Impact of patients’ primary language on stage of cancer at diagnosis

BY MARIYA E. SKUBE, MD, MPH; BRUCE H. ALEXANDER, PHD; GREG J. BEILMAN, MD; AND TODD M. TUTTLE, MD, MS

Non-English language (NEL) patients have been identified to have worse health outcomes and care utilization patterns compared to English-speaking patients. The objective of this study was to investigate if a patient’s primary language impacts the stage of cancer at the time of diagnosis. A retrospective review of the Minnesota Cancer Surveillance System’s database was conducted with incorporation of language data from an integrated health system’s data repository. Adult patients with cancer of the breast, colon and rectum, female genital system, male genital system, or respiratory system were included, and cancer stage at diagnosis was grouped as early (stages 0-2) or advanced (stages 3-4). Statistical analysis involved two-sided chi square tests of independence and multiple logistic regression modeling. From a cohort of 13,245 cancer patients spanning the years 1991-2017, 311 (2.3%) were NEL. Thirty-six percent of NEL individuals were diagnosed at an advanced stage compared to 25.8% of English speakers (P < 0.001; OR 1.62 [95% CI 1.27-2.04]). Analysis by site of cancer revealed that NEL patients with a male genital system cancer were 2.58 times more likely to be diagnosed at an advanced stage (95% CI 1.03-5.60; P = 0.022). NEL, male gender, and more recent year of diagnosis were significant predictors of advanced stage diagnosis in the logistic regression model. NEL patients are at risk of being diagnosed with cancer at a more advanced stage compared to English speakers. The NEL population warrants unique attention when it comes to cancer detection initiatives, and further study is warranted to delineate associated factors.

Introduction

In the United States, approximately 8.5% of the population or 25 million individuals have limited English proficiency.1 This population faces unique challenges in many aspects of daily life, and successfully interfacing with the American health care system can be a particular struggle. This reality demands that the health care community seek to better understand how health equity is affected by language preference.

Patients with a non-English language (NEL) preference have been identified to fall behind their English-speaking counterparts in a number of measures of health, including asthma control, diabetes care, and hypertension management.2-5 NEL preference has been implicated in concerning health utilization patterns such as higher risk of readmission to the hospital and early emergency department return visits.6-8

Cancer care is not immune to the effect of language preference on the provision and use of key services. NEL patients have been found to have decreased rates of cancer screening compared to English-speaking patients.2,9,10 The diagnosis of cancer at later stages generally portends a worse prognosis, so these findings may have troubling repercussions on cancer mortality for NEL individuals in the United States.

Previous studies have demonstrated that demographic characteristics such as race and socioeconomic characteristics can lead to delayed cancer diagnosis, but the impact of language preference has not been well studied.11-13 The objective of this study was to investigate the impact of a patient’s primary spoken language on the stage of cancer at the time of initial diagnosis. We hypothesized that the cancer stage would be more advanced in patients with an NEL preference.

Methods

After approval by the University of Minnesota’s Institutional Review Board, a retrospective observational study was performed involving secondary analysis of existing data. Data were obtained from the Minnesota Cancer Surveillance System (MCSS). The State of Minnesota has mandated the reporting of new cancer diagnoses to the MCSS since 1988, in accordance
with standards set forth by the North American Association of Central Cancer Registries. Variables acquired from the MCSS included 1) age at diagnosis, 2) gender, 3) year of diagnosis, 4) site of cancer, and 5) clinical stage at diagnosis. As MCSS does not include a patient language variable, language data were acquired by interfacing the MCSS patients with the Clinical Data Repository (CDR) of an integrated health care system (Fairview Health Services), which includes both community and academic institutions. The patients’ self-reported primary languages were abstracted from the CDR.

The study’s primary outcome, or dependent variable, was the stage of cancer at the time of diagnosis. Stage of cancer was grouped into “early” (stages 0 through 2) and “advanced” (stages 3 and 4) based on clinical staging. Adult patients (over 17 years old) with one of the following sites of cancer were included: breast, colon and rectum, female genital system, male genital system, and respiratory system. These five sites were selected because they represent the most common cancers in the United States and because most have established screening recommendations in place. Patients were excluded if data were incomplete (i.e., no language data available, inadequate staging entries) and if the primary language was not a spoken language (i.e., American Sign Language). If the patient had multiple entries in the MCSS for more than one cancer diagnosis, data only from the first cancer diagnosis were utilized.

