Complementary Therapies in Medicine (2012) xxx, xxx–xxx

Available online at www.sciencedirect.com
SciVerse ScienceDirect
journal homepage: www.elsevierhealth.com/journals/ctim

# Development of differential criteria on tongue coating thickness in tongue diagnosis☆

**Jinsung Kim[a,*], Ga-Jin Han[a], Bin-Hye Choi[a], Jae-Woo Park[a], Kyungmo Park[b], In-Kwon Yeo[c], Bong-Ha Ryu[a]**

[a] *Department of Gastroenterology, College of Oriental Medicine, Kyung Hee University, Seoul, Republic of Korea*
[b] *Department of Biomedical Engineering, School of Electronics and Information, Kyung Hee University, Yongin, Kyunggi-do, Republic of Korea*
[c] *Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea*

**KEYWORDS**
Tongue diagnosis;
Tongue coating;
Diagnostic criteria;
Digital tongue imaging system;
Proportional odds model

**Summary**
*Objective:* To develop differential criteria on the tongue coating thickness (TCT), and especially propose the standard for judgment on thin and thick coating in tongue diagnosis.
*Methods:* Sixty oriental medical doctors evaluated the TCT (none, thin, and thick coating) in 50 realistic tongue photographs revealing from tip-to-root. The photographs were obtained with a digital tongue imaging system (DTIS) which measured the percentages of tongue coating on the tongue surface. We calculated the match rate which is the ratio in which the assessor's judgment is consistent with the decision standard on the tongue coating, and then 24 assessors ($\geq$80% match rate) were selected to improve the reliability of the decision. The agreement level among 24 assessors was examined to assess the inter-rater reliability. The correlation between TCT judgments and DTIS-measured values was examined to ascertain the reliability of DTIS measurements. Finally, the assumption probability for the analysis of quantified characteristics of the tongue coating was calculated with a proportional odds model.
*Results:* The inter-rater reliability was assessed as moderate ($\kappa = 0.56$) among 24 assessors, the level of correlation between TCT judgments and DTIS measurements in 24 assessors was relatively high (0.76, $p < 0.01$). As the analysis of the proportional odds model, 29.06% was a cut-off point to separate no coating and thin coating, 63.51% was a cut-off point to separate thin and thick coating.
*Conclusions:* The differential criteria for TCT in tongue diagnosis were suggested, and particularly thick coating is defined as one that tongue coating which tongue body is invisible, occupy approximately more than two third areas on the tongue surface.


☆ This research was supported by a grant KHU-20091418 from the Kyung Hee University Program for the Young Researcher in Medical Science.
* Corresponding author at: Department of Gastroenterology, Kyung Hee University Oriental Medicine Hospital, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-702, Republic of Korea. Tel.: +82 2 958 8895; fax: +82 2 958 8895.
*E-mail address:* oridoc@khu.ac.kr (J. Kim).

0965-2299/$ — see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ctim.2012.03.004

## Introduction

Tongue diagnosis is a useful method to examine the physiologic function and pathologic changes of the human body in Oriental Medicine (OM).[1] It consists of visually inspecting the tongue body for color, shape, moisture, and movement and assessing tongue coating for color, thickness, distribution, and characteristics at the root.[2] The tongue coating reflects the nature and location of the pathogenic factors, and often indicates the state of some organs, particularly stomach function.[3] Therefore, fastidious observation on the alteration of the tongue coating is particularly required in the course of the diagnosis and the progress of gastrointestinal diseases.

The conventional way for tongue diagnosis, including the evaluation of the tongue coating, is to look at the tongue (from tip-to-root) directly by observers. The tongue coating thickness (TCT) is expressed as no coating, thin coating, and thick coating. No coating is literally the absence of tongue coating, and the dividing standard between thin and thick coating is whether or not the bottom is visible (thin coating, through the tongue coating, one can vaguely see the tongue body; thick coating: through the tongue coating, one cannot see the tongue body).[3] However, these judging criteria is insufficient and vague. Moreover, there is no distinct standard with respect to the distribution of the coating on the tongue surface. Therefore, in order to improve the low levels of inter- and intra-practitioner agreements, it is required that adequate operational definitions of the tongue characteristics including tongue coating.[4,5] The inter-practitioner consistency for the TCT was partially investigated in the previous studies,[4,6] but there is no precise operational definitions concerning the judging criteria for the TCT.

