Risk of Pancreatic cancer in Individuals With Celiac Disease in the United States: A Population-Based Matched Cohort Study

SUPPLEMENTARY METHODS

Data source:

TriNetX (Cambridge, MA) Research Network uses electronic health record(EHR) data collected from member healthcare organizations (HCOs) in the United States[1]. HCO is a large academic health center with data from most of its affiliates. TriNetX's cloud-based feature allows real-time access to de-identified clinical data and analytical tools. Act-compliant, longitudinal clinical data to member HCOs is provided on a cloud-based platform. All clinical data is de-identified and aggregated directly from the EHR of participating HCOs. A typical organization has a complex enterprise architecture where the data flow through several different databases, such as a data warehouse and a research data repository, in addition to EHR data available in a structured fashion (e.g., demographics, diagnoses, and laboratory test results).

Standardizing the terminology (mapping codes) and data quality check:

TriNetX has production capabilities that have been tested that map data extensively from each of these structures to the standard model within TriNetX and can extract facts of interest from the narrative text of clinical documents using natural language processing. TriNetX maps the data to a standard and controlled set of clinical terminologies, enforces a list of required fields (e.g., patient identifier), and rejects those records where the required information is missing. As the data are refreshed, the software monitors changes in volumes of data over time to ensure data validity.

Clinical fact and coding system to present Data:

TriNetX has a team of informaticists who map data from the data provider's local codes to master terminology within TriNetX. Laboratory test results, vitals, and findings are coded to logical observation identifiers names and codes (LOINC). Demographics are health level 7 (HL7), version 3 (administrative standards), and Diagnoses are represented by the International Classification of Diseases, Ninth and 10th Revisions, Clinical Modification (ICD-9-CM and ICD-10-CM), and diagnoses data

are enriched with the chronic condition indicator. Depending on the coding system used by an HCO, procedure data are coded in current procedural terminology (CPT) or ICD-10 Procedure Coding System.

Definition of study variables:

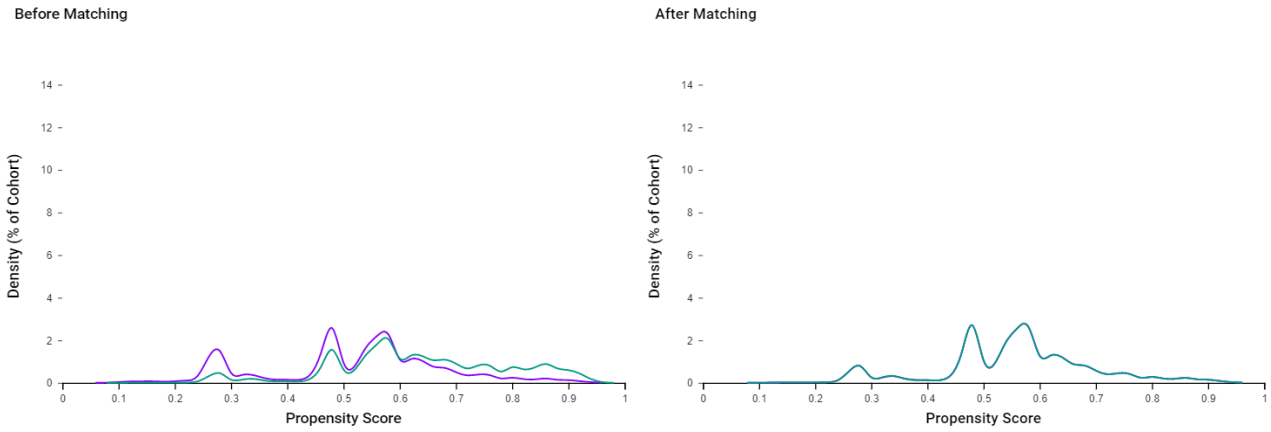
We defined Coeliac disease with the following inclusion criteria: ICD-10 codes of K90.0 (Celiac disease). Pancreatic cancer ICD-10: C25.X

Selection of patients with non-CD cohort:

We excluded the patients if they had a diagnosis of CD based on the above-mentioned codes that we used to select patients in the non-CD cohort.