### Statistical analysis
For descriptive statistics, normally distributed data were summarized using mean and standard deviation (SD). Two-sided chi square test of independence was applied to compare characteristics of NEL and English language patients as well as stage of cancer (early versus advanced) based on language preference, and odds ratios (OR) with 95% confidence intervals (CI) were calculated. Multiple logistic regression was employed to assess the impact of other demographic and situational variables—namely age, gender, and year of diagnosis—in addition to language on the stage of cancer at the time of diagnosis. Statistical significance was considered with an alpha level less than 0.05. Data analysis was conducted using SPSS Statistics 23 (IBM, Inc., Somers, NY) and the R statistical software package (R Foundation for Statistical Computing, Vienna, Austria).

### Results
#### Demographics
A total of 13,245 individuals met the inclusion criteria: 7,924 women (60.2%) and 5,271 men (39.8%) with a mean age of 61.5 years ± 12.9 (Figure 1, Table 1). Years of diagnosis ranged from 1991 to 2017 with the greatest representation in the current decade. Of the five included cancer sites, breast (n=3,974; 30.0%) and male genital system (n=3391; 25.6%) were the most common in the total cohort. For all five cancers, 72.9% (n=9,790) had an early stage diagnosis while 26.1% (n=3455) were diagnosed at an advanced stage.

#### Language and cancer stage
Three hundred and eleven patients (2.3%) had a non-English primary language compared to 12,934 (97.7%) with English as the primary language (Table 1). Women represented a greater proportion of the NEL population (n=205; 65.9%) compared to English-speaking patients (n=7769; 60.0%), and NEL patients were slightly younger on average (58.4 years ± 14.4) than English language patients (61.6 years ± 12.9). Thirty-one unique non-English languages were represented (Table 2) with the most common non-English languages being Spanish (n=60), Russian (n=44), and Vietnamese (n=37).

Thirty-six percent of NEL individuals were diagnosed at an advanced stage compared to 25.8% of English speakers (Table 1).
3). This result was statistically significant (P < 0.0001) with NEL patients being 1.62 times more likely than English-speaking patients to be diagnosed at an advanced stage (95% CI 1.27-2.04). Examining the specific cancer sites included in this study, NEL patients with a male genital system cancer were 2.58 times more likely to be diagnosed at an advanced stage (P = 0.0223; 95% CI 1.03-5.60) (Figure 2). NEL patients with cancers of the breast, female genital system, and respiratory system were also more likely to be diagnosed at an advanced stage compared to their English language counterparts; however, the differences did not reach statistical significance. The only cancer site that showed increased odds—albeit not statistically significant—of advanced stage diagnosis in the English-speaking group compared to the NEL population was colon and rectal cancer.

**Multiple logistic regression**

The logistic regression model—with stage at diagnosis as the binary dependent variable (early versus advanced) and age, gender, language, and year of diagnosis as the independent variables—was statistically significant (χ²(6) = 400.35, P < 0.001). The regression analysis revealed that in addition to language, patient gender and year of diagnosis (categorized by decade) contributed significantly to the model in predicting advanced stage at diagnosis (Table 4). NEL patients were more likely (P = 0.002; OR 1.47 [1.16-1.86]) to be diagnosed at an advanced stage. Furthermore, men were more likely than women to be diagnosed at an advanced stage (P = 0.024; OR 1.10 [1.01-1.19]), and more recent year of diagnosis was associated with an increased likelihood of advanced stage diagnosis. Patients diagnosed in the 2000s (P < 0.001; OR 1.86 [1.46-2.38]) and 2010s (P = 0.001; OR 3.98 [3.16-5.01]) were more likely to be diagnosed at an advanced stage than patients diagnosed in the 1990s.

The overall age variable (categorized as young adult [18-24 years], adult [25-64 years], and mature adult [>65 years]) did not contribute significantly to the regression model (P = 0.063) however there was a significant intra-category comparison. Adults (P = 0.05; OR 0.53 [0.28-1.0]) and mature adults (P = 0.074; OR 0.56 [0.29-1.06]) were less likely than young adults to be diagnosed with cancer at an advanced stage.

