

## Prevalence of Diagnosed Celiac Disease and Associated Conditions in the United States between 2012-2017: Results from a National Electronic Patient Database

Daniel B. Karb, MD, Emad Mansoor, MD, Sravanthi Parasa, MBBS, Gregory S. Cooper, MD

Daniel B. Karb, Emad Mansoor, Sravanthi Parasa, Gregory S. Cooper, Case Western Reserve University, University Hospitals Cleveland Medical Center, United States

**Conflict-of-interest statement:** The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

**Open-Access:** This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

**Correspondence to:** Daniel Burke Karb, Case Western Reserve University, University Hospitals Cleveland Medical Center, United States.

Email: [dbkarb@gmail.com](mailto:dbkarb@gmail.com)  
Telephone: +985432239031

Received: October 29, 2018

Revised: December 8, 2018

Accepted: December 10, 2018

Published online: February 21, 2019

### ABSTRACT

**BACKGROUND:** Celiac Disease (CD) is an autoimmune condition caused by the ingestion of gluten in genetically susceptible individuals. Previous screening studies of CD have shown wide-ranging prevalence estimates. We sought to describe the prevalence of diagnosed CD in the United States (US) using a large, electronic, population-based database, and identify associated conditions of interest.

**METHODS:** From a large commercial database (Explorys), we identified a cohort of patients with CD between April 2012 and April 2017, based on Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT). From this cohort, we calculated the prevalence of CD and were able to identify associations between CD cases

and other conditions.

**RESULTS:** Of the 35,854,260 individuals in the database, we identified 83,090 cases of CD, with an overall prevalence of 231.7/100,000 persons (0.23%). Prevalence was higher in females than males [odds ratio (OR) 2.40; 95% Confidence Interval (CI) 2.36-2.43,  $p < 0.0001$ ], whites versus non-whites (OR 2.25; 95% CI 2.20-2.30,  $p < 0.0001$ ) and adults ( $\geq 18$  years) versus children ( $< 18$ ) (OR) 2.60; 95% CI 2.53-2.67,  $p < 0.0001$ ). CD was found to be associated with many other medical conditions.

**CONCLUSIONS:** In this large study, we found that the estimated prevalence of diagnosed CD in the USA is considerably lower than previous screening estimates, suggesting a large burden of undiagnosed CD. We were also able to identify strong associations between CD and many other medical conditions.

**Key words:** Celiac disease; Prevalence; Epidemiology; Exploryst; Electronic database; SNOMED; EMR

© 2019 The Author(s). Published by ACT Publishing Group Ltd. All rights reserved.

Karb DB, Mansoor E, Parasa S, Cooper GS. Prevalence of Diagnosed Celiac Disease and Associated Conditions in the United States between 2012-2017: Results from a National Electronic Patient Database. *Journal of Gastroenterology and Hepatology Research* 2019; 8(1): 2793-2799 Available from: URL: <http://www.ghrnet.org/index.php/joghr/article/view/2454>

### Abbreviations

CD: Celiac Disease;  
US: United States;  
SNOMED: CT-Systematized Nomenclature of Medicine—Clinical Terms;  
OR: Odds Ratio;  
CI: Confidence Interval;  
EHR: Electronic Health Record;  
HDG: Health Data Gateway;  
HIPAA: Health Insurance Portability and Accountability Act;  
IRB: Institutional Review Board;  
GERD: Gastroesophageal Reflux Disease;

EoE: Eosinophilic Esophagitis;  
 DM: Diabetes Mellitus;  
 ICD: 9: International Classification of Diseases, Ninth Revision;  
 ACG: American College of Gastroenterology;  
 SLE: Systemic Lupus Erythematosus;  
 RA: Rheumatoid Arthritis;  
 PBC: Primary Biliary Cirrhosis.

## BACKGROUND AND AIMS

CD, or gluten-sensitive enteropathy, is a chronic autoimmune inflammatory condition caused by gluten exposure in genetically sensitive individuals. A wide variety of epidemiologic studies using a variety of screening and diagnostic methods have attempted to define the prevalence of CD, and several studies suggest an increase in worldwide CD prevalence<sup>[1,2]</sup>. Prior analysis has tended to focus on white Europeans and their descendants (the population in which the disease was originally described)<sup>[3-6]</sup>, but studies conducted in Africa, the Middle East, Asia, and South America support the notion that CD is a global phenomenon that has been underdiagnosed in these populations<sup>[1,6]</sup>. In addition to the increasing recognition of CD in non-Western cultures, several other factors are hypothesized to account for this increase: improvement in diagnostic techniques and disease awareness, changing dietary habits, and intercurrent disease triggers such as gastroenteritis, surgeries, and trauma <sup>[2]</sup>.

To our knowledge, there have been no studies attempting to define the prevalence of diagnosed CD utilizing a large electronic database. We sought to describe the prevalence of diagnosed CD in the USA and identify associated symptoms and disorders using a population-based database. The aims of the study were to identify cases of CD, identify CD-associated symptoms, identify other disorders associated with CD, and estimate overall prevalence of diagnosed CD in the USA and among different age-based, race-based, and gender-based subgroups. These data will help better define the epidemiology of CD in the United States.

## METHODS

### Database

We followed the methods established by Mansoor and Cooper<sup>[7]</sup>. In this study, the authors employed the Explorys database to determine the prevalence of Eosinophilic Esophagitis. The results of this study were found to fall within the same order of magnitude of prevalence rates established by prior studies, thus lending empiric validity to these methods<sup>[7]</sup>.

