Comparison of Alternative Tumor Size Classifications for Posterior Uveal Melanomas

Cassandra C. Skinner, James J. Augsburger, Bret D. Augsburger, and Zelia M. Correa

Department of Ophthalmology, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States

Correspondence: Zelia M. Correa, Department of Ophthalmology, Medical Science Building 5th Floor, 231 Albert Sabin Way, Cincinnati, OH 45267–0527, USA; correazm@uc.edu.

Submitted: August 4, 2016
Accepted: June 1, 2017
Citation: Skinner CC, Augsburger JJ, Augsburger BD, Correa ZM. Comparison of alternative tumor size classifications for posterior uveal melanomas. Invest Ophthalmol Vis Sci. 2017;58:3335–3342. DOI: 10.1167/iovs.16-20465

PURPOSE. Determine which posterior uveal melanoma (PUM) size classification with three categories has the best prognostic discrimination.

METHODS. Single-institution study of 424 consecutive patients with PUM. The tumor’s largest basal diameter (LBD), smallest basal diameter (SBD), and thickness (TH) were estimated by fundus mapping and ultrasonography. Tumors were assigned to “small,” “medium,” or “large” size categories defined by 11 different classifications (Linear LBD, Rectangular LBD × TH, Cubic LBD × SBD × TH, Warren Original, Warren Modified, Augsburger, COMS Original, COMS Revised, TNM 2002, and modified TNM 2010 classification [a,b]). Prognostic significance of classifications was evaluated by Kaplan-Meier event curves with computation of log rank test for trend statistic.

RESULTS. In six classification systems (Warren Original, Warren Modified, COMS Revised, TNM 2002, TNM 2010a, TNM 2010b) >50% of tumors fell within one subgroup. In the Warren Original classification, among three size categories was judged “excellent” in four classifications (Linear LBD, Cubic Volume, TNM 2010a, and TNM 2010b) and “very poor” in the Warren Original. Linear LBD was associated with the highest log rank statistic value. TNM 2010a, TNM 2010b, TNM 2002, Augsburger, and Cubic Volume classifications were also determined to be quite good.

CONCLUSIONS. Linear LBD classification was the best three-size category discriminator among low-, intermediate-, and high-risk subgroups. Considering our findings, it seems possible that the arduous work required to apply complex classifications, especially for three-category systems, for PUM may not be justified in routine clinical practice.

Keywords: melanoma, classifications, choroidal melanoma, prognosis
Table 1. Tumor and Population Descriptive Statistics.

<table>
<thead>
<tr>
<th>Variable, Units</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBD, mm</td>
<td>4.0</td>
<td>24.0</td>
<td>12.2</td>
<td>3.7</td>
</tr>
<tr>
<td>SBD, mm</td>
<td>3.5</td>
<td>20.0</td>
<td>10.5</td>
<td>3.5</td>
</tr>
<tr>
<td>TH, mm</td>
<td>1.5</td>
<td>16.2</td>
<td>5.8</td>
<td>3.1</td>
</tr>
<tr>
<td>Cross-sectional area, mm²</td>
<td>9.0</td>
<td>268.3</td>
<td>76.8</td>
<td>56.6</td>
</tr>
<tr>
<td>Cubic volume, mm³</td>
<td>35.4</td>
<td>4953.0</td>
<td>966.0</td>
<td>955.4</td>
</tr>
<tr>
<td>Age, y</td>
<td>15.5</td>
<td>99.2</td>
<td>62.0</td>
<td>13.9</td>
</tr>
</tbody>
</table>

2010 classification as follows: TNM 2010a combined T1 and T2 subgroups, and TNM 2010b combined T2 and T3 subgroups.

Tumor classification graphs demonstrating the boundary lines for the separation of small, medium, and large tumors were prepared for each of the 11 classifications (Fig. 1). Tumors were then plotted on each classification’s boundary lines and assigned a tumor size subgroup (i.e., small, medium, or large) according to each system. Tumors falling on a boundary line were assigned arbitrarily to the lower risk group.

