Author's response to reviews

Title: Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model

Authors:

Masahiro Takada (masahiro@kuhp.kyoto-u.ac.jp)
Masahiro Sugimoto (msugi@kuhp.kyoto-u.ac.jp)
Yasuhiro Naito (ynaito@sfc.keio.ac.jp)
Hyeong-Gon Moon (moonhgsurgi@hanmail.net)
Wonshik Han (hanw@snu.ac.kr)
Dong-Young Noh (dynoh@snu.ac.kr)
Masahide Kondo (mkondo@md.tsukuba.ac.jp)
Katsumasa Kuroi (kurochan@dd.ii4u.or.jp)
Hironobu Sasano (hsasano@patholo2.med.tohoku.ac.jp)
Takashi Inamoto (inamoto@tenriyorozu-hp.or.jp)
Masaru Tomita (mt@sfc.keio.ac.jp)
Masakazu Toi (toi@kuhp.kyoto-u.ac.jp)

Version: 2 Date: 28 April 2012

Author's response to reviews: see over
Dear Drs Morrey, L’italien, Sorace and Taylor,

Please find attached our revised manuscript entitled “Prediction of axillary lymph node metastasis in primary breast cancer patients with a decision tree-based model” (MS: 1410459146656059) and a summary of our responses to the reviewers’ comments.

We wish to thank the reviewers for their valuable comments, which have helped to improve our manuscript. We also thank you for giving us the opportunity to resubmit our manuscript to *BMC Medical Informatics & Decision Making*, and hope that it is now suitable for publication. We look forward to hearing from you at your earliest convenience.

Yours sincerely,

Masakazu Toi, MD, PhD
Corresponding author
Professor, Department of Breast Surgery
Graduate School of Medicine, Kyoto University
54 Kawaracho, Shogoin, Sakyō-ku
Kyoto 606-8507, Japan
Tel: +81-75-751-3660
Fax: +81-75-751-3616
E-mail: toi@kuhp.kyoto-u.ac.jp
Responses to Dr. Patrick Tighe’s comments

Thank you very much for the opportunity to review this manuscript. The authors should be commended on constructing a well-written manuscript that appropriately details an interesting project involving the use of an alternating decision tree to predict axillary lymph node metastasis in patients with primary breast cancer. The authors diligently describe a multi-stage approach, including their steps in preprocessing, model development using hold-out datasets, testing for accuracy and accounting for potential bias, and comparison with prior models.

Importantly, the authors should provide greater detail on the method of ADTree induction and Ensemble creation, and soften the strength of their conclusions. Specific concerns are listed as follows. I believe each of these items can be easily remedied via rephrasing of content or inclusion of additional methodologic detail.

Major Compulsory Revisions:

Background.
1. In the third paragraph, the authors describe their use of the alternating decision tree. Why did the authors choose the ADTree instead of, perhaps, a Bayesian network or support vector machine? If the purpose of this project was to specifically explore the use of the ADTree, please say so. As currently written, the authors correctly point out the limitations of logistic regression-based classifiers, but fail to pivot on this argument and defend their use of the ADTree.

We have described why we used the ADTree method in the Introduction (P6, L4–9), in the Discussion (P14, L2–8), and in the figure legends of Appendix C (Additional File 1).

Methods
2. The authors very nicely describe the use of bootstrapping to account for bias, as well as their inclusion of a sensitivity and missing value analyses. More detail is needed for each of these accountings. For instance, were there constrains on the range of imputed values? Were averages, max-min, or other methods employed? How many boostrap iterations were performed? Additionally, this reviewer is slightly confused on the total number of datasets generated; were 200 generated altogether, or 200 each for the bootstrap, sensitivity and missing value analyses for a total of 600?

As suggested by the reviewer, we have described the bootstrap iterations and the number of datasets generated in the Data analysis section of the Methods (P9, L15–22).
3. The authors need to expand their description of several important features of the ADtree. For instance, how was the Ensemble generated? How many boosting iterations were performed with each run of the ADTree? How many nodes were expanded with each trial? (I may be confusing these items for the author’s inclusion of the number of nodes and trees, this is a big unclear.) How did the authors choose n=10-20 for number of nodes, random seed 1-10, and number of trees 2-20? Specific details on the implementation within Weka would be of interest. Similarly, the authors should include more description of the types of statistical comparisons used within the Methods section, including the rationale for a Kruskal-Wallis test.

