Influence of Site Classification on Cancer Incidence Rates: An Analysis of Gastric Cardia Carcinomas

Douglas A. Corley, Ai Kubo

Background: Recent reports suggest that the incidences of cardia and gastroesophageal junction carcinomas have increased markedly. The influence of improvements in cancer site classification (i.e., from no specific site to a specific site) on these incidence rates is unknown. Methods: We analyzed data for all gastric cancers reported to the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) cancer registries from 1974 through 1998. We compared incidence figures adjusted for improvements in site classification with the standard unadjusted incidence rates traditionally reported from SEER data. All analyses used two-sided statistical tests. Results: Among white males, the proportion of gastric cancers with an unspecified location decreased from 38% from 1974 to 1976 to 14% in 1996 to 1998. Between 1974–1976 and 1996–1998, the adjusted cardia cancer incidence rate for white males was unchanged (5.3% increase, from 3.6 to 3.8 per 100 000 population/year, respectively; \( P = .59 \)), whereas the unadjusted cardia cancer incidence rate underwent a statistically significant increase (77% increase, from 1.9 to 3.4 per 100 000 population/year, respectively; \( P = .001 \)). During the same period, the adjusted noncardia gastric cancer incidence rate in white males decreased from 6.8 to 3.8 per 100 000 population/year (\( P < .001 \)), an absolute decrease more than twice as large as that seen using standard unadjusted SEER data (from 4.5 to 3.2 per 100 000 population/year; \( P < .001 \)). Similar findings were observed for black males. Conclusions: Improved specification of gastric cancer sites may largely account for the purported increase in cardia cancer incidence in recent decades. Noncardia gastric cancer incidence may be decreasing much more rapidly than previously appreciated. These results illustrate the potentially large influence of changes in site classification on some cancer incidence rates. [J Natl Cancer Inst 2004; 96:1383–7]

The tremendous research and clinical activities generated by reported changes in cancer incidence rates emphasize the need to thoroughly understand the potential biases contained in such rates. A cancer incidence rate contains multiple potential sources for error because it includes both a numerator (e.g., the number of new cancer cases, which is usually obtained from cancer registries) and a denominator (e.g., the size of the overall population, which is typically obtained from census estimates). Inaccurate census estimates of the underlying population or changes in cancer site classification, for example, may substantially influence reported cancer rates (8,9).

Therefore, the reported increase in the incidence of gastric cardia adenocarcinomas could result from any of several factors: 1) a true increase in cancer incidence; 2) improved site classification, whereby cancers previously assigned to no specific site (e.g., gastric cancer, site not otherwise specified) are now classified according to a specific location within the stomach (e.g., antrum, body, or cardia); 3) increased reporting bias, whereby cancers are now more likely to be differentially classified as being in the gastric cardia or esophagus (e.g., a given gastric or esophageal cancer is now more likely to be assigned to these locations than it used to be, even in the absence of improved classification); or 4) a combination of these factors.

The influence of improved site classification on cardia cancer incidence rates is unknown and potentially important. If, for example, 50% of the gastric cancers reported in 1974 were not designated as being at a specific site location within the stomach, then the reported incidence of cancer at any given gastric site in 1974 might be only half its actual value. If site classification then improved such that the proportion of cancers without a designated location decreased to 25% over subsequent years, the reported incidence of cancer at a designated location might increase by 50%, even though no real change in incidence had occurred.

We assessed the influence of improvements in site classification over time on the reported incidences of gastric cancers within the SEER cancer registry, a unified, high-quality reporting system covering a large portion of the U.S. population. This evaluation might provide new insights into cancer incidence data and information on potential biases in reports of cancer incidence trends.

Affiliations of authors: Northern California Kaiser Permanente Division of Research (DAC); Division of Gastroenterology, Department of Medicine, University of California, San Francisco (DAC); Mailman School of Public Health, Columbia University, New York, NY (AK).