The Ease Versus Difficulty of Applying the Classification to a Group of Tumors

The ease versus difficulty of applying a size classification to a group of tumors was rated from easiest (0 points) to most difficult (2 points) based on whether mathematical manipulation of the raw data and/or reference to a graph was required. If mathematical manipulation of raw data was not required, 0 points were given. If mathematical manipulation (e.g., multiplication of LBD × TH) was required, then 1 point was given. If reference to a graph of intergroup boundary lines was not required, 0 points were given. If reference to category boundary lines was required to determine the size category, then 1 point was given.

Conventional descriptive statistics (mean, standard deviation of mean, minimum, maximum) were computed for all the evaluated continuous numerical variables (LBD, SBD, TH, rectangular area [LBD × TH] of tumor, cubic volume [LBD × SBD × TH] of tumor, age of patient) and frequency distributions for all study variables, including categorical variables.

The principal outcome evaluated was development of metastasis during follow-up. No patient had evidence of metastasis at the time of initial treatment. Data analysis extended to December 31, 2015.

Kaplan-Meier event rate curves for metastasis according to tumor size category assigned by each classification were computed and plotted. The effectiveness of each classification for assigning patients to discrete ordered low-, medium-, and high-risk categories for metastasis was assessed by (1) inspection of the plotted event rate curves to determine how clearly separated the three curves were and whether the curves were ordered as expected (i.e., showing lowest risk for the cases categorized as small, highest risk for the cases categorized as large, and intermediate for the cases categorized as medium), and (2) computing a log rank test of trend statistic for the significance of the separation of the curves plotted for each classification. The null hypothesis was that the curves were substantially different from one another and ordered as expected. The greater the separation of the ordered event rate curves, the higher the value of the computed test statistic.

RESULTS

Baseline Patient and Tumor Characteristics

Features of the 424 patients and their PUM are summarized in Table 1. Seventy-two percent of studied tumors were choroidal (n = 306), 23.8% were ciliochoroidal (n = 101), and 4% were iridociliochoroidal (n = 17).

Frequency Distribution of Tumors According to Classification

The frequency distributions of tumors assigned to size categories by the 11 evaluated classifications are summarized in Table 2. Six classification systems (Warren Original, Warren Modified, COMS Revised, TNM 2002, TNM 2010a, TNM 2010b) had more than 50% of tumors falling within one subgroup and another classification system, the Warren Original, had fewer than 5% of tumors falling within one subgroup.

Tumor Size Category of the “Average Tumor” In the Group

The “average tumor” in our study (i.e., hypothetical tumor having the mean value of LBD, SBD, and TH of all tumors in this series [12.2 × 10.5 × 5.8 mm³]) was categorized as “small” by one classification (TNM 2010a), “medium” by eight classifications (Linear LBD, Rectangular Area, Cubic Volume, Augsburger, COMS Original, COMS Revised, TNM 2002, and TNM 2010b), and “large” by two classifications (Warren Original, Warren Modified).

Separation of Event Curves

Kaplan-Meier curves for the three tumor size subgroups by each evaluated classification are shown in Figure 2. Inspection of these curves shows considerable differences in discrimination between small, medium, and large tumors based on each individual classification. The separation between curves was rated as indicated in Table 3. Linear LBD, Cubic Volume, TNM 2010a, and TNM 2010b classifications were all rated “excellent,” whereas the Warren Original classification was rated “very poor.”

Ease of Classification Application

A scoring system from 0 to 2 is listed in Table 3 for the evaluated classification systems.

Prognostic Significance of Evaluated Classification Systems

Table 4 lists the evaluated classifications in rank order by the log rank test of trend statistic. The highest the log rank test statistic was associated with the Linear LBD classification, whereas the lowest log rank test statistic was associated with the Rectangular LBD × TH classification.

DISCUSSION

It is surprising that tumor size was not recognized to be a prognostic factor for metastasis and metastatic death in PUM until approximately the middle of the 20th century. Classic articles by Callender and colleagues in the 1930s through 1940s describe histopathologic features of PUM in emaciated eyes but do not mention tumor size. The earliest article mentioning tumor size as a prognostic factor that found in our


In 1974, Warren8 published his original classification system. This classification has only one category system. See the Methods section for an explanation of how the three-category TNM 2010a and TNM 2010b classifications were formed.