For the current analysis, we performed a retrospective analysis

of a large population-based, commercial database (Explorys Inc, Cleveland, OH). This database contains an aggregate of electronic health record (EHR) data from 26 major integrated healthcare systems spread over 50 states in the USA from 1999 to 2017. Explorys contains de-identified patient data from participating institutions and uses a health data gateway (HDG) server behind the firewall of each participating healthcare organization that collects de-identified data from various health information systems--EHR using billing inquiries. Data are then standardized and normalized by Explorys. Diagnoses, findings, and procedures are mapped into the SNOMED-CT hierarchy. Each participating healthcare institution has access to Explorys online (password protected), which provides for browsing of the data from all participating healthcare institutions. Explorys data are automatically updated at least once every 24 hours<sup>[8]</sup>. Explorys is a Health Insurance Portability and Accountability Act (HIPAA) compliant platform, and thus Institutional Review Board (IRB) approval is not required.

### Patient Selection

Using the Explorys search tool, we identified an aggregated patient cohort of eligible patients with CD. CD patients were defined as those having a SNOMED-CT diagnosis of “celiac disease” and “active within the last 5 years” at the time of data gathering (April 2012-April 2017).

### Associated Medical Conditions of Interest

We identified multiple medical conditions and medications associated with CD, as demonstrated by prior studies. Data on these conditions were extracted by using SNOMED-CT diagnostic terms for these

**Table 1** Demographic Characteristics of CD.

Total Number of CD Cases, n	83090
Female, n (%)	62070 (74.7)
Male, n (%)	21020 (23.3)
<b>Age Group</b>	
Children (< 18 years of age), n (%)	5680 (6.8)
Adults (18-65 years of age), n (%)	58270 (70.1)
Elderly (> 65 years of age), n (%)	19110 (23.0)
<b>Race</b>	
Caucasian, n (%)	73110 (88.0)
African American, n (%)	3060 (3.7)
Asian, n (%)	1040 (1.3)
Hispanic, n (%)	560 (0.7)
Unknown race, n (%)	2370 (2.9)

**Table 2** Prevalence of CD in the Explorys Database between April 2012 and April 2017.

Group	Source Population (% of Source)	CD Cases, n (% of CD Cases)	Prevalence (per 100,000)
Overall	35,854,260 (100)	83,090 (100)	231.7 (0.23% of total population)
Male	16,053,620 (44.8)	21,020 (25.3)	130.9
Female	19,800,640 (55.2)	62,070 (74.7)	313.5 (OR 2.40; 95% CI 2.36-2.43, <i>p</i> < 0.0001)
Children (< 18 years)	5,743,600 (16.0)	5,680 (6.8)	98.9
Adults (18-65 years)	22,809,140 (63.6)	58,270 (70.1)	255.5 (OR 2.60; 95% CI 2.53-2.67, <i>p</i> < 0.0001)
Elderly (> 65 years)	7,262,300 (20.3)	19,110 (23.0)	263.1
Caucasian	22,543,450 (62.9)	73,110 (88.0)	324.3 (OR 2.25; 95% CI 2.20-2.30, <i>p</i> < 0.0001)
African-American	3,980,880 (11.1)	3,060 (3.7)	76.9
Asian	690,170 (1.9)	1,040 (1.3)	150.7
Hispanic	439,170	560 (0.7)	127.5
Unknown Race	4,596,840 (12.7)	2,370 (2.9)	51.6

disorders. CD-associated symptoms and disorders that were identified in the Explorlys database can be found in Tables 3-4.

### Statistical Analysis

For patients with CD, demographics and associated diseases, were characterized by descriptive statistics. Univariate analysis was performed to assess the differences in associated medical conditions in patients with CD and patients without CD by using the Pearson chi-square test. Univariate analysis was also performed to assess the differences in associated medical features of children with CD (aged under 18 years) and adults with disease (aged 18 years and above) by

using the Pearson Chi-square test.

For calculation of overall period prevalence, we identified all patients in the database with CD from April 2012 and April 2017. We then divided this number by the total number of patients in the database (from April 2012 to April 2017) thus making sure that all patients in the denominator (population at risk) had an equal opportunity of being diagnosed with celiac if they had the disease. Similarly, age-based, gender-based, and race-based prevalence rates were calculated. The confidence intervals for prevalence rates were calculated using the Wald method for calculation of confidence intervals for single proportions<sup>91</sup>.

