**The Effect of Socioeconomic and Demographic Factors on the Diagnosis of Celiac Disease**

Shivan Mehta, MD
IRB Proposal- ICCR
PGY-2, Department of Medicine
Columbia University Medical Center

A. Study Purpose and Rationale

Celiac disease is an immune-mediated disease that causes an adverse response to gluten, the protein found in wheat, barley, and rye. While traditionally thought of as an enteropathy with gastrointestinal symptoms, it can also manifest with anemia, fatigue, abnormal liver tests, and osteoporosis. It is associated with the Human Leukocyte Antigen DQ2 and DQ8 haplotypes, but there are multiple environmental factors that affect the manifestation of the disease. Celiac disease was initially thought to afflict only those of northern European ancestry, but recent studies have demonstrated prevalence in a broader population. In fact, the prevalence in the United States is thought to be approximately 1% of the general population, despite a much lower diagnosis rate [1]. It is generally agreed that Celiac disease is under-diagnosed in this country, particularly compared to Europe, but rates of diagnosis have been increasing recently.

Since Celiac disease can have a variety of manifestations, factors that contribute to diagnosis include physician awareness, character of symptoms, and public knowledge about the disease. The traditional Celiac disease symptoms were cited as diarrhea, malabsorption, and weight loss. Based on more public and physician awareness, there is a trend towards more patients being diagnosed with non-gastrointestinal symptoms, in addition to the traditional symptoms [2,3]. In a retrospective cohort study of patients diagnosed before and after 1993, there was a reduction from 73% to 43% in patients with complaints of diarrhea [4]. The benefits of treating Celiac disease with a gluten-free diet have primarily been studied in the more symptomatic population, so it is not completely understood what the true effects of treating the population without. However, examining those patients that lack gastrointestinal symptoms will give a glimpse into the trends in diagnosis of the disease in the United States. The iceberg theory of Celiac disease discusses how the symptomatic patients that have already been diagnosed represent only a small portion of the total prevalence of the disease. Thus, one could extrapolate that those with gastrointestinal symptoms will be more likely to be diagnosed earlier. By comparing patients with gastrointestinal symptoms to all other patients, we could gain insight into the population that is more likely to be diagnosed in this country.

Given that celiac disease is likely under-diagnosed, it would be useful to figure out which factors contribute to patients not being diagnosed adequately. There are many studies in the United States examining the clinical manifestations of celiac disease, and how they are similar to Europe [5]. However, there are not any studies examining the socioeconomic factors that contribute to diagnosis of celiac disease. A Swedish paper does a literature review of all studies looking at socioeconomic variables affecting Celiac disease, and only finds correlation with urban status in two studies [6]. One would expect socioeconomic factors to play a larger role in diagnosis in the United States because there is a significantly lower rate of diagnosis as compared to European populations. In particular, due to the disparate regional and economic barriers to health care in this country, we hypothesize that geographic location and economic status will have a significant

correlation with the 'asymptomatic' versus the 'symptomatic' Celiac disease population. Since the asymptomatic patients are more difficult to diagnose, they are likely to be from urban areas and higher economic status. Specifically, they would likely go to physicians with more celiac disease experience and have a greater social awareness of celiac disease. Based on these findings, we could potentially target populations that are disproportionately under-diagnosed. A recent study in California has shown that an active case-finding strategy can significantly improve the diagnosis rate of celiac disease in targeted clinics [7]. This study could also drive studies of larger populations across the country to identify key barriers to diagnosis and treatment of Celiac disease.

B. Study Design and Statistical Procedures

This will be a retrospective cohort study of adult Celiac disease patients (>16yrs old) seen at Columbia University Medical Center between 1981 and 2006. These are patients that will have been followed over time at the Celiac disease center with varying degrees of follow-up. Only patients with biopsy proven celiac disease will be included in the study. Based on a previous study, approximately 50% of these patients have presented with gastrointestinal symptoms as their reason for diagnosis [8]. These patients will be designated as the 'symptomatic' patients, while all other patients will be designated as 'asymptomatic' patients for the purpose of this study. This study will examine how key socioeconomic, demographic, and clinical characteristics differ between these two cohorts of patients.

