University of Southampton

Faculty of Medicine

Human Development and Health

HbA1c as a predictor of pancreatic ductal adenocarcinoma

by

Adrian Wei Ern Cheang

BSc (Hons) MB BS

Thesis for the degree of Master of Philosophy

2nd December 2023

University of Southampton

Abstract

Faculty of Medicine

Human Development and Health

Thesis for the degree of Master of Philosophy

HbA1c as a predictor of pancreatic ductal adenocarcinoma

Adrian Wei Ern Cheang

Introduction

UK adults aged 60 or over with a new onset diabetes (NODM) diagnosis and recent weight loss are now being recommended for urgent investigations to rule out possible underlying pancreatic ductal adenocarcinoma (PDAC). Differently measured HbA1c concentration levels over time with repeat sampling may be a key approach to triggering referral pathways for further investigation.

Aim

The aim of this study was to investigate the relationship between HbA1c interpreted as single values and as grouped values according to clinical criteria for pre-diabetes and diabetes, and occurrence of PDAC in individuals diagnosed with NODM.

Methods

In this matched case-control study, 502,459 UK Biobank dataset participants were screened for incidental PDAC and HbA1c measurements at study baseline. Cox proportional hazards regression modelling with univariate and multivariate analysis generated hazard ratios (HR) for each of our chosen variables. Log-rank tests were performed for statistical significance. Receiver-operating characteristic (ROC) analysis and Youden index calculations were used to evaluate the performance of HbA1c as a test for detecting PDAC.

Results

HbA1c was not a useful standalone predictive marker for PDAC. However, newly elevated HbA1c in participants without a history of diabetes suggestive of emerging new pre-diabetes or new diabetes diagnosis was associated with a significantly higher risk of PDAC.

Conclusion

Interpretation of HbA1c measurements benefits from knowledge of prior diabetes status when predicting risk of PDAC. The limitations of our study with single HbA1c measurements suggest that to enhance our understanding of the relationship between HbA1c and PDAC, future studies may benefit from examining repeat HbA1c measurements over time.

1 Table of Contents

2 Research Thesis: Declaration of Authorship

Print name: Dr Adrian Wei Ern Cheang

Title of thesis: HbA1c as a predictor of pancreatic ductal adenocarcinoma

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at his university;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. None of this work has been published before submission;

Signature: Date:

3 Acknowledgements

I would like to express my utmost gratitude to my lead supervisor Professor Chris Byrne for his constant encouragement and unwavering support throughout my postgraduate degree, and to Mr Zaed Hamady for the relentless efforts he has put in to making this project possible. I would also like to extend these thanks to Professor Sarah Wild for her readiness to offer invaluable insights and guidance at times when it was much needed. Special thanks must also go to Dr Declan McDonnell for his thoughtful contributions and comradery as a fellow medical trainee navigating the world of academic research. I would also like to thank the British Association of Surgical Oncology for the research grant award that helped fund this project and allow me to conduct my thesis. Within the Faculty of Medicine and University of Southampton, I would like to thank Professor Dankmar Boehning, Mr Scott Harris and Mr Brian Yuen of the Faculty of Medicine Medical Statistics Group and Professor Nisreen Alwan, to all of whom I am indebted for the skills I have gained through their tutelage early in the programme. I am also incredibly grateful to our university's librarians who taught me the wide range of technical skills needed to facilitate my work. I cannot begin to express my thanks to Dr Nicola Englyst, who provided moral and emotional support during the hardest parts of the COVID-19 pandemic. By extension, I also wish to thank Dr Ashley Pringle, Dr Reuben Pengelly, and Dr Tracey Newman for the incredible patience and understanding they have shown in helping me to accomplish my goals as a postgraduate research student under their oversight. Lastly, I would like to express my endless gratitude to my wife for her endless encouragement and support throughout my studies.

This research has been conducted using the UK Biobank Resource under Application Number 17749

4 List of abbreviations

AMP	adenosine monophosphate
ANA	American Diabetes Association
BMI	body mass index
cfDNA	circulating fluid cell-free DNA
DM	diabetes
DF	Data Field
DPP-4i	dipeptidyl peptidase-4 inhibitors
FAMMM	Familial atypical multiple mole melanoma syndrome
FAP	Familial adenomatous polyposis
FBG	fasting blood glucose
GLP-1	glucagon-like peptide 1
GLP-1a	glucagon-like peptide agonist
GNI	gross national income
HDI	human development index
HDL	high density lipoprotein
HR	hazard ratio
IDDM	insulin-dependent diabetes mellitus
IPMN	intraductal papillary mucinous neoplasms
LSDM	long-standing diabetes mellitus
MetS	metabolic syndrome
mPanIN	mouse pancreatic intraepithelial neoplasms
NAFPD	non-alcoholic fatty pancreas disease
NICE	(UK) National Institute for Health and Care Excellence
NIDDM	non-insulin-dependent diabetes mellitus
NODM	new-onset diabetes mellitus
OAM	oral antidiabetic medication
OGGT	oral glucose tolerance test
OR	odds ratio
PanC4	Pancreatic Cancer Case-Control Consortium
PanIN	pancreatic intraepithelial neoplasms

List of abbreviations

PC	Pancreatic cancer
PDAC	pancreatic ductal adenocarcinoma
PDAC-DM	PDAC-associated new onset diabetes mellitus
PRS	polygenic risk scoring
PPP$	purchasing power parity
RCGP RSC	UK Royal College of General Practitioners Research and Surveillance Centre
SU	sulphonylurea
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
T3cDM	Type 3c diabetes mellitus
TDI	Townsend deprivation index
TZD	Thiazolidinediones
UK	United Kingdom
WHO	World Health Organisation