Supplementary Figure 1:

Figure 1A. Propensity score density graph before and after matching among CD patients (purple) and Non-CD controls (blue).



Supplementary Table 1: Laboratory findings of the study cohort

Variables	Before propensity matching			After propensity matching		
	Celiac disease (N=155877)	**Non-CD controls** (N=234103)	**SMD**	**Celiac disease** (N=134680)	**Non-CD controls** (N=134680)	**SMD**
Gliadin antibodies, mean ± SD						
Serum Gliadin peptide IgA Ab [Units/volume] by Immunoassay	22.1 ± 33.6	7.75 ± 16.1	0.5429	21 ± 33	8.63 ± 19	0.4591
Serum Gliadin IgG Ab [Units/volume] by Immunoassay	15.6 ± 17.8	4.94 ± 6.09	0.7276	15.1 ± 17.4	5.47 ± 7.02	0.6336
Metabolic profile, mean ± SD						
HbA1C	6.47 ± 2.07	6.22 ± 1.78	0.1267	6.31 ± 1.95	6.24 ± 1.89	0.0322
Blood glucose (mg/dL)	104 ± 50.2	105 ± 38.9	0.0140	103 ± 47.9	103 ± 37.4	0.0058
Creatinine (mg/dL)	0.815 ± 1.22	1.01 ± 2.05	0.1143	0.833 ± 1.11	0.927 ± 1.68	0.0661
Total protein (g/L)	6.97 ± 1.05	7.08 ± 1.16	0.0936	6.97 ± 1.04	7.04 ± 1.21	0.0612
BUN (mmol/L)	13.6 ± 7.04	15.8 ± 10.5	0.2465	13.8 ± 7.1	15.2 ± 9.5	0.1690
Digestive enzymes, *mean ± SD*						
Serum lipase (*U/L*)	70.5 ±	70.8 ± 302	0.00	66.2 ±	69.2 ± 224	0.006

	714		05	629		4
Serum amylase (*U/L*)	66.3 ± 92.8	70.3 ± 89.2	0.0437	65.6 ± 68.7	66.5 ± 75.2	0.0124
Inflammatory markers, mean ± SD						
CRP(mg/L)	11 ± 29.1	16.9 ± 37.4	0.1752	11.1 ± 29.7	15.8 ± 35.6	0.1439
ESR(mm/h)	16.3 ± 18.4	23.7 ± 23.8	0.3478	17 ± 18.9	20.7 ± 21.7	0.1853
Lipid profile, *mean ± SD*						
CH(mg/dL)	177 ± 43.8	179 ± 44.5	0.0392	178 ± 43.8	179 ± 43.7	0.0265
TG(mg/dL)	121 ± 97.5	130 ± 95.8	0.0937	122 ± 97.7	129 ± 99	0.0753
LDL(mg/dL)	102 ± 35.2	102 ± 36.5	0.0007	102 ± 35.2	102 ± 35.7	0.0147
HDL (mg/dL)	51.8 ± 19.6	51.4 ± 19.6	0.0193	51.8 ± 19.7	51.9 ± 20	0.0059
Others, mean ± SD						
Hemoglobin (g/dL)	13.1 ± 1.84	12.8 ± 2.07	0.1664	13.1 ± 1.86	12.9 ± 1.98	0.0884
HCT (%)	39.3 ± 5.21	38.5 ± 5.99	0.1323	39.3 ± 5.26	38.8 ± 5.78	0.0910
Serum Calcium	9.27 ± 0.64	9.27 ± 0.609	0.0033	9.26 ± 0.637	9.29 ± 0.596	0.0408

Abbreviations: CD, celiac disease; SD, standard deviation; Ad, antibodies; HbA1C, Hemoglobin A1C; BUN, blood urea nitrogen; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; CH, cholesterol; TG, total cholesterol; LDL, low-density lipoprotein; HDL, high-density lipoprotein; HCT, hematocrit.

Reference:

1 . TriNetX: Longitudinal Real World Data. TriNetX; 2021. Available at: https://www.trinetx.com/trinetx-research.