In response to this comment, we have provided more details regarding the development of the ADtree models and the statistical tests applied in this study in the Data analysis section of the Methods (P9, L15–P10, L2).

Results
4. I would encourage the authors to use a bit softer language in describing their results. “The model successfully discriminated….at statistically significant levels” refers back to Figure 3. While Figure 3 does suggest statistically-significant differences in the probability of LN metastases in the LN-versus LN+ groups, there is considerable overlap in the probability distribution. This sentence should thus be reworded. Critically, the type of statistical test needs to be included, preferably with the mean and 95% confidence interval (or better yet, the mean difference with confidence intervals) for each comparison in this Figure.

In response to this comment, we have toned down the Results (P10, L19–P11, L2) and described the statistical tests applied in the Data analysis section of the Methods (P9, L23–P10, L2) and the legend of Figure 3 (P26). We have also added the median and 95% confidence intervals of the predicted probabilities to the legend of Figure 3 (P26).

Discussion
5. "...using the ADTree ensemble technique successfully predicted AxLN metastasis...” may be a bit too strong. I would consider softening this to read “...technique improved upon older models such as the MSKCC nomogram”

As suggested by the reviewer, we have toned down this sentence in the Discussion (P12, L11–13).
Minor Essential Revisions:
6. On page 12, “mammary grand” should read “mammary gland” I believe.

Thank you for this observation. We have corrected this term in the Results (P10, L14) and Discussion (P13, L18).

7. In Figure 3, each plot should be titled with its location as indicated in the

As recommended by the reviewer, we have added titles to each graph shown in Figure 3.

Figure Legend.
8. Likewise, in Appendix D, each plot should be titled with its respective location.

As recommended by the reviewer, we have added titles to each graph shown in Appendix D.

Discretionary Revisions:
Background
9. In the first paragraph, 4th sentence ending with “...and could avoid this procedure” should be rephrased. Perhaps specify that it is the patients who could avoid this procedure?

In accordance with this comment, we have suggested that oncologists may wish to avoid this procedure in elderly patients and in patients with complications (P5, L7–8).

10. In the second paragraph, the authors state that “tolerance against missing values” is “an important feature of the prediction model”. Which prediction model? The MSKCC nomogram? All classification systems? This paragraph would benefit from improved transitions and greater specificity of terms.

To clarify our meaning, we have rewritten this sentence in the Background (P5, L17–18).

11. In paragraph 3, the authors may wish to use the term "machine learning" or "data mining", but perhaps not both.
As suggested, we have deleted “data mining” from this sentence in the Background (P5, L21).

Methods
12. The authors included sample populations from three distinct locations, but also from two distinct time frames. Differences in location were used for training, validation and testing. The authors may wish to mention whether there were any differences in screening, incidence, or testing over the two time intervals. Such differences may partially explain the fall-off in model accuracy with the test-set.

As described by the reviewer, the data were collected at different times for each dataset. However, all data were collected after establishing a methodology for SLN biopsy, and several meta-analyses have found no significant differences in accuracy of SLN biopsy. To clarify our meaning, we have revised this paragraph in the Patients section of the Methods (P7, L4–6).

13. For readers who may not have a oncologic background, the authors may wish to reiterate that the “pathological findings from the surgical specimens” was the outcome variable which labeled each instance for classification purposes.

In accordance with this comment, we have revised the section Data collection and sentinel lymph node biopsy in the Methods (P7, L17–21).

14. Within the Data Collection and SLN Biopsy section, the authors may want to comment on the possibility of different sensitivities for SLN detection across the three methods described. If the sensitivities differ significantly, the authors may wish to include an additional analysis whereby the location of surgery is included within the feature set, and a 10-fold cross-validation is conducted using all 465 subjects.

Although the methodology for SLN differed among the three datasets, several clinical trials have no significant differences in these methods. Therefore, we have mentioned this in the Data collection and sentinel lymph node biopsy section of the Methods (P8, L13–15) with appropriate citations.