**Table 3** CD-Associated Symptoms and Disorders.

Diagnosis	CD Cases, n (%)	Controls, n (%)	Odds Ratio (95% CI)	P
Malabsorption Syndrome	82990 (99.9)	305350 (0.9)	96617.1 (79408.52-117555.02)	< 0.0001
Diarrheal Disorder	7210 (8.7)	85030 (0.2)	40.0 (38.98-40.99)	< 0.0001
Malnutrition of Mild Degree (75% to less than 90% of standard weight)	320 (0.4)	30960 (0.1)	4.5 (4.01-5.00)	< 0.0001
Malnutrition of Moderate Degree (60% to less than 75% of standard weight)	580 (0.7)	62120 (0.2)	4.1 (3.73-4.40)	< 0.0001
Severe Protein-Calorie Malnutrition (less than 60% of standard weight)	590 (0.7)	52530 (0.1)	4.9 (4.49-5.29)	< 0.0001
Anemia of Chronic Disease	2800 (3.4)	363680 (1.0)	3.4 (3.28-3.53)	< 0.0001
Anemia due to B12 Deficiency	1910 (2.3)	121900 (0.3)	6.9 (6.59-7.22)	< 0.0001
Iron Deficiency Anemia	9890 (11.9)	788660 (2.2)	6.0 (5.88-6.14)	< 0.0001
Vitamin B Deficiency	5310 (6.4)	426480 (1.2)	5.7 (5.51-5.83)	< 0.0001
Vitamin D Deficiency	16820 (20.2)	2067920 (6.1)	4.1 (4.08-4.22)	< 0.0001
Osteomalacia	170 (0.2)	10220 (0.0)	7.2 (6.18-8.37)	< 0.0001
Osteoporosis	12960 (15.6)	1009370 (2.9)	6.4 (6.26-6.50)	< 0.0001
Depressive Disorder	29950 (36.0)	3985500 (12.5)	4.5 (4.44-4.57)	< 0.0001
Epilepsy	2520 (3.0)	447770 (1.3)	2.5 (2.38-2.57)	< 0.0001
Migraine	15440 (18.6)	1422730 (4.1)	5.5 (5.43-5.62)	< 0.0001
Anxiety Disorder	21520 (25.9)	2878890 (8.7)	4.0 (3.94-4.07)	< 0.0001
Carpal Tunnel Syndrome	3260 (3.9)	562510 (1.6)	2.6 (2.47-2.65)	< 0.0001
Osteoarthritis	28450 (34.2)	4180340 (13.2)	3.9 (3.89-4.00)	< 0.0001
Arthritis	24010 (28.9)	2764390 (8.4)	4.9 (4.79-4.94)	< 0.0001
Arthropathy	46890 (56.4)	9368110 (35.4)	3.7 (3.61-3.71)	< 0.0001
Dermatitis Herpetiformis	1050 (1.3)	4490 (0.0)	4563.5 (4422.50-4708.92)	< 0.0001
Type 2 DM	18720 (22.5)	3220520 (9.9)	2.9 (2.90-3.00)	< 0.0001
Liver Disease	19260 (23.2)	1454560 (4.2)	7.1 (7.02-7.25)	< 0.0001
Chronic Liver Disease	4200 (5.1)	571330 (1.6)	3.3 (3.19-3.39)	< 0.0001
GERD	30550 (36.8)	4111300 (13.0)	4.5 (4.43-4.55)	< 0.0001
Eosinophilic Esophagitis	460 (0.6)	22690 (0.1)	8.8 (8.01-9.64)	< 0.0001
Disorder associated with menopause and/or menstruation	12240 (14.7)	1641500 (4.8)	3.6 (3.53-3.67)	< 0.0001
Infertility	1000 (1.2)	144540 (0.4)	3.0 (2.83-3.20)	< 0.0001
Myocardial Disease	9660 (11.6)	1422310 (4.1)	3.2 (3.12-3.25)	< 0.0001
Myocardial Dysfunction	440 (0.5)	93270 (0.3)	2.0 (1.86-2.24)	< 0.0001
Cardiomyopathy	2210 (2.7)	458750 (1.3)	2.1 (2.02-2.20)	< 0.0001
Ischemic Heart Disease	9790 (11.8)	1385430 (4.0)	3.3 (3.25-3.39)	< 0.0001
Atrophic Gastritis	3270 (3.9)	181870 (0.5)	8.0 (7.76-8.32)	< 0.0001
Glossitis	340 (0.4)	33460 (0.1)	4.4 (3.95-4.90)	< 0.0001
Pancreatitis	13090 (15.8)	266480 (0.7)	25.0 (24.50-25.45)	< 0.0001
Disorder of Pancreas	14330 (17.2)	388970 (1.1)	19.0 (18.66-19.35)	< 0.0001
Peripheral Neuropathy	3640 (4.4)	470780 (1.3)	3.4 (3.33-3.56)	< 0.0001
Cerebellar Ataxia	90 (0.1)	9490 (0.0)	4.1 (3.33-5.04)	< 0.0001
Autism	3300 (4.0)	74210 (0.2)	19.9 (19.24-20.67)	< 0.0001
Colitis	21500 (25.9)	1438240 (4.2)	8.4 (8.22-8.48)	< 0.0001
Turner Syndrome	50 (0.1)	2770 (0.0)	7.8 (5.89-10.31)	< 0.0001
Down Syndrome	460 (0.6)	24730 (0.1)	8.1 (7.35-8.85)	< 0.0001
Common Variable Immunodeficiency	190 (0.2)	8030 (0.0)	10.2 (8.86-11.82)	< 0.0001
IgA Nephropathy	0 (0.0)	370 (0.0)	0.6 (0.04-9.33)	0.702

The odds ratio (OR), its standard error, and 95 % confidence interval were calculated according to Altman using the MedCalc Statistical Software using a case-control design<sup>[10,11]</sup>. It should be mentioned that as a measure to protect the identities of patients, Explorys rounds cell counts to the nearest 10 and treats all cell counts < 10 as equivalent to zero.

## RESULTS

A total of 35,854,260 individuals in the database from April 2012 to April 2017 made up the source population. Of these, 83,090 had at least one SNOMED-CT diagnosis of CD, with an overall age-adjusted prevalence of 231.7/100,000 persons (0.23%). In the CD case group, the majority were female (74.7%), Caucasian (88.0%), and adult (age 18-65; 70.13%). Prevalence was higher in females than males [odds ratio (OR) 2.40; 95% CI 2.36-2.43,  $p < 0.0001$ ], whites versus non-whites (OR 2.25; 95% CI 2.20-2.30,  $p < 0.0001$ ) and adults ( $\geq 18$  years) versus children ( $< 18$ ) (OR 2.60; 95% CI 2.53-2.67,  $p < 0.0001$ ) (Tables 1, 2). The age distribution by gender of the source population and the CD populations is shown in Figures 1 and 2, respectively.

Individuals with CD were more likely than controls to have a variety of signs, symptoms, and associated disorders. A statistically significant association existed between CD and each associated disorder queried except for IgA nephropathy<sup>[12,13]</sup> (Table 3).

In addition, individuals with CD were more likely than controls to have a variety of autoimmune disorders. (Table 4) Though reported in the literature, there was no statistically significant association between CD and PBC<sup>[14,15]</sup>.

