Digestive and Liver Disease xxx (xxxx) xxx

Contents lists available at ScienceDirect

Digestive and Liver Disease

journal homepage: www.elsevier.com/locate/dld

Oncology

Analysis of exposome and genetic variability suggests stress as a major contributor for development of pancreatic ductal adenocarcinoma

Giulia Peduzzi[a], Alessio Felici[a], Roberto Pellungrini[b], Francesca Giorgolo[c], Riccardo Farinella[a], Manuel Gentiluomo[a], Andrea Spinelli[c], Gabriele Capurso[d], Anna Monreale[e], Federico Canzian[f], Marco Calderisi[c], Daniele Campa[a,*]

[a] Department of Biology, University of Pisa, Via Luca Ghini, 13, 56126 Pisa, Italy
[b] Classe di scienze, Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126 Pisa, Italy
[c] Kode s.r.l., Lungarno Galileo Galilei 1, 56125 Pisa, Italy
[d] Pancreato-Biliary Endoscopy and Endosonography Division, Pancreas Translational & Clinical Research Center, IRCCS San Raffaele Scientific Institute, Via Olgettina 60, 20132 Milan, Italy
[e] Department of Computer Science, University of Pisa, Largo Bruno Pontecorvo 3, 56127 Pisa, Italy
[f] Genomic Epidemiology Group, German Cancer Research Center, (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 3 July 2023
Accepted 18 October 2023
Available online xxx

Keywords:
Cancer
Gene-environment interaction
Polygenic risk score
SNPs
UK Biobank

ABSTRACT

Background: The current knowledge on pancreatic ductal adenocarcinoma (PDAC) risk factors is limited and no study has comprehensively tested the exposome in combination with the genetic variability in relation to the disease susceptibility.
Aim: The aim of this study was to analyze the exposome and its interaction with known genetic susceptibility loci, in relation to PDAC risk.
Methods: A case-control study nested in UK Biobank cohort was conducted on 816 PDAC cases and 302,645 controls. A total of 347 exposure variables, and a polygenic risk score (PRS) were analyzed through logistic regression. Gene-environment interaction analyses were conducted.
Results: A total of 52 associations under the Bonferroni corrected threshold of $p < 1.46 \times 10^{-4}$ were observed. Known risk factors such as smoking, pancreatitis, diabetes, PRS, heavy alcohol drinking and overweight were replicated in this study. As for novel associations, a clear indication for length and intensity of mobile phone use and the stress-related factors and stressful events with increase of PDAC risk was observed. Although the PRS was associated with PDAC risk ($P = 2.09 \times 10^{-9}$), statistically significant gene-exposome interactions were not identified.
Conclusion: In conclusion, our results suggest that a stressful lifestyle and sedentary behaviors may play a major role in PDAC susceptibility independently from the genetic background.

1. Background

Pancreatic ductal adenocarcinoma (PDAC) is the 8th most common form of cancer in Europe and has a survival rate that approaches 11% at 5 years after diagnosis [1,2]. PDAC is a complex disease that arises from the interplay of many independent variables. Several lifestyle and environmental exposure such as smoking and obesity have been identified to be risk factors, and conditions such as type 2 diabetes mellitus and chronic pancreatitis also play a substantial role in the development of PDAC [3,4]. Moreover, several common low penetrance genetic mutations, the majority of which single nucleotide polymorphisms (SNPs) increase the risk of developing the disease [5]. Most of the loci have been discovered through genome-wide association studies (GWAS) [6–11], while a small number have been identified though large multicentric candidate region or candidate gene approaches [12–20]. Additionally, other markers such as telomere length or mitochondrial copy number variation have also been suggested to play a role [21,22]. However, in comparison with more common cancers our knowledge is limited on both genetic and non-genetic factors. Gene-environment interactions studies have also been attempted but only considering small numbers of SNPs or environmental risk factors [23]. The identification of additional risk factors, either external or endogenous, will be instrumental in better understanding the disease and

* Corresponding author.
E-mail address: daniele.campa@unipi.it (D. Campa).

https://doi.org/10.1016/j.dld.2023.10.015

in the long term to establish preventive and screening programs to identify high risk individuals.

However, considering the complexity of the disease, to focus only on a small number of exposures will hardly be informative on the individual risk of developing PDAC. Instead, a more comprehensive approach must be considered.

The aim of this study was to comprehensively analyze the exposome, defined as the measures of all exposures to risk factors, such as dietary habits, smoking and alcohol consumption, physical activity, and the effect of environmental pollution in relation to PDAC risk. Furthermore, an additional aim was to identify possible gene-exposome interactions.

2. Materials and methods

2.1. UK Biobank (UKBB) and outcomes of interest

UKBB is a large population-based prospective cohort, where around 500,000 participants aged 37–73 have been recruited from 2006 to 2010 across the United Kingdom. Information on the participants lifestyle and health data at baseline or follow-up assessments were collected via touchscreen questionnaires, physical measurements, and through the analysis of biological samples. The study protocol has been described elsewhere in detail [24]. All participants provided informed written consent before data collection. The UKBB study was approved by the Northwest Multicenter Research Ethical Committee (MREC). Lifestyle, environmental and genetic data of UKBB participants were obtained from UKBB (project ID 66591).

Considering that they represent the vast majority, only subjects identified as having a white ethnic background were included: specifically, codes 1 (white), 1001 (British), 1002 (Irish) or 1003 (any other white background) of the field 21,000 were used, for a total of 472,622 subjects. PDAC cases were selected using data from the United Kingdom cancer registry (field 40,006) where diagnoses are coded according to the International Classification of Disease version 10 (ICD-10). The ICD codes were converted from version 10 to version 11 using a reference provided by the World Health Organization (WHO), accessible at https://icd.who.int/browse11/l-m/en. The code used to select subjects with exocrine pancreatic cancer is 2C10. In addition, the fields 'histology of cancer tumour (40,011) and' 'behaviour of cancer tumour' (40,012) were used in combination with 'type of cancer - ICD10' (40,006). These two fields (provided by the UK cancer registry) contain the numeric codes describing the histology and behaviour of the tumour. From 40,011, we selected only codes identifying PDAC (8140, 8211, 8260, 8440, 8452, 8472, 8480, 8481, 8490, 8500, 8503, 8550 and 8571); whereas from 40,012, we selected only codes identifying malignant and microinvasive tumours (3 and 5, respectively). In addition to field 40,006, UKBB uses also a field for self-reported tumors (field 20,001) and hospital inpatients diagnosis (field 41,270). These two fields were not used in this study to select PDAC cases, but all individuals that had no value in any of the three fields (40,006, 20,001, 41,270) where defined as controls. Following these criteria, a total of 816 PDAC cases and 302,645 controls were used in this study.

The selection of cases and controls was done using KNIME, a free and open-source data analysis tool.

2.2. Exposome variables

A total of 347 exposure variables, grouped in 28 categories were analyzed (a complete list of variables used, alongside the category they belong, and the measurement unit used is reported in Supplementary Table 1). Correlation matrixes were used to calculate intra-category correlation to identify the number of independent variables. Continuous and categorical variables were kept separate to calculate correlation using Pearson and Cramer's V tests, respectively. The threshold to declare the variables to be independent was $r = 0.90$. When variables were correlated among them only one was used in the regression analysis (keeping the variable that expressed a more general value, for example for time spend outdoors was kept instead of time spend outdoors in summer). Six variables (Apolipoprotein A, LDL-direct, severity of manic/irritable episode, direct bilirubin, greenspace percentage and time spent outdoors in summer) were thus dropped from the list. The correlation matrixes are shown in Supplementary File 1.

The implementation of correlation matrices was carried out in the python programming language using the following libraries: pandas, numpy, and matplotlib.