Literature search divided the PUMs into “smaller” and “larger” subgroups at the mean of the cubic tumor volume of the evaluated tumors in a series of enucleated eyes (LBD² × TH = 1344 mm³).22 The authors indicated that metastasis and metastatic death were substantially higher in the subgroup with larger tumors.

The first three-category (i.e., small, medium, and large) size classification of PUM was developed in the 1950s.20 This classification, the Linear LBD, has been used in multiple publications since that time.4–6 This classification has only one size parameter, making it the easiest among the tumor size classifications evaluated in this study to use. Although we could find no explanation why LBD of 10 and 15 mm were selected as boundary lines for this classification, our study could find no explanation why LBD of 10 and 15 mm were classified as "large". Unfortunately, despite the ease of using this system, the prognostic significance was relatively low compared with the other systems we evaluated.

In 1975, Davidorf and Lang24 defined small PUM as having LBD × TH ≤ 30 mm² and all other tumors as large. In 1979, Thomas and colleagues21 used this same definition for small versus large PUMs. Both groups reported that patients having a small PUM had an extremely favorable prognosis for metastasis-free survival. As previously mentioned, these two “two-category” classifications were combined to form an evaluable three-category “Rectangular Area” classification with boundary lines at 30 mm² and 100 mm². By using this system, tumors within our cohort were fairly well distributed among the subgroups. Unfortunately, despite the ease of using this system, the prognostic significance was relatively low compared with the other systems we evaluated.

In 1980, Gass20 devised a classification system using tumor volume (LBD³). This classification adequately distributed the cohort among the subgroups and provided satisfactory discrimination among low-, intermediate-, and high-risk cases.

In 1984, Warren4 published his original classification system for PUM that combined LBD and TH. In his classification,
tumors were categorized as small if LBD was ≤10 mm and TH was ≤2 mm, large if LBD was >15 mm or TH was >5 mm, and medium if dimensions were between those two extreme categories. Warren8 justified his selection of his boundary lines by treatment guidelines used at the time. His system is relatively simple to use, having only two parameters and a relatively straightforward graphical representation. One drawback of the original Warren classification is the somewhat restrictive small and medium criteria categorizing most tumors as large. Furthermore, this classification does not provide strong prognostic significance based on subgroups generated.

By the mid to late 1970s, several ophthalmologists were using a modified Warren classification in which the thickness boundary line between small and medium tumors was changed from 2 mm to 3 mm.9–12 Individual authorship and rationale for this modification has not been established in the peer-reviewed literature, making it impossible to determine why this change was suggested. Like the original Warren classification, this
modified system still assigned the average tumor to the large category principally based on tumor thickness.

In 1993, Augsburger proposed an alternative system to the modified Warren classification attempting to decrease the impact of tumor TH in the categorization of PUM. He observed on a scatterplot of TH by LBD that PUM of a typical PUM was approximately half the LBD. This finding influenced the change in boundaries used by the Warren classification but chose 5.0 and 7.5 mm as TH boundaries. He retained 10 and 15 mm as LBD boundary lines. Although this classification succeeded in creating nearly equivalent patient subgroups, its prognostic significance was not as good as that of the Linear LBD system.

In 1988, the COMS planning group devised a tumor size classification used to direct patients into different treatment arms of that study. The original boundary lines for the small, medium, and large categories in this study are shown in Figure 1 (COMS original classification). In 1990 the COMS group modified this classification by moving the tumor thickness boundaries from 3 and 8 mm to 2.5 and 10 mm (Fig. 1, COMS modified classification).
Revised classification. Both classifications used LBD of 16 mm as the dividing boundary between large and smaller tumors. Additionally, both COMS classifications defined nevi as melanocytic choroidal tumors \( \leq 5\) mm in LBD and \( \leq 1\) mm in TH. Bear in mind that the COMS classifications were not developed for prognostic purposes but rather for therapeutic decision-making. In spite of this, these classifications perform reasonably well in discriminating patients as low, medium, and high risk of metastasis. Both classifications place most patients in the medium size category.