**TABLE 2**

<table>
<thead>
<tr>
<th>Specific primary languages</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>12934</td>
<td>97.7%</td>
</tr>
<tr>
<td>Non-English</td>
<td>311</td>
<td>2.3%</td>
</tr>
<tr>
<td>Amharic</td>
<td>6</td>
<td>1.9%</td>
</tr>
<tr>
<td>Arabic*</td>
<td>16</td>
<td>5.1%</td>
</tr>
<tr>
<td>Bengali</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Bosnian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Central Khmer</td>
<td>13</td>
<td>4.2%</td>
</tr>
<tr>
<td>Chinese*</td>
<td>26</td>
<td>8.4%</td>
</tr>
<tr>
<td>Gujarati</td>
<td>3</td>
<td>1.0%</td>
</tr>
<tr>
<td>Hmong*</td>
<td>16</td>
<td>5.1%</td>
</tr>
<tr>
<td>Indonesian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Italian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Japanese</td>
<td>3</td>
<td>1.0%</td>
</tr>
<tr>
<td>Javanese</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Karen languages</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Korean</td>
<td>8</td>
<td>2.6%</td>
</tr>
<tr>
<td>Lao</td>
<td>10</td>
<td>3.2%</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Nepali</td>
<td>2</td>
<td>0.6%</td>
</tr>
<tr>
<td>Oromo</td>
<td>5</td>
<td>1.6%</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>4.5%</td>
</tr>
<tr>
<td>Panjabi; Punjabi</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Persian</td>
<td>3</td>
<td>1.0%</td>
</tr>
<tr>
<td>Polish</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Portuguese</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Romanian</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Russian*</td>
<td>44</td>
<td>14.1%</td>
</tr>
<tr>
<td>Somali*</td>
<td>30</td>
<td>9.6%</td>
</tr>
<tr>
<td>Spanish; Castilian*</td>
<td>60</td>
<td>19.3%</td>
</tr>
<tr>
<td>Tagalog</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Tigrinya</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Turkish</td>
<td>1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Vietnamese*</td>
<td>37</td>
<td>11.9%</td>
</tr>
</tbody>
</table>

*Languages spoken by at least 5% of the non-English language population.

**Discussion**

This retrospective analysis of the diagnosis of cancer in Minnesota yielded a few key findings. Overall, for the five examined cancers, NEL patients were diagnosed at an advanced stage more frequently than English language patients. Language remained a significant factor in regression analysis as well. This finding adds to the growing body of literature highlighting the importance of a patient’s preferred language on health outcomes. In regard to cancer diagnosis, the association of advanced stage diagnosis with worse prognosis is particularly disconcerting in light of our study’s main finding. Both patient care and public health initiatives must incorporate strategies to address the disparities present secondary to the socioeconomic determinants of healthcare, including a patient’s language. A recent publication by Genoff et al. (2016) reviewed studies involving interventions related to cancer screening for NEL patients based on the “patient navigator” model and found that targeted interventions were able to increase screening rates in this patient population. Such models should be considered when planning and providing care for NEL patients. It must be acknowledged that the health outcomes of the various language groups (i.e. Spanish versus Somali) are impacted to varying degrees. For example, Minnesota’s annual Health Equity of Care Report reveals that patients with a NEL preference frequently fall behind on various measures of adequate healthcare, such as optimal diabetes care and colorectal cancer screening. The trends, however, vary by preferred language. Although our study is not powered to analyze each language independently, it will be important...
for future work to appreciate and to investigate differences among the various non-English languages.

Secondly, our data show that stage at diagnosis stratified by language preference varies based on the site of cancer. For example, we found that 18% of NEL patients diagnosed with a male genital system cancer had an advanced stage at diagnosis compared to 8% of English speakers. This is a particularly interesting finding in light of the significance of gender in the regression model with male patients having an increased likelihood of advanced diagnosis. NEL male patients may be especially at-risk of a late cancer diagnosis.

A surprising finding in our study was that cancer was more likely to be diagnosed at an advanced stage in more recent years. Advanced stage diagnoses increased from 10% of the cohort in the 1990s to 18% in the 2000s to 31% in the 2010s. A closer examination of our patient cohort reveals that English language patients followed a nearly identical trajectory (10% to 17% to 31%) while NEL patients displayed greater variation (33% to 27% to 38%) with each subsequent decade. A potential contributor to this finding is that clinical staging in more recent years may actually be better able to identify regionalized and metastatic disease compared to earlier decades due to improved diagnostic tools such as positron emission tomography and computed tomography. The full explanation for this finding of increasing advanced stage diagnoses warrants further investigation.

Finally, our regression analysis found that other factors contribute to disparities in cancer diagnosis, namely gender and year of diagnosis. Factors not included in our analysis have been also implicated, including race, socioeconomic status, and insurance status. Determining the most crucial factors remains challenging, however are a variety of patient characteristics play key roles. Addressing language inequity is a reasonable approach as there are a number of interventions that can potentially modify patient and clinician behavior (i.e. ensuring adequate interpreter services, providing NEL patients with patient navigators, marketing screening campaigns to NEL populations).