A digital tongue imaging system (DTIS) was designed to obtain a digital image of the tongue surface, segment this into regions and extract the areas corresponding to the tongue coating.[7] We showed the new possibility that the tongue coating can be quantitatively evaluated using the DTIS.[8] By starting from the percentage of the tongue coating area among the tongue surface analyzed by the DTIS, we attempted to suggest the differential criteria on the TCT compared with three diagnostic endpoints (no coating, thin coating, and thick coating). First, we verified the 60 assessors as to how well they evaluated the TCT using the conventional method of tongue diagnosis, we selected reliable assessors and investigated inter-assessor reliability, then confirmed the correlation between the TCT judgments and the DTIS-measured values (percentages of tongue coating). Finally, a proportional odds model was applied to calculate the assumption probability for the analysis of quantified characteristics of the tongue coating.[9] The aim of this study was to develop differential criteria for the TCT, and especially propose the standard for judgment on thin coating and thick coating with the amount of tongue coating in tongue diagnosis.

## Materials and methods

### Tongue photographs and DTIS

We prepared 50 tongue photographs revealing the entire tongue surface from tip-to-root. The photographs had high resolution which selected based on the clarity and reality of the image, depiction of different diagnostic characteristics, and diverse aspects of the tongue coating distribution. We used the DTIS to acquire the tongue photographs under the same conditions and to measure the tongue coating quantitatively.[8]

The DTIS consists of the exterior part with an interface for the patient's face, image acquisition part with a high-resolution digital camera, illumination part, and software part (Fig. 1). In particular, the interface of the exterior part was designed to block out outside light, and the DTIS software controls image acquisition, locating an appropriate region of the tongue, isolating this from the background using a snake algorithm, and extracting the areas of the tongue coating from the tongue images.[7,10,11] The tongue surface on the DTIS tongue images were divided into six sections, as the method of Winkel et al.,[12] and calculated the percentage of the tongue coating area for each section, and for the tongue as a whole (Fig. 2). Informed consent was obtained from each patient before taking their tongue photographs, as approved by the Institutional Review Board of the Kyung Hee University Oriental Medicine Hospital.

### Assessors and procedures

Sixty assessors were recruited to evaluate the tongue coating in the study. They graduated from 6-year programs of Oriental medical colleges in Korea, and were working for the Kyung Hee University Oriental Medicine Hospital as clinicians. They were reminded about the evaluation method such as how to differentiate the TCT and educated on how to divide the tongue surface into 6 areas before participating in the study. The procedures were performed as two sessions. In the first session, the assessors rated the tongue coating as no coating = 0, thin coating = 1, and thick coating = 2, and rated it again in each of 6 areas with the same method in the second session (the results in the 2nd session will be used in the subsequent study on the relationship between the decision for TCT and the regions of the tongue coating). All data collection was conducted at the same location under identical conditions over 50 min. These photographs were presented with a picture viewer software on the LCD monitor (47″ HD, Samsung, Korea). All assessments were independent, and no assessors were color blind.

### Data processing and assessors selection

The features of the TCT judgments on the 50 tongue photographs were computed as frequency and mode according to the classification of tongue coating and the DTIS-measured values. We regarded the highest number of the assessors' choice (mode) as the decision standard. In order to select reliable assessors, we calculated the match rate, which is the ratio that the assessor's judgment is consistent with the decision standard on the TCT (number of TCT judgments/50 photographs [%]). The reliable assessors were considered to be $\geq$80% match rates. The agreement level of the assessors' judgment for each tongue photograph was calculated as an agreement rate (number of TCT judgments/reliable assessors [%]), and investigated the inter-rater reliability among the selected reliable

Please cite this article in press as: Kim J, et al. Development of differential criteria on tongue coating thickness in tongue diagnosis. *Complement Ther Med* (2012), doi:10.1016/j.ctim.2012.03.004

