Celiac disease (CD) is a gastrointestinal inflammatory disorder caused by gluten ingestion in genetically susceptible individuals and is treated by adhering to a gluten free diet. Approximately 1% of the western population has CD, though many individuals remain undiagnosed. Few studies have determined the prevalence of CD outside of the United States (US) and Europe, though some evidence suggests that CD is common in North Africa, the Middle East, and India and rare in China and sub-Saharan Africa.

Even less is known about the prevalence of nonceliac gluten sensitivity (GS), a disorder without a biomarker or established pathogenesis, in which symptoms are thought to be triggered by gluten. Although many people with GS avoid gluten in order to alleviate gastrointestinal symptoms, others do so to alleviate symptoms as widespread as tiredness, mood disorders, and respiratory problems. As most people with GS have not been diagnosed by a health care professional and begin a gluten-free diet on their own, the prevalence of GS is elusive. The gluten free diet has been growing in popularity; in a national survey, the proportion of gluten avoiders without CD has significantly increased in the United States from 2009 to 2014, though the reasons for gluten avoidance were not characterized.

Given the perceived health benefits of the gluten-free diet, it is possible that individuals following this diet may also avoid other foods that are perceived as unhealthy or commonly cause gastrointestinal symptoms. Unfortunately, food avoidance patterns in individuals with GS or CD have been understudied.

In this study, we assessed the relative proportion of CD and GS across 15 world regions and 4 US regions, as well as food avoidance patterns among participants in a web-based questionnaire.

**Methods**

**Data Source and Inclusion Criteria**

We analyzed questionnaire responses provided by visitors to a web site detailing a newly developed portable gluten detection device (Nima) between August 2014 and May 2016. On email signup participants were instructed to complete questions about food avoidance reasons and food avoidance patterns. Participants’ residences were inferred from computer IP addresses. Of 12,346 respondents, 937 were excluded because of identifiable regions of residence. Of the remaining 11,409 participants, only those who identified as having CD or GS were included in the study (n = 9071), with responses of “celiac disease” and “celiac disease—diagnosed medically” characterized as CD and “I feel better not eating gluten—never been diagnosed” and “nonceliac gluten intolerance—medically diagnosed” characterized as GS.

**Geographic Analysis**

Countries that contained responses from at least 40 participants were used in our analysis, whereas countries that contained fewer than 40 participants were grouped into
regions before analysis. The relative distribution of CD and GS was determined for 8 individual countries: Argentina, Canada, Germany, Italy, the Netherlands, Spain, the United Kingdom, and the United States; and 7 regions: Africa/the Middle East (containing Burkina Faso, Egypt, Iran, Israel, Jordan, Kuwait, Lebanon, Morocco, Qatar, Saudi Arabia, South Africa, Turkey, and the United Arab Emirates), Asia (containing Hong Kong, India, Japan, Korea, the Philippines, Singapore, Taiwan, Thailand, and Vietnam), Central America/ the Caribbean (containing Aruba, Barbados, Belize, Costa Rica, Curacao, the Dominican Republic, Guatemala, Mexico, Panama, and Puerto Rico), Eastern Europe (containing Bosnia, Bulgaria, Croatia, the Czech Republic, Greece, Hungary, Poland, Romania, Russia, Serbia, Slovenia, and the Ukraine), Western Europe (containing Andorra, Austria, Belgium, Cyprus, Denmark, Finland, France, Gibraltar, Ireland, Luxembourg, Norway, Portugal, Sweden and Switzerland), and South America (containing Brazil, Columbia, Chile, Ecuador, Peru, Paraguay, Venezuela, and Uruguay). As Australia had 133 participants and New Zealand had 26, we combined Australia and New Zealand into 1 category.

Within the United States the relative distribution of CD and GS was determined in 4 regions, as has been previously described, defined as follows: Northeast (Connecticut, Maine, Massachusetts, New Hampshire, New Jersey New York, Pennsylvania, Rhode Island, Vermont), Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin), South (Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, West Virginia, Virginia), and West (Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington).

Food Avoidances
We also determined the prevalence of food avoidances among respondents in each US region and we compared these avoidances among participants with CD versus GS. The following food avoidance categories were solicited: eggs, milk and dairy products, peanuts, tree nuts, soy, shellfish, and other.

Statistical Analysis
We used the $\chi^2$ and Fisher exact tests to compare proportions. All $P$-values are 2-sided. We used SAS version 9.4 (Cary, NC) for all analyses. This research was deemed “nonhuman subjects research” by the Columbia University Medical Center’s Institutional Review Board as patient data were deidentified before being provided to the investigators.

