Reviewer's report

Title: Risk Group Defined by Recursive Partitioning Analysis of Patients with Colorectal Adenocarcinoma Treated with Colorectal Resection

Version: 1 Date: 6 September 2011

Reviewer: Galal Elgazzaz

Reviewer's report:

The authors present their large retrospective series to define different prognostic groups of surgical colorectal adenocarcinoma patients derived from recursive partitioning analysis (RPA). 10494 patients with colorectal adenocarcinoma underwent colorectal resection from Taiwan Cancer Database during 2003 to 2005 were included for this study. For the definition of risk group, the method of classification and regression tree was performed. Main primary outcome was 5-year cancer-specific survival.

The authors identified six prognostic factors for cancer specific survival, resulting in eleven terminal nodes. Four risk groups were defined as following: Group1 (mild risk, 5329 patients), Group2 (intermediate risk, 4517 patients), Group3 (high risk, 281 patients) and Group 4 (very high risk, 367 patients). The 5-year cancer-specific survival for Group 1, 2, 3, and 4 was 86.6%, 62.7 %, 55.9%, and 36.6 %, respectively.

They concluded that RPA offers an alternative grouping method that can predict the survival of patients with colorectal adenocarcinoma underwent surgery.

This is a well-written, retrospective analysis of a large data set, and adds to literatures.

I have a few questions:

1- Were any pathological subgroup done in this series? If so, they should be included.

2- The description of the chemotherapy is incomplete.

3- The authors state that 49% of patients received chemotherapy. There is no data provided regarding whether they were treating before or after resection.

4- Were any patients in the resection group re-resected? Any patients underwent combined colorectal resection?

5- How long did these patients follow-up in this study?

6- The authors stated that they excluded patients without information of lymph node; however in the result section and Table 1 they reported only 9924 out of 10494 patients. Please clarify why these missing patients.

7- Statistical analysis: How was incomplete follow-up handled in your survival analysis? What was your definition for "censoring" for determination of DFS? for OS? The survival plots should show the censoring times on the curves, and the n
of patients at risk at key time intervals needs to be specified also-e.g. at 5 years?

8- Figure 2 is a bit confusing especially regarding CCI. The data is not consistent and agree with the text. Which is correct? Please revised and needs to be corrected appropriately, then the table and narrative should be clear and consistent. Also % is confusing and needs to be explained in the figure or at the figure footnote what does it mean?

9- In result section second paragraph the sentence started with “Among these patients, most patients….” Needs to be rephrased.

10- what does mean” process” in third paragraph of discussion section.

11- General manuscript edits: There are numerous obvious errors in the narrative, in terms of grammar, punctuation, spelling, etc. These problems are too numerous for this reviewer to list here. This needs to be corrected. As a few examples (not an exclusive list): Introduction section: line 2 from bottom: "rearding" should be "regarding"; In statistical analysis and results sections “spitted” should be “split” discussion and conclusion: "arbituary, regreesion, seffect,..." are not correct. Etc.

**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:**

declare that I have no competing interests