2.3. Polygenic risk score

A PRS was computed using all known PDAC susceptibility SNPs, identified through GWAS and/or with candidate gene/region approaches. All SNPs showed $p < 5 \times 10^{-8}$ for association with PDAC risk in the original publication where they were reported. The method has been described in detail by Galeotti and colleagues [25]. Briefly, the weighted score was generated multiplying the number of risk alleles for the beta reported in the literature by GWAS on PDAC susceptibility. Subsequently, from the sum of each product a weighted score for each individual was generated. The computed score was used as a categorical variable, calculating the quintiles based on the distribution in controls as reported by Galeotti and colleagues [25].

The PRS was calculated using the python programming language, specifically using the following libraries: pandas and numpy.

2.4. Statistical analysis

A total of 341 independent exposome variables were analyzed to test their association with PDAC risk. Unconditional logistic regression analysis, adjusted for age and sex was used, calculating Odds Ratios (OR) and the relative 95% confidence intervals (95%CI). For categorical variables the lowest category was set as reference (*e.g.,* nondrinkers for alcohol consumption). Considering multiple testing the threshold used to declare statistical significance was $p = 0.05/341 = 1.46 \times 10^{-4}$. Additionally, a multivariable model including all variables with <25% of missing values and a statistically significant association was also carried out.

All logistic regression analyses including the multivariable model were calculated using RStudio software.

2.5. GxE analysis

Gene – environment interaction analyses (GxE) were performed for the UKBB variables that showed an association with PDAC risk with p-values lower than the Bonferroni correction. Unconditional logistic regression models were carried out, adjusting for age and sex, but an interaction term was introduced in each model:

$$logit\,(PDAC) = \beta_0 + \beta_1 Age + \beta_2 Sex + \beta_3 E + \beta_4 PRS + \ \beta_5 E * PRS$$

In the formula above, PDAC is the binary outcome modelled as 0 (control) or 1 (case), β is the coefficient for every predictor included in the model, E is one of the UKBB exposures, and PRS is the continuous PRS. GxE analyses were performed with RStudio software.

3. Results

To test their association with PDAC risk, 341 exposome variables and a PRS were analyzed in 816 PDAC cases and 302,645 controls

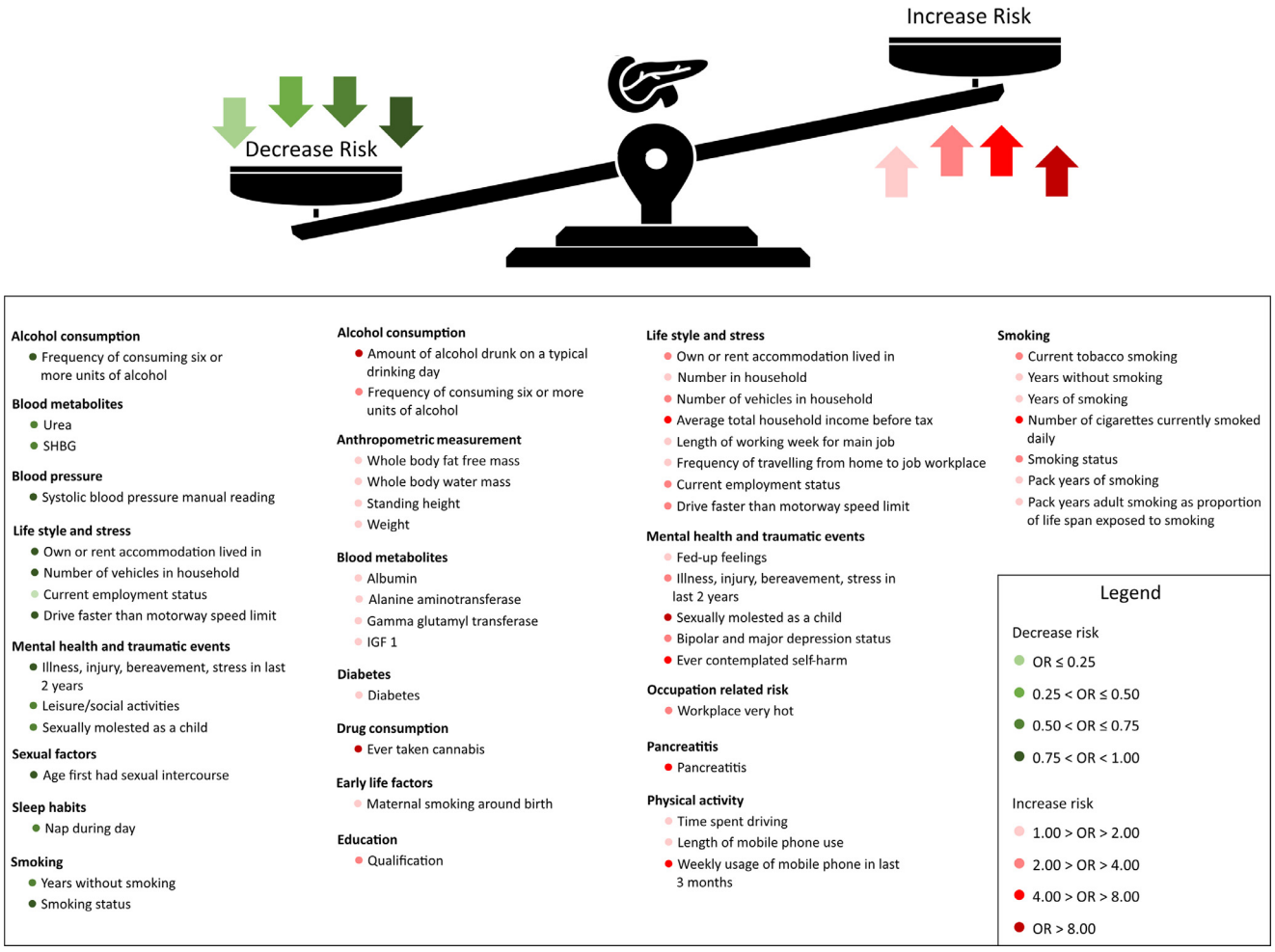

Fig. 1. General overview of pancreatic ductal adenocarcinoma risk categories. The figure shows the categories and the variables that have a significant association below the Bonferroni-corrected threshold with increased or decreased risk of developing PDAC. In red are the categories that increase the risk of developing pancreatic ductal adenocarcinoma and in green those that decrease it.

from the UKBB cohort. A total of 52 associations under the Bonferroni corrected threshold of $p < 1.46 \times 10^{-4}$ were observed (Fig. 1). Several known associations with PDAC risk, such as smoking, pancreatitis, type 2 diabetes mellitus, the PRS, heavy alcohol drinking, overweight and high percentage of fat mass in body composition, were confirmed in this study Specifically, smoking showed a strong association with PDAC risk. In particular, a variable labeled as "pack years adult smoking as proportion of life span exposed to smoking" as proposed by Tobin was used in the analysis. This variable considers for each individual the pack years smoked, only when they were smokers and better captures smoking intensity. This variable showed a strong increase in risk for individuals that smoked many cigarettes in a short period of time. The analysis indicates an almost two-fold increase of risk for each unit increase of the variable, expressed as pack years / age at recruitment – 16 (OR = 1.83, 95%CI: 1.55–2.16, $p = 8.12 \times 10^{-13}$). Maternal smoking around birth was also associated with increased risk (OR = 1.47, 95%CI: 1.24–1.73, $p = 4.95 \times 10^{-6}$). Pancreatitis and type 2 diabetes mellitus were also associated with increased risk of developing PDAC (OR = 5.85, 95%CI: 3.93–8.42, $p = 7.97 \times 10^{-20}$ and OR = 1.47, 95%CI: 1.20–1.79, $p = 1.44 \times 10^{-4}$ respectively). The PRS analysis showed also that individual in the highest quintile (*i.e.*, individuals with the highest number of risk alleles) had over two-fold increased risk of developing PDAC compared to individuals in the lowest quintile (OR = 2.25, 95%CI: 1.73–2.95 95%CI, $p = 3.70 \times 10^{-9}$).