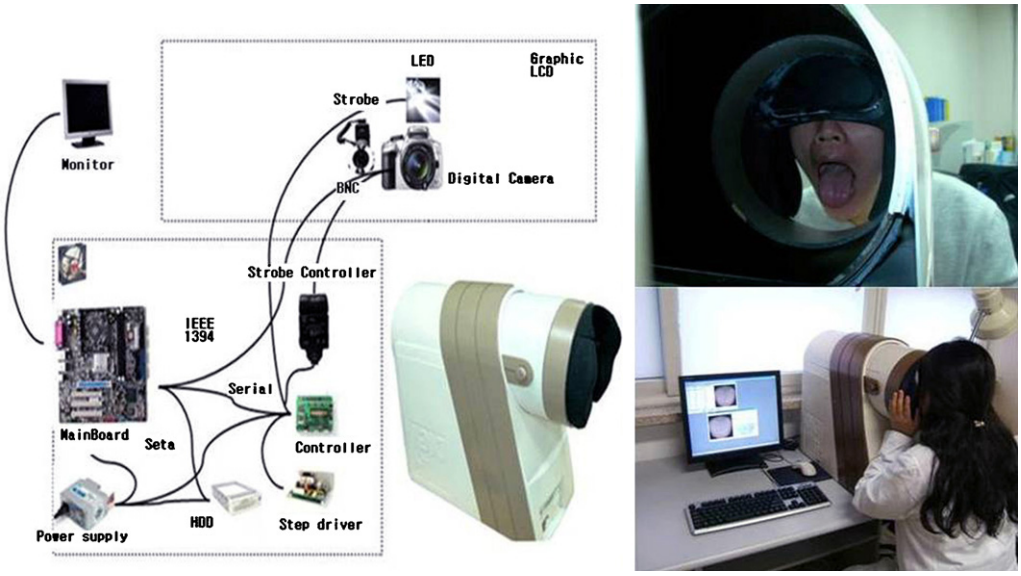


**Figure 1** Overview of the digital tongue imaging system (DTIS) and scenes of measurement. In the left part, the lower right side shows the exterior part of DTIS with interface for patient's face designed for easy presentation of patient's tongue to digital camera. The upper left and right sides show the internal composition of DTIS, which contains computer system and software, and parts for image acquisition and illumination. The interface of the exterior part was designed to block out outside light (right part).

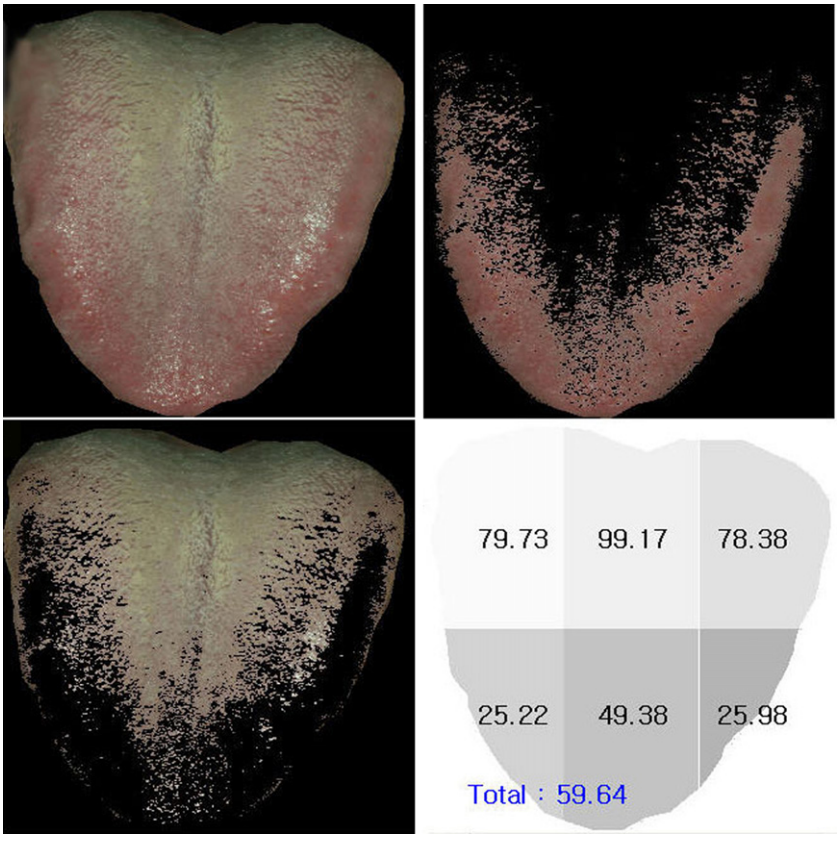


**Figure 2** Digital tongue imaging system program display showing acquired tongue image (upper left), extracted tongue substance image (upper right), extracted tongue coating image (lower left), and percentage score diagram for tongue coating (lower right).

