Reviewer's report

Title: Classification tree analysis of second neoplasms after childhood cancer

Version: 1 Date: 21 August 2006

Reviewer: Phyllis Gimotty

Reviewer's report:

General
The authors have presented a classification of patients based on the risk of second neoplasms. Except for having had radiation therapy for the first primary cancer, none of the other factors identified appear to define groups with significantly differential risk. One explanation for the lack of identifying significant differential risk is that the explanatory variables may not be predictive of the primary outcome; this is not discussed by the investigators. The identified risk groups need to be validated in another dataset.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
1. Page 6; lines 20-23. What was the justification for the choice of the three misclassification costs?

2. Page 7 lines 10-12; misclassification costs need to be selected based on actual or estimated costs, otherwise it is a somewhat artificial classification and highly subjective. Do these costs relate to “real” misclassification costs?

3. Page 7 line 23-Page 8. The rates for those with and without radiation have confidence intervals that don’t overlap. However, confidence intervals for the rates for the second split into HD versus others overlap considerably. It is likely that confidence intervals for groups further down the tree will also overlap. What is the justification for statements about differential risk?

4. The authors present three scenarios that depend on misclassification costs. There was no discussion of the implication of this choice on the interpretation of the resulting risk and/or risk groups. one can always set the cost high enough so that the misclassification of “cases” is minimized, but is it meaningful.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
1. Page 6; line 14. This sentence implies that the low rate of second neoplasms (SN) is responsible for the observed tree where all cases were misclassified. This is not always the case with low event rates. Another explanation is that the factors used to develop the tree are not strong predictors.

2. Table 2. “three-pruning 0.01” is not defined for the first scenario or corresponding characteristic for the other two scenarios.

Discretionary Revisions (which the author can choose to ignore)
none

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

Statistical review: No

Declaration of competing interests:
I declare that I have no competing interests