Heavy drinkers (more than ten drinks per day on average), compared with nondrinkers showed a very strong effect (OR = 9.73, 95% CI: 3.02–26.74, $p = 4.40 \times 10^{-5}$). However, this association arise from an analysis conducted on a small number of individuals. Weight was also strongly associated with a 2% increase in risk for each kg, OR 1.02 95%CI: 1.01–1.02 $p = 3.38 \times 10^{-10}$.

Focusing on analysis for which information was available for at least half of the PDAC patients, we observed several novel associations. For example, the use of mobile phones both as duration (length of mobile phone use) and intensity (weekly usage of mobile phone in last 3 months) were strongly associated with increased PDAC risk (OR = 1.88 95%CI: 1.53–2.31, $p = 1.81 \times 10^{-9}$ and OR = 4.70 95%CI: 2.71–7.64, $p = 3.93 \times 10^{-9}$, respectively for the analyses comparing the highest with the lowest categories). Several blood metabolites showed statistically significant association with PDAC risk, such as IGF-1 levels (OR = 1.47, 95%CI: 1.22–1.76, $p = 3.80 \times 10^{-5}$) and SHBG levels (OR = 0.59 95%CI: 0.48–0.71, $p = 9.35 \times 10^{-8}$).

Additionally, eight exposures related to lifestyle, work duration and income (own or rent accommodation lived in, number of people living in the household, number of vehicles in the household, average total household income before tax, length of working week for main job, frequency of travelling from home to job workplace, current employment status and driving faster than motorway speed limit) showed a statistically significant association with PDAC. The results of the statistically significant associations with

Fig. 2. Manhattan plot of all variables analyzed. The x-axis shows the 28 categories analysed, while the y-axis shows the logarithm in base ten of the p-value. The solid red line represents a *p*-value threshold of 0.05, while the dashed red line represents the p-value threshold using the Bonferroni correction (p-value $= 1.46 \times 10^{-4}$). For each category, the name of the variable with the highest significance that exceeded the Bonferroni correction was shown in the figure.

Bonferroni correction are reported in Table 1. The results of all the analysis are reported in Fig. 2 and Supplementary Table 2.

Most of the associations observed in the univariable analysis were substantially confirmed by the results of the multivariable analysis. The variables that showed a statistically significant association are: the PRS, urea, gamma glutamyltransferase and SHBG concentration, systolic blood pressure, hand grip strength, accommodation lived in, number of vehicles in household, average total household income before tax, current employment status, illness/injury/bereavement/stress in last 2 years, pancreatitis, type 2 diabetes mellitus, age of the first sexual intercourse, current tobacco smoking, pack years adult smoking as proportion of life span exposed to smoking, pack years of smoking and time spent outdoors. It should be noted, however, that the multivariable analysis was only carried out in a small subset of individuals (248 cases and 93,615 controls) and for those variables with a number of missing values lower than 25% (number of variables=39). The results of the multivariable analysis are reported in Table 1.

The gene-environment interaction showed three interesting associations. Specifically, an interaction was observed between the PRS and two stress-related variables, current employment status ($OR_{interaction} = 0.74$, 95%CI: 0.56–0.99, $p = 4.16 \times 10^{-2}$) and fed-up feeling ($OR_{interaction} = 1.42$, 95%CI: 1.08–1.88, $p = 1.31 \times 10^{-2}$). In addition to these two interactions, one was also observed between the PRS and time spent outdoor ($OR_{interaction} = 1.13$, 95%CI: 1.04–1.22, $p = 2.31 \times 10^{-3}$). However, none remained significant after Bonferroni correction ($p = 9.09 \times 10^{-4}$).

4. Discussion

This study represents the first attempt to comprehensively analyze the exposome in combination with genetic variability expressed as a PRS in relation to PDAC risk. We observed 52 associations (51 exposome variables + the PRS) that remained significant after correction for multiple testing. Alongside known associations such as smoking, heavy alcohol intake, body weight, type 2 diabetes mellitus and chronic pancreatitis we identified 29 new ones [26]. Of particular interest is the association between long mobile phone use and increased risk of PDAC. Individuals who used a mobile phone for more than eight years have an almost two-fold increase of risk compared to individuals that never used it ($p = 1.81 \times 10^{-9}$). Additionally, individuals who use the mobile phone for more than 6 h weekly have an almost five-fold risk increase compared to individuals who do not use it ($p = 3.93 \times 10^{-9}$). It is unlikely that the mobile phone use per se would be the cause of the association, but it could represent instead a very good proxy for a low level of physical activity.

In the past, the majority of the time spent in sedentary activities was represented by watching television [26–28], but it is possible that now this has been replaced by other activities such as using smartphones or tablets. This association has never been reported in the literature, probably because the majority of cohorts studies have not included this variable in the questionnaires administered to the subjects.

Other association that are particularly striking are those related to income and the level of education obtained from the participants. Higher income, living in own house without paying mortgage, having a college or university degree, and longer working weeks are all associated with increased risk of developing PDAC. All these associations are highly statistically significant, with p-values ranging from 1.85×10^{-8} to 2.00×10^{-16} and have a strong effect on the disease risk with ORs ranging from 1.77 to 4.18 for the categorical variables and an increase of 2% risk associated to each working hour increment. It is highly unlikely that these variables are directly associated to the risk of developing PDAC, instead a possible explanation is that they are all proxies for stress, and therefore that stress plays a major role in pancreatic carcinogenesis. To support this hypothesis, we also observed that stress-

Table 1
Results of statistically significant associations with *P*-value under the Bonferroni correction.