Please cite this article in press as: Kim J, et al. Development of differential criteria on tongue coating thickness in tongue diagnosis. *Complement Ther Med* (2012), doi:10.1016/j.ctim.2012.03.004

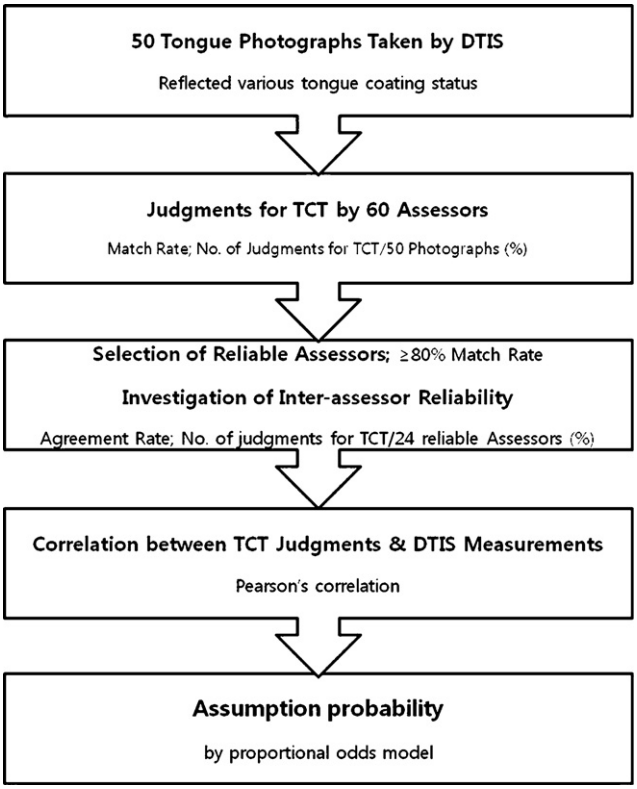


**Figure 3** Flow chart of data processing. Firstly, 50 tongue photographs were analyzed as tongue coating percentages with DTIS. Secondly, 60 assessors evaluated TCT as none, thin and thick coating, and match rate was calculated to select reliable assessors. Thirdly, agreement rate was obtained to investigate inter-assessor reliability among assessors who have ≥80% match rates. Fourthly, correlation between TCT judgments and tongue coating percentages were compared. Lastly, proportional odds model was applied to calculate assumption probability for analysis of quantified characteristics of tongue coating. DTIS; digital tongue imaging system. TCT; tongue coating thickness.

assessors. Then, we calculated the correlation between the TCT judgments and the DTIS-measured values to ascertain the reliability of the DTIS measurements. Fig. 3 shows all data processing in the study.

## Assumption probability for differential criteria on TCT

The proportional odds model is a useful extension of the logistic regression model which allows more than two ordered response categories. The model may be represented by a series of logistic regressions with common regression parameters reflecting the proportional odds assumption.[9]

We can treat the DTIS-measured values as the explanatory variable and the TCT judgments (0 = no coating, 1 = thin coating, and 2 = thick coating) as the response variable. Let $Y$ be the response variable and let $\pi_i(x)$ be the probability that $Y$ takes the value $j$, $j = 0, 1, 2$ when the DTIS-measured value is $x$. Because the TCT judgments are ordinal responses, a statistical way to model the relationship between $Y$ and $x$

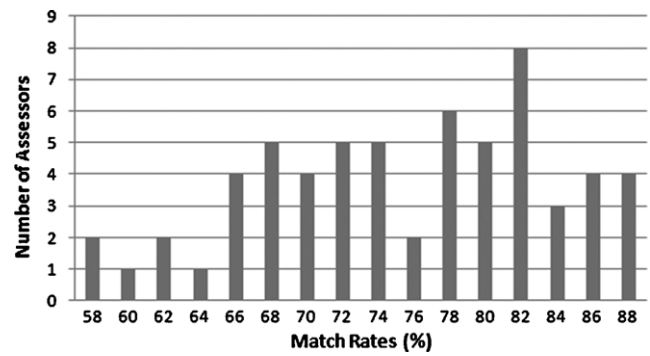


**Figure 4** Distribution of match rate, which is the ratio that assessor's judgment is consistent with decision standard on tongue coating in 50 tongue photographs; number of judgments/50 photographs (%). Decision standard is the highest number of the assessors' choice (mode). Those who have ≥80% match rates were regarded as reliable assessors.