## DISCUSSION

In this study, we evaluated the prevalence of diagnosed CD over 5 years in the Explorys database between 2012 and 2017. To our knowledge, our sample size of over 35 million individuals is, by orders of magnitude, the largest population ever used to define CD prevalence in the United States at the national level. This is also the first large study to describe race-, age-, and sex-based prevalence of CD in the United States from national-level data. Finally, this is the largest study to identify conditions associated with CD, both classical and non-classical, and the large cohort enabled identification of rare associations.

We used SNOMED-CT to identify cases of CD, and our study is

the first to use either International Classification of Diseases, Ninth Revision (ICD-9) or SNOMED-CT to define the prevalence of CD. At least three prior studies have attempted to use ICD-9 codes to help further define the epidemiology of CD<sup>[2,16,17]</sup>, and one study has used SNOMED-CT<sup>[18]</sup>, though none has used these methods to define prevalence. While ICD-9 and SNOMED-CT are both medical terminology systems for recording medical diagnoses and concepts, SNOMED-CT has many more concepts to be coded per clinical document than ICD-9<sup>[19]</sup>, which makes it more accurate in terms of enlisting pertinent clinical information<sup>[7]</sup>. The Explorys methodology employed here has been used in the past to identify the prevalence for a variety of conditions with empirically valid results<sup>[7,20]</sup>.

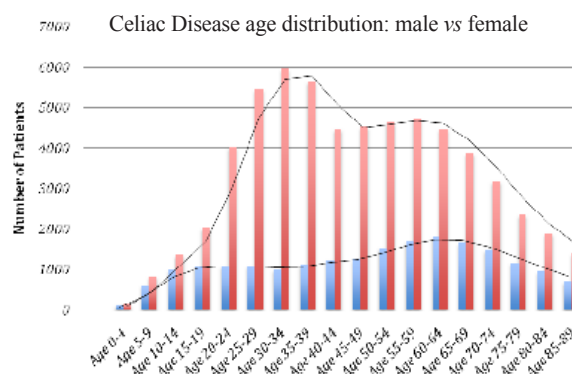


Figure 1 Source Population Age Distribution.

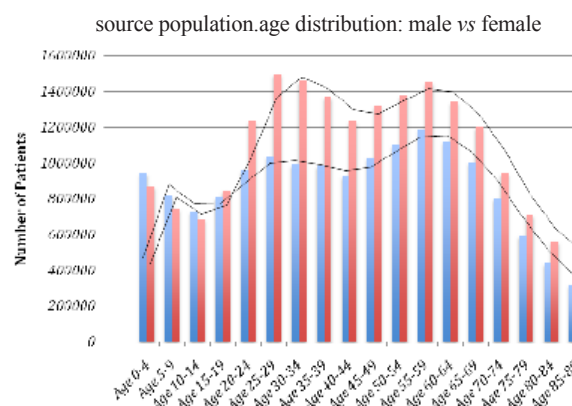


Figure 2 Celiac Disease Age Distribution.

Table 4 CD-Associated Autoimmune Disorders.

Diagnosis	CD Cases, n (%)	Controls, n (%)	Odds Ratio (95% CI)	P
Type 1 DM	11660 (14.0)	363320 (1.0)	15.9 (15.63-16.27)	< 0.0001
Crohn's Disease	7310 (8.8)	140820 (0.4)	24.5 (23.87-25.07)	< 0.0001
Ulcerative Colitis	6850 (8.2)	122310 (0.3)	26.2 (25.59-26.92)	< 0.0001
Graves' Disease	90 (0.1)	8400 (0.0)	4.6 (3.76-5.70)	< 0.0001
Hashimoto's Thyroiditis	3020 (3.6)	111830 (0.3)	12.1 (11.62-12.51)	< 0.0001
Autoimmune Thyroiditis	3260 (3.9)	124560 (0.3)	11.7 (11.31-12.14)	< 0.0001
Myasthenia Gravis	190 (0.2)	19030 (0.1)	4.3 (3.74-4.98)	< 0.0001
Psoriasis	1610 (1.9)	237500 (0.7)	3.0 (2.82-3.11)	< 0.0001
Systemic Lupus Erythematosus	9920 (11.9)	136570 (0.4)	35.5 (34.70-36.23)	< 0.0001
Addison's Disease	90 (0.1)	4850 (0.0)	8.0 (6.51-9.87)	< 0.0001
Rheumatoid Arthritis	11620 (14.0)	378910 (1.1)	15.2 (14.92-15.53)	< 0.0001
Vitiligo	300 (0.4)	32770 (0.1)	4.0 (3.53-4.44)	< 0.0001
Alopecia Areata	210 (0.3)	28450 (0.1)	3.2 (2.79-3.66)	< 0.0001
Primary Biliary Cholangitis	90 (0.1)	5650 (0.0)	1.0 (0.81-1.23)	0.996

CD can be diagnosed at any age, with a peak at early childhood and at the fourth and fifth decade of life for women and men, respectively and our results support this<sup>[21]</sup>. The white predominance observed in our study is consistent with prior epidemiological data, which show CD is a disease primarily of whites (though research increasingly suggests that it affects all races)<sup>[1-6]</sup>. As with many other autoimmune diseases, CD is more common in women, and our results are consistent with prior observed female to male ratio between 2:1 and 3:1<sup>[21,22]</sup>. Further genetic, environmental, and behavioral studies are needed to understand the higher prevalence in females and Caucasians.

The overall prevalence of 231.7/100,000 persons (0.23%) identified in our study is significantly lower than the rate identified in the US through screening. These studies have found CD prevalence rates of approximately 1-2% in Europe and up to 0.95% in the USA<sup>[4,21,23,24]</sup>.