Variables	Units of measure	N° subjects (case/control)	% missing value	Univariate analysis		Multivariable analysis	
				OR (95%CI)[A]	*P*-value[B]	OR (95%CI)[A]	*P*-value[B]
Amount of alcohol drunk on a typical drinking day	Unit of alcohol	100,627 (126/100,501)	33.15%	–	–	–	–
0–0		7905 (12/7893)		Ref	–	–	–
1–1 or 2		46,487 (63/46,424)		0.85 (0.47–1.82)	7.20E−01	–	–
2–3 or 4		25,652 (24/25,628)		1.14 (0.54–2.53)	7.41E−01	–	–
3–5 or 6		11,536 (15/11,521)		1.84 (0.78–4.44)	1.66E−01	–	–
4–7, 8 or 9		6287 (6/6281)		2.53 (0.83–7.05)	8.43E−02	–	–
5–10 or more		2760 (6/2754)		9.37 (3.02–26.74)	4.40E−05	–	–
Frequency of consuming six or more units of alcohol	–	92,885 (114/92,771)	69.39%	–	–	–	–
1 - Never		44,954 (60/44,894)		Ref	–	–	–
2 - Less than monthly		23,449 (31/23,418)		2.25 (1.36–3.67)	1.26E−03	–	–
3 - Monthly		8544 (3/8541)		0.93 (0.22–2.60)	9.04E−01	–	–
4 - Weekly		12,718 (16/12,702)		3.84 (1.98–7.09)	3.22E−05	–	–
5 - Daily or almost daily		3220 (4/3216)		2.16 (0.56–6.13)	1.97E−01	–	–
Whole body fat free mass	kg	297,995 (796/297,199)	1.80%	1.06 (1.04–1.07)	2.39E−22	1.00 (0.64–1.55)	9.95E−01
Whole body water mass	kg	298,020 (796/297,224)	1.79%	1.08 (1.06–1.09)	6.85E−22	1.05 (0.58–1.90)	8.80E−01
Standing height	cm	302,426 (813/301,613)	0.34%	1.04 (1.03–1.06)	1.21E−14	1.02 (0.99–1.05)	2.70E−01
Weight	kg	302,225 (813/301,412)	0.41%	1.02 (1.01–1.02)	3.38E−10	0.99 (0.97–1.02)	5.64E−01
Albumin	g/L	260,818 (710/260,110)	14.05%	–	–	–	–
Tertile 1 (20.67–44.23)		86,421 (247/86,174)		Ref	–	Ref	–
Tertile 2 (44.24–46.37)		86,002 (269/85,733)		1.44 (1.20–1.72)	5.80E−05	1.32 (0.95–1.83)	9.88E−02
Tertile 3 (46.38–59.8)		88,395 (194/88,203)		1.38 (1.14–1.68)	9.25E−04	1.34 (0.93–1.91)	1.14E−01
Alanine aminotransferase	U/L	284,681 (759/283,924)	6.19%	–	–	–	–
Tertile 1 (3.01–16.89)		93,952 (208/93,744)		Ref	–	Ref	–
Tertile 2 (16.9–24.29)		93,935 (258/93,677)		1.07 (0.89–1.30)	4.59E−01	0.8 (0.55–1.16)	2.44E−01
Tertile 3 (24.3–491.03)		96,794 (293/96,503)		1.54 (1.28–1.86)	5.26E−06	1.06 (0.71–1.56)	7.84E−01
Urea	mmol/L	284,595 (760/283,837)	6.22%	–	–	–	–
Tertile 1 (0.81–4.75)		94,641 (205/94,436)		Ref	–	Ref	–
Tertile 2 (4.76–5.76)		93,355 (246/93,109)		0.79 (0.65–0.95)	1.45E−02	0.83 (0.58–1.20)	3.30E−01
Tertile 3 (5.77–38.2)		96,599 (309/96,292)		0.64 (0.53–0.76)	1.11E−06	0.67 (0.47–0.96)	2.99E−02
Gamma glutamyl transferase	U/L	284,641 (758/283,885)	6.20%	–	–	–	–
Tertile 1 (0.995–20.4)		94,280 (165/94,115)		Ref	–	Ref	–
Tertile 2 (20.5–33.5)		93,595 (255/93,340)		1.14 (0.93–1.40)	1.98E−01	1.3 (0.86–1.97)	2.09E−01
Tertile 3 (33.6–1162.1)		96,766 (338/96,430)		1.52 (1.25–1.85)	3.86E−05	2.16 (1.41–3.31)	3.87E−04
IGF-1	nmol/L	283,309 (757/282,554)	6.64%	–	–	–	–
Tertile 1 (1.909–19.03)		93,547 (296/93,251)		Ref	–	Ref	–
Tertile 2 (19.031–23.578)		93,502 (249/93,253)		1.22 (1.02–1.45)	2.56E−02	0.95 (0.68–1.32)	7.45E−01
Tertile 3 (23.579–124.818)		96,260 (212/96,050)		1.47 (1.22–1.76)	3.80E−05	1.31 (0.92–1.86)	1.31E−01
SHBG	nmol/L	258,279 (705/257,576)	14,89%	–	–	–	–
Tertile 1 (0.39–36.28)		85,247 (236/85,011)		Ref	–	Ref	–
Tertile 2 (36.29–55.9)		85,257 (251/85,006)		0.66 (0.55–0.79)	9.34E−06	0.68 (0.48–0.95)	2.54E−02
Tertile 3 (55.91–241.58)		87,775 (218/87,559)		0.59 (0.48–0.71)	9.35E−08	0.59 (0.39–0.87)	8.63E−03
Systolic blood pressure (manual reading)	mmHg	281,959 (769/281,190)	7.09%	0.99 (0.98–0.99)	2.67E−10	0.99 (0.98–0.99)	5.88E−04
Diabetes	–	303,461 (816/302,645)	0%	–	–	–	–
0 - No		285,863 (693/285,170)		Ref	–	Ref	–
1 - Yes		17,598 (123/17,475)		1.47 (1.20–1.79)	1.44E−04	0.28 (0.14–0.55)	2.75E−04
Ever taken cannabis	–	100,772 (126/100,646)	66.79%	–	–	–	–
0 - No		77,400 (97/77,303)		Ref	–	–	–
1 - Yes, 1–2 times		9826 (13/9813)		2.75 (1.31–5.24)	3.86E−03	–	–
2 - Yes, 3–10 times		5941 (5/5936)		3.38 (1.10–8.15)	1.52E−02	–	–
3 - Yes, 11–100 times		4729 (8/4721)		13.91 (5.48–30.42)	1.05E−09	–	–
4 - Yes, more than 100 times		2876 (3/2873)		11.88 (2.68–35.92)	1.16E−04	–	–
Maternal smoking around birth	–	261,788 (685/261,103)	13.73%	–	–	–	–
0 - No		182,369 (467/181,902)		Ref	–	Ref	–
1 - Yes		79,419 (218/79,201)		1.47 (1.24–1.73)	4.95E−06	1.25 (0.93–1.68)	1.44E−01
Qualifications	–	298,772 (801/297,971)	1.55%	–	–	–	–
7 (ex −7) - None of the above		46,927 (182/46,745)		Ref	–	Ref	–
1 - College or university degree		100,439 (234/100,205)		1.78 (1.46–2.17)	1.45E−08	1.58 (0.99–2.53)	5.60E−02
2 - Advanced (A)levels/ Advanced Subsidiary (AS) levels or equivalent		35,013 (65/34,948)		1.43 (1.07–1.90)	1.50E−02	0.74 (0.38–1.46)	3.88E−01
3- Ordinary (O) levels/ General Certificate of Secondary Education (GCSEs) or equivalent		64,826 (172/64,654)		1.48 (1.19–1.83)	3.34E−04	0.97 (0.59–1.60)	9.17E−01
4 - Certificate of Secondary Education (CSEs) or equivalent		17,548 (25/17,523)		2.55 (1.62–3.84)	2.03E−05	1.57 (0.6–4.14)	3.61E−01
5 - National Vocational Qualifications (NVQ) or Higher National Diplomas (HND) or Higher National Certificates (HNC) or equivalent		19,284 (71/19,213)		1.77 (1.32–2.34)	8.36E−05	1.57 (0.87–2.82)	1.34E−01

(*continued on next page*)

Table 1 (*continued*)

