Reviewer's report

Title: External validation of risk classification in patients with docetaxel-treated castration-resistant prostate cancer

Version: 2

Date: 26 November 2013

Reviewer: Andrew J Armstrong

Reviewer's report:

This is a small validation cohort study of Japanese men with mCRPC treated with docetaxel, in which a number of validated risk factors were examined for survival at Jichi Medical University. The authors validated the Armstrong 3-risk group prognostic model for survival and also identified additional prognostic factors that were important for outcome in these men. Several major issues merit consideration, and several minor issues:

MAJOR

1) This is a very small single institution cohort (78 men) as compared with the TAX327 multinational trial of 1006 men with mCRPC. It is thus really not suitable for validation of a multivariate prognostic model. The inclusion of additional patients across institutions would be preferred. However, the fact that the ARC was highly prognostic is interesting even in this small sample size, illustrating external validation to some degree. The authors should report on some discriminatory index for this model such as the c-index: how good was the ARC at estimating prognosis? Hazard ratios with 95% CIs for each risk group for OS should be provided.

2) The lack of association of visceral metastases with OS is perplexing and makes me concerned about the data in this study. Visceral metastases are a validated risk factor for death in CRPC in nearly EVERY study conducted to date. Perhaps the authors should look back on how visceral disease was defined in their own database before concluding that a patient has visceral disease. The number of men with liver, lung, and other sites of visceral disease should be reported. Nonspecific lung nodules, liver cysts, liver hemangiomas, etc may confound this and lead to misclassification of visceral status. How confident are the authors in the assessment of visceral status? How many of these men had CT scans with contrast? It appears from table 1 that only 4 men had liver/lung visceral disease yet in table 2 it says that 31 men had visceral disease. What were the sites of visceral spread in the other 27 men? This is highly unusual. I am concerned that the authors are really unable to validate the ARC because they essentially have a different non-visceral population.

3) The progression by bone scan was explicitly defined in TAX327 and was a strong risk factor for death. The authors should explain how this was determined in their database. Did they have prior bone scans on their patients that allowed for this assessment accurately? If not, their ability to validate this model is
compromised.

4) For the multivariate analysis, levels of certain biomarkers are arbitrarily set and this should be discouraged given that this will lead to overfitting of their model. Examining which model leads to the best c-index may be a better way to go about this. I agree with the authors that the validation population is quite different than the TAX327 population in terms of estramustine exposure, lower PSA, lack of visceral spread, and the lower number of DTX cycles given. The discussion should be revised to expand on the major limitations of this study: small sample size, different population of men, etc.

MINOR:

1) Discussion. Recommendations to switch to cabazitaxel in poor risk men should be removed as this is highly speculative. It is reasonable to recommend clinical trial participation in these poor risk men, however, given the poor outcomes.

2) Would examine the ability of the TAX327 nomogram (Armstrong et al CCR 2007) to predict OS in this cohort, which might have better performance than the more limited 4 risk factor model from the same dataset. In the introduction, it is mentioned that the ARC is superior as a prognostic model because of its ease. It may be superior for this reason but it has less prognostic discrimination than the overall nomogram/model and this should be mentioned.

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:

only COI is the fact that I developed this risk model.