Reviewer's report

Title: Adherence to hormone therapy among women with breast cancer

Version: 4  Date: 17 March 2014

Reviewer: Arier Lee

Reviewer's report:

Major Compulsory Revisions:

1. Table 1, 2 and 3 looked into the distribution of each individual variables compared between adherent and non-adherent patients using Chi square test. A significant p-value from a Chi square test tells us the distribution of the variable tested is not the same between the adherent and non-adherent patients. That p-value does not test to see which is the significantly different group, and certainly does not test if the two means are significantly different. For example, p.7 'The group of adherent women was slightly older than the group of non-adherent women, with mean ages of 58.0 and 56.0 years, respectively (p <0.0001)'. It is misleading to associate the two means with the Chi-square test p-value. Chi-square test doesn't tell us which category has lower or higher likelihood of adherence. It is especially difficult to conclude from Chi-square test the least or highest likelihood group when the 'no information' group was also included. This appears in the interpretation of all Table 1, 2 and 3 results.

2. Page 8 '...logistic regression model that identified the independent effects of the explanatory variables...'. It looks like table 4 displays result from a multiple logistic regression model. If this is the case, it should be labeled clearly. Also the explanatory variables from a multiple logistic regression model are not 'independent effects'. The explanatory variables are analysed and adjusted for all other explanatory variables presented in the model.

3. The display of Table 4 and also the interpretation of Table 4 result on page 8 requires revision. Confidence interval of odds ratio should be presented together with OR in the text. It is not clear what were used as reference level for all of the explanatory variables. OR is a type of comparison, although stated in the text regarding the direction of effect, without knowing what reference levels were used it is confusing. It is especially confusing because Table 1, 2 and 3 all included 'no information category'. It is not clear if data with 'No information' were still included or excluded from logistic regression analysis displayed in Table 4.

It is difficult to see which rows in Table 4 are for the same variable. For example, are 'Completed 2nd grade' and 'Higher education' two categories of a three-categories Education variable? If so, what is the reference level for this 'Education variable'? It is not directly comparable to Table 1 'Education' variable, where there are 4 categories including a 'No information' category. What is 'Alcohol consumption' variable? Is this a continuous variable representing the amount of alcohol consumption? Or a categorical variable indicating 'Consumed
any alcohol' as opposed to 'No alcohol consumption'? If so, is 'No alcohol consumption' the reference level? What about the 'No information' category in Table 1? DATS (tests) is another problematic variable. In Table 3 there were 4 categories, if taking out the reference level, there should still be 3 categories in Table 4. I wonder if Table 4 is an analysis with complete cases only. If so, does that mean the 'No information' were excluded from the logistic regression in Table 4? But for variable such as DATS, if none means no tests were done, it may not mean the same as 'No information'.

Minor Essential revisions:

1. On page 5 it was stated that 'If the initial hormone treatment date was # 3 months earlier than the date of diagnosis, the diagnosis and the initiation of hormone treatment were assumed to coincide (i.e., the difference was equal to 0)'. It is not clear that how the dates were corrected. Was the initial hormone Rx date updated to the date of diagnosis or the date of diagnosis updated to the initial hormone Rx date?

2. The adherence was defined as 'the number of doses dispensed in relation to the dispensing period'. What about the doses dispensed but not consumed by the patients? If MPR accounts for only the doses dispensed by not adjusted for doses dispensed but not consumed by the patients, there could be an overestimate of adherence rate.

3. In 'Results' section. means are often represented with + - a number. It is not clear what the number is. Is it standard deviation? For example, on page 7 '...with a mean age of 57.5 (±13.6) years...' What does the number 13.6 represent?

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.