RESULTS
Of 9071 respondents, CD was reported by 6474 (71.4%) and GS by 2597 (28.6%). Among the 7756 US respondents, 69.8% reported having CD and 30.2% reported having GS. The rates of CD and GS by US region are shown in Table 1 and Figure 1. The highest ratio of CD to GS was in the Northeastern United States (CD: 74.3%; GS: 25.7%) and the lowest ratio in the Western United States (CD: 67.1%; GS: 32.9%) ($P < 0.0001$).

Internationally, compared with the United States, most other countries had lower proportions of respondents with GS (Table 2). Compared with the United States (CD: 69.8%; GS: 30.2%), the lowest proportion of GS was in Argentina (CD: 94.7%; GS: 5.3%; $P < 0.0001$). The region with the highest proportion of GS was Asia (35.6%), though the total number of respondents from that region was small (n = 45) and showed no difference compared with the United States (P = 0.4). Although Eastern (P = 0.2) and Western Europe (P = 0.4), the Netherlands (0.68), and Germany (0.75) all showed similar ratios of CD to GS compared with the United States, Italy (CD: 89.8%; $P < 0.0001$), the United Kingdom (CD: 81.3%; $P = 0.01$) and Spain (CD: 81.8%; $P = 0.02$) all showed higher rates of CD compared with the United States (CD: 69.8%).

Food avoidances other than gluten were reported by 25% of all respondents (Table 3). Overall, dairy products were the most avoided food category, reported by 13.8% of participants, followed by soy, reported by 10.6% of participants. Avoiding 1 food class in addition to gluten was reported by 10.8% of CD respondents compared with 21.3% with GS and avoiding ≥ 2 foods was reported in 7.3% of CD respondents compared with 21.9% with GS ($P < 0.0001$). Food avoidance patterns in GS compared with CD is shown in Table 3. In all categories food avoidances were more common in GS than in CD ($P < 0.0001$ per each category).

Food avoidance patterns in each US region are shown in Table 4. 26.8% US respondents reported ≥ 1 food avoidance, other than gluten, with dairy products the most avoided food category (14.4%), followed by soy (11.5%). Avoidance of dairy was highest in the Western United States and lowest in Northeast (West: 12.6% vs. Midwest: 9.8%; $P < 0.0001$), whereas soy avoidance was highest in the West and lowest in the Midwest (West: 12.6% vs. Midwest: 9.8%; $P = 0.008$). Avoidance of eggs, shellfish, peanuts, tree nuts, and other foods did not differ across US region. Dairy and soy avoidance per US region is shown in Figures 2 and 3.

In a post hoc analysis, we plotted the percentages of respondents with GS in each country by gross domestic product per capita shown in Figure 4. There was a strong correlation between gross domestic product per capita and percent of GS ($r = 0.86$).

DISCUSSION
There has been a recent rise in interest in the perceived health benefits of a gluten-free diet and an increased adoption of these diets, especially in the United States. Our study, which included participants from over 70 countries, attests to the growing international awareness of the gluten free diet and sensitivity to gluten. In our study, more than a quarter (28.6%) of gluten avoiders had self-reported GS,
with marked differences in the worldwide proportions, ranging from 5% (Argentina) to 36% (Asia).

The prevalence of GS has been estimated variously, from 13% in the United Kingdom to 7.3% in Australia, to 6.2% in the Netherlands, to 0.8% in the United States. However, the true prevalence of GS has been difficult to estimate as up to 44% of individuals who adhere to a gluten-free diet initiate the diet on their own without physician contact. Furthermore, many individuals with self-adopted gluten avoidance may instead have another disease that has not been diagnosed, such as fructose intolerance or small intestinal bacterial overgrowth. In addition, a study found that 69% of people who self-initiated the gluten free diet had not been tested for CD. In our study, individuals with GS, who were never medically diagnosed, came from all world regions, and made up 29% of all gluten avoiders. There is concern that a significant proportion of individuals with GS initiated a gluten-free diet in the absence of medical consultation, and possibly in the absence of CD exclusion.

Though some studies have suggested that the prevalence of GS is the same or larger than that of CD, our study found a higher relative proportion of CD in all countries and regions. The reason for this finding is unclear. It could reflect a larger percentage of individuals with CD among gluten