is to use logits of cumulative probabilities $\eta_j = P[Y \le j|x] = \pi_1(x) + \cdots + \pi_j(x), \quad j = 0, \quad 1, \quad 2$. In order to estimate the probabilities $\pi_i$'s, we used the proportional odds model in which the cumulative probabilities are written as

$$\eta_j(x) = \frac{\exp(\alpha_j + \beta_x)}{1 + \exp(\alpha_j + \beta_x)} \quad \text{for} \quad j = 0, \quad 1 \quad \text{and} \quad \eta_3(x) = 1$$

where $\alpha_j$'s and $\beta$ are parameters to be estimated. Setting $\eta_{-1}(x) = 0$, we obtain the probability $\pi_j(x) = \eta_j(x) = \eta_{j-1}(x)$ and assign the TCT judgments to the coating degree whose probability is maximum at $x$.

## Statistical analysis

Fleiss' $\kappa$ was used as a measure of inter-rater reliability of the decision on the TCT among the assessors. It measures the overall agreements between all the assessors. Conventionally, a $\kappa$ of <0.2 is considered poor agreement; 0.21–0.4, fair; 0.41–0.6, moderate; 0.61–0.8, strong; and more than 0.8 near complete agreement.[13] Pearson's correlation was used to ascertain the relationship between the TCT judgments and the DTIS-measured values. The proportional odds model was applied to calculate the assumption probability for the quantified analysis for the characteristics of the TCT.[9] Analyses were performed using SAS 9.1 software. The significance level was set at $p$-values equal to or less than 0.05.

# Results

## Characteristics of TCT judgments and DTIS measurements

The characteristics of the TCT judgments and the DTIS measurements according to the classification of the tongue coating are shown Table 1. The match rates in 60 all assessors are presented Fig. 4, and the reliable assessors who had ≥80% match rates were 24 persons. The inter-rater reliability was assessed as moderate ($\kappa = 0.56$) among 24 assessors (Table 2), the frequency of the TCT judgments evaluated

**Table 1** Characteristics of TCT judgments and DTIS measurements according to classification of tongue coating on 50 tongue photographs in 60 assessors.

| Classification of TCT[a] | | No coating | Thin coating | Thick coating |
|---|---|---|---|---|
| Frequency of TCT | | 8 | 30 | 12 |
| | Mean (%) | 31.07 | 45.97 | 70.50 |
| DTIS measurements[b] | Median (%) | 31.61 | 47.99 | 71.06 |
| | Standard deviation | 12.61 | 11.27 | 8.27 |

[a] Tongue coating thickness.
[b] Percentage of tongue coating measured by digital tongue imaging system.

**Table 2** Inter-rater reliability of judgments on tongue coating thickness among 24 reliable assessors.

| Variable (tongue coating thickness) | Observation (number of photographs) | Endpoints | Overall $\kappa$ of result | Standard error (95% confidence interval) | Assessment of outcome |
|---|---|---|---|---|---|
| Among 24 assessors | 50 | 3 | 0.5580 | 0.0065 (0.5452–0.5708) | Moderate |

by the selected reliable assessors are presented in Fig. 5. The DTIS measurements ranged from 14.8 to 83.0% (average, 49.5%), the level of correlation between the TCT judgments and DTIS-measured values in 24 assessors was relatively high (0.76, $p < 0.01$).

## Differential criteria on TCT calculated by proportional odds model

Table 1 showed that the thicker degree of the tongue coating is, the higher the percentage of the tongue coating is, and the majority of the thick group definitely has a higher percentage than the 75th percentile of the thin classification. Table 3 shows the analysis of maximum likelihood estimates based on the proportional odds model.

