Interobserver Variability of Mitotic Index and Utility of PHH3 for Risk Stratification in Gastrointestinal Stromal Tumors

Ahmad Alkhasawneh, MD,1 John D. Reith, MD,1 Tania Zuluaga Toro, MD,1 Ayed O. Ayed, MD,2 Xiaomin Lu, PhD,3 Thomas J. George, MD,2 and Lizette Vila Duckworth, MD1

From the 1Department of Pathology, Immunology and Laboratory Medicine, 2Department of Medicine, Division of Hematology and Oncology, and 3Department of Biostatistics, University of Florida College of Medicine, Gainesville.

Key Words: Anti-phosphohistone H3; Gastrointestinal stromal tumors; Ki-67; Mitoses; Soft tissue

ABSTRACT

Objectives: Accurate grading of gastrointestinal stromal tumors (GISTs), based on mitotic index, can be problematic.

Methods: In this study, we compared interobserver variability in detecting mitosis on H&E with PHH3 immunohistochemistry (IHC). In addition, we examined the correlation between H&E mitosis and Ki-67 and the association of PHH3 and Ki-67 with overall survival. Four pathologists independently reviewed 50 GIST cases.

Results: Intraclass correlation coefficients showed good interobserver variability for mitotic counts on both H&E (0.918; 95% confidence interval [CI], 0.874-0.950) and PHH3 IHC (0.923; 95% CI, 0.882-0.953). Nineteen (38%) cases were graded higher and five (10%) cases were downgraded by at least one observer using PHH3 compared with H&E. Using receiver operating characteristic curve analysis, a PHH3 cutoff of seven or more mitoses was associated with worse overall survival (P = .028). Ki-67 showed poor correlation with H&E mitotic counts and overall survival (P = .077).

Conclusions: PHH3 may thus be a valuable adjunct for risk stratification in GISTs.

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal neoplasms of the gastrointestinal tract. The biologic behavior of GISTs is determined by three parameters: location, size, and mitotic index, as outlined in the risk assessment guidelines by Miettinen and Lasota in 2006. GISTs are characterized based on mitotic count into low grade (≤5 mitoses/5 mm²) and high grade (>5 mitoses/5 mm²) on H&E-stained slides. The current edition of the AJCC Cancer Staging Manual introduced a staging system for GISTs based on four key prognostic parameters, including tumor size (T), lymph node status (N), metastasis (M), and histologic grade (G).

Mitotic index can be subject to intra- and interobserver variability. The reliability of mitotic counts on H&E slides is limited by several factors, including the ability to find the “hot spots” with highest mitotic rates and the ability to differentiate true mitotic figures from apoptotic bodies, pyknotic cells, and karyorrhectic debris. On the other hand, having an overly rigid definition for mitoses may result in underestimating the mitotic counts in certain cases, resulting in the classification of a GIST as low grade when it actually behaves aggressively. Proliferation markers such as Ki-67 are not routinely used in the assessment of GISTs and may overestimate the mitotic activity of the tumor, although some studies have shown Ki-67 to be an independent marker of poorer prognosis in GISTs.

Recently, immunohistochemistry (IHC) for the phosphorylated form of histone 3 (PHH3), which is present during early prophase, has been shown to be a reliable mitosis-specific marker. PHH3 has been useful in the assessment of the proliferation index in gliomas, meningiomas, melanomas, uterine smooth muscle tumors, and pulmonary neuroendocrine carcinomas. It has also been shown...
to decrease interobserver variability in detecting mitosis in melanomas.\textsuperscript{17} PHH3 should be interpreted as positive only in cells with a staining pattern that highlights a mitotic figure.

In this study, we compared interobserver variability of mitotic activity in GISTs using H&E- and PHH3-stained slides. We also evaluated the correlation between H&E mitotic counts and Ki-67 labeling index and the association of PHH3 and Ki-67 with overall survival. In addition, we attempted to define a PHH3 cutoff value that best correlated with overall survival. To our knowledge, there are no published reports to date on the appropriate PHH3 cutoff value for grading/risk stratification in GISTs.