TABLE 2. Rates of Celiac Disease Versus Gluten Sensitivity by Country

<table>
<thead>
<tr>
<th>Countries/Region</th>
<th>Celiac Disease (n = 6474)</th>
<th>Gluten Sensitivity (n = 2597)</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>5416 (69.8)</td>
<td>2340 (30.2)</td>
<td>[reference]</td>
</tr>
<tr>
<td>Africa/Middle East</td>
<td>48 (76.2)</td>
<td>15 (23.8)</td>
<td>0.27</td>
</tr>
<tr>
<td>Argentina</td>
<td>107 (94.7)</td>
<td>6 (5.3)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Asia</td>
<td>29 (64.4)</td>
<td>16 (35.6)</td>
<td>0.43</td>
</tr>
<tr>
<td>Australia/New Zealand</td>
<td>127 (79.9)</td>
<td>32 (20.1)</td>
<td>0.006</td>
</tr>
<tr>
<td>Canada</td>
<td>147 (79.5)</td>
<td>38 (20.5)</td>
<td>0.005</td>
</tr>
<tr>
<td>Central America/Caribbean</td>
<td>33 (78.6)</td>
<td>9 (21.4)</td>
<td>0.24</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>50 (76.9)</td>
<td>15 (23.1)</td>
<td>0.21</td>
</tr>
<tr>
<td>Germany</td>
<td>27 (67.5)</td>
<td>13 (32.5)</td>
<td>0.75</td>
</tr>
<tr>
<td>Italy</td>
<td>132 (89.8)</td>
<td>15 (10.2)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>32 (72.7)</td>
<td>12 (27.3)</td>
<td>0.68</td>
</tr>
<tr>
<td>South America</td>
<td>51 (87.9)</td>
<td>7 (12.1)</td>
<td>0.002</td>
</tr>
<tr>
<td>Spain</td>
<td>72 (81.8)</td>
<td>16 (18.2)</td>
<td>0.015</td>
</tr>
<tr>
<td>Western Europe</td>
<td>116 (73.0)</td>
<td>43 (27.0)</td>
<td>0.40</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>87 (81.3)</td>
<td>20 (18.7)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

*Compared with United States.

TABLE 3. Food Avoidance in Celiac Disease Versus Gluten Sensitivity

<table>
<thead>
<tr>
<th>Food Avoidance</th>
<th>Celiac Disease (n = 6474)</th>
<th>Gluten Sensitivity (n = 2597)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shellfish</td>
<td>146 (2.3)</td>
<td>152 (5.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Eggs</td>
<td>177 (2.7)</td>
<td>211 (8.1)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Dairy</td>
<td>588 (9.1)</td>
<td>667 (25.7)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Other</td>
<td>230 (3.6)</td>
<td>231 (8.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Peanuts</td>
<td>212 (3.3)</td>
<td>234 (9.0)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Tree nuts</td>
<td>171 (2.6)</td>
<td>184 (7.1)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Soy</td>
<td>431 (6.7)</td>
<td>529 (20.4)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
avoiders worldwide. CD is considered a relatively common condition especially in the United States and Europe, with a prevalence of 0.8% in the United States\textsuperscript{16} and 1.0% in Europe.\textsuperscript{2} Another possibility lies in the connection between socioeconomic status and CD diagnosis.\textsuperscript{18} All respondents in our study had access to the internet and had an interest in purchasing or being emailed information about a gluten detection device and therefore were likely to be in a high socioeconomic bracket. As evidence suggests that individuals of higher socioeconomic status are more likely to be diagnosed with CD,\textsuperscript{18} it follows that our study’s sample would therefore contain a higher rate of individuals with CD. We did not have access to individual indicators of socioeconomic status in our survey to test this hypothesis. Another possibility is that the avoidance of inadvertent gluten exposure, which is the motivation for use of the device, may be more of a concern among those with CD than those with GS.

It is unclear if the different international rates of relative distributions of GS in our study reflect true international variations in GS prevalence. The availability of gluten-free food varies by location, with greater availability in markets that cater to wealthier clientele: regular grocery stores carried the least amount of gluten-free products compared with health stores and upscale markets in a US study\textsuperscript{19} and budget supermarkets carried less gluten-free products compared with regular supermarkets in a UK study.\textsuperscript{20} Both studies reported that gluten-free products are considerably more expensive than their gluten-based complements.\textsuperscript{19,20} It is therefore possible that international differences in our reported rates of GS reflect the cost and ability to afford this diet. This likely explains our finding that countries with high gross domestic product per capita such as the United States and Germany had high rates of GS compared with countries with lower gross domestic product per capita such as Argentina and Italy (Fig. 4). However, this finding could also represent international differences in interest in and awareness of the gluten detection device. Future research should focus on the relationship between socioeconomic status and prevalence of GS, as well as the reasons for differing international rates of GS.

We also found that the relative proportions of CD and GS among gluten-free individuals varied within the United States, with the highest CD rate in the Northeast. To our
knowledge, only 1 prior study, performed by Unalp-Arida et al., investigated regional differences of GS and CD in the United States. Although our study differed from that prior study in that we examined relative proportions of CD and GS and they measured the regional prevalence of both conditions, the prior study also found the highest percentage of CD in the Northeast. Differences in the regional prevalence of CD are likely related to an interplay of genetic and environmental factors. Both dietary patterns of gluten exposure and inheritance of genetic variations in the HLA-DQA1 and HLA-DQB1 genes are known to modulate CD risk. Sunlight and UVB radiation exposure may be 1 such environmental factor that contributes to the increased rates of CD found in the North. Low UVB radiation predisposes to vitamin D deficiency, which has been linked to the frequency of autoimmune diseases.