Correspondence to: Douglas A. Corley, MD, MPH, PhD, Kaiser Permanente Division of Research, 2000 Broadway, Oakland, CA 94612-2304 (e-mail: douglas.corley@kp.org).

See “Notes” following “References.”

DOI: 10.1093/jnci/djh265
Journal of the National Cancer Institute, Vol. 96, No. 18, © Oxford University Press 2004, all rights reserved.
METHODS

Data Sources

We analyzed SEER cancer registry data for the years 1974 through 1998. During this time, the SEER program collected data from nine population-based tumor registries throughout the United States and used U.S. census data to calculate incidence rates (10).

Definitions

For anatomic site and histology definitions, we used the following site codes from the International Classification of Disease for Oncology, second edition (ICD-O-2) (11): gastric cardia (defined as cardioesophageal junction, esophagogastric junction, and gastroesophageal junction) (C16.0); gastric fundus (C16.1); overlap cancers (i.e., cancers that overlap more than one cancer site) and distal gastric cancers (i.e., those of the body, antrum, and pylorus) (C16.2–8); and gastric cancers of unspecified location (C16.9).

Statistical Analysis

Annual incidence rates per 100 000 population were age adjusted to the 1970 U.S. standard population (10). To maximize the stability of our rate estimates, we compared incidence rates averaged from 3 years of data for 1974 through 1976, 1986 through 1988, and 1996 through 1998 by using the two-tailed Z statistic (12). Analyses were performed by using the Stata statistical package (release 8; Stata, College Station, TX) and the SEER Stat 4.1 program (National Cancer Institute, Bethesda, MD).

We adjusted the incidence data for gastric cancers of unspecified location (i.e., location “NOS” or “not otherwise specified”) by using the following equation: adjusted incidence for location \( Y = \text{reported incidence for location } Y + \left(\frac{\text{incidence of cancers at location } Y}{\text{incidence of all cancers assigned a specified location}}\right)\times\left(\text{incidence of cancers at location } Y\right)\times\left(\text{incidence of all cancers assigned a specified location}\right)\). This adjustment method assigned the unspecified cancers a location and added them to the cancers that had a reported location. We assumed that cancers without a site specification were distributed similarly to those with a reported location. Examples of this equation’s application are provided below (see “Sensitivity Analysis”). We used the \( \Delta \) method (13) to calculate the variance for the adjusted incidence data.

Sensitivity Analysis

We created four models to examine the influence of different assumptions about the cancer site distributions on cancer incidence. The base case model used the gastric cancer site distribution observed in the 1986–1988 SEER data. For example, for white males, 43% of all gastric cancers assigned a site in 1986–1988 were cardia cancers; thus, the model assigned 43% of all gastric cancers of unknown site to the cardia for each time interval (see equation above). We selected this model as the base case model \( a \ priori \) for several reasons. First, the 1986–1988 interval came before the widespread reporting of increased cardia cancer incidence. We hypothesized that widespread reporting of increased cardia cancer incidence rates heightened interest in this carcinoma in the 1990s, possibly increasing the likelihood that a gastric cancer would be classified to a cardia location; thus, we felt that the 1986–1988 interval would be less susceptible to reporting biases than later periods (1,2). Second, this time interval was intermediate in time and in the proportion of cardia cancers to the other time intervals and thus potentially provided intermediate (rather than extreme) values of cancer site distribution. Finally, during this period, the proportion of cases with an unspecified location was relatively low compared with that in 1974–1976; site assignments were available for 83% of gastric cancers in 1986–1988 but for only 62% of cancers in 1974–1976. The 1986–1988 data, therefore, provided substantially greater precision about cancer locations than data from earlier time periods.

Models 1 and 3 distributed the cancers of unknown location according to the site distributions observed in SEER data in 1974–1976 and in 1996–1998, respectively. The 1974–1976 model for white males, for example, assumed that 29% of cancers not assigned a site were cardia cancers, and the 1996–1998 model assumed that 51% of cancers not assigned a site were cardia cancers.