### Materials and Methods

#### Case Selection

This study was approved by the Institutional Review Board of the University of Florida. Initially, we identified 70 consecutive cases (≥2 cm) with a diagnosis of GIST resected between January 2000 and August 2011 at our institution. Histologic sections were reviewed by a senior gastrointestinal and soft tissue pathologist (J.D.R.) to ensure accuracy of prior diagnosis. On the basis of the original pathology reports, we included all cases with more than five mitoses per 5 mm\(^2\) (14 cases) to ensure a fair representation of the less common but more aggressive tumors. Thirty-six of the remaining 56 cases with five or fewer mitoses per 5 mm\(^2\) were randomly selected, for a total of 50 cases. All slides from each case were reviewed to identify the slide with the most “hot spots” (areas with readily identifiable mitotic figures). In cases lacking prominent mitoses, slides with hypercellular areas and/or nuclear pleomorphism were selected.

#### Histologic Grade on H&E

H&E slides were available for all 50 cases and were reviewed independently by four pathologists: observer 1: pathologist-in-training (A.A.), observer 2: gastrointestinal pathology fellow (L.V.D.), observer 3: junior gastrointestinal and soft tissue pathologist (T.Z.T.), and observer 4: senior gastrointestinal and soft tissue pathologist (J.D.R.). The mitotic index was obtained from counting mitotic figures in a 5-mm\(^2\) area (Olympus microscope; 21 high-power fields [hpf] at ×400). Positive PHH3 staining was defined as strong chromatin staining with characteristic morphologic features of mitotic figures. Prophase nuclei, which are PHH3 labeled, were excluded from mitotic counts. To minimize any intraobserver bias, each pathologist performed examination of the H&E and PHH3 slides separately (ie, H&E slides were reviewed first, and only after these slides had been passed off to another pathologist and results submitted to the principal investigator were the PHH3 slides reviewed by the pathologist).

#### IHC for PHH3 and Ki-67

IHC studies were performed on the same block as the corresponding H&E slide using freshly cut 4-μm slides. We used anti-PHH3 monoclonal antibody (prediluted; Cell Marque, Rocklin, CA) with a BenchMark automated immunohistochemical stainer (Ventana Medical Systems, Tucson, AZ) for PHH3 IHC and MIB-1 monoclonal antibody (1:250 dilution; DAKO, Carpinteria, CA) with a DAKO Autostainer Plus for Ki-67 IHC according to established protocols.

The PHH3 mitotic index was obtained manually by counting the number of PHH3-labeled mitotic figures in a 5-mm\(^2\) area (Olympus microscope; Olympus, Pittsburgh, PA; 21 high-power fields [hpf] at ×400). The histologic grade and clinical stage were determined according to the seventh edition of the AJCC Cancer Staging Manual.\textsuperscript{8}

#### Table 1

Prognosis of Gastrointestinal Stromal Tumor Based on Long-Term Follow-up\textsuperscript{a}

