Histological Classification of Gastric Adenocarcinoma for Epidemiological Research: Concordance between Pathologists

Atsuko Shibata,² Teri A. Longacre, Balaram Puligandla, Julie Parsonnet, and Laurel A. Habel

Abstract
Epidemiology of gastric adenocarcinoma suggests that intestinal-type and diffuse-type cancers develop through distinct causal pathways. To examine the differences in risk factors and molecular changes between the histological types, reliable data on histological typing are essential. We evaluated the concordance between two pathologists in assessment of 95 gastric adenocarcinomas for Laurén classification and tumor grade. Two pathologists, each blinded to the other’s assessment, reviewed H&E-stained slides of gastric tumor. The responses of the two pathologists for histological type were considered as concordant if they fell on one of the three categories (intestinal type, diffuse type, or other). Tumor grade was classified into three categories (well, moderately, or poorly differentiated). The pathologists agreed on the classification of histological type for 71 of 92 (77%) tumors. $\kappa$ coefficient was 0.59 (95% confidence interval, 0.44–0.73). Concordance for tumor grade was 87%, with a $\kappa$ coefficient of 0.72 (95% confidence interval, 0.57–0.87). Both observed concordance and $\kappa$ coefficient for histological type and tumor grade were similar across three calendar periods of study. Interobserver agreement was virtually identical between tumors with biopsy specimens only and those with surgical specimens. Although the level of disagreement for histological type observed in this study is comparable with that in other studies, the resulting misclassification would lead to the reduction in observed differences in prevalence and odds ratio estimates between two histological types.

Introduction
The most common form of stomach cancer is adenocarcinoma (1). Laurén (2) classified gastric adenocarcinoma into two histological types, intestinal and diffuse, according to morphological features of tumor. The decrease in intestinal-type tumors largely accounts for the declining trends of stomach cancer incidence (3), suggesting larger influences of environmental risk factors on intestinal-type cancer than diffuse-type cancer. Preneoplastic lesions, such as intestinal metaplasia, are often associated with intestinal-type cancer but are infrequently associated with diffuse-type cancer. Both observations suggest that the two histological types develop through distinct pathways (4).

Investigators have examined risk factors and molecular changes that might be differentially associated with various histological types. In such investigations, consistent and reliable classification of tumors is essential. As a part of our molecular epidemiological study of stomach cancer, we examined the concordance between two pathologists in assessment of Laurén classification and tumor grade. We also assessed the effects of misclassification in histological typing on the observed differences in prevalence and OR estimates between intestinal-type- and diffuse-type tumors.

Materials and Methods
Study Subjects. Study subjects were identified for a study of Helicobacter pylori serology and molecular changes in gastric cancer from the members of the KPMC of Northern California who had participated in the MPHC program between 1964 and 1969. Diagnoses of gastric cancer between the date of participation in the MPHC program and June 1996 were ascertained by linkage of the MPHC database with the KPMC’s local tumor registry and computer-stored hospitalization records. The study protocol has been approved by the human subjects panels of Stanford University and the Kaiser Foundation Research Institute.

We first screened pathology reports and H&E-stained slides to confirm the eligibility of subjects. The following tumors were excluded from this study: adenocarcinoma of the gastroesophageal junction; gastric neoplasm other than adenocarcinoma; metastatic tumor to the stomach from another organ; and metastatic cancer of gastric origin without concomitant histological material from primary tumor in the stomach. The present analysis includes 95 cases of gastric adenocarcinoma, and tumor specimens of these cases were reviewed by two study pathologists before August 1999.

1 The abbreviations used are: CI, confidence interval; IHC, immunohistochemistry; KPMC, Kaiser Permanente Medical Care Program; MPHC, multiphasic health checkup; OR, odds ratio.
Short Communication: Concordance of Gastric Cancer Typing

Assessment of Interobserver Agreement. Responses of the two pathologists regarding histological type were considered as concordant if their responses both fell on one of the three categories (i.e., intestinal type, diffuse type, or other). Observed concordance was calculated as a percentage of tumors with concordant responses (i.e., tumors that fell on the diagonal cells in a $3 \times 3$ table, such as Table 1).

<table>
<thead>
<tr>
<th>Pathologist 1</th>
<th>Pathologist 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intestinal type</td>
<td>42</td>
<td>5</td>
</tr>
<tr>
<td>Diffuse type</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>32</td>
</tr>
</tbody>
</table>

*Observed concordance $= (42 + 27 + 2)/92 = 77\%$. $k$ coefficient $= 0.59$ (95% CI, 0.44–0.73).