Model 4 adjusted the incidence rates by using the proportion of cardia cancers in each time interval; e.g., the 1974–1976 data used the proportion of cardia cancers in 1974–1976, and the 1986–1988 data used the proportion of cancers in 1986–1988. Although it seems intuitive that this model might provide a good estimate of the base case incidence, the large proportion of cancers of unknown location in the 1974–1976 incidence estimates created substantial uncertainty in site distribution, and the 1996–1998 estimates were potentially influenced by reporting bias (see above). We therefore compared results from these different models to provide a range of potential values for cancer incidence.

RESULTS

Cancers Without a Designated Location

The assignment of gastric cancers to specific sites improved substantially from 1974 through 1998. For white males, the proportion of all gastric cancers without a designated location decreased by 74%, from 38% of all gastric cancers in 1974–1976 to 17% in 1996–1998 (from 4.0 to 1.1 per 100 000 population/year; \( P<.001 \)) (Fig. 1); most of the improvement in site classification occurred between the 1974–1976 and 1986–1988
time intervals, the decade with the greatest increase in cardia cancer incidence (see below). The results were similar for black males; the proportion of all gastric cancers without a designated location decreased by 68%, from 41% of all gastric cancers in 1974–1976 to 16% in 1996–1998 (from 7.5 to 2.4 per 100 000 population/year; P < .001) (Fig. 2). Similar trends were observed for gastric cancers in white and black females (data not shown because of the relatively low cardia cancer incidence rates in these groups).

Cardia Cancer Incidence

Improved reporting of the gastric cancer sites largely accounted for the purported increase in cardia cancer incidence from 1974 through 1998. The adjusted cardia cancer incidence rate for white males did not increase statistically significantly between 1974–1976 and 1996–1998 (5% increase, from 3.6 to 3.8 per 100 000 population/year; P = .59). By contrast, the conventional unadjusted (i.e., SEER) cardia cancer incidence rate for white males increased by 77% between 1974–1976 and 1996–1998 (from 1.9 to 3.4 per 100 000 population/year; P < .001) (Fig. 3). Similarly, for black males, the adjusted cardia cancer incidence rates were stable during this time (4% increase, from 1.9 to 2.0 per 100 000 population/year between 1974–1976 and 1996–1998, respectively; P = .9), whereas the unadjusted SEER incidence rates increased statistically significantly by 77% (from 1.0 to 1.7 per 100 000 population/year; P = .04) during this interval (Fig. 4).

Noncardia Gastric Cancer Incidence

Adjustment for improvements in site classification suggested a more rapid decrease in the incidence of noncardia gastric cancer than was previously appreciated (Fig. 5). For white males, the incidence of noncardia gastric cancer, adjusted using the base case model, decreased 44% between 1974–1976 and 1996–1998 (from 6.8 to 3.8 per 100 000 population/year; P < .001); the absolute decrease in the adjusted incidence rate was more than twice as large as that observed using standard unadjusted SEER data for the same period (from 4.5 to 3.2 per 100 000 population/year; P < .001). For black males, the adjusted noncardia cancer incidence also decreased statistically significantly between the 1974–1976 and 1996–1998 time intervals (from 16.5 to 12.5 per 100 000 population/year; P = .005); this decrease was in contrast to the stable incidence rates suggested by the unadjusted SEER data for this period (from 10.0 to 10.6 per 100 000 population/year; P = .67).