One strength of this study is that the data contained in Explorys is culled from over 300 health institutions across all 50 states and spanning the East, Midwest, South, Central, and West divisions of the USA<sup>[8]</sup>, thus providing a broad regional and climatic distribution of source population. In this way, the patients included in this study are broadly representative of the general US population that utilizes healthcare services. However, a study of CD prevalence in the US utilizing the National Health Examination Survey (NHANES) observed stark regional differences in 22,000 cases of CD identified by questionnaire and sequential serology<sup>[24]</sup>. The overall US prevalence identified by this study was 0.7%, but prevalence increased from 0.2% to 0.6% to 1.2% as latitude increased from < 35°North to 35° to < 40°North to 40°North, respectively, with the highest prevalence located in the northernmost US latitude., with the highest prevalence located in the northernmost US latitude<sup>[24]</sup>. Because Explorys uses aggregated, de-identified data from institutions across the US, it is not possible to determine whether regional differences in prevalence exist, and we are unable to corroborate the regional disparities observed in NHANES data. However, due to the broad geographic distribution of the data contained in Explorys, the authors feel that regional differences cannot explain the lower prevalence observed in our study, which suggest that the nationwide prevalence of diagnosed CD is roughly equivalent to that found in the southernmost portion of the US in NHANES (the least prevalent region).

A more likely explanation of the discrepancy is the difficulty of CD diagnosis, and the differences between tests used for screening and those used for diagnosis. It should be noted that the results of this study show the prevalence of *diagnosed* CD, while screening studies show the prevalence of *diagnosed plus undiagnosed* CD. As with any disease, diagnosis of CD depends first on identifying which individuals to test diagnostically. Classic malabsorption symptoms (e.g. bloating, diarrhea, weight loss, etc.) are relatively straightforward testing indications<sup>[25]</sup>. However, wide clinical discretion is required to test for atypical CD, which may present with a variety of non-malabsorption symptoms, and is thought to comprise as much as 50% of all diagnosed patients<sup>[26]</sup>. Because of this, we sought to identify the prevalence of a multitude of non-classic conditions associated with CD in order to identify possible indications for CD testing. We were able to find strong associations between CD and many non-classic symptoms.

In addition to the above, for any disease with a wide variation of phenotypic expressivity and a serologic diagnostic test, of which CD is one, the screening of asymptomatic individuals yields a higher proportion of diseased individuals than is found in the population<sup>[27]</sup>.

A wide range of clinical severity exists among individuals with both classic CD and atypical variants<sup>[28]</sup>. Thus, individuals with silent or mild CD may satisfy diagnostic criteria when screened but fail to show clinical signs of CD that would otherwise warrant diagnostic testing in the clinical setting<sup>[28]</sup>. Unlike with screening tests, all cases identified in Explorys are diagnosed CD. The results of this study, therefore, are consistent with NHANES data that suggest that up to 83% of CD cases go undiagnosed, with a considerable morbidity burden of undiagnosed disease<sup>[5]</sup>. Indeed, undiagnosed CD is associated with a nearly 4-fold increased risk of death<sup>[29]</sup>. Therefore, we recommend liberal use of CD diagnostic testing in patients that present with the atypical symptoms described here, especially in the presence of unexplained, refractory, or malabsorptive qualities, and especially in those patients who also have other autoimmune conditions.

This study is subject to the same limitations affecting all epidemiological research collected from electronic databases. Because the data contained in electronic databases were not collected as part of a designed study, many variables of interest to research remain unrecorded<sup>[30]</sup>. Explorys data contain no information about socioeconomic status, geographic data on patient population, endoscopic abnormalities, histology reports, severity of disease, and specific indications for medications prescribed<sup>[7]</sup>. Moreover, although Explorys uses a master-patient identifier to match the same patient across different healthcare institutions and combine the data<sup>[31]</sup>, a few patients may have received care in multiple institutions within Explorys healthcare partners and could have been counted multiple times. However, the healthcare systems involved in Explorys are large, integrated networks, and while it is possible that some patients may travel between healthcare systems, it is not clear that any error in this area would disproportionately affect the CD group more than the control group<sup>[31]</sup>.

Although evidence suggests reliable recording of principal diagnosis codes for many disorders (e.g. myocardial infarction), many other medical conditions are coded with limited accuracy<sup>[32]</sup>. We could have under- or overestimated the true prevalence of CD since the SNOMED-CT diagnostic code for CD in Explorys was not validated clinically, and individuals might have been misclassified; validation of the SNOMED-CT diagnostic code for CD was not possible since the patient information in the database is de-identified. Furthermore, providers sometimes omit single diagnosis codes when more serious conditions push milder diagnoses off the diagnosis list or discharge abstract<sup>[30]</sup>. For example, studies of hospital discharge abstract data have found that secondary diagnoses such as diabetes, angina, and history of myocardial infarction are paradoxically associated with improved in-hospital mortality, as their inclusion in discharge diagnosis codes implies less serious primary diagnoses<sup>[33]</sup>. There also appears to be an under-coding of chronic comorbidities in patients who die in the hospital<sup>[30]</sup>. The chronic and often indolent nature of CD lends itself to such under-diagnosis.

Finally, though the major strength of this study is its large sample size, one potential consequence is the power to find statistically significant associations which may not be clinically significant. More research may be required to make this differentiation.

In summary, using the large commercial database, Explorys, we have analyzed the largest cohort of diagnosed CD cases to date. From this sample, the overall prevalence of diagnosed CD in the US is estimated to be 231.7/100,000 persons (0.23%), which is lower than estimates derived from screening methodology. These results support past studies which suggest a large burden of undiagnosed CD, which is at least partially due to unidentified atypical disease. Because of

this, we suggest the liberal use of CD diagnostic testing in patients presenting with unexplained or refractory CD-associated symptoms, which we have identified here. We hope this study may provide guidance for the clinical recognition of classical and non-classical CD.