Histological Review of Tumor Specimens. H&E-stained slides from endoscopic biopsies and gastrectomies were obtained from pathology departments and a storage facility of the KPMCP. Biopsy specimens alone were available for 42 of 95 (44%) subjects. For each subject, one pathologist reviewed all of the available slides (mean = 4.7 slides/subject, including slides that did not contain tumor) and selected slides containing most representative tumor sections. The selected slides (mean = 1.8 slides/subject) were forwarded to the second pathologist. The two pathologists, each of whom was blinded to the other’s assessment, recorded their assessment for histological type and grade of tumor on an evaluation form. For each tumor, pathologists were instructed to identify a type [intestinal type, diffuse type, or other (undifferentiated or borderline)] and indicate the degree of differentiation (well, moderately, or poorly differentiated; Ref. 5).

Initial review of the 95 tumors was conducted between November 1997 and July 1999. The pathologists held joint review sessions in July 1998 and April 1999 to resolve disagreements between their assessments. Data analyzed in this study came from the initial assessments made by each pathologist before these joint reviews.

Concordance of Tumor Grade. Agreement of tumor grade assessment was examined similarly, using three categories (well, moderately, and poorly differentiated).

In addition to an overall estimate of agreement among all of the cases reviewed, we examined concordance by calendar period of review (three periods separated by the two joint review meetings) and by the availability of surgical specimens. Because of the time lag between initial assessments by two pathologists, as well as a delay between the initial and joint reviews of discordant typing, there were some unavoidable overlaps in the timing of each pathologist’s review between adjacent calendar periods.

Concordance of Tumor Grade

Responses on tumor grade were available for 91 tumors. Poorly differentiated tumors were the most common (67% and 66% by two pathologists, respectively). Overall agreement for tumor grade was 87%, with a $k$ coefficient of 0.72 (95% CI, 0.57–0.87; Table 2). The observed concordance for the three calendar periods was 93%, 84%, and 89%, with $k$ coefficients of 0.85 (95% CI, 0.64–1.00), 0.64 (95% CI, 0.42–0.86), and 0.73 (95% CI, 0.37–1.00). Interobserver agreement was virtually identical between cases with biopsy specimens only (concordance, 88%; $k$, 0.70) and those with surgical specimens (concordance, 86%; $k$, 0.72).

Effects of Misclassification on Prevalence and OR Estimates

Example 1: Prevalence of p53 IHC-positive Tumors. If 10% of 100 intestinal-type tumors are misclassified as diffuse type, that a certain percentage (10% or 15%) of intestinal-type tumors was misclassified as diffuse type and that the same percentage of diffuse-type tumors was misclassified as intestinal type, independent of p53 IHC and $H. pylori$ status (i.e., non-differential misclassification). The two percentages used in the examples were chosen to represent the level of disagreement between pathologists observed in this study. For the purpose of these illustrations, we assumed, although it is unrealistic, that there was no measurement error in p53 IHC and $H. pylori$ status.

For the example of p53 IHC, we assumed that the true prevalence of p53-positive tumors was 60% in intestinal-type tumors and 30% in diffuse-type tumors (reasonable estimates based on work by A. S.4). We estimated the difference between the two percentages observed in a study including 200 gastric tumors (100 intestinal-type and 100 diffuse-type tumors) when a specified level of misclassification occurred.

For the example of $H. pylori$ and gastric cancer risk, we assumed that the true ORs were 5.0 for intestinal-type tumors and 10.0 for diffuse-type tumors. Assuming that 60% of control subjects were $H. pylori$ positive, these ORs are translated into the $H. pylori$ prevalence of 88% and 94% in subjects with intestinal-type and diffuse-type tumors, respectively. We estimated observed ORs for intestinal-type and diffuse-type tumors if the tumors were misclassified at a specified level, in a hypothetical case-control study with 200 cases (100 intestinal-type and 100 diffuse-type tumors) and 200 controls.

### Table 1 Overall concordance in histological classification between pathologists

<table>
<thead>
<tr>
<th>Pathologist 1</th>
<th>Intestinal type</th>
<th>Diffuse type</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologist 2</td>
<td>42</td>
<td>5</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>Diffuse type</td>
<td>7</td>
<td>27</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>32</td>
<td>10</td>
<td>92</td>
</tr>
</tbody>
</table>

*Observed concordance $= (42 + 27 + 2)/92 = 77\%$. $k$ coefficient $= 0.59$ (95% CI, 0.44–0.73).
and 10% of 100 diffuse-type tumors are misclassified as intestinal type. 100 tumors classified as intestinal type include 90 “true” intestinal-type and 10 “true” diffuse-type tumors. If 60% of “true” intestinal-type and 30% of “true” diffuse-type tumors are p53 IHC positive (Table 3A), the number of p53 IHC-positivity between intestinal-type and diffuse-type tumors is 21%. Similarly, if 15% of tumors are misclassified, the difference in prevalence of p53 IHC tumors between the two types is 57%–33% = 24%, which is smaller than the true difference of 30%. Similarly, if 15% of tumors are misclassified, the difference in p53 IHC positivity between intestinal-type and diffuse-type tumors is 21%.