According to the analysis, the cumulative probabilities $\eta_0(x)$ and $\eta_1(x)$ are respectively estimated by

$$\hat{\eta}_1(x) = \frac{\exp(5.0708 - 0.1747x)}{1 + \exp(5.0708 - 0.1747x)} \quad \text{and}$$

$$\hat{\eta}_1(x) = \frac{\exp(11.0995 - 0.1747x)}{1 + \exp(11.0995 - 0.1747x)}$$

**Table 3** Analysis of maximum likelihood estimates based on proportional odds model.

| Parameter | DF | Estimate | Standard error | $p$-Value |
|---|---|---|---|---|
| $\alpha_1$ | 1 | 5.0708 | 1.4233 | <.0001 |
| $\alpha_2$ | 1 | 11.0995 | 2.4208 | <.0001 |
| $\beta$ | 1 | −0.1747 | 0.0397 | <.0001 |

Fig. 6 showed that $\hat{\pi}_0(x) < \hat{\pi}_1(x)$ for $x \geq 29.06$ and $\hat{\pi}_1(x) < \hat{\pi}_2(x)$ for $x \geq 63.51$. This implied that the tongue coating were classified as no coating, thin coating, and thick coating when the percentages of the tongue coating were as follows: [0,29.06); [29.06, 63.51); and [63.51, 1], respectively.

## Discussion

We suggested that the differential criteria for the TCT judgments, and especially the standard for judgment on thin and thick coating. In this study, we selected 24 assessors who had ≥80% match rates to increase the reliability of the

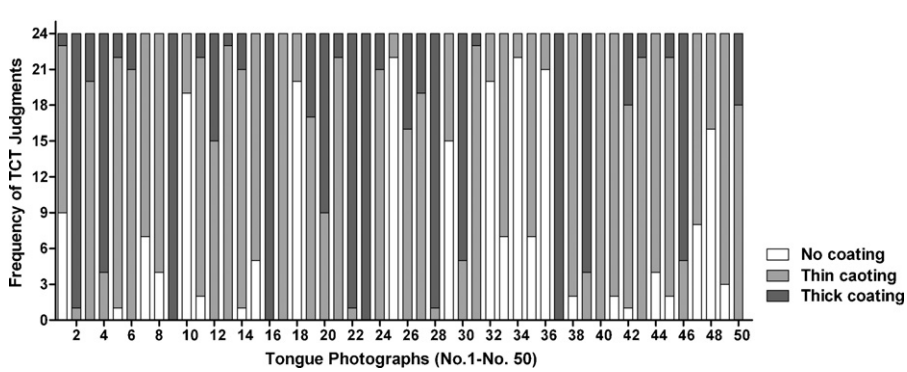


**Figure 5** Frequency of TCT judgments according to classification of tongue coating evaluated by 24 reliable assessors on 50 tongue photographs. This figure shows assessors' agreement levels of TCT judgments on each tongue photograph. TCT; tongue coating thickness.

Please cite this article in press as: Kim J, et al. Development of differential criteria on tongue coating thickness in tongue diagnosis. *Complement Ther Med* (2012), doi:10.1016/j.ctim.2012.03.004

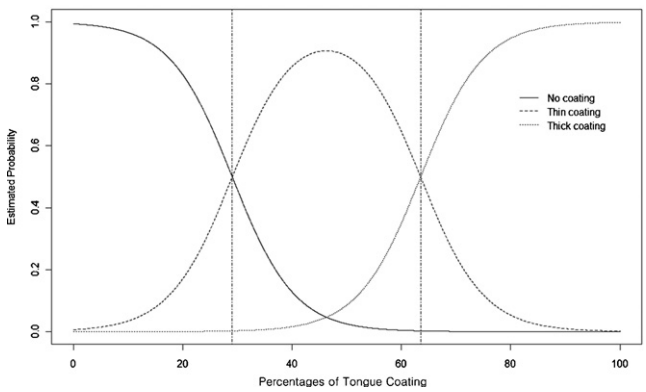


**Figure 6** Assumption probability of each classification of tngue coating at DTIS measurements. The estimates of the probabilities that TCT is $j$, $j = 0, 1, 2$ (0 = no coating, 1 = thin coating, and 2 = thick coating), $\pi_j(x)$. DTIS; digital tongue imaging system. TCT; tongue coating thickness.

decision, referring to the study that the authors used ≥80% for the inter-practitioner agreement of presence of tongue coating.[4] The level of agreement among the 24 assessors on the TCT judgments was moderate ($\kappa = 0.56$), which was relatively high compared with the result of O'Brien et al. (fair, $\kappa = 0.22$).[6] In fact, the inter-practitioner consistency for TCT judgments was partially investigated in the previous studies. 30 OM practitioners had a high level of agreement (≥80%) for the presence of tongue coating (yes/no choice) in all 10 slides,[4] and the level of agreement for TCT (4 endpoints) between 3 practitioners was fair (Cohen's $\kappa = 0.22$) in the 45 healthy Australians with hypercholesterolemia.[6] However, the former just used two endpoints, the latter investigated the agreements only between 3 practitioners.