In 2002, the American Joint Committee on Cancer (AJCC) released a new staging system for PUM. This staging system was derived empirically from a collaborative database of more than 7000 patients with uveal melanoma. Because our study evaluated only classifications that divided patients into three size subgroups, two modifications of the TNM 2010 classification were generated by combining specified TNM 2010 tumor size categories. Although the frequency distribution of tumors was skewed to the small category for TNM 2010a and to the medium category for TNM 2010b, both systems exhibited excellent discrimination between the event curves and reasonably good prognostic significance compared with most other systems we evaluated.

In 2010, the AJCC released the seventh edition of its cancer staging manual. In this revision, tumor size was subdivided into four rather than three size categories (see Fig. 1, TNM 2010 classification). The boundary lines for LBD and TH for the four T categories were derived empirically from a collaborative database of more than 7000 patients with uveal melanoma. Because our study evaluated only classifications that divided patients into three size subgroups, two modifications of the TNM 2010 classification were generated by combining specified TNM 2010 tumor size categories. Although the frequency distribution of tumors was skewed to the small category for TNM 2010a and to the medium category for TNM 2010b, both systems exhibited excellent discrimination between the event curves and reasonably good prognostic significance compared with most other systems we evaluated.

### Table 3: Comparison of Tumor Size Classifications of PUMs Evaluated in This Study

<table>
<thead>
<tr>
<th>Classification</th>
<th>Category Assigned to Average Tumor</th>
<th>Proportional Size of Tumor Subgroups Assigned by Classification</th>
<th>Separation Between Event Rate Curves for Subgroups</th>
<th>Ease of Application of the Classification</th>
<th>Prognostic Significance of the Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear LBD</td>
<td>M</td>
<td>Unsatisfactory†‡</td>
<td>Excellent</td>
<td>0</td>
<td>Very good</td>
</tr>
<tr>
<td>Rectangular Area</td>
<td>M</td>
<td>Satisfactory</td>
<td>Good</td>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>Cubic Volume</td>
<td>M</td>
<td>Outstanding</td>
<td>Excellent</td>
<td>2</td>
<td>Good</td>
</tr>
<tr>
<td>Warren Original</td>
<td>L</td>
<td>Very unsatisfactory‡</td>
<td>Very poor</td>
<td>1</td>
<td>Poor</td>
</tr>
<tr>
<td>Warren Modified</td>
<td>L</td>
<td>Very unsatisfactory‡</td>
<td>Poor</td>
<td>1</td>
<td>Poor</td>
</tr>
<tr>
<td>Augsburger</td>
<td>M</td>
<td>Outstanding</td>
<td>Good</td>
<td>1</td>
<td>Good</td>
</tr>
<tr>
<td>COMS Original</td>
<td>M</td>
<td>Satisfactory</td>
<td>Good</td>
<td>1</td>
<td>Good</td>
</tr>
<tr>
<td>COMS Revised</td>
<td>M</td>
<td>Very unsatisfactory‡</td>
<td>Poor</td>
<td>1</td>
<td>Good</td>
</tr>
<tr>
<td>TNM 2002</td>
<td>M</td>
<td>Very unsatisfactory‡</td>
<td>Good</td>
<td>1</td>
<td>Good</td>
</tr>
<tr>
<td>TNM 2010a</td>
<td>S</td>
<td>Very unsatisfactory‡</td>
<td>Excellent</td>
<td>1</td>
<td>Very good</td>
</tr>
<tr>
<td>TNM 2010b</td>
<td>M</td>
<td>Very unsatisfactory‡</td>
<td>Excellent</td>
<td>1</td>
<td>Very good</td>
</tr>
</tbody>
</table>