<table>
<thead>
<tr>
<th>Tumor Parameters</th>
<th>Size, cm</th>
<th>Gastric</th>
<th>Duodenum</th>
<th>Jejunum/Ileum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitotic Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5 per 50 high-power fields (hpf)</td>
<td>≤2</td>
<td>None (0)</td>
<td>None (0)</td>
<td>None (0)</td>
</tr>
<tr>
<td></td>
<td>&gt;2 to ≤5</td>
<td>Very low (1.9)</td>
<td>Low (8.3)</td>
<td>Low (4.3)</td>
</tr>
<tr>
<td></td>
<td>&gt;5 to ≤10</td>
<td>Low (3.6)</td>
<td>Insufficient data</td>
<td>Moderate (24)</td>
</tr>
<tr>
<td></td>
<td>&gt;10</td>
<td>Moderate (10)</td>
<td>High (34)</td>
<td>High (52)</td>
</tr>
<tr>
<td>&gt;5 per 50 hpf</td>
<td>≤2</td>
<td>None/insufficient data</td>
<td>Insufficient data</td>
<td>High/nsufficient data</td>
</tr>
<tr>
<td></td>
<td>&gt;2 to ≤5</td>
<td>Moderate (16)</td>
<td>High (60)</td>
<td>High (73)</td>
</tr>
<tr>
<td></td>
<td>&gt;5 to ≤10</td>
<td>High (65)</td>
<td>Insufficient data</td>
<td>High (85)</td>
</tr>
<tr>
<td></td>
<td>&gt;10</td>
<td>High (86)</td>
<td>High (88)</td>
<td>High (90)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Adapted from Miettinen and Lasota.\textsuperscript{4}
Clinical Characteristics

The following clinical information was recorded from the medical record: age, sex, location, tumor size, clinical stage, associated pathologic findings, therapy received (adjuvant and/or neoadjuvant), and survival status. Death was determined through the National Death Index query performed on July 2, 2013.

Statistical Analysis

Pairwise interobserver and intermethod reliabilities were evaluated using Pearson correlation coefficients in SAS 9.2 (SAS Institute, Cary, NC). The corresponding 95% confidence interval was computed based on Fisher transformation.

The overall interobserver reliability among the four observers was analyzed in SPSS version 20 (SPSS, Chicago, IL). For each method, an averaged intraclass correlation coefficient (ICC) was obtained based on a two-way random-effect model with Spearman-Brown correction. Receiver operating characteristic (ROC) curve analysis was used to determine the appropriate PHH3 cutoff value that best correlated with overall survival based on mitotic count of the observer with the largest area under the curve (AUC). Survival curves were compared using the log-rank test, and the survival rate was estimated using the Kaplan-Meier method. All ROC curve and survival analyses were performed using SAS 9.2.

Results

Intra- and Interobserver Variability

Results for PHH3 and Ki-67

Pearson correlations showed very good intraobserver variability between H&E and PHH3 mitotic counts for each
The number of cases graded differently based on PHH3 vs H&E (using the established criteria of \( \leq 5 \) mitoses/5 mm\(^2\) for low grade and \( >5 \) mitoses/5 mm\(^2\) for high grade) for each observer is summarized in Table 3. Overall, a higher grade was obtained in 19 (38%) cases using PHH3 IHC vs H&E mitoses, and a lower grade was obtained in five (10%) cases using PHH3 IHC. Based on tumor site, 15 (42%) of 36 gastric GISTs were upgraded using PHH3 by at least one observer, and five (14%) of 36 gastric GISTs were downgraded. In the small intestine GISTs, four (29%) of 14 cases were upgraded using PHH3 by at least one observer, and none were downgraded. Based on tumor size, most of the gastric cases that were upgraded using PHH3 were 2 to observer. The intraobserver variability results of mitotic counts using H&E compared with PHH3 immunohistochemistry (A-D). 

\[ R = 0.80 \quad (A), \quad R = 0.92 \quad (B), \quad R = 0.71 \quad (C), \quad R = 0.72 \quad (D). \]

Comparison of Histologic Grading Using H&E Mitoses vs PHH3 IHC

The mean mitotic count of all tumors was 8.0 mitoses/21 hpf using H&E and 11.4 mitoses/21 hpf using PHH3.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Observer & Mitosis & PHH3 & Mitosis & PHH3 \\
\hline
1 & 27.5 & 38.5 & 49.5 & \\
2 & 52.5 & 73.5 & 94.5 & \\
3 & 60 & 84 & 108 & \\
4 & 88 & 24 & 40 & 56 & 72 & 88 \\
\hline
\end{tabular}
\caption{Comparison of mitotic counts using H&E and PHH3 IHC.}
\end{table}
5 cm (11 of 15 cases [73%]), and most of the small bowel cases that were upgraded were 5 to 10 cm (three of four cases [75%]).