Variables	Units of measure	N° subjects (case/control)	% missing value	Univariate analysis		Multivariable analysis	
				OR (95%CI)[A]	*P*-value[B]	OR (95%CI)[A]	*P*-value[B]
6 - Other professional qualifications		14,735 (52/14,683)		1.20 (0.87–1.63)	2.52E−01	1.34 (0.73–2.46)	3.53E−01
Age high blood pressure diagnosed	Years	67,251 (272/66,979)	77.84%	–	–	–	–
Tertile 3 (57–70)		19,127 (107/19,020)		Ref	–	–	–
Tertile 2 (48–56)		25,241 (102/25,139)		2.40 (1.81–3.19)	1.40E−09	–	–
Tertile 1 (18–47)		22,883 (63/22,820)		2.81 (2.02–3.88)	4.51E−10	–	–
Hand grip strength (right)	kg	301,866 (810/301,056)	0.53%	1.04 (1.03–1.05)	8.34E−14	1.03 (1.01–1.05)	1.26E−03
Own or rent accommodation lived in		300,587 (808/300,234)	0.95%	–	–	–	–
1 - Own outright (by you or someone in your household)		151,050 (553/150,952)		Ref	–	Ref	–
2 - Own with a mortgage		120,391 (175/120,216)		2.91 (2.43–3.48)	2.00E−16	2.34 (1.67–3.28)	8.85E−07
3 - Rent - from local authority, local council, housing association		15,759 (60/15,699)		1.72 (1.30–2.25)	9.79E−05	1.79 (0.8–4.04)	1.59E−01
4 - Rent - from private landlord or letting agency		8920 (12/8908)		1.04 (0.54–1.78)	9.07E−01	0.66 (0.15–2.85)	5.75E−01
5 - Pay part rent and part mortgage (shared ownership)		859 (1/858)		0.78 (0.04–3.52)	8.03E−01	NA	NA
6 - Live in accommodation rent free		3608 (7/3601)		1.29 (0.55–2.54)	5.09E−01	0.62 (0.08–4.85)	6.48E−01
Number in household	Peoples	301,882 (810/301,072)	0.52%	1.07 (1.03–1.10)	3.94E−05	1.08 (1.00–1.17)	5.71E−02
Number of vehicles in household	–	301,707 (811/300,896)	0.58%	–	–	–	–
1 - None		24,224 (80/24,144)		Ref	–	Ref	–
2 - One		122,651 (353/122,298)		0.79 (0.62–1.02)	6.64E−02	1.72 (0.73–4.06)	2.12E−01
3 - Two		118,624 (297/118,327)		1.16 (0.90–1.50)	2.68E−01	2.38 (0.98–5.77)	5.60E−02
4 - Three		27,411 (62/27,349)		2.03 (1.43–2.86)	6.45E−05	3.08 (1.14–8.30)	2.64E−02
5 - Four or more		8797 (19/8778)		2.36 (1.37–3.87)	1.12E−03	1.90 (0.51–7.01)	3.38E−01
Average total household income before tax	£	261,483 (683/260,800)	13.83%	–	–	–	–
1 - Less than 18,000		53,118 (185/52,933)		Ref	–	Ref	–
2 - 18,000 to 30,999		63,795 (206/63,589)		1.34 (1.10–1.65)	4.48E−03	1.63 (1.06–2.51)	2.72E−02
3 - 31,000 to 51,999		70,482 (158/70,324)		2.02 (1.62–2.51)	3.89E−10	1.28 (0.77–2.13)	3.37E−01
4 - 52,000 to 100,000		58,247 (104/58,143)		3.39 (2.63–4.36)	2.00E−16	2.16 (1.24–3.78)	6.58E−03
5 - Greater than 100,000		15,841 (30/15,811)		4.18 (2.74–6.16)	3.49E−12	1.98 (0.92–4.25)	8.05E−02
Length of working week for main job	Hours	184,471 (336/184,135)	39.21%	1.03 (1.02–1.03)	2.85E−11	–	–
Frequency of travelling from home to job workplace	Times	183,706 (328/183,378)	39.46%	1.03 (1.01–1.05)	9.45E−06	–	–
Current employment status	–	302,554 (815/301,739)	0.30%	–	–	–	–
1 - In paid employment or self-employed		187,005 (336/186,669)		Ref	–	Ref	–
2 - Retired		89,842 (413/89,429)		0.18 (0.15–0.21)	2.00E−16	0.22 (0.16–0.31)	2.00E−16
3 - Looking after home and/or family		8493 (13/8480)		0.93 (0.50–1.60)	8.15E−01	1.05 (0.33–3.38)	9.34E−01
4 - Unable to work because of sickness or disability		8916 (35/8881)		3.63 (2.47–5.19)	8.30E−12	5.22 (2.3–11.87)	8.05E−05
5 - Unemployed		4714 (13/4701)		3.83 (2.04–6.58)	6.04E−06	5.61 (1.91–16.46)	1.71E−03
6 - Doing unpaid or voluntary work		1395 (3/1392)		0.30 (0.07–0.82)	4.62E−02	NA	NA
7 - Full or part-time student		709 (0/709)		N/A	N/A	NA	NA
8 - None of the above		1480 (2/1478)		0.23 (0.04–0.72)	3.91E−02	0.46 (0.06–3.76)	4.67E−01
Drive faster than motorway speed limit	–	295,191 (788/294,546)	2.68%	–	–	–	–
1 - Never/rarely		120,864 (342/120,522)		Ref	–	Ref	–
2 - Sometimes		110,572 (302/110,270)		1.34 (1.14–1.58)	4.02E−04	0.87 (0.63–1.20)	4.03E−01
3 - Often		30,204 (66/30,138)		1.87 (1.41–2.45)	9.59E−06	1.12 (0.69–1.84)	6.45E−01
4 - Most of the time		13,596 (21/13,575)		2.16 (1.33–3.33)	9.85E−04	0.75 (0.33–1.70)	4.84E−01
5 - Do not drive on the motorway		20,098 (57/20,041)		0.98 (0.73–1.30)	9.05E−01	1.18 (0.64–2.19)	5.95E−01
Fed-up feelings	–	297,191 (801/296,390)	2.07%	–	–	–	–
0 - No		176,193 (490/175,703)		Ref	–	Ref	–
1 - Yes		120,998 (311/120,687)		1.34 (1.16–1.55)	8.25E−05	1.05 (0.78–1.42)	7.50E−01
Illness, injury, bereavement, stress in last 2 years	–	299,668 (798/298,870)	1.25%	–	–	–	–
0 - None of the above		170,082 (442/169,640)		Ref	–	Ref	–
1 - Serious illness, injury or assault to yourself		22,057 (105/21,952)		2.09 (1.67–2.60)	4.94E−11	1.45 (0.90–2.33)	1.25E−01
2 - Serious illness, injury or assault of a close relative		31,411 (80/31,331)		1.52 (1.18–1.94)	7.62E−04	1.17 (0.73–1.86)	5.20E−01
3 - Death of a close relative		47,686 (118/47,568)		1.17 (0.95–1.43)	1.42E−01	0.99 (0.65–1.49)	9.45E−01
4 - Death of a spouse or partner		2672 (9/2663)		0.83 (0.40–1.53)	5.94E−01	0.89 (0.21–3.72)	8.70E−01
5 - Marital separation/divorce		6222 (15/6207)		2.90 (1.62–4.80)	1.08E−04	2.86 (1.05–7.79)	3.98E−02
6 - Financial difficulties		19,538 (29/19,509)		1.74 (1.16–2.49)	4.67E−03	0.51 (0.21–1.26)	1.46E−01
Leisure/social activities	–	302,652 (814/301,838)	0.25%	–	–	–	–
0 - None of the above		91,817 (257/91,560)		Ref	–	Ref	–
1 - Sports club or gym		93,159 (226/92,933)		0.91 (0.76–1.09)	3.14E−01	1.04 (0.73–1.49)	8.15E−01
2 - Pub or social club		55,849 (167/55,682)		0.89 (0.73–1.09)	2.59E−01	0.82 (0.53–1.26)	3.67E−01
3 - Religious group		26,392 (68/26,324)		0.54 (0.41–0.71)	1.15E−05	0.71 (0.42–1.19)	1.96E−01
4 - Adult education class		8500 (27/8473)		0.75 (0.49–1.11)	1.76E−01	0.60 (0.26–1.42)	2.48E−01
5 - Other group activity		26,935 (69/26,866)		0.60 (0.46–0.79)	2.83E−04	0.73 (0.43–1.24)	2.46E−01

(*continued on next page*)

Table 1 (*continued*)