In a previous study,[8] the agreement between WTCI ratings and DTIS assessments was relatively high at $r = 0.56$ for one examiner and $r = 0.77$ for the other ($p < 0.01$). In this study, the degree of correlation between the TCT judgments and the DTIS-measured values in 24 assessors was 0.76 ($p < 0.01$). Thus, we decided that we utilized DTIS measurements to develop the differential criteria on the TCT.

The standard model for quantitative criteria is required because improved consensus on OM diagnostic criteria results in an increased agreement level of diagnosis,[14] so the proportional odds model was used for developing the diagnostic standard of the TCT. 29.06% was the cut-off point to divide into no coating and thin coating, and 63.51% was the cut-off point to separate the thin coating from the thick coating (Fig. 6). Among these results, the cut-off point between no coating and thin coating was relatively high, because it is likely to be reflected the tongue coating at the root of tongue. Most people generally have tongue coating on the surroundings of circumvallate papillae. If the root of tongue is examined closely, we can see a little tongue coating on it in spite of no coating. Fig. 7 is the tongue photographs which have the approximate values of 29.01% and 63.51% in 50 tongue photographs.

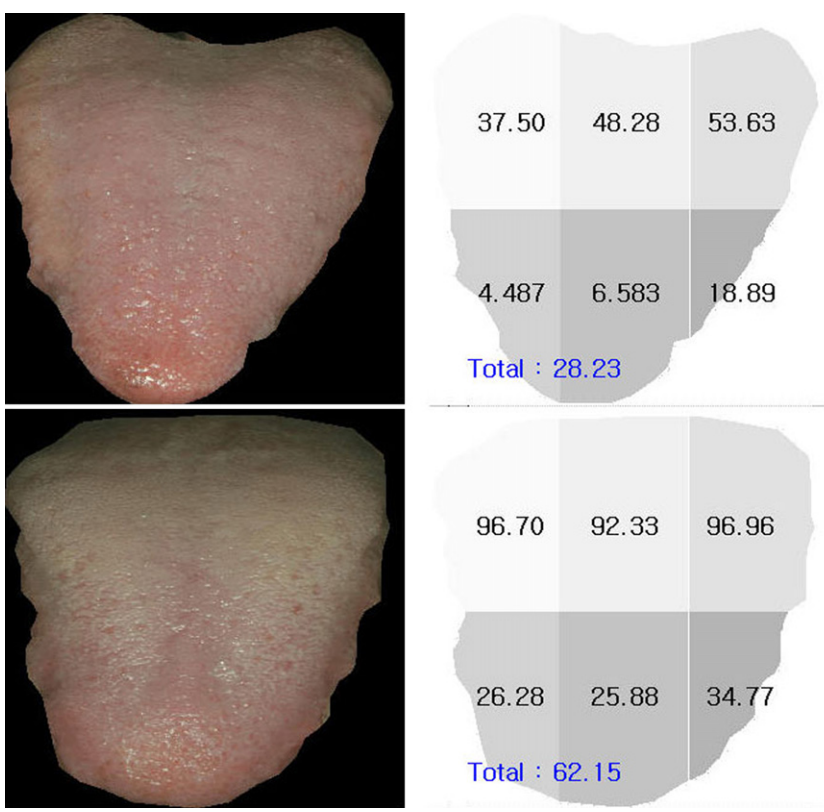


**Figure 7** Tongue photographs and percentage score diagram of tongue coating for differential criteria obtained with digital tongue imaging system. They are similar to borderline of no coating and thin coating; 29.01% (upper) and thin coating and thick coating; 63.51% (lower) in 50 tongue photographs.