Clinical Characteristics and Survival

A total of 50 GIST cases were evaluated from 19 men and 31 women with a mean age of 62.5 years (range, 33-85 years). The mean tumor size was 7.3 cm (range, 2.1-27 cm). The location of tumors was as follows: 36 in the stomach, four in the duodenum, and 10 in the ileum or jejunum. Due to the low number of duodenal cases and similar prognosis to other small bowel GISTs (ie, jejunal and ileal), all small bowel GISTs were grouped as one category for analysis. Nineteen patients received adjuvant therapy, and one patient received neoadjuvant therapy with a tyrosine kinase inhibitor. Overall survival rate at the time of investigation (mean follow-up, 4.7 years; range, 9 months to 13.4 years) was 82% (41/50). On the basis of the histologic grade established from H&E-stained slides, three (8.3%) of 36 patients with low-grade tumors and six (42.9%) of 14 patients with high-grade tumors died of disease.

ROC Curve Analysis and PHH3 Cutoff

ROC curves were analyzed to determine the prognostic significance of PHH3 and Ki-67 mitotic indices in relation to survival. ROC analysis of each observer showed PHH3 to have strong association with overall survival, with observer 3 having the largest AUC (0.836; P < .0001). Using ROC data of observer 3 revealed a PHH3 of seven or more mitoses to be the most appropriate cutoff value in relation to overall survival (sensitivity, 0.778; specificity, 0.756)

Discussion

Accurate assessment of mitotic activity in GISTs is essential in predicting the risk of disease progression (metastasis and/or tumor-related death).1-7 GISTs are graded based on mitotic count into low grade (≤5 mitoses/5 mm²) and high grade (>5 mitoses/5 mm²) on standard H&E-stained slides, which can be subject to intra- and interobserver variability. This may result in errors and inconsistencies in assessing mitotic count, especially when the index is close to five. Accurate histologic grading is needed for proper clinical staging and medical treatment decisions.
In clinical practice, risk categorization using mitotic indices is critical for the appropriate patient selection for adjuvant tyrosine kinase inhibitor use. Using PHH3 in this study, most of the upgraded gastric GIST cases were 2 to 5 cm, which places these patients in the intermediate risk category (stage II instead of stage Ia). Most of the upgraded small bowel GIST cases were 5 to 10 cm, which places these patients in the high-risk category (stage IIIb instead of stage II). Interestingly, none of the small bowel GIST cases in our series were downgraded by PHH3, which is not surprising given that most of these tumors are already considered high risk based on tumor size and location.

The duration of imatinib therapy in GISTs may vary according to histologic grade. According to the 2012 National Comprehensive Cancer Network guidelines, 3 years of postoperative imatinib therapy is recommended for high-risk patients (tumor size >5 cm and >5 mitoses/5 mm²), which is based on the Scandinavian Sarcoma Group XVIII trial that showed improved overall survival and recurrence-free survival for patients treated for 3 years.

### Table 3

<table>
<thead>
<tr>
<th>Tumor Location</th>
<th>Histologic Grade vs H&amp;E</th>
<th>Size, cm</th>
<th>No. of Cases</th>
<th>Total No. of Cases With a Change in Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach</td>
<td>Higher grade using PHH3</td>
<td>&gt;2 to ≤5</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;5 to ≤10</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Lower grade using PHH3</td>
<td>&gt;2 to ≤5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;5 to ≤10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Small bowel</td>
<td>Higher grade using PHH3</td>
<td>&gt;2 to ≤5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;5 to ≤10</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Lower grade using PHH3</td>
<td>&gt;2 to ≤5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;5 to ≤10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;10</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

* Using criteria of low grade (≤5 mitoses/5 mm²) and high grade (>5 mitoses/5 mm²).

