

81484\_Auto\_Edited-check.docx

2

**Name of Journal:** *World Journal of Gastrointestinal Oncology*

**Manuscript NO:** 81484

**Manuscript Type:** ORIGINAL ARTICLE

*Retrospective Cohort Study*

**Risk of** pancreatic cancer in individuals with celiac disease in the United States: A population-based matched cohort study

Krishnan A *et al.* Risk of PC in CD

## **Abstract**

### **BACKGROUND**

Celiac disease (CD) has been associated with gastrointestinal malignancies. However, the magnitude of the risk of pancreatic cancer (PC) associated with CD is much less clear, and risks have not been estimated from large populations.

### **AIM**

To assess the risk of PC in CD patients.

### **METHODS**

We conducted a population-based, multicenter, propensity score-matched cohort study with consecutive patients diagnosed with CD using the TriNeTx research network platform. We examined the incidence of PC in patients with CD compared with a matched cohort of patients without CD (non-CD, controls). Each patient in the main group (CD) was matched to a patient in the control group using 1:1 propensity score matching to reduce confounding effects. The incidence of PC was estimated using a Cox proportional hazards model with hazard ratio (HR) and 95% confidence interval (CI).

### **RESULTS**

A total of 389980 patients were included in this study. Among them, 155877 patients had a diagnosis of CD, and the remaining 234103 individuals without CD were considered a control cohort. The mean duration of follow-up for patients in the CD and control cohorts was  $5.8 \pm 1.8$  and  $5.9 \pm 1.1$  years. During the follow-up, 309 patients with CD developed PC, whereas 240 patients developed PC in the control group (HR = 1.29; 95%CI: 1.09-1.53). In the secondary analyses in the first year after diagnosis of CD, patients with CD were at a significant increase in risk for PC; 151 patients with CD had an incidence of PC compared with 96 incidences of PC among the patients in the non-CD control group (HR = 1.56; 95%CI: 1.20-2.01) and sensitivity analysis showed similar magnitude to the one generated in the primary and secondary analysis.

## CONCLUSION

Patients with CD are at increased risk of PC. Risk elevation persists beyond the first year after diagnosis to reference individuals without CD from the general population.

**Key Words:** Celiac disease; Cancer; Epidemiology; Pancreas; Pancreatic cancer; Malignancy; Carcinoma

Krishnan A, Hadi YB, Shabih S, Mukherjee D, Patel RA, Patel R, Singh S, Thakkar S. Risk of pancreatic cancer in individuals with celiac disease in the United States: A population-based matched cohort study. *World J Gastrointest Oncol* 2023; In press

**Core Tip:** Celiac disease (CD) is an immune-mediated disorder precipitated by the ingestion of gluten in genetically predisposed individuals. Population-based studies have shown the risk of cancer among patients with CD; however, the magnitude of the risk of pancreatic cancer (PC) in association with CD is much less clear. Therefore, In this multicenter retrospective cohort study, we analyzed the risk of PC in patients with CD. We examined the incidence of PC in patients with CD compared with a propensity-matched cohort of patients without CD. We found that patients with CD were at increased risk of PC. Risk elevation persisted beyond the first year after diagnosis to reference individuals without CD from the general population.

## INTRODUCTION

Celiac disease (CD) is an immune-mediated disorder with a worldwide prevalence approaching 1%. A higher risk of death<sup>[1]</sup>, and an increased risk of cancer, particularly lymphomatous malignancies<sup>[2]</sup> and gastrointestinal malignancies, including small bowel adenocarcinoma<sup>[3]</sup>, among others, have been reported in patients with CD. However, most of the reports of cancer risk in CD suffer from the small sample size and are based on data derived from the pre-serologic testing era, and thus likely selected for

more severe phenotypic cases. Therefore, cancer risk in CD remains a subject of current interest.

The magnitude of the risk of pancreatic cancer (PC) in association with CD is much less clear<sup>[1,4]</sup>. Nevertheless, despite the magnitude of the risk remaining debatable, the unequivocal association between CD and PC remains. In the two most recent matched studies that included data on PC risk in CD, a higher risk of PC diagnosis was observed in CD patients; however, discordant results were obtained regarding whether PC risk persists long-term after CD diagnosis<sup>[4,5]</sup>. While Elfström *et al*<sup>[4]</sup> noted no persistent increased risk of PC after the first year of CD diagnosis, Lebwohl *et al*<sup>[5]</sup> in their recent study, have noted a persistently elevated risk of PC in patients with CD after 1 year of diagnosis. Hence, there is considerable, but not definitive, evidence that strict compliance to CD is associated with increased risk for the development of PC. Furthermore, owing to the lack of control for known risk factors of PC, any independent associations between these diseases cannot be assessed based on current data; given the high incidence of CD, and the poor outcomes associated with PC, any such potential association warrants further investigation. To address the abovementioned knowledge gaps, we conducted a propensity-matched retrospective cohort study on a large multi-institutional electronic healthcare registry to assess the risk of PC in patients with CD.