* Tumor size is categorized as S (small), M (medium), or L (large).
† Proportional size of the tumor size subgroups in classification is rated as follows: Excellent: no subgroup smaller than 25% or larger than 40%, Satisfactory: no subgroup smaller than 20% or larger than 50%, Unsatisfactory: one subgroup smaller than 20% or larger than 50%, Very unsatisfactory: one subgroup smaller than 20% and one subgroup larger than 50%.
‡ Separation of the actuarial event rate curves for subgroups in the classification is rated as follows: Excellent: all curves are well separated throughout the follow-up period, Good: all curves are distinct throughout the follow-up period, Poor: two curves overlap substantially, Very poor: two curves overlap substantially, curve for a larger tumor size subgroup is lower than curve for a smaller tumor size subgroup.
§ Ease versus difficulty of application of the classification (see Discussion section).
|| Prognostic significance of the tumor size classification is categorized as follows: Very good: Log rank test for trend statistic \( >3.5\), Good: Log rank test for trend statistic \( 2.5 \) but \( <3.5\), Poor: Log rank test for trend statistic \( <2.5\).
‡ Although this classification is rated as “unsatisfactory” because the proportion of tumors categorized by it as small was 19.4%, it barely missed being categorized as satisfactory.

### Table 4: Ranking of Tumor Size Classifications by Log Rank Test of Trend Statistic From Kaplan-Meier Event Rate Curves Applied to 424 Cases of PUM

<table>
<thead>
<tr>
<th>Rank</th>
<th>Tumor Size Classification</th>
<th>Log Rank Test of Trend Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear LBD</td>
<td>51.9</td>
</tr>
<tr>
<td>2</td>
<td>TNM 2010a</td>
<td>42.4</td>
</tr>
<tr>
<td>3</td>
<td>TNM 2010b</td>
<td>40.5</td>
</tr>
<tr>
<td>4</td>
<td>TNM 2002</td>
<td>34.9</td>
</tr>
<tr>
<td>5</td>
<td>COMS Revised</td>
<td>31.9</td>
</tr>
<tr>
<td>6</td>
<td>Cubic Volume (LBD × SBD × TH)</td>
<td>28.1</td>
</tr>
<tr>
<td>7</td>
<td>Augsburger</td>
<td>25.7</td>
</tr>
<tr>
<td>8</td>
<td>Warren Modified</td>
<td>22.5</td>
</tr>
<tr>
<td>9</td>
<td>COMS Original</td>
<td>22.2</td>
</tr>
<tr>
<td>10</td>
<td>Warren Original</td>
<td>21.7</td>
</tr>
<tr>
<td>11</td>
<td>Rectangular Area (LBD × TH)</td>
<td>18.8</td>
</tr>
</tbody>
</table>

mm in TH, and medium (T2) tumors as intermediate in size. This system categorized the “average tumor” into the medium (T2) subgroup but assigned only 10.6% of our studied patients to the small (T1) category. The prognostic strength of the TNM 2002 was moderate when compared with other systems evaluated.

In 2010, the AJCC released the seventh edition of its cancer staging manual. In this revision, tumor size was subdivided into four rather than three size categories (see Fig. 1, TNM 2010 classification). The boundary lines for LBD and TH for the four T categories were derived empirically from a collaborative database of more than 7000 patients with uveal melanoma. Because our study evaluated only classifications that divided patients into three size subgroups, two modifications of the TNM 2010 classification were generated by combining specified TNM 2010 tumor size categories. Although the frequency distribution of tumors was skewed to the small category for TNM 2010a and to the medium category for TNM 2010b, both systems exhibited excellent discrimination between the event curves and reasonably good prognostic significance compared with most other systems we evaluated.

When evaluating and comparing tumor size classifications of PUM, several features of each classification must be taken into account, including those described in the following sections.

### The Size Category (Small, Medium, or Large) to Which the Classification Assigns an “Average Tumor”

As long as the patient group and tumors being evaluated in a study are representative of the entire size spectrum of PUM and not just a limited tumor size subgroup, a satisfactory classification should always assign the “average tumor” (i.e., hypothetical tumor having the mean LBD, SBD, and TH for the entire group by that classification) to the medium size category.
It is important to point out that several of the classifications evaluated (Table 3) did not achieve this goal.

The Proportion of Tumors Assigned to Small, Medium, and Large Categories by the Classification

In our opinion, a satisfactory classification of PUM should assign nearly equivalent numbers of patients to each of the tumor size categories. In this study, we regarded a tumor size classification as satisfactory if no subgroup contained <20% of the cases, or >50% of the cases. Six of the evaluated classifications were judged very unsatisfactory (Table 3) by this aspect of its separation of cases in this study.