In clinical practice, risk categorization using mitotic indices is critical for the appropriate patient selection for adjuvant tyrosine kinase inhibitor use. Using PHH3 in this study, most of the upgraded gastric GIST cases were 2 to 5 cm, which places these patients in the intermediate risk category (stage II instead of stage Ia). Most of the upgraded small bowel GIST cases were 5 to 10 cm, which places these patients in the high-risk category (stage IIIb instead of stage II). Interestingly, none of the small bowel GIST cases in our series were downgraded by PHH3, which is not surprising given that most of these tumors are already considered high risk based on tumor size and location.

The duration of imatinib therapy in GISTs may vary according to histologic grade. According to the 2012 National Comprehensive Cancer Network guidelines, 3 years of postoperative imatinib therapy is recommended for high-risk patients (tumor size >5 cm and >5 mitoses/5 mm²), which is based on the Scandinavian Sarcoma Group XVIII trial that showed improved overall survival and recurrence-free survival for patients treated for 3 years.

### Figure 2

Receiver operating characteristic curve using PHH3 mitotic count in relation to overall survival for gastrointestinal stromal tumors.

### Figure 3

Kaplan-Meier survival curve for gastrointestinal stromal tumors based on H&E histologic grade (A) and PHH3 (B) at a cutoff point of seven mitoses. \( P = .0058 \) (A) and \( P = .0278 \) (B).
as opposed to only 1 year of therapy. In our study, four (27%) of 15 patients with gastric tumors greater than 5 cm were upgraded using PHH3 IHC by any one observer and may have benefited from longer duration of imatinib therapy. In contrast, three (20%) of 15 patients with gastric tumors greater than 5 cm were downgraded with PHH3. A potential bias in our study, however, is that some cases predated the introduction of modern imatinib therapy for GISTs, which could influence recurrence-free and overall survival.25

A limitation of our study is that we did not include tumors less than 2 cm to ensure adequate field areas to be examined to reach 5 mm². Although change in histologic grade in gastric tumors less than 2 cm may not affect overall outcome, a change in grading of small bowel GISTs less than 2 cm may affect patient prognosis. Also, while we attempted to correlate H&E and PHH3 mitotic counts and Ki-67 labeling index with survival outcome, our sample size and follow-up data are too small to allow for definitive prognostic assessment. Our sample size of small bowel GISTs was too small to investigate if there is a different PHH3 cutoff that should be used for prognostication at this site. Future studies with a larger series of patients, including a greater number of small bowel GISTs, and longer follow-up data are necessary for comprehensive risk assessment in these patients.

Additional studies may also be warranted to examine the prognostic significance of PHH3 in extragastrointestinal stromal tumors (EGISTs), which, despite having a similar morphology and immunoprofile to GISTs, follow a more aggressive clinical course and have a lower mitotic threshold (>2 mitoses/50 hpf) for malignancy than gastric or small bowel stromal tumors.26 Analyzing our data using the same cutoff value for EGISTs (>2 mitoses) reveals a change in mitotic index in 30 of 50 cases using PHH3 compared with H&E and suggests the grading of sarcomas might change using PHH3 for the mitotic index.

In summary, this study compared interobserver variability for mitotic counts using H&E and PHH3 IHC in GISTs and examined the role of Ki-67 labeling index. Our results demonstrated a tendency to undergrade GISTs based on H&E compared with PHH3, which alters the stage, risk of disease progression, and treatment recommendations. In addition, we attempted to correlate PHH3 data with clinical outcomes in GISTs and found that a PHH3 cutoff value of seven mitoses or higher was associated with worse overall survival in our series. Ki-67 proved to have poor correlation with H&E mitosis and was not prognostic of overall survival compared with PHH3. Revision of the current grading system may be warranted to include a mitosis-specific marker such as PHH3 for overall risk stratification in GISTs; however, larger studies are needed to validate the prognostic value of PHH3 in GISTs.

References