Results
15. I would rephrase the sentence “This indicates that these variables were more important than the other variables...”. The sensitivity analysis simply suggests that the model was more sensitive to changes in these attributes, not necessarily that the variables themselves were more important. I state this because the
sensitivity analysis altered each attribute independently, and it is unknown how each attribute correlated with other attributes in each, or the overall, sample.

In accordance with this comment, we have rewritten this paragraph in the Results (P11, L13–15).

Discussion
16. On page 13-14, the authors make a great point; the MSKCC nomogram and Russells Hall Hospital scoring systems, like many decision support tools generated from multiple logistic regression techniques, are considered “good” because they use a restricted number of variables. This also makes them “easy” for clinicians to apply by themselves. With the advent of electronic medical record systems and improved classification algorithms, however, the use of high-dimensional data may offer substantial improvements in accurate classifications. The authors may wish to expand on this difference.

In response to this comment, we have expanded our discussion on this interesting topic and potential implications for clinical practice (P15–16).
Responses to Dr. Fengxi Su’s comments

Major Compulsory Revisions
--The author might have missed a predictive model for SLN metastasis that was developed by Fabien Reyal (Reyal F, Rouzier R, Depont-Hazelzet B, Bollet MA, Pierga J-Y, et al. (2011) The Molecular Subtype Classification Is a Determinant of Sentinel Node Positivity in Early Breast Carcinoma. PLoS ONE 6(5): e20297. doi:10.1371/journal.pone.0020297). This model should be incorporated in your study since it suggested that molecular subtype is involved in predictive of SLN metastasis. On the other hand, the MSKCC model does not include molecular subtype or ER/PR/HER2 status as predictors. The author should have a discussion about the role of molecular subtype in the prediction model for axillary LN mets.

In accordance with this comment, we have added a paragraph describing Fabien Reyal’s model and the MSKCC model in the Discussion (P15, L10–19).

--In a multicenter study conducted by Rouzier's team, (J Clin Oncol. 2009 Jun 10;27(17):2800-8. Epub 2009 Apr 6.), RP-ROC and CART were used for model construction. Are there any differences between your "tree" model and their RP-ROC or CART models? If there are some differences, please clarify why you choose your model rather than theirs. Besides, in their study, they performed an unreliability test and calculated Eaver and Emax and obtained a calibration curve. Do you think it is necessary to perform this analysis in your calibration plot?

In accordance with this comment, we have briefly described the differences between the ADTree model and the models used by Rouzier et al. in the Background section (P6, L2–9). The reviewer is correct that Rouzier et al used Eaver and Emax, but they also stated these were not adequate alternatives to tree-based models (J Clin Oncol. 2009;27:2800).

Minor Essential Revisions
--In table 1, the information about clinical stage (TNM) should be clarify as that would make it clear whether the tumor burden was similar or not among the three dataset. Especially for the LN metastasis, the among of patients with pN0, pN1, pN2 etc should be clarified.

The 86 patients (30%) in Tokyo and Kyoto datasets received neoadjuvant chemotherapy before surgery. For these patients, we predicted lymph node status before neoadjuvant chemotherapy. Therefore, the pathological stage (pTNM) determined at surgery was affected by prior therapy and the data could not be used in
our study. Instead, we included clinical TNM classification in Table 1.

--In appendix C, if the patients has F=5, should the score of F be +0.5 rather than -0.3?

Thank you for this observation. We have corrected Appendix C.

Discretionary Revisions
--The author should have a brief introduction about the rational and purpose of the pruning analysis since most clinical physician may not familiar with this analysis.

In accordance with this comment, we have revised this statement in the Background to more clearly describe the pruning analysis (P9, L12–14).
Responses to the editorial requests

Editorial Requests

Consent
Please can you clarify your consent statement. We interpret the statement to mean that if researchers use anonymous clinical data for epidemiological research then they just need to send details on the study to everyone who's data may be used rather than getting individual consent. We would not usually expect a statement of consent in an manuscript using anonymised data from a database.

We used anonymous clinical data. We have revised this statement in the Patients section of the Methods (P7, L8–9).