This study is subject to a number of limitations. Although reporting of new cancer diagnosis to the MCCS is mandatory, some new cancer diagnoses will inevitably be missed. Also, as MCCS does not include a primary language variable, patient language had to be acquired from a secondary source (Fairview Health Services CDR). This limits the patient cohort to those with contact with this health system; the entire Minnesota population is not represented. Fairview Health Services serves a wide variety of patients including rural and urban populations across the state in both community-based and academic-affiliated centers, providing a reasonable cross-section of the population at large. The Minnesota experience may not be reflective of national trends; similar studies should be carried out at the national level. Additionally, defining a patient’s language preference can be elusive. The variable within the CDR is based on patient self-reporting, but comprehension of English even with a self-reported preferred NEL population is wide-ranging.
Language preference or reported primary language is not necessarily equivalent to facility with English. Given that just over 2% of our patient cohort was categorized as NEL, some of the subgroup analyses have low sample sizes and so may be underpowered to detect true differences. There also are patient characteristics not able to be assessed in the current study that also play a role in health-seeking behaviors and may contribute to advanced stage diagnosis, such as income and insurance status.

**Conclusions**

NEL patients are at risk of being diagnosed with cancer at a more advanced stage than are English speakers, which can impact ultimate disease prognosis. The NEL population warrants unique attention when it comes to cancer detection initiatives. A number of overlapping patient characteristics also may influence advanced stage diagnosis, so further study is warranted to delineate associated and contributing factors. Health disparities secondary to language preference must be accounted for while planning and implementing the care of patients. MM

Mariya E. Skube, MD, MPH, is a general surgery resident in the Department of Surgery, University of Minnesota. Bruce H. Alexander, PhD, is professor in Environmental Health Sciences, University of Minnesota. Greg J. Beilman, MD, is the Owen H. and Sarah Davidson Wangenstein Chair of Surgical Research in the Department of Surgery, University of Minnesota. Todd M. Tuttle, MD, MS, is the vice chair of Global Surgery and Disparities in the Department of Surgery, University of Minnesota. Greg J. Beilman, MD, is the Owen H. and Sarah Davidson Wangenstein Chair of Surgical Research in the Department of Surgery, University of Minnesota. Todd M. Tuttle, MD, MS, is the vice chair of Global Surgery and Disparities in the Department of Surgery, University of Minnesota.

Mariya Skube, MD, MPH, is supported by NIH/NIDDK T32DK108733.

The authors appreciate the collaboration of the staff of the Clinical and Translational Science Institute of the University of Minnesota for their contribution to this work.

**TABLE 4**

<table>
<thead>
<tr>
<th>Multiple logistic regression model</th>
<th>EARLY STAGE (N=9790)</th>
<th>ADVANCED STAGE (N=3455)</th>
<th>P</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young adult (18-24 years)</td>
<td>26 (61.9)</td>
<td>16 (38.1)</td>
<td>0.063</td>
<td>Reference</td>
<td>--</td>
</tr>
<tr>
<td>Adult (25-64 years)</td>
<td>5592 (74.6)</td>
<td>1900 (25.4)</td>
<td>0.050</td>
<td>0.53</td>
<td>0.28-1.00</td>
</tr>
<tr>
<td>Mature adult (&gt;65 years)</td>
<td>4172 (73.1)</td>
<td>1539 (26.9)</td>
<td>0.074</td>
<td>0.56</td>
<td>0.29-1.06</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5886 (73.8)</td>
<td>2088 (26.2)</td>
<td>--</td>
<td>Reference</td>
<td>--</td>
</tr>
<tr>
<td>Male</td>
<td>3904 (74.1)</td>
<td>1367 (25.9)</td>
<td>0.024</td>
<td>1.10</td>
<td>1.01-1.19</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>9591 (74.2)</td>
<td>3343 (25.8)</td>
<td>--</td>
<td>Reference</td>
<td>--</td>
</tr>
<tr>
<td>Non-English</td>
<td>199 (64.0)</td>
<td>112 (36.0)</td>
<td>0.002</td>
<td>1.47</td>
<td>1.16-1.86</td>
</tr>
<tr>
<td><strong>Year of diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>740 (89.8)</td>
<td>84 (10.2)</td>
<td>&lt;0.001</td>
<td>Reference</td>
<td>--</td>
</tr>
<tr>
<td>2000s</td>
<td>2860 (82.5)</td>
<td>605 (17.5)</td>
<td>&lt;0.001</td>
<td>1.86</td>
<td>1.46-2.38</td>
</tr>
<tr>
<td>2010s</td>
<td>6190 (69.1)</td>
<td>2766 (30.9)</td>
<td>&lt;0.001</td>
<td>3.98</td>
<td>3.16-5.01</td>
</tr>
</tbody>
</table>

Data presented as n (%). OR, odds ratio. CI, confidence interval.

**REFERENCES**


