variability of the distribution.

- Minor Essential Revisions

The reporting of the results is general difficult to follow and sometimes confusing. Both the texts and tables for results require improvement.

1. It is not clear by the first sentence in the statistical analysis section when they wish to compare ECFR overexpression what do they consider as positive and negative tumors.

2. Table 1, 2, 4, 5: It is confusing that the table lists all the categories but p-value were come from the statistical tests based on combined categories (e.g., 0, 1+,
2+, 3+, versus overexpression (2+, 3+) versus not (0, 1+)). And different ways of combining categories gave different p-values. Alternative presentation of the table/results should be explored. For example, one can present the combined categories with associated p-values in a table, or present all categories in table, but in the text, present p-values along with the N or percentage of the combined categories.

3. Footnotes for Table 1, 2, 4: the word “difference” is not specific and could be misleading. Depending what statistical tests are used, one can replace it as “association”, “independence”, or “correlation”. The two footnotes for Table 2 are duplicated.

4. Table 3, the p-value would never be 0. Replace 0.0000 with a small number or express it as <0.001 or <0.0001.

5. More explanation on samples and the sampling process is needed. How many patients were eligible to be selected for the period of 2000-2005 for this study? How representative of the patient population seen at the Department of Surgery in the Peking Union Medical College Hospital as compare to the Chinese population that seek their treatment in other parts of country? Is there evidence that the random selection has produced a sample as expected?

6. How the sample size of 100 was determined? What is the minimum association (e.g., odds ratio, correlation) can be detected with 100 subjects for 80% or 90% power? This information can inform the readers that a small or significant p-value is not simply a consequence of a large sample size.

- Discretionary Revisions

1. For all tables, it should be simple and clear enough to use “Total” in replace of “Total number”. In addition to column total, row total can be listed as well.

2. Move up the result for the correlation with clinicopathologic variables, section 3.9, as this was the first analysis the authors proposed to perform.

**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.