The Ordered Separation Between Event Rate Curves of the Patients Assigned to the Small, Medium, and Large Categories by the Classification

A satisfactory tumor size classification should separate tumors into subgroups that yield distinct low, intermediate, and high-risk event rate curves. A classification that provides overlapping curves for any of the tumor size subgroups (Warren classifications, COM Revised classification) or a lower event curve for a larger size subgroup (original Warren classification) than for a smaller size subgroup cannot be regarded as an effective classification.

The Ease of Application of the Classification System

A linear boundary line classification separating tumors based on a single dimension (e.g., Linear LBD classification) is clearly easiest to apply. Such a classification does not necessitate reference to a graph or mathematical manipulation of the original data. Nonoverlapping “single-step” boundary line classifications (e.g., Warren Original, Warren Modified, Augsburger, TNM 2002) that subdivide tumors on the basis of both LBD and TH with discrete nonoverlapping small-medium and medium-large boundary lines are more difficult to apply given the need to reference a graph. Nonoverlapping “multiple-step” classifications (e.g., TNM 2010a and TNM 2010b) and classifications that have overlapping boundary lines (e.g., COMS Original and COMS Revised) that subdivide tumors on the basis of both LBD and TH are also more difficult to apply given the need for graph reference. These classifications require plotting of each tumor individually on predetermined category boundary lines. Curvilinear boundary line classifications separating tumors based on the product of two tumor dimensions (rectangular area classification, LBD × TH) or three tumor dimensions (cubic volume classification, LBD × SBD × TH) are most difficult to apply given the required mathematical manipulation followed by reference to a graph. In these cases, the original data need to be computed before these values can be used to separate the cases. In general, an easy-to-apply classification that also provides good proportional separation of the cases and good prognostic discrimination of the event curves for those subgroups is more effective in clinical situations.

The Prognostic Value of the Classification With Regard to the Event Rates of Interest

The clinical significance of a tumor size classification (i.e., its prognostic value) takes into account the magnitude of the separation among the relevant event curves for the tumor size subgroups as well as the proportional sizes of the subgroups. This significance was evaluated in this study using the logrank test of trend. This statistical method evaluates both the separation of the curves and whether they are ordered as expected. The larger the test statistic provided by the classification, the stronger the classification as a prognostic variable (Table 4).

Caveats

In our application of the TNM 2010 tumor size classification, we combined tumor size subgroups in ways that were not intended by the individuals who developed this classification. Readers should understand that our purpose was to make the TNM 2010a and TNM 2010b classifications directly comparable to the other classifications we evaluated. The sub classifications developed may not fully encompass the prognostic ability of this staging system compared with other classifications evaluated. Similarly, the Rectangular Area classification evaluated in this study was devised by different authors.19,24 The new three-part classification we generated may not reflect the original intent of the authors. We also recognize that the results of a study such as ours could possibly differ if applied to other PUM populations.

In conclusion, our study does not “anoint” any individual tumor size classification as the “optimal” or “best” system or provide sufficient evidence to replace all other classifications by a single system. However, it points out the limitations of many classifications and provides sufficient evidence to suggest that some of them are insufficiently discriminating to be in use. Although some classifications contain the best combinations of favorable ratings for each evaluated factor, overall the Linear LBD classification had the highest predictive value for PUM metastasis-free survival. Considering the Linear LBD system’s ease of application and prognostic value determined by this study of three-category classification systems, it seems debatable whether more intricate classification systems are really necessary in routine clinical practice.

Acknowledgments

Supported in part by an unrestricted grant from Research to Prevent Blindness, Inc., New York, NY, USA, to the Department of Ophthalmology, University of Cincinnati College of Medicine; the Augsburger Ocular Oncology Fund of the University of Cincinnati College of Medicine; the E. Vernon and Eloise C. Smith Chair of Ophthalmology Fund of the University of Cincinnati; and the Dr. E. Vernon and Eloise C. Smith Chair of Ophthalmic Pathology and Ocular Oncology Fund of the University of Cincinnati.

Disclosure: C.C. Skinner, None; J.J. Augsburger, None; B.D. Augsburger, None; Z.M. Correa, Castle Biosciences, LLC (C, R)

References