## ACKNOWLEDGMENTS

The results of this study were presented as a poster at the World Congress of Gastroenterology at American College of Gastroenterology, Orlando, Fla. October 17, 2017 entitled Who Should We Screen for Celiac Disease? Prevalence and Risk Factors for Celiac Disease in USA: A Large Population-Based Study of 83,090 Patients with Celiac Disease.

## REFERENCES

1. Abdo A, Meddings J, Swain M. Liver Abnormalities in Celiac Disease. *Clinical Gastroenterology and Hepatology*. 2004; **2**(2): 107-112. [PMID: 15017614]; [DOI: 10.1016/S1542-3565(03)00313-6]
2. Badenhop K, Dieterich W, Segni M, Hofmann S, Hüfner M, Usadel KH, Hahn EG, Schuppan D. HLA DQ2 and/or DQ8 is Associated with Celiac Disease-Specific Autoantibodies to Tissue Transglutaminase in Families with Thyroid Autoimmunity. *Am J Gastroenterol*. 2001; **96**(5): 1648. [PMID: 11374722]; [DOI: 10.1111/j.1572-0241.2001.03821.x]
3. Bardella MT, Quatrini M, Zuin M, Podda M, Cesarini L, Velio P, Bianchi P, Conte D. Screening Patients with Celiac Disease for Primary Biliary Cirrhosis and Vice Versa. *Am J Gastroenterol*. 1997; **92**(9): 1524-6. [PMID: 9317077]
4. Bardella MT, Vecchi M, Conte D, Del Ninno E, Fraquelli M, Pacchetti S, Minola E, Landoni M, Cesana BM, De Franchis R. Chronic Unexplained Hypertransaminasemia May be Caused by Occult Celiac Disease. *Hepatology*. 1999; **29**(3): 654-657. [PMID: 10051464]; [DOI: 10.1002/hep.510290318]
5. Carlsson A, Axelsson I, Borulf S, Bredberg A, Forslund M, Lindberg B, Sjöberg K, Ivarsson SA. Prevalence of IgA-Antigliadin Antibodies and IgA-Antiendomysium Antibodies Related to Celiac Disease in Children with Down Syndrome. *Pediatrics*. 1998; **101**(2): 272-275. [PMID: 9445503]; [DOI: 10.1542/peds.101.2.272]
6. Cataldo F, Marino V, Bottaro G, Greco P, Ventura A, Pediatrca C. Celiac Disease and Selective Immunoglobulin A Deficiency. *J Pediatr*. 1997; **131**(2): 306-308. [PMID: 22476042]; [DOI: 10.1097/MCG.0b013e31824b2277]
7. Collin P, Maki M, Keyrilainen O, Hallstrom O, Reunala T, Pasternack A. Selective IgA Deficiency and Coeliac Disease. *Scand J Gastroenterol*. 1992; **27**(5): 367-371. [PMID: 1529270]; [DOI: 10.3109/00365529209000089]
8. Collin P, Vilksa S, Heinonen PK, Hallstrom O, Pikkarainen P. Infertility and Coeliac Disease. *Gut*. 1996; **39**(3): 382-384. [PMID: 8949641]
9. Counsell CE, Taha A, Ruddell WS. Coeliac Disease and Autoimmune Thyroid Disease. *Gut*. 1994; **35**(6): 844-846. [PMID: 802081]
10. Curione M, Barbato M, De Biase L, Viola F, Russo LL, Cardi E. Prevalence of Coeliac Disease in Idiopathic Dilated Cardiomyopathy. *The Lancet*. 1999; **354**(9174): 222-223. [PMID: 10421311]; [DOI: 10.1016/S0140-6736(99)01501-9]
11. Dickey W, McMillan SA, Callender ME. High Prevalence of Celiac Sprue Among Patients with Primary Biliary Cirrhosis. *J Clin Gastroenterol*. 1997; **25**(1): 328-329. [PMID: 9412913]
12. Duggan J, Duggan A. Systematic Review: The Liver in Coeliac Disease. *Aliment Pharmacol Ther*. 2005; **21**(5): 515-518. [PMID: 15740533]; [DOI: 10.1111/j.1365-2036.2005.02361.x]
13. Einarsdottir E, Koskinen LL, Dukes E, Kainu K, Suomela S, Lappalainen M, Zibera F, Korponay-Szabo IR, Kurppa K, Kaukinen K, Adány R, Pocsai Z, Széles G, Färkkilä M, Turunen U, Halme L, Paavola-Sakki P, Not T, Vatta S, Ventura A, Löfberg R, Torkvist L, Bresso F, Halfvarson J, Mäki M, Kontula K, Saarialho-Kere U, Kere J, D'Amato M, Saavalainen P. IL23R in the Swedish, Finnish, Hungarian and Italian Populations: Association with IBD and Psoriasis, and Linkage to Celiac Disease. *BMC Medical Genetics*. 2009; **10**(1): 8. [PMID: 19175939]; [DOI: 10.1186/1471-2350-10-8]
