Author’s response to reviews

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

Authors:

Kazuhiko Nakano (nknkzhk@jichi.ac.jp)
Kenji Komatsu (komatsu@jichi.ac.jp)
Taro Kubo (t.kubo@jichi.ac.jp)
Shinsuke Natsui (natsui@jichi.ac.jp)
Akinori Nukui (nukui@jichi.ac.jp)
Shinsuke Kurokawa (s-kuro@jichi.ac.jp)
Minoru Kobayashi (minoruk@jichi.ac.jp)
Tatsuo Morita (moritatu@jichi.ac.jp)

Version: 3
Date: 17 January 2014

Author’s response to reviews: see over
We thank both referees and the editorial team for their quick and thoughtful suggestions and comments. We have revised the manuscript according to the referees’ critiques. The revised phrases and sentences that were suggested by referees 1 and 2 are highlighted in red and green, respectively.

Responses to referee 1

TITLE: No issues

(1) ABSTRACT: In the US there are a few treatment options for CRPC. Have these options come to Japan too?

No. Among the drugs discussed in this report, only docetaxel (DTX) is currently available in Japan. We added the following sentence to the first paragraph of the Background section: “In Japan, the currently available drugs are limited even further.”

(2) ARC has 3 classes good, intermediate and poor risk. Is this risk of developing complications with DTX?

In this cohort, DTX-induced adverse events with grades $\geq 3$ included neutropenia (47%), leucopenia (33%), decreased hemoglobin levels (4%), and febrile neutropenia (4%), among others. However, there were no significant associations between these adverse events and the ARC risk groups.

(3) Are responses graded according to RECIST criteria?

In measurable lesions, tumor response was evaluated according to the World Health Organization (WHO) criteria. PSA response was evaluated according to the Recommendations of the Prostate Cancer Clinical Trial Working Group (PCWG). This study mainly validated the usefulness of the ARC for classifying the OS and PSA responses of CRPC patients. Thus, we included the following sentence as the first sentence of the first paragraph in the Methods-Assessment section: “According to the recommendations of the Prostate Cancer Clinical Trial Working Group,[18] PSA responses were demonstrated in a waterfall plot of decreasing PSA rates for each patient.”
(4) PSA response was not different between risk groups, correct? However OS was
different with poor risk group having the worse OS. On multivariate analysis, ARC
was not significant?

This is correct. The PSA responses did not significantly differ between the risk groups,
in contrast to the OS. As you have indicated, we performed a multivariate analysis that
included the ARC as another factor and proved that the ARC was an independent
prognostic factor. Consequently, we added the following sentences: “A multivariate
analysis revealed that the ARC and PSA doubling time were independent prognostic
factors” (Abstract-Results section), “A multivariate analysis of OS was performed to
compare prognostic factors by a Cox proportional hazard analysis” (Methods-Statistical
analysis section), “A multivariate analysis with these 4 factors revealed that the ARC
and PSADT were independent prognostic factors (Table 3)” (Results-Subgroup Analysis
section), and “A multivariate analysis revealed that the ARC was an independent
prognostic factor” (first paragraph of the Discussion section).

INTRODUCTION: No issues

(5) MATERIALS AND METHODS: Define EMB (EMP?) the first time that it is used.

Because EMP was used for the first time, we added the full name (estramustine
phosphate) in the last sentence of the first paragraph of the Background section.

(6) So visceral mets only would be 1 on this ARC system? And visceral mets plus
anemia would be 2?

Yes. The ARC determines the intended CRPC patient’s status with regard to 4 items
(risk factors), including visceral metastases, bone scan progression, significant pain, and
anemia, and classifies the patient into 1 of 3 groups according to the sum of these risk
factors. These details were defined as follows in the first sentence of the first paragraph
in the Methods-Armstrong Risk Classification section: “The patients were classified as
good-, intermediate-, and poor-risk according to the ARC, which included the following
4 risk factors: visceral metastases, bone scan progression, significant pain, and anemia
(Hg level < 13 g/dL).[10] Patients with 0 or 1, 2, and 3 or 4 risk factors were classified
as good-, intermediate-, and poor-risk, respectively.”

(7) Why did you not use the pain classification as previously reported with ARC?
Although the risk factor of significant pain was defined as a Present Pain Intensity score (PPI) ≥ 2 and/or an analgesic score (AS) ≥ 10, these scores were not measured in the CRPC patients in this study. Thus, we designated the use of some types of analgesic at DTX initiation as a surrogate measurement of significant pain. We described this as follows: “Although significant pain was defined as a Present Pain Intensity score (PPI) ≥ 2 and/or an analgesic score (AS) ≥ 10 in the ARC,[10] we defined the use of some types of analgesic at DTX initiation as a surrogate measurement of significant pain because the PPI and AS of the patients in this study were not measured” in the fifth sentence of the first paragraph in the Methods-Armstrong Risk Classification section and “a different definition of significant pain as a risk factor” in the first sentence of the fourth paragraph in the Discussion section.