## **MATERIALS AND METHODS**

### ***Methods***

We conducted a retrospective cohort study using TriNetX multi-institutional research network (Cambridge, MA, United States). TriNetX is a federated multicenter research network that provides real-time access to an anonymized data set from participating healthcare organizations' electronic health records (EHR). Patients were included using standard International Classification of Diseases (classification, Tenth Revision, Clinical Modification) (ICD-10-CM), and Current Procedural Terminology classification terminologies. Details of the data source, quality checks, and diagnosis codes used for patient selection are detailed in the supplementary. TriNetX has received a waiver from

Western Institutional Review Board as it only provides de-identified information. At our institution West Virginia University, the Clinical and Translational Science Institute manages the TriNetX platform and provides access to the end-users.

### *Study participants*

All patients with the diagnosis of CD from January 1, 2005, to December 31, 2021, were included. Follow-up of these patients ended on September 31, 2022. The diagnosis of CD was on the ICD-10 diagnostic code. During the same study period, patients who didn't have a diagnosis of CD were classified as non-CD comparators (control group). We excluded CD patients and controls with a history of pancreatic diseases, gallbladder diseases, procedures such as endoscopic retrograde cholangiopancreatography (ERCP) or cholecystectomy, and any malignancies, including PC preceding the index period (defined as the diagnosis of CD) or the corresponding date for controls.

### *Matching process*

Each patient in the CD group was matched to a patient in the control (non-CD comparators) group using 1:1 propensity score matching (PSM) to reduce confounding effects. Covariates in the propensity score model were adjusted for apriori-identified potential confounders: age, sex, race/ethnicity (Hispanic, non-Hispanic white, non-Hispanic black, or non-Hispanic other), nicotine dependence, alcohol-related disease, body mass index (BMI), type 2 diabetes mellitus (T2DM), hypertension, hyperlipidemia, type 1 diabetes, autoimmune thyroid disease, rheumatoid arthritis, inflammatory bowel disease, disease of gallbladder, diseases of the pancreas, hypercholesterolemia, hypercalcemia, familial hypercholesterolemia, family history of primary malignant neoplasm, ERCP, cholecystectomy, cholesterol level, low-density lipoprotein, and serum triglyceride level, hemoglobin A1C (HbA1C) and genomics (Table 1). Logistic regression on these input matrices was used to obtain propensity scores for each patient in both cohorts. Logistic regression was performed in Python 3.6.5 (Python Software Foundation) using standard libraries NumPy and sklearn. The same analyses were also

performed in R 3.4.4 software (R Foundation for Statistical Computing, Vienna, Austria) to ensure outputs match. After calculating propensity scores, matching was performed using a greedy nearest-neighbor matching algorithm with a caliper of 0.1 pooled standard deviations. The balancing of potential confounding variables between the CD and the control group after PSM was evaluated using standardized differences (SMD) with a threshold set a priori at 0.10. SMD was used to measure the magnitude of difference between the groups rather than the *P*-value because of their insensitivity to sample size. The order of the rows in the covariate matrix can affect the nearest neighbor matching; therefore, the order of the rows in the matrix was randomized to eliminate this bias.

### ***Outcomes***

Our primary outcome was the incidence of PC. We defined PC according to relevant ICD codes (details depicted in the Supplementary).

### ***Statistical analysis***

All statistical analyses were performed in real-time using TriNetX. Categorical variables were compared using the Pearson chi-square test, and continuous variables using an independent-sample *t*-test. Continuous variables are expressed as mean  $\pm$  standard deviation (SD). Categorical variables were presented as frequency and percentage. The balancing of potential confounding variables between the BS and the control group after PSM was evaluated using SMD with a threshold set a priori at 0.10. SMD was used to measure the magnitude of difference between the groups rather than the *P*-value because of their insensitivity to sample size. For each outcome, Cox proportional hazards models were used to estimate the hazard ratio (HR). HR and confidence interval (CI), along with tests for proportionality, were calculated using R's Survival package v3.2-3. Numbers were then validated by comparing them with the output from SAS version 9.4. Kaplan-Meier survival analyses were also used to estimate the risk of PC at the end of 7 years after the index event. Patients were censored when the time

window ended or the day after the last fact in their record. Hypothesis testing for Kaplan-Meier survival curves was conducted by using the log-rank test. A priori-defined 2-sided alpha of  $< 0.05$  was used for statistical significance.