14. Farthing MJ, Rees LH, Edwards CR, Dawson AM. Male Gonadal Function in Coeliac Disease: 2. Sex Hormones. *Gut*. 1983; **24**(2): 127-135. [PMID: 6682819]; [DOI: 10.1136/gut.23.7.608]
15. Frustaci A, Cuoco L, Chimenti C, Pieroni M, Fioravanti G, Gentiloni N, Maseri A, Gasbarrini G. Celiac Disease Associated with Autoimmune Myocarditis. *Circulation*. 2002; **105**(22): 2611-2618. [PMID: 12045166]; [DOI: 10.1161/01.CIR.0000017880.86166.87]
16. Gale L, Wimalaratna H, Brotodiharjo A, Duggan JM. Down's Syndrome is Strongly Associated with Coeliac Disease. *Gut*. 1997; **40**(4): 492-496. [PMID: 9176077]; [DOI: 10.1136/gut.40.4.492]
17. Gasbarrini A, Torre ES, Trivellini C, De Carolis S, Caruso A, Gasbarrini G. Recurrent Spontaneous Abortion and Intrauterine Fetal Growth Retardation as Symptoms of Coeliac Disease. *The Lancet*. 2000; **356**(9227): 399-400. [PMID: 10972376]; [DOI: 10.1016/S0140-6736(00)02535-6]
18. Glas J, Stallhofer J, Ripke S, Wetzke M, Pfennig S, Klein W, Eppelen JT, Griga T, Schiemann U, Lacher M, Koletzko S, Folwaczny M, Lohse P, Göke B, Ochsenkühn T, Müller-Myhsok B, Brand S. Novel Genetic Risk Markers for Ulcerative Colitis in the IL2/IL21 Region are in Epistasis with IL23R and Suggest a Common Genetic Background for Ulcerative Colitis and Celiac Disease. *Am J Gastroenterol*. 2009; **104**(7): 1737-1744. [PMID: 19455118]; [DOI: 10.1038/ajg.2009.163]
19. Jandaghi E, Hojatinia M, Vahedi H, Shahbaz-Khani B, Kolahdoozan S, Ansari R. Is the Prevalence of Celiac Disease Higher than the General Population in Inflammatory Bowel Disease? *Middle East J Dig Dis*. 2015; **7**(2): 82-87. [PMID: 26106467]
20. Kaukinen K, Halme L, Collin P, Färkkilä M, Mäki M, Vehmanen P, Partanen J, Höckerstedt K. Celiac Disease in Patients with Severe Liver Disease: Gluten-Free Diet May Reverse Hepatic Failure. *Gastroenterology*. 2002; **122**(4): 881-888. [PMID: 11910339]; [DOI: 10.1053/gast.2002.32416]
21. Kieft-de Jong JC, Jaddoe VW, Uitterlinden AG, Steegers EA, Willemsen SP, Hofman A, Hooijkaas H, Moll HA. Levels of Antibodies Against Tissue Transglutaminase During Pregnancy are Associated with Reduced Fetal Weight and Birth Weight. *Gastroenterology*. 2013; **144**(4): 726-735.e2. [PMID: 23313966]; [DOI: 10.1053/j.gastro.2013.01.003]
22. Kingham JG, Parker DR. The Association Between Primary Biliary Cirrhosis and Coeliac Disease: A Study of Relative Prevalences. *Gut*. 1998; **42**(1): 120-122. [PMID: 9518232]; [DOI: 10.1136/gut.42.1.120]
23. Kocsis D, Tóth Z, Csontos ÁA, Miheller P, Pák P, Herszényi L, Tóth M, Tulassay Z, Juhász M. Prevalence of Inflammatory Bowel Disease Among Coeliac Disease Patients in a Hungarian Coeliac Centre. *BMC Gastroenterology*. 2015; **15**(1): 141. [PMID: 26481725]; [DOI: 10.1186/s12876-015-0370-7]
24. Kumar A, Meena M, Begum N, Kumar N, Gupta RK, Aggarwal S, Prasad S, Batra S. Latent Celiac Disease in Reproductive Performance of Women. *Fertil Steril*. 2011; **95**(3): 922-927. [PMID: 21106194]; [DOI: 10.1016/j.fertnstert.2010.11.005]
25. Lähteenoja H, Toivanen A, Viander M, Maki M, Irjala K, Raihaä I, Syrjänen S. Oral Mucosal Changes in Coeliac Patients on a Gluten-Free Diet. *Eur J Oral Sci*. 1998; **106**: 899-906. [PMID: 9786318]; [DOI: 10.1046/j.0909-8836.1998.eos106501.x]
26. Leslie C, Mews C, Charles A, Ravikumara M. Celiac Disease and Eosinophilic Esophagitis: A True Association. *J Pediatr Gastroenterol Nutr*. 2010; **50**(4): 397-399. [PMID: 19841598]; [DOI:

- 10.1097/MPG.0b013e3181a70af4]
27. Ludvigsson JF, de Faire U, Ekbom A, Montgomery SM. Vascular Disease in a Population-Based Cohort of Individuals Hospitalized with Coeliac Disease. *Heart*. 2007; **93(9)**: 1111-1115. [PMID: 17277354]; [DOI: 10.1136/hrt.2006.097097]
  28. Ludvigsson JF, Elfstrom P, Broom E U, Ekbom A, Montgomery SM. Celiac Disease and Risk of Liver Disease: A General Population-Based Study. *Clinical Gastroenterology and Hepatology*. 2007; **5(1)**: 63-69. e1. [PMID: 17161656]; [DOI: 10.1016/j.cgh.2006.09.034]
  29. Ludvigsson JF, James S, Askling J, Stenestrand U, Ingelsson E. Nationwide Cohort Study of Risk of Ischemic Heart Disease in Patients with Celiac Disease. *Circulation*. 2011; **123(5)**: 483-490. [PMID: 21262996]; [DOI: 10.1161/CIRCULATIONAHA.110.965624]
  30. Ludvigsson JF, Montgomery SM, Ekbom A. Risk of Pancreatitis in 14,000 Individuals with Celiac Disease. *Clinical Gastroenterology and Hepatology*. 2007; **5(11)**: 1347-1353. e2. [PMID: 17702659]; [DOI: 10.1016/j.cgh.2007.06.002]
  31. Meini A, Pillan NM, Ugazio AG, Villanacci V, Monafò V, Plebani A. Prevalence and Diagnosis of Celiac Disease in IgA-Deficient Children. *Annals of Allergy, Asthma & Immunology*. 1996; **77(4)**: 333-336. [PMID: 8885812]; [DOI: 10.1016/S1081-1206(10)63329-7]
  32. Meloni GF, Dessole S, Vargiu N, Tomasi PA, Musumeci S. The Prevalence of Coeliac Disease in Infertility. *Hum Reprod*. 1999; **14(11)**: 2759-2761. [PMID: 10548618]; [DOI: 10.1093/humrep/14.11.2759]
  33. Nachman F, Vázquez H, González A, Andrenacci P, Compagni L, Reyes H, Sugai E, Moreno ML, Smecuol E, Hwang HJ, Sánchez IP, Mauriño E, Bai JC. Gastroesophageal Reflux Symptoms in Patients with Celiac Disease and the Effects of a Gluten-Free diet. *Clinical Gastroenterology and Hepatology*. 2011; **9(3)**: 214-219. [PMID: 20601132]; [DOI: 10.1016/j.cgh.2010.06.017]
  34. Nørgård B, Fonager K, Sørensen HT, Olsen J. Birth Outcomes of Women with Celiac Disease: A Nationwide Historical Cohort Study. *Am J Gastroenterol*. 1999; **94(9)**: 2435-2440. [PMID: 10484005]; [DOI: 10.1111/j.1572-0241.1999.01370.x]
  35. Ooi CY, Day AS, Jackson R, Bohane TD, Tobias V, Lemberg DA. Eosinophilic Esophagitis in Children with Celiac Disease. *J Gastroenterol Hepatol*. 2008; **23(7pt1)**: 1144-1148. [PMID: 18070017]; [DOI: 10.1111/j.1440-1746.2007.05239.x]
  36. Oxford EC, Nguyen DD, Sauk J, Korzenik JR, Yajnik V, Friedman S, Ananthakrishnan AN. Impact of Coexistent Celiac Disease on Phenotype and Natural History of Inflammatory Bowel Diseases. *Am J Gastroenterol*. 2013; **108(7)**: 1123-1129. [PMID: 23419379]; [DOI: 10.1038/ajg.2013.20]
  37. Pascual V, Dieli-Crimi R, Lopez-Palacios N, Bodas A, Medrano LM, Núñez C. Inflammatory Bowel Disease and Celiac Disease: Overlaps and Differences. *World Journal of Gastroenterology*. 2014; **20(17)**: 4846-4856. [PMID: 24803796]; [DOI: 10.3748/wjg.v20.i17.4846]
  38. Pastore L, Muzio LL, Serpico R. Atrophic Glossitis Leading to the Diagnosis of Celiac Disease. *N Engl J Med*. 2007; **356(24)**: 2547-2547. [PMID: 17568041]; [DOI: 10.1056/NEJMc070200]
  39. Quaglietta L, Coccorullo P, Miele E, Pascarella F, Troncone R, Staiano A. Eosinophilic Oesophagitis and Coeliac Disease: Is There an Association? *Aliment Pharmacol Ther*. 2007; **26(3)**: 487-493. [PMID: 17635383]; [DOI: 10.1111/j.1365-2036.2007.03388.x]
  40. Reunala T, Kosnai I, Karpati S, Kuitunen P, Torok E, Savilahti E. Dermatitis Herpetiformis: Jejunal Findings and Skin Response to Gluten Free Diet. *Arch Dis Child*. 1984; **59(6)**: 517-522. [PMID: 6742871]; [PMCID: PMC1628792]
  41. Rubio-Tapia A, Murray JA. The Liver in Celiac Disease. *Hepatology*. 2007; **46(5)**: 1650-1658. [PMID: 17969053]; [DOI: 10.1002/hep.21949]
  42. Sadr-Azodi O, Sanders DS, Murray JA, Ludvigsson JF. Patients with Celiac Disease Have an Increased Risk for Pancreatitis. *Clinical Gastroenterology and Hepatology*. 2012; **10(10)**: 1136-1142. e3. [PMID: 22801059]; [DOI: 10.1016/j.cgh.2012.06.023]
  43. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C, Heap GA, Dubois PC, Clayton DG, Hunt KA, van Heel DA, Todd JA. Shared and Distinct Genetic Variants in Type 1 Diabetes and Celiac Disease. *N Engl J Med*. 2008; **359(26)**: 2767-2777. [PMID: 19073967]; [DOI: 10.1056/NEJMoa0807917]
  44. Tata L, Card T, Logan R, Hubbard R, Smith C, West J. Fertility and pregnancy-related events in women with celiac disease: A Population-Based Cohort Study. *Gastroenterology*. 2005; **128(4)**: 849-855. [PMID: 15825068]; [DOI: 10.1053/j.gastro.2005.02.017]
  45. Thompson JS, Leibold B, Reilly NR, Talley NJ, Bhagat G, Green PH. Increased Incidence of Eosinophilic Esophagitis in Children and Adults with Celiac Disease. *J Clin Gastroenterol*. 2012; **46(1)**: e6-e11. [PMID: 21778897]; [DOI: 10.1097/MCG.0b013e318221aefd]
  46. Volta U, De Franceschi L, Lari F, Molinaro N, Zoli M, Bianchi FB. Coeliac Disease Hidden by Cryptogenic Hypertransaminasaemia. *The Lancet*. 1998; **352(9121)**: 26-29. [PMID: 9800742]; [DOI: 10.1016/S0140-6736(97)11222-3]
  47. Wei L, Spiers E, Reynolds N, Walsh S, Fahey T, MacDonald T. The Association Between Coeliac Disease and Cardiovascular Disease. *Aliment Pharmacol Ther*. 2008; **27(6)**: 514-519. [PMID: 18162081]; [DOI: 10.1111/j.1365-2036.2007.03594.x]