Variables	Units of measure	N° subjects (case/control)	% missing value	Univariate analysis		Multivariable analysis	
				OR (95%CI)[A]	*P*-value[B]	OR (95%CI)[A]	*P*-value[B]
Bipolar and major depression status	–	73,120 (182/72,935)	75.90%	–	–	–	–
0 - No Bipolar or Depression		53,118 (134/52,984)		Ref	–	–	–
1 - Bipolar I Disorder		492 (1/491)		1.65 (0.09–7.59)	6.21E−01	–	–
2 - Bipolar II Disorder		441 (2/439)		2.94 (0.48–9.68)	1.39E−01	–	–
3 - Probable Recurrent major depression (severe)		5166 (21/5145)		2.55 (1.55–3.98)	9.92E−05	–	–
4 - Probable Recurrent major depression (moderate)		9013 (17/8996)		1.28 (0.74–2.07)	3.51E−01	–	–
5 - Single Probable major depression episode		4890 (10/4880)		1.04 (0.50–1.91)	9.15E−01	–	–
Ever contemplated self-harm	–	100,529 (126/100,403)	66.87%	–	–	–	–
0 - No		85,317 (107/85,210)		Ref	–	–	–
1 - Yes, once		7615 (9/7606)		1.67 (0.71–3.4)	1.93E−01	–	–
2 - Yes, more than once		7597 (10/7587)		4.13 (1.90–7.99)	9.30E−05	–	–
Sexually molested as a child	–	99,829 (126/99,703)	67.10%	–	–	–	–
0 - Never true		91,499 (115/91,384)		Ref	–	–	–
1 - Rarely true		4446 (5/4441)		0.75 (0.23–1.83)	5.80E−01	–	–
2 - Sometimes true		3016 (3/3013)		0.54 (0.09–1.97)	4.35E−01	–	–
3 - Often		476 (1/475)		3.20 (0.18–15.71)	2.60E−01	–	–
4 - Very often true		392 (2/390)		19.16 (2.98–69.14)	1.07E−04	–	–
Workplace very hot	–	76,348 (105/76,243)	74.84%	–	–	–	–
0 - Rarely/never		43,810 (58/43,752)		Ref	–	–	–
1 - Sometimes		29,116 (35/29,081)		1.20 (0.76–1.86)	4.30E−01	–	–
2 - Often		3422 (12/3410)		3.67 (1.8–6.91)	1.36E−04	–	–
Pancreatitis	–	303,461 (816/302,645)	0.00%	–	–	–	–
0 - No		302,121 (782/301,339)		Ref	–	Ref	–
1 - Yes		1340 (34/1306)		5.85 (3.93–8.42)	7.97E−20	7.6 (3.13–18.45)	7.31E−06
Time spent driving	Hours/day	299,232 (801/298,431)	1.39%	1.13 (1.06–1.19)	7.87E−05	1.05 (0.93–1.19)	4.54E−01
Length of mobile phone use	Years	299,908 (798/299,110)	1.17%	–	–		
0 - Never used mobile phone at least once per week		43,116 (156/42,960)		Ref	–	Ref	–
1 - One year or less		7864 (24/7840)		1.13 (0.71–1.71)	5.93E−01	1.14 (0.28–4.61)	8.59E−01
2 - Two to four years		52,419 (144/52,275)		1.34 (1.06–1.69)	1.38E−02	0.68 (0.19–2.37)	5.40E−01
3 - Five to eight years		93,087 (220/92,867)		1.48 (1.20–1.83)	2.54E−04	0.60 (0.17–2.06)	4.17E−01
4 - More than eight years		103,422 (254/103,168)		1.88 (1.53–2.31)	1.81E−09	0.58 (0.17–1.95)	3.81E−01
Weekly usage of mobile phone in last 3 months	Minutes	297,648 (796/296,852)	1.92%	–	–	–	–
−1 - Never		43,116 (156/42,960)		Ref	–	Ref	–
0 - Less than 5 min		52,253 (166/52,087)		1.19 (0.95–1.49)	1.22E−01	1.26 (0.37–4.22)	7.13E−01
1 - 5–29 min		99,041 (289/98,752)		1.60 (1.32–1.96)	3.52E−06	1.60 (0.5–5.19)	4.31E−01
2 - 30–59 min		44,294 (94/44,200)		1.76 (1.35–2.29)	2.42E−05	1.15 (0.34–3.85)	8.26E−01
3 - 1–3 h		36,754 (54/36,700)		1.87 (1.35–2.55)	1.20E−04	0.9 (0.26–3.14)	8.74E−01
4 - 4–6 h		10,922 (20/10,902)		3.24 (1.93–5.14)	2.32E−06	1.33 (0.31–5.78)	7.02E−01
5 - More than 6 h		11,268 (17/11,251)		4.70 (2.71–7.64)	3.93E−09	NA	NA
Polygenic risk score (PRS) weighted	–	237,537 (627/236,910)	21.72%	–	–	2.09 (1.66–2.64)	6.61E−10
PRS weighted - Q1		47,463 (81/47,382)		Ref	–	–	–
PRS weighted - Q2		47,478 (96/47,382)		1.17 (0.87–1.58)	3.03E−01	–	–
PRS weighted - Q3		47,493 (111/47,382)		1.37 (1.02–1.83)	3.59E−02	–	–
PRS weighted - Q4		47,535 (153/47,382)		1.90 (1.45–2.51)	4.61E−06	–	–
PRS weighted - Q5		47,568 (186/47,382)		2.25 (1.73–2.95)	2.09E−09	–	–
Age first had sexual intercourse	Years	268,137 (714/267,423)	11.64%	0.94 (0.92–0.96)	3.70E−09	0.93 (0.89–0.98)	2.20E−03
Nap during day	–	303,285 (816/302,469)	0.06%	–	–	–	–
1 - Never/rarely		176,533 (426/176,107)		Ref	–	Ref	–
2 - Sometimes		112,077 (330/111,747)		0.75 (0.64–0.86)	1.03E−04	0.91 (0.68–1.22)	5.29E−01
3 - Usually		14,675 (60/14,615)		0.70 (0.52–0.92)	1.22E−02	0.94 (0.54–1.65)	8.31E−01
Current tobacco smoking	–	303,311 (815/302,496)	0.05%	–	–	–	–
0 - No		272,352 (690/271,662)		Ref	–	Ref	–
1 - Yes, on most or all days		22,620 (99/22,521)		2.70 (2.16–3.34)	7.13E−19	NA	NA
2 - Only occasionally		8339 (26/8313)		2.03 (1.32–2.96)	5.60E−04	2.80 (1.18–6.69)	2.02E−02
Years without smoking	Years	239,421 (625/238,796)	21.10%	–	–	–	–
1 - Never smoker		169,197 (383/168,814)			–	Ref	–
2 - less than 10 years		20,611 (83/20,528)		1.88 (1.46–2.39)	5.96E−07	1.59 (0.73–3.47)	2.39E−01
3 - between 10 and 25 years		25,516 (75/25,441)		1.06 (0.81–1.36)	6.81E−01	1.24 (0.54–2.82)	6.16E−01
4 - more than 25 years		24,097 (84/24,013)		0.74 (0.58–0.94)	1.51E−02	0.40 (0.13–1.19)	9.87E−02
Years of smoking	Years	261,461 (723/260,738)	13.84%	–	–	–	–
1 - Never smoker		169,197 (383/168,814)		Ref	–	Ref	–
2 - less than 20 years		33,690 (84/33,606)		1.04 (0.81–1.32)	7.60E−01	2.22 (0.82–5.98)	1.16E−01
3 - between 20 and 30 years		23,333 (75/23,258)		1.33 (1.03–1.71)	2.60E−02	1.12 (0.51–2.43)	7.79E−01
4 - more than 30 years		35,241 (181/35,060)		1.45 (1.20–1.74)	8.42E−05	NA	NA
Number of cigarettes currently smoked daily (current cigarette smokers)	Cigarettes/day	20,894 (88/20,806)	93.11%	–	–	–	–

(*continued on next page*)

Table 1 (*continued*)

