Author's response to reviews

Title: Prognostic parameters for recurrence of papillary thyroid microcarcinoma

Authors:

Tae Yong Kim (tykim@amc.seoul.kr)
Suck Joon Hong (sjhong2@amc.seoul.kr)
Jung Min Kim (benaiyah@hanmail.net)
Won Gu Kim (blessing19@hanmail.net)
Gyungyub Gong (gygong@amc.seoul.kr)
Jin Sook Ryu (jsryu2@amc.seoul.kr)
Won Bae Kim (kimwb@amc.seoul.kr)
Sung-Cheol Yun (ysch97@amc.seoul.kr)
Young Kee Shong (ykshong@amc.seoul.kr)

Version: 4 Date: 28 July 2008

Author's response to reviews:

July 28th, 2008

Dear Editor.

We are grateful for the opportunity to respond to these comments and believe our manuscript has been clearly improved by this revision. We hope that the changes are satisfactory and that the manuscript is now suitable for publication in BMC Cancer.

Kind and helpful comments from the referees are highly appreciated. We have carefully considered each of the issues raised by the three reviewers (including two additional reviewers) and have addressed to each concerns in the attached letter and the revised manuscript.

Revision or Insertions are shown with the following attributes and color: Bold, Red.

Thank you for your kind consideration in advance and I am looking forward to hearing good news.

Sincerely yours,

Young Kee Shong, M.D., Ph.D.
Department of Internal Medicine, Asan Medical Center,
University of Ulsan College of Medicine
388-1 Pungnap-dong, Songpa-gu, Seoul 138-736, Korea
[Response to Reviewer 1]

Comment 1] I found only one part which may be terminologically misunderstanding. In the results section on page 12, multivariate logistic analysis and odds ratio were used as statistical terminologies to describe the results. But, this study used Cox’s proportional hazard model for the analyses. Appropriate terminologies are multivariate (Cox’s proportional hazard) analysis and hazard ratio (risk)?

: Thanks for your comment. We contacted professional statistician and corrected entire part of manuscript using statistics. We have modified the terminologies as Cox’s proportional hazard model and hazard ratio in revised manuscript.

[Response to Reviewer 2]

Comment 1] The article is much better now. Should be publicised!

: Thank you very much.

[Response to Reviewer 3]

Overall comment] On the one hand, it is possible that the statistical analyses were done correctly. But on the other hand, I see several inconsistencies and deficiencies that make me suspect that the statistical analyses for Table 3 and Table 4 may have been done incorrectly. I cannot tell for sure. Moreover, I found strong indications that the multivariate analysis was done using an incorrect method. My recommendation for the authors is that they find a professional statistician to add as a co-author, and have the statistician re-analyze their data using appropriate methodology, in order to see how much of their analysis results can be confirmed.

: Thanks for your thoughtful comment. We have contact a professional statistician, professor Yun SC, and he thoroughly review our original raw data and revise our manuscript. We also added him as a co-author of this paper.

Comment 1] “Definition of Clinical Outcome”: Having a Methods paragraph with this title is a great idea, and more writers should follow the authors’ example. However, for survival analysis, the authors’ current paragraph is incomplete. In addition to the definition of “Recurrence”, we also need the definition of “Time to Recurrence”, namely, the duration of time from when follow-up began to when the patient either had a recurrence or received their last follow-up evaluation (whichever came first). The definition of Time to Recurrence should include whether follow-up began at diagnosis, at treatment, or at some other time. For
this paper, Time to Recurrence is (or should be) the true clinical-outcome variable, while Recurrence is (or should be) just the “status indicator” (i.e., the censoring variable) that accompanies Time to Recurrence and helps define it.

: Thanks for your comment. The time to recurrence was defined duration between day of first operation and time of detection of recurrence by diagnostic procedure (for who recur) or last show-up date to our hospital (who do not recur). We added the definition of ‘time to recurrence’ in revised manuscript. (Methods: page 7, Definition of clinical outcome)

Comment 2] In the Statistical Analysis paragraph of their Methods section, the authors say that “Cox’s proportional hazard model and the forward stepwise method were used to analyze the relative importance of various prognostic factors for prediction of clinical recurrence.” But in the Results section of the Abstract, they say that “multivariate logistic regression analysis” was used for this, and in the Multivariate Analyses paragraph of their Results, they say that “multivariate logistic analysis” showed the prognostic value of cervical node metastasis. So already, there is an inconsistency between the Methods and the Results as to what method was used. Moreover, because their “Definition of Clinical Outcome” paragraph defined Recurrence but not Time to Recurrence, I am almost certain that the authors used multivariate logistic regression with Recurrence as the outcome variable. If so, logistic regression is not the correct multivariate method to use. The authors should have instead used multivariate Cox regression with Time to Recurrence as the outcome variable. The authors’ future statistician co-author will probably want to construct a table of results from the multivariate Cox regression.

: Thanks for your comment. We revised our whole manuscript that makes you confused. The Kaplan-Meier method including the log rank test was used to compare recurrence. Univariate analyses were performed separately for each of the variables using Cox's proportional hazard model. Variables for which the P value was <.2 in univariate analysis were included in a Cox's proportional hazard model. A backward exclusion process was used, and adjusted hazard ratio (HR) and 95% confidence intervals (CI) were calculated. P value of <.05 was considered significant. Please see Abstract and Statistical analysis section in Methods.

Comment 3] Tables 3 and 4: The footnotes say that “the Kaplan-Meier method with the log-rank test was used to compare recurrence”, and that would be the correct method to use. However, the tables calculate rates of recurrence as the number of recurrences divided by the number of patients; this would be an appropriate summarization for logistic-regression methods or chi-square testing methods, but is not appropriate for survival-analysis methods. What the authors need to do instead is, (1) present for each analysis group the “person-years” of total observation time, then (2) calculate the recurrence rates for each group as number of recurrences divided by number of person-years. The resulting recurrence rates will be “hazard rates”, will have units of proportion (or percent)
*per* *year*, and will more accurately reflect the group differences in risk of recurrence than the authors’ current calculation method. The authors’ future statistician co-author will know how to do this effectively and efficiently.