Distribution of Gastric Cancer

Among white males, the proportion of gastric cancers with a reported site that were classified as cardia cancer increased from 29% during 1974–1976 to 51% during 1996–1998. For black males, these proportions were 9% and 14%, respectively. If the site distribution was, in fact, stable over time, and the apparent increase in cardia cancers was due only to changes in site classification, we might expect to see a disproportionate decrease in the proportions of cancers in adjacent sites (because cancers at sites adjacent to the cardia may be the ones most likely to be reclassified as cardia cancers) (Fig. 6). To evaluate
whether proximal gastric cancers might have been disproportionately reclassified as cardia cancers, we examined the incidence, over time, of cancers of the gastric fundus (an anatomic site immediately adjacent to the gastric cardia). The incidence of gastric fundic cancers in white males decreased 25% between 1974–1976 and 1996–1998 (from 0.4 to 0.3 per 100 000 population/year; *P* = .03); this decrease was smaller than the 37% decrease in the incidence of more distal gastric cancers in the antrum and pylorus that occurred between 1974–1976 and 1996–1998 (from 1.9 to 1.2 per 100 000 population/year; *P* < .001). Similar findings were observed for black males: There was essentially no change in the incidence of fundic cancer over the same time interval (from 0.6 to 0.7 per 100 000 population/year; *P* = .91). These results suggest that proximal cancers (such as those in the fundus) were not disproportionately reclassified as cardia cancers.

### Sensitivity Analysis

The change in the adjusted cardia cancer incidence between 1974–1976 and 1996–1998 ranged from a 1% decrease to a 27% increase, depending on which model was used to adjust for cases of unknown anatomic site (Table 1). All estimates of the change in adjusted incidence were substantially lower than the 77% increase suggested by the traditional unadjusted incidence figures from the SEER database (Table 1). The incidence of overlap cancers (cancers that overlap more than one cancer site) reported in SEER was low throughout this time interval and did not substantially influence the other results (data not shown).

#### DISCUSSION

Our results suggest that a substantial portion of the recent purported increase in cardia cancer incidence is due solely to improvements in site classification of gastric carcinomas and that the incidence rate of noncardia gastric cancers is decreasing more rapidly than was previously appreciated.

Our results expand and supplement prior reports of esophageal and cardia cancer incidence. Those reports (1,2,6,7) suggested that the incidence rates for both cardia and esophageal cancers increased rapidly in the last three decades, with a flattening of the cardia incidence curve in the 1990s. Our results, however, suggest that the increase in cardia cancer incidence that occurred in the 1970s and 1980s was due primarily to improvements in site classification that occurred before 1990 and that the subsequent flattening of the incidence curve was due to the lack of substantial additional improvement in site classification during the 1990s.

Why did the proportion of gastric cancers assigned to the gastric cardia increase? One potential explanation is that there was an increased tendency to classify proximal cancers (e.g., gastric fundic cancers) as cardia cancers. If this were the case, however, we would have expected to see a decrease in the incidence of cancers at sites adjacent to the cardia, such as the

---

Table 1. Sensitivity analysis, cardia cancer incidence

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjustment method*</th>
<th>% allocation†</th>
<th>Mean 3-y incidence per 100 000 population (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Unadjusted</td>
<td>N/A</td>
<td>1.89 (1.71 to 2.07)</td>
</tr>
<tr>
<td>1</td>
<td>Adjusted to 1975 distribution</td>
<td>29</td>
<td>3.07 (2.80 to 3.35)</td>
</tr>
<tr>
<td>2</td>
<td>Adjusted to 1987 interval (base case)</td>
<td>43</td>
<td>3.61 (3.33 to 3.89)</td>
</tr>
<tr>
<td>3</td>
<td>Adjusted to 1997 distribution</td>
<td>53</td>
<td>3.94 (3.65 to 4.24)</td>
</tr>
<tr>
<td>4</td>
<td>Adjusted by each interval’s distribution</td>
<td>Varied by interval</td>
<td>3.07 (2.80 to 3.35)</td>
</tr>
</tbody>
</table>

*Adjustment to 1975 distribution = sites assigned in proportion to the site distribution in the 1974–1976 interval; adjustment to 1987 distribution = sites assigned in proportion to the site distribution in the 1986–1988 interval; adjustment to 1997 distribution = sites assigned in proportion to the site distribution in the 1996–1998 interval; adjustment by each interval’s distribution = sites assigned to the unknown cases in proportion to the site distribution for each time interval. See “Methods” for details.