### *Secondary analyses*

We performed a prespecified secondary analysis. To assess the possible duration of the disease and risk of PC, we performed a follow-up time-specific risk of PC. For these time-dependent analyses, we estimated time-specific HRs and HRs in which follow-up time commences 1-year after CD diagnosis.

### *Sensitivity analysis*

Due to the heterogeneous nature of cancer and to assess the robustness of our findings. We performed a sensitivity analysis by estimating risk after excluding patients with outcomes within 1-year after the index event. In this analysis, the outcome was the first and new diagnosis of PC that developed after the start of follow-up, as some cases of PC have been reported to occur shortly after diagnosis of CD.

## **RESULTS**

### *Baseline characteristics*

A total of 389980 patients were included in this study. Among them, 155877 patients had a diagnosis of CD and were thus placed in the main cohort (CD) (Figure 1). The remaining 234103 individuals did not have CD and were considered as a control cohort (non-CD comparators). The mean duration of follow-up for patients in the CD and control cohorts was  $5.8 \pm 1.8$  and  $5.9 \pm 1.1$  years, respectively. The baseline characteristics of the study cohorts are depicted in Table 1. Patients with CD were generally younger ( $38.8 \pm 22.0$  vs  $47.2 \pm 19.0$ ) than the control cohort group. In both groups, the majority of the patients were females and white race. There were differences in comorbidities between the groups at baseline before PSM (Table 1). Type 1 diabetes, autoimmune thyroiditis, Crohn's disease, and ulcerative colitis were slightly more



common in patients with CD than in controls. However, compared with the CD cohort, the non-CD cohort had a higher prevalence of comorbidities, including T2DM, hyperlipidemia, and hypercholesterolemia. In addition, 76% of the CD cohort had positive for tissue transglutaminase immunoglobulin A serology. After PSM, both cohorts (134680 patients each) were well-matched and balanced (Supplementary Figure 1). In contrast, compared with the CD cohort, the control cohort had a higher prevalence of rheumatoid arthritis (1.8% *vs* 5.9%). No significant difference was observed between other comorbidities, procedures, genomic profile, family history of primary malignant neoplasm, or hypercholesterolemia between both cohorts after propensity matching.

### *Study outcome*

During the follow-up, 309 patients with CD developed PC, whereas 240 patients developed PC in the control group. Patients with CD <sup>4</sup> had a significantly higher risk of PC (HR = 1.29; 95%CI: 1.09-1.53) than the non-CD controls (Figure 2). In the secondary analyses for the incidence of PC, when starting follow-up at 1-year after diagnosis of CD, 151 patients with CD had an incidence of PC compared with 96 incidences of PC among the patients in the non-CD control group. <sup>7</sup> CD was associated with a significantly higher risk of PC (HR = 1.56; 95%CI: 1.20-2.01). Kaplan-Meier survival analysis showed a cumulative incidence of PC at 7 years among patients with CD compared to non-CD controls (log-rank test  $P < 0.001$ , Figure 3).

### *Sensitivity analyses*

The sensitivity analyses led to consistent results when starting follow-up at 1 year after diagnosis of CD by excluding the outcomes within 1 year after the index event. CD <sup>3</sup> generated a hazard ratio of similar magnitude to the one generated in the primary analysis (HR = 1.22; 95%CI: 1.01-1.49) (Figure 2).

## **DISCUSSION**

This was a population-based, multicenter retrospective study analysis that used a large sample size from a nationally representative database and statistical adjustments with apriori-identified potential confounders to balance the critical variables at the baseline by PSM. Patients with CD were found to have a modest excess risk of PC. The association of CD with PC remained unchanged in the secondary and sensitivity analyses, where a lag period of 1 year was applied. Our study confirmed that the risk for PC is significantly greater in patients with CD than in the unaffected population.