Variables	Units of measure	N° subjects (case/control)	% missing value	Univariate analysis		Multivariable analysis	
				OR (95%CI)[A]	*P*-value[B]	OR (95%CI)[A]	*P*-value[B]
Tertile 1 (0.5–10)		7438 (15/7423)		Ref	–	–	–
Tertile 2 (11–20)		10,433(50/10,383)		2.56 (1.46–4.76)	1.62E−03	–	–
Tertile 3 (21–120)		3023 (23/3000)		4.77 (2.46–9.53)	5.19E−06	–	–
Smoking status	–	302,497 (812/301,685)	0.32%	–	–	–	–
0 - Never		169,197 (383/168,814)		Ref	–	Ref	–
1 - Previous		102,341 (304/102,037)		0.94 (0.81–1.1)	4.50E−01	NA	NA
2 - Current		30,959 (125/30,834)		2.45 (1.98–3.01)	7.31E−17	NA	NA
Pack years of smoking	Pack years	25,7261 (697/256,564)	15.22%	1.01 (1.00–1.01)	2.31E−05	1,070,257 (94,470–12,124,953)	<2.00E−16
Pack years adult smoking as proportion of life span exposed to smoking	Pack years/Age-16	25,7261 (697/256,564)	15.22%	1.83 (1.55–2.16)	8.12E−13	0.75 (0.70–0.79)	<2.00E−16
Time spent outdoors	Hours	28,8443 (778/287,665)	4.95%	0.91 (0.87–0.95)	4.06E−05	0.90 (0.83–0.99)	2.59E−02
Age	Years	303,461 (816/302,645)	0.00%	1.51 (1.48–1.54)	<2.00E−16	1.85 (1.76–1.93)	<2.00E−16
Sex	–	303,461 (816/302,645)	0.00%	–	–	–	–
Male		142,022 (390/161,049)		Ref	–	Ref	–
Female		161,439 (426/142,022)		0.79 (0.69–0.92)	1.15E−03	0.21 (0.10–0.42)	1.60E−05

[A] OR (95% CI): Odds ration (95% confidence intervals).
[B] *P*-value with Bonferroni correction.

ful events, either past or recent, have a strong effect on the disease; having fed-up feelings, having experienced extremely traumatic events such as having been sexually molested during childhood or mild stressful events such as marital separation or divorce all increase PDAC risk. On the contrary, leisure and social activities such as spending time outside in the summer or attending meetings of religious groups decrease the risk of developing the disease. Additionally, retirement also decreases the risk of PDAC. This might seem counterintuitive, since retirement could be considered a proxy of age, however retirees have lower exposure to work-related stress.

High income has been reported to increase the risk of pancreatic cancer in Chinese population [29]. However, in Western countries, high income levels were also reported to decrease the risk of pancreatic cancer in men [30]. High income was also reported to increase the risk of colorectal [29], breast and prostate cancer [30]. Additionally, being unemployed, which is a major stress factor in our modern society [31], was also associated with PDAC risk. These two socioeconomic factors (high income and unemployment) may represent the two faces of the same coin, both increasing stress and therefore PDAC risk. The association between high levels of IGF-1 and increased risk of PDAC is known and has been studied providing evidence that high levels of IGF-1 increase the basal growth rate of cancer cells since IGF-1 is necessary for progression in the cell cycle [32]. SHBG has been associated with a wide range of health outcomes, including risk of prostate [33], endometrial [34], liver, gastric and colorectal cancer [35], but not with pancreatic cancer. The literature on SHBG levels and PDAC is limited to a single study by Peila and colleagues, which found no association between high SHBG levels and the risk of developing PDAC, probably due to the small number of subjects analysed [36].

Dietary habits were also associated with PDAC risk. For example, processed meat intake, fresh fruit intake, dried fruit intake, beef intake, poultry intake, lamb or mutton intake showed associations that were significant at the conventional $p < 0.05$ but did not reach the Bonferroni corrected threshold, Supplementary Table 2. There are several lines of evidence associating asthma and allergies with a decreased risk of developing PDAC [37,38], but in the present study these association were not statistically significant (OR = 1.25, 95%CI: 0.99–1.56, $p = 5.75 \times 10^{-2}$) probably due to the low number of cases who had these data ($N = 88$). Physical activity was also not associated with PDAC risk, however the number of subjects for which this information was available is high ($n = 289{,}629$). These results are in agreement with the majority of what present in the current literature on the topic [39,40].

Our study has several strengths, such as the comprehensive analysis of the exposome, the integration of genetic data in the form of PRS, and the homogeneity of the measures across the cohort. We are aware of possible limitations. This is an observational study that highlights several associations between exposure variables and risk of developing PDAC, but these associations could be influenced by residual confounders that make it difficult to link cause and effect. In UKBB there are thousands of variables, but with many missing data, leading to a considerable reduction in statistical power, and very large estimates, that may be on overestimation of the real effect. Additionally, in UKBB there are no data on pancreatic cancer family history. The lack of this information may represent a confounding factor for some of the association. Furthermore, some of the variables analyzed refer to self-reported data collected through questionnaires, this could introduce an evaluation error given the subjective nature of the perception of exposure. Additionally, all the results have been obtained considering only white British people and therefore it is not possible to generalize the effect to other ethnicities or other countries. Finally, the multivariable analysis was only performed on a small subset of individuals and for variables with less than 25% missing values. This reduces the statistical power and increases the chances of false negatives. Therefore, the results have to be taken with caution.

In conclusion, our results suggest that a stressful lifestyle and sedentary behaviors may play a major role in PDAC susceptibility suggesting that changes in lifestyle could be beneficial to reduce the risk of the disease. Other studies are warranted to confirm and better understand the role of these potential new risk factors that we have highlighted.

Authors' contributions

Giulia Peduzzi and Daniele Campa were responsible for study concept and design, acquisition of data, statistical analysis, interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, and approval of the final draft manuscript.

Manuel Gentiluomo and Federico Canzian were responsible for the acquisition of data, interpretation of data, critical revision of the manuscript for important intellectual content, and approval of the final draft manuscript.

Alessio Felici, Roberto Pellungrini, Francesca Giorgolo, Riccardo Farinella, Andrea Spinelli, Gabriele Capurso, Anna Monreale and Marco Calderisi were responsible for the interpretation of data, critical revision of the manuscript for important intellectual content, and approval of the final draft manuscript.

Ethics approval and consent to participate

All participants have provided their informed consent to participate. Studies based on UK Biobank were performed in accordance with the Declaration of Helsinki. Ethical approval and consent were granted by the North West-Haydock NRES multicentre ethics committee, REF: 16NW/ 0274.

Funding information

The present research was funded with intramural funding of the University of Pisa.

Conflict of interest

All the authors have no conflicts of interest to declare.

Data availability

This research has been conducted using data from UK Biobank, a major biomedical database (https://www.ukbiobank.ac.uk/). Approval for the study and permission to access the data was granted by UK Biobank under Application Number 66591. Bona fide researchers can access the UK Biobank dataset by registering and applying at http://ukbiobank.ac.uk/register-apply/".

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 66591. The authors want to thank the UK Biobank participants for their contribution. This article is based upon work from COST Action "Identification of biological markers for prevention and translational medicine in pancreatic cancer (TRANSPAN)", CA21116, supported by COST (European Cooperation in Science and Technology). This work has been supported by the European Union under ERC-2018-ADG G.A. 834756 (XAI), and by the Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", It has been realized also thanks to computational (data and algorithms) resources of "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu), G.A.No.871042 and by NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Project: "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR000001 3 - Avvison. 3264 del 28/12/2021.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dld.2023.10.015.