: Thanks for your comment. We revised Table 3 and 4 to be appropriately represented.

Comment 4] The Statistical-analysis paragraph in Methods has a couple of deficiencies in addition to the previously noted issue with Cox regression. This paragraph has a sentence that says “values are presented as means ± standard deviations”, but only one quantity (tumor size) is ever presented in this manner. Follow-up time was presented as median and range, and all the other quantities I could find were presented as number and/or percent with a characteristic. The footnote in Table 1 says that baseline characteristics were compared by Fisher’s exact test for two-group analysis variables, and by chi-square test for multi-group analysis variables; this information is missing from the Statistical-analysis paragraph, and needs to be included. Point of Clarification:

: Thank you for your quality-improving review. Categorical variables are presented as numbers and percentages, and were compared using chi-square or Fisher’s exact test. Continuous variables are presented as mean # SD and range. We corrected the statistical-analysis paragraph of our revised manuscript that makes you confused.

Comment 5] In the Definition of Clinical Outcome paragraph, Recurrence is defined as “persistence and/or reappearance of disease after initial operation”. How many recurrences were reappearances, and how many were persistent disease? (I haven’t been able to figure this out from the text.) If all recurrences were reappearances, and none were persistent disease, then I don’t see the need to include persistent disease in the definition of Recurrence. On the other hand, if there were cases of persistent disease among the recurrences, then those patients should appear in the Kaplan-Meier curves as recurrences on Day 0, and I don’t see any Day-0 recurrences in the curves. So are the persistent-disease cases inadvertently being excluded from the analysis, or were there none to begin with?

: Thanks for your comment. The definition of recurrence in well differentiated thyroid cancer is somewhat arbitrary. Thyroid cancer is very slow-progressing tumor, and we think that most of recurrence is progression of tiny residual disease, which could not be removed at initial surgery. However, clinically we cannot differentiate these two entities, persistence and reappearance of disease, using current many cancer guidelines (from American thyroid association, from European thyroid association, and from British thyroid association) for follow-up. Thus, we and many other research group practically defined these two entities (persistence and reappearance) as a whole, not separately, as “persistent and/or reappearance of disease after initial operation”.
[Response to Reviewer 4]

Comment 1] Table 1 - comparison of baseline clinical characteristics.

The statistical significance tests in Table 1 are presumably based on the null hypothesis that the distributions of the numbers of patients with given characteristics various be equal/proportional between the categories compared. But is such a null hypothesis justifiable in terms of clinical expectations? The authors should state why they think this is an appropriate null hypothesis to test.

: Thank you for your comment. We added table 1 of this manuscript according to many surgeon’s recommendation. There is no guideline for surgical extent of PTMC. Thus, they (mostly thyroid surgeons) want to know that differences in surgical extent and/or multifocality affect clinical outcome of PTMCs. Statistically, table 1 might be confusing and with inappropriate null hypothesis but we think that table 1 give many useful information for surgeons, who will read this manuscript.

Comment 2] Why does the number of 307 patients referred to at the beginning of the Results section reduce to 306 at the start of the section headed "Clinical characteristics of patients with recurrence".

: Thanks for your comment. Recurrence could not be evaluated in patient with distant metastasis. As one patients showed distant metastasis at initial presentation, we excluded this patients for subsequent evaluation in recurrence. This explanation is described in methods section (page 4, paragraph 3, in revised manuscript)

Commnet 3] And why were only 293 of the these 306 patients "eligible for analysis" - please state reason why 14 patients were excluded from analysis.

: Thank you for your comment. The follow-up for thyroid cancer after initial treatment require thyroid ultrasonography, stimulated Tg measurement, and/or diagnostic scan. Some patient did not show-up on subsequent follow-up and we cannot assess recurrence status in these patients. That why we excluded these 14 patient in subsequent analysis.

Comment 4] Again in Table 3 we need an explanation of why the null hypothesis – which is presumably that equal recurrence is expected in the categories of patient that are defined, is a reasonable thing to test for these categories. Why, for example, split age into <=45/>45 yrs?

: Thanks for your comment. Our null hypothesis was that each clinicopathological parameters was not associated with clinical outcome in Table3. The clinicopathological parameters in Table 3, 4 were based on previous studies and AJCC/UICC TNM staging system of differentiated thyroid carcinoma. We split age into <=45/>45 yrs because TNM staging system takes into account age (cut
off 45 years). For example, patients with distant metastasis were classified in stage II (age<=45) or stage IV (age>45) according to age. We revised Table3 and 4 to be appropriately represented.

Comment 5] Also, the Kaplan-Meier method compares "time-to-event", so to put K-M p-values alongside simple tabulations of rates of recurrence is potentially misleading - the p-value is comparing time-to-recurrence and not rate of recurrence. So perhaps this table should be amended to show median times to recurrence as well as rates of recurrence, with a footnote to emphasise that the p-value refers to a comparison of times and not rates. Similar considerations apply to Table 4. Generally - avoid confusion between "rate of recurrence" and "time-to-recurrence".

: Thanks for your comment. We contacted professional statistician and corrected entire part of manuscript using statistics. We revised our Table 3 and Table 4 to be appropriately represented.

Comment 6] A summary table of odds-ratios for the logistic analysis should be presented. The simple discussion in the text is not sufficient

: Thank you for your comment. We added Hazard ratio information in revised (table 3 and 4) and added table (table 5 and 6).