Previous retrospective cohort studies have noted an increase in overall cancer risk and gastrointestinal cancer risk in patients with CD<sup>[4,6,7]</sup>. However, as most of the reported data on cancer risk in CD is derived from patients diagnosed in the 20<sup>th</sup> century, concerns have been raised about its current validity as these CD cases likely represent a more severe disease phenotype<sup>[4]</sup>. Some recent reports have noted that the overall excess risk of malignancy diagnosis in CD may be elevated only in the short term following CD diagnosis (about 1 year), and the risk elevation was noted to subside after the first year of CD diagnosis<sup>[4,5,8]</sup>. Two hypothetical explanations have been proposed in the literature to explain this finding: (1) The risk elevation may be mitigated over time due to the introduction of a gluten-free diet after diagnosis<sup>[9,10]</sup>; or (2) The elevated risk of malignancy in the short term may be a reflection of the increased surveillance and diagnostic testing that these patients may be undergoing in the peridiagnostic period due to presentation with weight loss and other mutual symptoms that may result in the discovery of CD in malignancy cases and vice versa, and thus not be a true elevation in risk<sup>[10,11,12]</sup>. Thus any excess overall malignancy risk in CD is still unclear. However, an excess risk of certain individual malignancies in CD, *e.g.*, small bowel adenocarcinoma, lymphoproliferative malignancies, and liver cancer, have been strongly noted in CD and have been reported to persist in the longer term beyond 1 year<sup>[3]</sup>. More recently, Lebwohl *et al*<sup>[5]</sup> in their retrospective cohort analysis, noted an increase in the incidence of PC in a large Swedish CD cohort; the risk elevation persisted at a longer duration of follow-up, contrary to the earlier report by Elfström *et al*<sup>[4]</sup>. Nevertheless, data on PC risk in CD patients is limited, and these prior studies

were not designed to study PC specifically; thus, control for PC-specific risk factors was not adequately performed.

Our results are largely consistent with prior observations of elevated PC risk in CD<sup>[5,9]</sup>. We were able to confirm and strengthen this prior observed elevated risk of PC due to robust PSM and control for PC risk factors. This study has one major advantage over those that have preceded it: Our propensity score-matched analysis with covariables that are directly correlated with the risk of PC. The present study includes an expansion of clinical outcomes addressed in previous studies and an adjustment for confounders such as BMI, race, ethnicity, disease of gallbladder, diseases of the pancreas, hypercholesterolemia, hypercalcemia, familial hypercholesterolemia, family history of primary malignant neoplasm, ERCP, cholecystectomy, HbA1C, and genomics. Further experimental data and large multicenter cohort studies are required to ensure better comprehension of the association between CD and PC.

Most importantly, we conducted a sensitivity analysis where only PC diagnosed after 1 year of lag time following CD diagnosis was included as the outcome, and the study conclusion remained unchanged. When taken in totality, these data point towards a persistently elevated risk of PC in CD patients. One important observation to note is that the absolute number of events (PC cases) captured in our cohort was much higher than in previous studies due to the large population included in our analysis, as well as the long duration of follow-up. We believe that the high number of events coupled with available follow-up time increased the study's power and allowed us to deduce the persistence of risk elevation. In the reports by Lebwohl *et al*<sup>[5]</sup> and Landgren *et al*, an adjusted HR of 2.29 (1.84-2.84) and 2.27 (1.22-4.23) was described; the risk noted by Lebwohl *et al*<sup>[5]</sup> beyond the first year was 1.66 (1.32-2.10), and we noted HRs of 1.29 and 1.56 for PC in our cohort of CD patients. The overall risk elevation that we noted was thus slightly lower than these earlier reports.

Previously, some authors have hypothesized that an increased incidence of pancreatitis, impaired exocrine pancreatic function, as well as malnutrition, and resulting pancreatic atrophy may be the cause for the elevation in PC risk in patients

with CD<sup>[9]</sup>. The mechanism for the development of malignancies in patients with CD is still unknown. Autoimmunity has also been proposed as the etiopathologic culprit, as autoantibodies targeting the pancreas have been noted in CD patients<sup>[13]</sup>. In addition, chronic inflammation, chronic antigenic stimulation, the release of proinflammatory cytokines, increased intestinal permeability of environmental carcinogens, immune surveillance problems, and the gluten-free diet or nutritional deficiencies due to the disease have been suggested to be contributory<sup>[14,15]</sup>. Owing to the design of our study, we cannot deduce the pathophysiological basis of our observation of increased PC risk in CD.

### *Strengths and limitations*

The major strengths of this study include the robust control, and comprehensive organ-specific assessment focused on the pancreas. In addition, we adjusted for various lifestyle risk factors and baseline and potential confounders that may have confounded the association of CD and PC risk. The large sample in the propensity-matched analyses resulted in narrow confidence intervals. It allowed us to capture a significant number of outcomes, which lends strength to the conclusions that we have derived. Lastly, including a sensitivity analysis lends strength to our study and its findings.