Based on the hypothesis, the two main characteristics to be examined will be location and economic status. Based on zip codes provided by the patient for contact information, patients will be designated to be urban and non-urban in location. Zip codes and address will also be used to determine the median income of that patient's home address census tract area as a measure of economic status. In previous epidemiological studies, zip code and census tract area can both serve as strong predictors of health outcomes [9]. All patients from a particular census tract area will be assigned the median income for that area. For location, a chi-squared test will be performed for the categorical variable. It is hypothesized that more symptomatic patients will be from urban locations, so the study will be powered ($p<.05$, 80% power) to detect an absolute increase in urban prevalence of 10% for the asymptomatic group (70% urban in the symptomatic group, 80% urban in the asymptomatic group). This power analysis requires 313 patients in each group for a total of 626 patients (http://www.biomath.info/crc/index.html). For income, an unpaired t-test will be performed to detect a difference of median income for area between the two populations. In order to detect a meaningful effect size between the groups of 1/4 the standard deviation, 257 patients need to be included in each group for a total of 514 patients ($n = 1 + 16 (\text{std dev}/ \text{effect})^2$). Based on the power analysis for both income and location, at least 626 patients will be needed from the celiac database after all exclusion criteria. Additional secondary outcomes that will be analyzed will be family history, age, gender, insurance type, and race. Since there are approximately 1000 patients in the CUMC database, there should be adequate number of patients to adequately power the study.

After the basic analysis comparing the asymptomatic and symptomatic patients, a logistic regression will also be performed with asymptomatic celiac disease as the dependant variable. The independent variables will be a combination of continuous and binary variables including: age, gender, insurance, urban location, median income. This analysis will examine variable

effect and the interaction between different socioeconomic factors to give more robust data about specific population characteristics.

C. Study Procedures
There are no procedures to be performed in this study.

D. Study Drugs or Devices
There are no drugs or devices involved in this study.

E. Study Questionnaires
This study involved secondary data analysis, so there are no primary questionnaires.

F. Study Subjects.
The study with include adult Celiac disease patients (>16yrs old) seen at Columbia University Medical Center between 1981 and 2006. Only biopsy-proven patients who fulfilled strict criteria based on small intestinal biopsy and response to gluten-free diet will be included.

G. Recruitment
Since this is a retrospective study of patients who have already presented to a Celiac disease clinic, subjects will not need to be contacted or recruited.

H. Confidentiality of Study Data
Data collected will be identified only by an encoded patient identification number and will be maintained on a password-protected computer.

I. Potential Risks
The only potential risk of this study is the minimal chance of loss of confidentiality.

J. Potential Benefits
This study could identify key socioeconomic factors and limitations to diagnosing Celiac disease in the United States.

K. Alternatives
There is no therapy involved in this study.

L. References

1. Fasano et al. Prevalence of Celiac Disease in At-Risk and Not-At-Risk Groups in the United States. Arch Int Med 2003; 163: 286-292.
2. Rampertab et al. “Trends in the Presentation of Celiac Disease.” The American Journal of Medicine (2006) 119, 355.e9-355.e14.
3. Lanzini A et al. “Epidemiological, clinical and histopathologic characteristics of celiac disease: Results of a case-finding population-based program in an Italian community.” Scandinavian Journal of Gastroenterology, 40:8, 950 - 957
4. Lo W et al. “Changing Presentation of Adult Celiac Disease.” Digestive Diseases and Sciences, Vol. 48, No. 2 (February 2003), pp. 395–398.
5. Green PHR et al. “Characteristics of Adult Celiac Disease in the USA: Results of a National Survey.” The American Journal of Gastroenterology. Vol. 96, No. 1, 2001: 126-131.
6. Ludvigsson, J. F. (2005) “Socio-economic characteristics in children with coeliac disease.” Acta Paediatrica, 94:1, 107-113.
7. Catassi C et al. “Detection of Celiac Disease in Primary Care: A Multicenter Case-Finding Study in North America.” Am J Gastroenterol 2007;102:1454–1460.
8. Bai D et al. “Effect of gender on the manifestations of celiac disease: Evidence for greater malabsorption in men.” Scandinavian Journal of Gastroenterology, 2005; 40: 183-187.
9. Thomas AJ et al. “ZIP-Code-based versus Tract-based Income Measures as Long-Term Risk-adjusted Mortality Predictors.” Am J Epidemiol 2006;164:586–590.