†Percentage of unclassified cases allocated to the cardia location by each model. N/A = not applicable.

‡CI = confidence interval.
funic cancer, but we did not observe a decrease in the incidence of fundic cancer. A second potential explanation is that the previously unclassified cancers were disproportionately located in the cardia. In support of this possibility, we found that the primary increase in the assignment of gastric cancers to the cardia occurred between 1974–1976 and 1986–1988, coincident with the main decline in gastric cancers of unknown location. Given that the incidence of cancer at other gastric sites did not increase, a disproportionate number of these unclassified carcinomas may have been cardia tumors. A third explanation is that distal esophageal adenocarcinomas may have been misclassified as cardia cancers. If this possibility entirely explained the results, we would expect cardia cancer incidence trends to parallel those of esophageal adenocarcinoma; however, esophageal adenocarcinoma incidence rates continued to increase throughout the 1990s, while cardia cancer incidence rates were stable during this period. A fourth explanation is that this redistribution of gastric cancers to more proximal locations may represent either a true relative increase in cardia cancer incidence or a combination of some or all of the previous explanations.

Our study has several strengths. First, we used the SEER cancer registry, a comprehensive population-based cancer registry that covers a diverse population, possesses a high case ascertainment rate, and uses consistent quality assurance procedures. Second, the results were robust to different assumptions about cancer site distributions and were consistent between ethnicities. Third, the results were consistent over time—when classification stopped improving, the incidence of cardia carcinoma stopped increasing.

Our study also has several potential weaknesses. First, we were unable to determine the true site distribution for the carcinomas of unknown location; it is likely that these cancers were not distributed exactly proportionately to those with a known site specification. We evaluated this uncertainty by performing a sensitivity analysis of different distributions of gastric cancer, and all methods suggested that the increase in cardia cancer incidence was accounted for largely (or entirely) by the decrease in the proportion of carcinomas without a specified location. Second, it is possible that other differences in site classification between geographic regions, between ethnicities, or over time may also contribute to the variability that we observed. For example, inaccuracies in regional SEER incidence rates from inaccurate census estimates have been reported (9); however, no studies have documented inaccuracies in the pooled national data (used in this analysis) that would be sufficient to account for our results. Third, we did not evaluate changes in the classification of esophageal adenocarcinomas. Reports of esophageal adenocarcinoma incidence are less susceptible to site reporting bias than reports of gastric cancer incidence because esophageal cancers are usually reported in the literature by their general location in the esophagus rather than by subsite (14,15). However, compared with gastric cancer incidence, esophageal cancer incidence is more susceptible to changes in histologic classification over time because esophageal cancers are divided into two different dominant histological types: squamous-cell carcinomas and adenocarcinomas.

In conclusion, our results provide several insights into recent gastric cancer incidence trends. First, the purported increase in cardia cancer incidence over the last three decades may be attributable mainly to improvements in site classification rather than to a true increase in incidence rates. Second, the relative distribution of gastric cancers may have changed, with a relative decrease in the proportion of distal gastric cancers. Third, although the adjusted incidence of cardia cancers remained stable over time, the incidence of the more distal gastric cancers appears to be decreasing much more rapidly than previously appreciated. Our results emphasize the importance of evaluating changes in site classification when assessing trends in raw cancer incidence rates over time. Evaluations of risk factor profiles for cardia versus distal gastric carcinomas may further clarify the biologic differences between these sites.

References


Notes

1Editor’s note: SEER is a set of geographically defined, population-based central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

Supported by a United States National Institutes of Health Mentored Clinical Scientist Development Award (K08 DK02697). The funding organization did not participate in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript. Manuscript received February 24, 2004; revised July 8, 2004; accepted July 28, 2004.