However, our study is limited by its retrospective design and its analysis of EHR-based data, which can be subject to documentation and coding errors, resulting in some selection bias. Some residual confounding can exist, and we cannot exclude any possible confounding factors not controlled for in our study that may have an effect on the risk of PC. Due to the retrospective cohort nature of this study, any potential underlying pathophysiological mechanisms explaining the observed association of CD with PC cannot be explored. Furthermore, histological data on villous atrophy and compliance with a gluten-free diet was not studied. Whether villous atrophy and its persistence over time pose any elevation of risk for PC needs to be explored. Also, whether mitigation of the excess risk is observed with the normalization of villous

architecture or in patients who strictly adhere to a gluten-free diet should be explored in further studies.

## **CONCLUSION**

This nationwide study with large propensity-matched analysis found an increased risk of PC in patients with CD; this risk elevation persisted beyond the first year after diagnosis.

## **ARTICLE HIGHLIGHTS**

### ***Research background***

Studies have shown an increased risk for various malignancies, particularly lymphomas, small intestinal adenocarcinoma, and other gastrointestinal malignancies, in celiac disease (CD) patients. However, the magnitude of the risk of pancreatic cancer (PC) in association with CD is much less clear. Nevertheless, despite the magnitude of the risk remaining debatable, the unequivocal association between CD and PC remains.

### ***Research motivation***

Although malignancy occurring in the setting of CD has been well recognized; however, there is considerable, but not definitive, evidence that strict compliance to CD is associated with increased risk for the development of PC. Furthermore, owing to the lack of control for known risk factors of PC, any independent associations between these diseases cannot be assessed based on previous studies; given the high incidence of CD, and the poor outcomes associated with PC, any such potential association warrants further investigation.

### ***Research objectives***

This study aimed to assess the risk of PC in patients with CD.

**1**

### ***Research methods***



A population-based, multicenter, propensity score-matched cohort study included 155877 patients with CD and 234103 patients without CD (non-CD, controls). To reduce confounding effects, we performed a 1:1 propensity score matching with each patient in the main group to a patient in the control group. The incidence of PC was estimated using a Cox proportional hazards model with hazard ratio (HR) and 95% confidence interval (CI).

### **Research results**

During the follow-up, 309 patients with CD developed PC, whereas 240 patients developed PC in the control group (HR = 1.29; 95%CI: 1.09-1.53). In the secondary analyses in the first year after diagnosis of CD, patients with CD were at a significant increase in risk for PC; 151 patients with CD had an incidence of PC compared with 96 incidences of PC among the patients in the non-CD control group (HR = 1.56; 95%CI: 1.20-2.01) and sensitivity analysis showed similar magnitude to the one generated in the primary and secondary analysis.

### **Research conclusions**

This multicenter, propensity score-matched cohort study reveals that patients with CD are at increased risk of PC. Risk elevation persists beyond the first year after diagnosis to reference individuals without CD from the general population.

### **Research perspectives**

We still know little about the risk factors and mechanisms contributing to developing malignancies among individuals. Further experimental data and long-term follow-up studies are required to elucidate the pathogenic mechanisms and to ensure better comprehension of the association between PC and CD.

## ORIGINALITY REPORT

7%

SIMILARITY INDEX

## PRIMARY SOURCES

- |   |  |                 |
|---|--|-----------------|
| 1 | Arunkumar Krishnan, Ruhee A Patel, Yousaf Bashir Hadi, Diptasree Mukherjee et al. "Clinical characteristics and outcomes of COVID-19 in patients with autoimmune hepatitis: A population-based matched cohort study", World Journal of Hepatology, 2023<br><small>Crossref</small> | 118 words — 3%  |
| 2 | f6publishing.blob.core.windows.net<br><small>Internet</small>  | 17 words — < 1% |
| 3 | www.bmj.com<br><small>Internet</small>   | 15 words — < 1% |
| 4 | www.dovepress.com<br><small>Internet</small>   | 15 words — < 1% |
| 5 | arthritis-research.biomedcentral.com<br><small>Internet</small>  | 14 words — < 1% |
| 6 | hdl.handle.net<br><small>Internet</small>  | 14 words — < 1% |
| 7 | www.science.gov<br><small>Internet</small>   | 13 words — < 1% |
| 8 | Catassi, C.. "Association of celiac disease and intestinal lymphomas and other cancers", Gastroenterology, 200504  | 12 words — < 1% |

9

Lisandro Pereyra, Raquel Gonzalez, Adriana Mohaidle, Carolina Fischer et al. "Risk of colorectal neoplasia in patients with celiac disease: A multicenter study", Journal of Crohn's and Colitis, 2013

Crossref

12 words — &lt; 1%

10

academic.oup.com

Internet

12 words — &lt; 1%

11

journals.lww.com

Internet

12 words — &lt; 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES

EXCLUDE MATCHES

&lt; 12 WORDS

&lt; 12 WORDS