(8) In assessment paragraph last sentence, remove sentence in place under the stats subheading.

Per your suggestion, we have moved this sentence to the third sentence of the first paragraph in the Methods-Statistical Analysis section.

RESULTS: See abstract, Tables and Figures.

DISCUSSION: No issues

REFERENCES: No issues

TABLES:

Table 1 no issue

(9) Table 2 too complicated with lines. Please remove.

We have removed the problematic lines per your request.

(10) Please add ARC to analysis.

We have accordingly added the ARC to the analysis and noted that the ARC was proven to be an independent prognostic factor.

FIGURES:

Figure 1 no issue

(11) Figure 2 p value is for good vs. intermediate, intermediate vs. poor? Good vs.
poor? What?

The P value is for the global OS comparison. In addition, the OS comparisons for each pair that you mentioned were calculated and added to Figure 2.
Responses to referee 2

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.

After confirming and correcting the risk factors indicated in the following comments (2 and 3), the discriminatory ability of the ARC for OS was shown to be superior. As suggested, we have calculated the c-index (0.60) as a discriminatory index and have added this information to the Methods-Statistical Analysis section, Results-Armstrong Risk Classification Assessment section, and the first paragraph of the Discussion section. We have also calculated the OS hazard ratios with 95% CIs for each risk group and have added that information to Table 3. As the number of existing risk factors in each CRPC patient changed, the patient distributions of PSA responses and OS among the risk groups also changed; thus, we have corrected that information in the Results and Discussion sections.

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.
As you indicated, we mistakenly recognized lymph node metastasis as visceral metastasis. In addition, when confirming the patient registry data, we identified some patients in whom nonspecific lung nodules and liver cysts had been mistakenly registered as lung and liver metastases, respectively. After correcting and revalidating these data, the risk factor of visceral metastases was also proven to be significantly associated with OS in our cohort. We have added the definition of the visceral metastases risk factor to the Methods-Armstrong Risk Classification section and have corrected the data related to that risk factor in the Results-Armstrong Risk Classification Assessment section and Table 2. Moreover, we have removed the following sentence from the third paragraph of the Discussion section: “In fact, the presence of visceral metastases did not become a significant independent risk factor in the validation group.”

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.

Per your suggestion, we have added the definition of the risk factor of bone scan progression to the Methods-Armstrong Risk Classification section and removed those patients who had not performed comparable prior bone scans from that risk factor analysis. As a result, the risk factor of bone scan progression was also proven to associate significantly with OS in our cohort. We have corrected the data related to that risk factor in the Results-Armstrong Risk Classification Assessment section and Table 2.

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.

As indicated, we also considered the identification of new biomarkers from small cohorts such as ours to be an issue when comparing these with the ARC risk factors that were identified in a very large cohort in the same dimension. To compare the risk factors in a multivariate analysis, we have canceled the aim of identifying new biomarkers and limited our analysis to a simple comparison between the ARC and the following major patient characteristics that are often determined in clinical practice: age, PSA level, PSA doubling time, performance status, and Gleason score. As a result, the
ARC (and PSA doubling time) was proven to be an independent prognostic factor. We have corrected and added this information in the Methods-Statistical Analysis section, the Results-Subgroup Analysis section, and the third paragraph of the Discussion section. In addition, we have removed the following sentence from the third paragraph of the Background section: “and comparing the risk factors in the ARC with those that we identified.”

*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.*

Those indicated limitations have been mentioned and added to the fourth paragraph of the Discussion section.

**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.

As you pointed out, we have removed the phrase “switch to cabazitaxel” and corrected the sentence as follows: “However, CRPC patients who are classified as poor-risk should be recommended for clinical trial participation or other treatments because given the poor outcomes, these patients would be expected to experience a limited prognosis despite the use of DTX.”

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.

As suggested, we have applied our cohort to the TAX327 nomogram and calculated a c-index of 0.66. The TAX327 nomogram has a potentially better prognostic
discrimination ability than the ARC in our cohort. However, our cohort had missing values of TAX327 nomogram investigation items. The objective of this study was the external validation of the ARC, and a comparison of the ARC with the TAX327 nomogram in a retrospective study with a small cohort, such as ours, could be misleading. Thus, we omitted any mention of the TAX327 nomogram validation results in the revised manuscript. After collecting more CRPC patients, we would like to analyze and report the external validation of the TAX327 nomogram. However, we have corrected the following sentences: “some reports have demonstrated the usefulness of superior nomograms for predicting prognosis in CRPC patients” in the first paragraph of the Background section and “indicated the superior but insufficient discriminatory abilities” in the second paragraph of the Background section.