Although our study found the lowest rate of CD and the highest rate of GS in the West, Unalp-Arida and colleagues reported the lowest percentage of both CD and GS in the South and the highest percentage of GS

FIGURE 3. The percent of soy avoidance among respondents in United States regions. Regional P-value = 0.008.

FIGURE 4. International percentages of gluten sensitivity by gross domestic product per capita. R = 0.86.
in the Northeast. This difference is likely because of the methodology of each respective study: we reported the ratio of CD to GS, whereas they reported the prevalence in each region. Still, our finding of increased GS in the western United States along with increased avoidance of dairy and soy may reflect regional preferences for dietary trends perceived to be healthy. One study found that people residing in the western United States had the highest percentage of fruit and vegetable consumption and overall healthy lifestyle, indicating that people residing in the western United States may be more influenced by diets they believe are healthy. Differences in regional rates of GS may also be linked to availability of gluten-free foods. Gluten-free products are widely available in Westchester, New York and Portland, Oregon (Northeast and West) and limited in Atlanta, Georgia and Chicago, Illinois (South and Midwest), though availability may have changed in the past decade. There also seem to be regional variation in Google searches for different diets within the United States, again reflecting regional variation in diet trends. More research is needed to determine the reason for regional US differences in gluten avoidance, especially in individuals without CD.

Other food avoidances were commonly reported in our sample, with GS respondents reporting more food avoidances than CD respondents. Food avoidance besides gluten, such as soy and dairy, has been reported in individuals with GS, though ours is the first to compare food avoidance patterns in people with GS and CD. Somewhat contrary to our finding, a study performed in the Netherlands found no statistically significant differences between adults with self-reported GS compared with healthy controls in regards to nut allergy, egg allergy, and lactose intolerance. This difference is likely because of our comparison of individuals with GS to CD, and not healthy controls, our inclusion of individuals from many different countries, and the fact that our study asked about food avoidances and not diagnosed allergies.

Just as GS is usually self-diagnosed, adults who self-report food allergies usually do so without a physician diagnosis. Reasons why individuals are self-reporting food allergies and sensitivities are not known. It is possible that GS individuals avoid foods other than gluten because they believe these foods are contributing to their gastrointestinal symptoms, but do not necessarily consider themselves as having an allergy. The symptoms of irritable bowel syndrome, a common functional gastrointestinal illness, may be wrongly attributed to both gluten sensitivity and lactose intolerance. One study found that some individuals with nonceliac GS may benefit from a diet that excludes fermentable, poorly absorbed, short-chain carbohydrates (components of the FODMAPs diet), instead of a diet that only excludes gluten. Another study found that abdominal discomfort related to FODMAP containing foods was higher in individuals with self-reported GS compared with controls (P < 0.001). Even patients with inflammatory bowel disease frequently adopt restrictive diets including a gluten-free diet. Another possibility is that individuals with self-diagnosed GS may perceive a number of food groups as deleterious to their health. A study in the United States found that lay people including college students, tend to dichotomize foods as good or bad and often adopt simplifying strategies that are not scientifically accurate in order to make sense of food and health information. To that effect, a New Zealand study of children who avoided gluten found that doctor-diagnosed lactose intolerance predicted gluten avoidance. This suggests a possible conceptual link between gluten and dairy, despite any evidence that gluten is harmful for individuals with lactose intolerance in the absence of comorbid CD. The health implications of these restricted diets and reasons for diet adoption need to be further studied.

Our study has several limitations. Firstly, we could only report the relative rates of CD and GS in each region and did not have the necessary information to calculate the prevalence in each region. Secondly, the true percentage of CD may be overestimated as individuals self-reported their diagnosis and did not indicate whether they had biopsy-proven disease. Thirdly, as almost 70% of our sample resided in the United States, the rates of CD and GS in other regions may be underreported. In addition, as we did not have access to socioeconomic indicators or adherence to the FODMAP diet we could not analyze the relationship between GS and these factors. Lastly, as this study is a web-based questionnaire, it is limited by selection bias and may not be generalizable to other populations.

In conclusion, our study highlights a significant difference in the relative rates of CD and GS, both within the United States and worldwide, among avoiders of gluten. Our study is also among the first to discuss food avoidance patterns in this population. Future research should focus on the causes of these international and US regional differences and the significance of other food avoidances.

REFERENCES