**Example 2: OR for Gastric Cancer Associated with H. pylori Infection.** When 10% of case subjects are misclassified, the observed OR for intestinal-type tumors increases from 5.0 to 5.4 (Table 3B), whereas the observed OR for diffuse-type tumors decreases from 10.0 to 8.9, resulting in a smaller ratio of the two ORs (1.6 rather than 2). If the misclassification occurs in 15% of tumors, the ORs for intestinal-type and diffuse-type tumors are even closer to each other.

**Discussion**

Frequently used histological classifications for gastric adenocarcinoma (intestinal-type and diffuse-type carcinoma) are attributed to Laurén (2). Laurén himself observed a stronger male preponderance and older age at diagnosis for intestinal-type cancer than for diffuse-type cancer (2), which was corroborated by Hanai et al. (7).

Histological classification of cancers, as is the case with all measurements, is prone to error. Histological typing and grading are inherently subjective because they depend on a pathologist’s judgement in the absence of objective “gold standard” measures. Pathologists would be more likely to disagree for tumors with overlapping histological patterns, which consist of as high as 15% of all gastric cancers (1, 8). Given the absence of a “gold standard,” examining the level of agreement between two or more pathologists is a feasible alternative to estimate the reliability of histological data.

Our study showed an overall agreement for Laurén classification between two pathologists in 77% of the 92 gastric adenocarcinomas; a κ coefficient of 0.59 is considered as “fair to good” agreement (6). Both observed concordance and κ coefficient for histological type were consistent across the three calendar periods and regardless of whether surgical specimens of tumor were available in addition to biopsy specimens. The level of agreement was higher for tumor grade, which was somewhat surprising to us because we had expected that determination of tumor grade would be more subjective and thus more variable between pathologists. For example, in a study of 17 gastric cancers reviewed by 14 pathologists (8), agreement with accepted reference pathology was higher for Laurén classification (89%) than for tumor grade (63%).

Other studies also examined the reliability of histological typing and grading for gastric cancers. Palli et al. (9) observed agreement among six pathologists in 70–80% of the 100 gastric cancers reviewed for histological classification. Agreement between repeat readings with an up to 3-year interval by the same pathologist was also high (95%) (Ref. 9). In a study by Hansson et al. (10), one pathologist repeated a review of gastric cancer specimens after a 2-year interval. Overall agreement and κ coefficient, respectively, for Laurén classification was 79% and 0.63 for 67 resection materials and 83% and 0.70 for 88 biopsy materials. These results are comparable with our findings.

Distinct pathways of carcinogenesis for the two histological types of Laurén were suggested by more frequent observations of precancerous lesions in gastric mucosa adjacent to intestinal-type tumors (2, 9). Subsequently, many studies examined risk factors (e.g., *H. pylori* infection) (Ref. 11) and molecular markers (e.g., p53 abnormalities) for intestinal-type and diffuse-type tumors separately. Lack of reliability in histological assessment in these investigations would lead to misclassification of tumors, potentially resulting in inconclusive or erroneous findings. In our study, two pathologists disagreed on histological type for 24% of tumors; *i.e.*, an average of 12% of tumors classified as intestinal type by one pathologist was classified as diffuse type by the other pathologist, and *vice versa*. As our hypothetical examples show, this level of misclassification would result in reduced differences in prevalence and OR estimates between histological types, such that those differences might be discounted. It would be difficult to assess the degree and direction of bias for actual studies, in which additional measurement error occurs in other variables, and histological misclassification might not be independent of other marker measurements.

Extrapolation of our findings to other settings may be limited due to some aspects of study design. First, a second pathologist reviewed only a subset of slides that a first pathologist considered as most representative, which may have resulted in overestimated κ coefficients. Second, the joint review by two pathologists may have influenced subsequent typing of tumors, although we observed no increase in concordance over the calendar periods.

In conclusion, we observed reasonably good agreement...
between pathologists for Laurén classification and grade of gastric cancers. A central review of pathology material and resolution of disagreements between pathologists will be practical alternatives to improve the reliability of histological data in epidemiological studies.

Acknowledgments

We thank Maria Isabel Garcia, Lucy Bonds, Bonnie Bell, and Bill Frank for technical assistance; the pathology departments of Kaiser Permanente Medical Centers in Northern California for tumor specimens; and Dr. Gary Friedman for access to the KPMCP databases.

References

Histological Classification of Gastric Adenocarcinoma for Epidemiological Research: Concordance between Pathologists

Atsuko Shibata, Teri A. Longacre, Balaram Puligandla, et al.

Cancer Epidemiol Biomarkers Prev 2001;10:75-78.

Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/10/1/75

Cited articles
This article cites 5 articles, 1 of which you can access for free at:
http://cebp.aacrjournals.org/content/10/1/75.full.html#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
/content/10/1/75.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.