Reviewer’s report

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

Version: 1 Date: 18 July 2011

Reviewer: Masayuki Kurosaki

Reviewer’s report:

Chang et al. used the method of recursive partitioning analysis (RPA) to identify different prognostic groups of patients who received surgical resection of colorectal carcinoma. Using six prognostic factors, 4 groups of patients were identified with various rate of 5-year survival ranging from 36.6% to 86.6%. The number of patients involved in this study is sufficient and the result is interesting and is clearly presented as a flow chart. However, several major points should be addressed.

Major Compulsory Revisions

#1. The selection of prognostic factors is an automatic process and is totally dependent on the statistical algorithm of RPA. The authors ‘forced’ tumor location variable in the model as the first split variable (Results, 2nd page, line 16), which is usually not acceptable for this kind of analysis. If done so, the RPA tree should be shown separately for colonic cancer and rectal cancer, respectively.

#2. The reproducibility of the decision tree model created by RPA needs to be validated. If not tested by validation, there remain the possibility that the rules identified by this analysis fit only to the original cohort and not applicable to other cohorts. External validation using an independent cohort is the ideal way of validation, but this is not easily available. Internal validation could be an alternative: the decision tree model may be built using randomly selected 1/2 or 2/3 of the whole cohort, and the validity may be tested by the remaining 1/2 or 1/3 of the cohort. This kind of internal validation have been reported: ‘A Predictive Model of Response to Peginterferon Ribavirin in Chronic Hepatitis C using Classification and Regression Tree Analysis’ by Kurosaki et al. (Hepatol Res 2010;40:251-260), or ‘Development of a decision tree to assess the severity and prognosis of stable COPD’ by Esteban et al. (Eur Respir J. 2011. Epub ahead of print). Since there are a large number of patients in the present study, the authors are suggested to perform an internal validation.

#3. Two specific questions regarding the statistical method should be answered. How did the authors set the minimal p value for the splitting rule as <0.0001, or how did the authors set the maximum difference in risk as 1. Could this be done automatically by the PASW statistics 18 software? This point needs explanation.

#4. The comment of the authors (discussion, 1st paragraph, lines 3-4) is incorrect. This study is not the first one to use RPA or classification and
regression tree (CRT) analysis for outcome prediction of colorectal cancer. There is a previous study using this analysis to predict rectal tumor response to preoperative radiotherapy: ‘A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods’ by Zolbec et al. (Clin Cancer Res, 2005).

**Level of interest:** An article of importance in its field

**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.