Please cite this article in press as: Kim J, et al. Development of differential criteria on tongue coating thickness in tongue diagnosis. *Complement Ther Med* (2012), doi:10.1016/j.ctim.2012.03.004

The diagnostic criteria and the standard photographs of the tongue coating will be able to be used in the education of students and practitioners to improve the agreement level for the TCT judgment in tongue diagnosis. King et al.[15] provided an example of this in their development of a standardized pulse-taking procedure and precise operational definitions that resulted in ≥80% agreement between 2 practitioners on 10 of 16 pulse categories. Of course, this is not the absolute criteria and needs to be verified in a range of clinical settings. However, our suggestions were produced by credible assessors, so we can at least establish a guideline to differentiate the TCT. A few limitations of this study need to be acknowledged. First, we used the largest number of the assessors' choice (mode) as the decision value in the tongue coating assessment, thus the reliability of the decision declined in case of the low agreement level among the assessors. However, it was inevitable because there is no absolute standard to evaluate the TCT. Nevertheless, we could not discard the photographs which had a low agreement level because the diverse status of the tongue coating should be reflected. Second, although the DTIS can discriminate between thin and thick analyzed by color difference, it cannot distinguish heavy (thicker) from thick coating. Besides, it can be very susceptible to differences in lighting when the image is taken. Thus, further work is needed to improve the algorithm and device for evaluating the tongue coating to ensure accurate assessments.

The present study is the first trial for the differential criteria of the TCT, The data in our study was obtained from relatively high reliable assessors selected from well-educated and well-trained OM doctors and the DTIS verified in the previous study. However, our criteria are just suggestions based on the probability, so they should be verified by more reliable assessors in the clinical settings.

In conclusion, the thick coating is the case in which tongue coating that pinkish tongue body is invisible, occupy approximately more than two third areas on the tongue surface. This will contribute to one part of the standardization and objectification in tongue diagnosis.

## Conflict of interest statement

The authors have no personal or financial conflicts of interest associated with this work.

## References

1. Anastasi JK, Currie LM, Kim GH. Understanding diagnostic reasoning in TCM practice: tongue diagnosis. *Altern Ther Health Med* 2009;**15**(3):18—28.
2. Chen Jia-xu. *Chinese medicine study guide: diagnostics*. Beijing: People's medical publishing house; 2007. p. 40—1.
3. Fei Zhao-fu, Gu Yi-di. *Mirror of health: tongue diagnosis in Chinese medicine*. Beijing: People's Medical Publishing House; 2007. p. 15—7.
4. Kim M, Cobbin D, Zaslawski C. Traditional Chinese medicine tongue inspection: an examination of the inter- and intrapractitioner reliability for specific tongue characteristics. *J Altern Complement Med* 2008;**14**(5):527—36.
5. O'Brien KA, Birch S. A review of the reliability of traditional East Asian medicine diagnoses. *J Altern Complement Med* 2009;**15**(4):353—66.
6. O'Brien KA, Abbas E, Zhang J, Guo ZX, Luo R, Bensoussan A, et al. Understanding the reliability of diagnostic variables in a Chinese Medicine examination. *J Altern Complement Med* 2009;**15**(7):727—34.
7. Eo YH, Kim JG, Yoo HS, Kim JY, Park KM. Quantitative study on tongue images according to exterior, interior, cold and heat patterns. *J Kor Oriental Med* 2006;**27**(2):134—44.
8. Kim J, Jung Y, Park K, Park JW. A digital tongue imaging system for tongue coating evaluation in patients with oral malodour. *Oral Dis* 2009;**15**(8):565—9.
9. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990;**46**(4):1171—8.
10. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vis* 1987;**1**(4):321—31.
11. Pang B, Wang K, Zhang D, Zhang F. On automated tongue image segmentation in Chinese medicine. In: *Proceedings 16th international conference on pattern recognition, 2002, vol. 1*. 2002. p. 616—9.
12. Winkel EG, Roldan S, Van Winkelhoff AJ, Herrera D, Sanz M. Clinical effects of a new mouthrinse containing chlorhexidine, cetylpyridinium chloride and zinc-lactate on oral halitosis. A dual-center, double-blind placebo-controlled study. *J Clin Periodontol* 2003;**30**(4):300—6.
13. Fleiss J, Nee J, Landis J. Large sample variance of kappa in the case of different sets of raters. *Psycho Bull* 1979;**86**:974—7.
14. Zhang GG, Singh B, Lee W, Handwerger B, Lao L, Berman B. Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: effect of training. *J Altern Complement Med* 2008;**14**(4):381—6.
15. King E, Cobbin D, Walsh S, Ryan D. The reliable measurement of radial pulse characteristics. *Acupunct Med* 2002;**20**(4):150—9.

Please cite this article in press as: Kim J, et al. Development of differential criteria on tongue coating thickness in tongue diagnosis. *Complement Ther Med* (2012), doi:10.1016/j.ctim.2012.03.004