References

[1] Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021;149(4):778–89.
[2] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209–49.
[3] Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol 2021;18(7):493–502.
[4] Nodari Y, Gentiluomo M, Mohelnikova-Duchonova B, et al. Genetic and non-genetic risk factors for early-onset pancreatic cancer. Dig Liver Dis 2023;55(10):1417–25 In Press. doi:10.1016/j.dld.2023.02.023.
[5] Gentiluomo M, Canzian F, Nicolini A, Gemignani F, Landi S, Campa D. Germline genetic variability in pancreatic cancer risk and prognosis. Semin Cancer Biol. 2022;79:105–31.
[6] Childs EJ, Mocci E, Campa D, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet 2015;47(8):911–16.
[7] Wolpin BM. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer HHS public access author manuscript. Nat Genet 2014;46(9):994–1000.
[8] Klein AP, Wolpin BM, Risch HA, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun 2018;9(1):556.
[9] Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 2009;41(9):986–90.
[10] Zhang M, Wang Z, Obazee O, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget 2016;7(41):35.
[11] Petersen GM, Amundadottir L, Fuchs CS, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet 2010;5(3):224–8.
[12] Campa D, Pastore M, Gentiluomo M, et al. Functional single nucleotide polymorphisms within the cyclin-dependent kinase inhibitor 2A/2B region affect pancreatic cancer risk. Oncotarget 2016;7(35):57011–20.
[13] Gentiluomo M, Lu Y, Canzian1 F, Campa D. Genetic variants in taste-related genes and risk of pancreatic cancer. Mutagenesis 2019;34(5–6):391–4.
[14] Campa D, Rizzato C, Stolzenberg-Solomon R, et al. TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. Int J Cancer 2015;137(9):2175–83.
[15] Yang W, Liu H, Duan B, et al. Three novel genetic variants in NRF2 signaling pathway genes are associated with pancreatic cancer risk. Cancer Sci 2019;110(6):2022–32.
[16] Corradi C, Gentiluomo M, Gajdán L, et al. Genome-wide scan of long noncoding RNA single nucleotide polymorphisms and pancreatic cancer susceptibility. Int J Cancer 2021;148(11):2779–88.
[17] Corradi C, Lencioni G, Gentiluomo M, et al. Polymorphic variants involved in methylation regulation: a strategy to discover risk loci for pancreatic ductal adenocarcinoma. J Med Genet 2023;60(10):980–6 jmg-2022-108910.
[18] Pistoni L, Gentiluomo M, Lu Y, et al. Associations between pancreatic expression quantitative traits and risk of pancreatic ductal adenocarcinoma. Carcinogenesis 2021;42(8):1037–45.
[19] Giaccherini M, Farinella R, Gentiluomo M, et al. Association between a polymorphic variant in the CDKN2B-AS1/ANRIL gene and pancreatic cancer risk. Int J Cancer 2022:373–9.
[20] Campa D, Gentiluomo M, Stein A, et al. The PANcreatic Disease ReseArch (PANDoRA) consortium: ten years' experience of association studies to understand the genetic architecture of pancreatic cancer. Crit Rev Oncol Hematol 2023;186(May):104020.
[21] Gentiluomo M, Katzke V, Kaaks R, et al. Mitochondrial DNA copy number variation and pancreatic cancer risk in the prospective EPIC cohort. Pancreatology 2021;21(3):S62–3.
[22] Campa D, Matarazzi M, Greenhalf W, et al. Genetic determinants of telomere length and risk of pancreatic cancer: a PANDoRA study. Int J Cancer 2019;144(6):1275–83.
[23] Malats N. Gene-environment interactions in pancreatic cancer. Pancreatology 2001;1(5):472–6.
[24] Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.
[25] Galeotti AA, Gentiluomo M, Rizzato C, et al. Polygenic and multifactorial scores for pancreatic ductal adenocarcinoma risk prediction. J Med Genet 2020:1–9.
[26] Nguyen LH, Liu PH, Zheng X, et al. Sedentary behaviors, TV viewing time, and risk of young-onset colorectal cancer. JNCI Cancer Spectr 2018;2(4):1–8.
[27] Friberg E, Mantzoros CS, Wolk A. Physical activity and risk of endometrial cancer: a population-based prospective cohort study. Cancer Epidemiol Biomark Prevent 2006;15(11):2136–40.
[28] Sanchez-Bayona R, Gardeazabal I, Romanos-Nanclares A, et al. Leisure-time physical activity, sedentary behavior, and risk of breast cancer: results from the SUN ('Seguimiento Universidad De Navarra') project. Prev Med (Baltim) 2021;148:106535.
[29] Pang Y, Kartsonaki C, Guo Y, et al. Socioeconomic status in relation to risks of major gastrointestinal cancers in Chinese adults: a prospective study of 0.5 million people. Cancer Epidemiol Biomark Prevent 2020;29(4):823–31.
[30] Larsen IK, Myklebust TÅ, Babigumira R, Vinberg E, Møller B, Ursin G. Education, income and risk of cancer: results from a Norwegian registry-based study. Acta Oncol (Madr) 2020;59(11):1300–7.
[31] Lawes M, Hetschko C, Schöb R, Stephan G, Eid M. Unemployment and hair cortisol as a biomarker of chronic stress. Sci Rep 2022;12(1):21573. Available from: https://www.nature.com/articles/s41598-022-25775-1 .
[32] Mutgan AC, Besikcioglu HE, Wang S, Friess H, Ceyhan GO, Demir IE. Insulin/IGF-driven cancer cell-stroma crosstalk as a novel therapeutic target in pancreatic cancer. Mol Cancer 2018;17(1):66. Available from: https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-018-0806-0 .
[33] Gann PH, Hennekens CH, Ma J, Longcope C, Stampfer MJ. Prospective study

of sex hormone levels and risk of prostate cancer. JNCI J Natl Cancer Inst 1996;88(16):1118–26. Available from: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/88.16.1118 .
[34] Mullee A, Dimou N, Allen N, O'Mara T, Gunter MJ, Murphy N. Testosterone, sex hormone-binding globulin, insulin-like growth factor-1 and endometrial cancer risk: observational and Mendelian randomization analyses. Br J Cancer 2021;125(9):1308–17. Available from: https://www.nature.com/articles/s41416-021-01518-3 .
[35] Liu Z, Zhang Y, Lagergren J, et al. Circulating sex hormone levels and risk of gastrointestinal cancer: systematic review and meta-analysis of prospective studies. Cancer Epidemiol, Biomark Prevent 2023;32(7):936–46. Available from: https://aacrjournals.org/cebp/article/32/7/936/727317/Circulating-Sex-Hormone-Levels-and-Risk-of .
[36] Peila R, Arthur RS, Rohan TE. Association of sex hormones with risk of cancers of the pancreas, kidney, and brain in the UK biobank cohort study. Cancer Epidemiol, Biomark Prevent 2020;29(9):1832–6. Available from: https://aacrjournals.org/cebp/article/29/9/1832/72371/Association-of-Sex-Hormones-with-Risk-of-Cancers .
[37] Gomez-Rubio P, Zock J-P, Rava M, et al. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. Gut 2017;66(2):314–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26628509 .
[38] Gandini S, Lowenfels AB, Jaffee EM, Armstrong TD, Maisonneuve P. Allergies and the risk of pancreatic cancer: a meta-analysis with review of epidemiology and biological mechanisms. Cancer Epidemiol Biomark Prev 2005;14(8):1908–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16103436 .
[39] O'Rorke MA, Cantwell MM, Cardwell CR, Mulholland HG, Murray LJ. Can physical activity modulate pancreatic cancer risk? a systematic review and meta-analysis. Int J Cancer 2010;126(12):2957–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19856317 .
[40] Bao Y, Michaud DS. Physical activity and pancreatic cancer risk: a systematic review. Cancer Epidemiol Biomark Prev 2008;17(10):2671–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1884309 .