5 Introduction

There is growing recognition of pancreatic ductal adenocarcinomas (PDAC) as a genetic disease that offers new opportunities for disease prediction. Individuals who are at high risk of developing sporadic pancreatic cancers (PC) have no effective screening tool for the early detection of this deadly disease. As such, the discovery of biomarkers and new testing techniques for risk modelling remains the primary goal for most researchers in this field. Identification of sporadic PDAC in high-risk individuals using techniques such as polygenic risk scoring (PRS) [1] and targeted methylation analysis on circulating cell-free DNA (cfDNA) [2] offer new ways to enhance currently proposed risk models for detection of early-stage PDAC which rely primarily on clinical features alone. However, these genetic screening tools are still relatively novel, not widely accessible, and require clinical validation prior to their use in routine clinical practice [2]. Given these limitations, the optimal use of these techniques for early-stage sporadic PDAC risk prediction models as they currently exist may be as secondary screening tools. These have the potential to be used to identify or enrich a high-risk cohort of individuals first identified based on clinical indicators for further monitoring or investigations.

Due to the relative rarity of sporadic PDAC in the general population, identifying high-risk individuals in this cohort remains a significant challenge. To date, diabetes is one of the few known clinical high-risk factors associated with sporadic PDAC. It is widely accepted that the timing of diabetes onset plays a significant role in the likelihood of developing PDAC and early recognition of new onset diabetes (NODM) may improve prognosis by triggering earlier investigations for PDAC. In December 2021, the UK National Institute for Health and Care Excellence (NICE) updated their guidance for the detection of pancreatic cancer symptoms in primary care to include the combination of NODM with weight loss and age over 60 as a criteria for urgent referral for further investigation and management [3]. The recommendation for management requires clinicians to consider an urgent direct access CT scan (to be done within 2 weeks), or urgent ultrasound scan if CT is not available [3]. The inclusion of NODM as a recognised risk factor for PDAC in the UK NICE guidelines is an important first step towards the development of a screening algorithm that improves overall prognosis for this disease. The relationship between diabetes and pancreatic cancer appears to be bi-directional. As such developing an effective screening strategy for detection of early-stage PDAC still requires better understanding of the interaction between modifiable and

non-modifiable risk factors, genetic risk factors, and utility of investigative tools and screening markers associated with both of these highly complex diseases.

6 Literature Review

6.1 Pancreas Anatomy

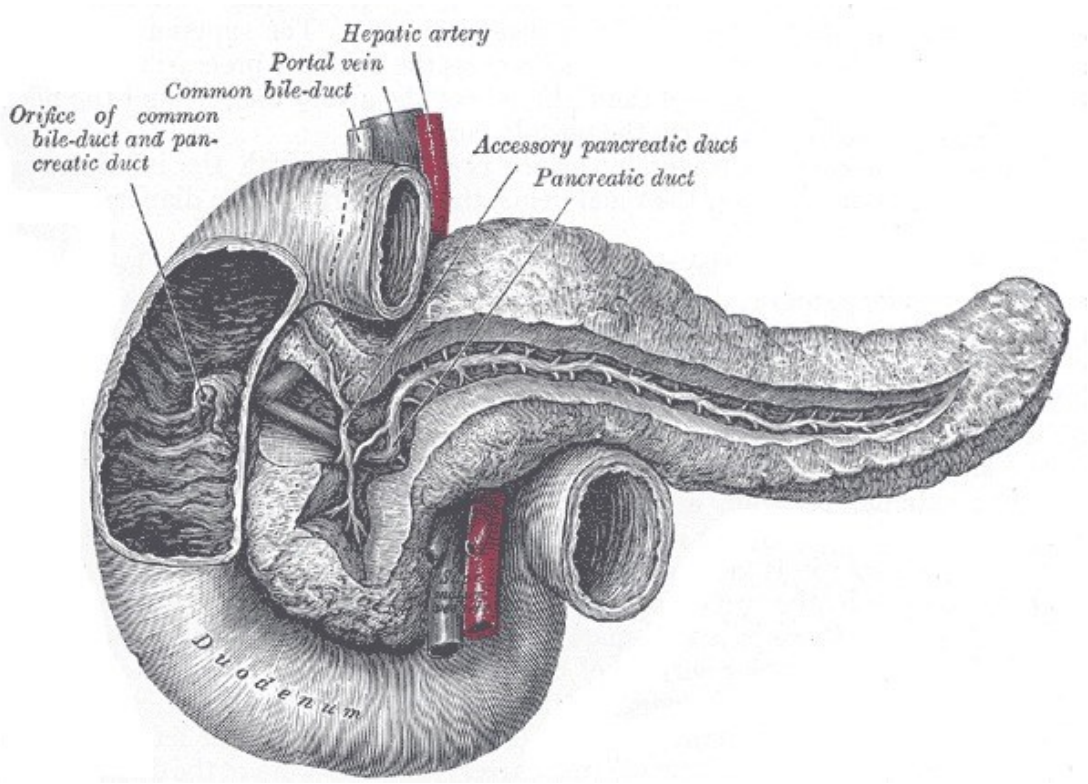


Figure 1. The Pancreas, The pancreatic duct, Orifice of common bile-duct and pancreatic duct, Accessory pancreatic duct. Contributed by Gray's Anatomy Plates [4]. (Republished, with permission from Statspearls Publishing LLC)

The pancreas is an accessory digestive gland that is found retroperitoneally on the posterior abdominal wall that crosses the L1 and L2 vertebra [5]. It lies in the upper abdomen between the duodenum on the right (Figure 1.) and spleen on the left, and divided anatomically into the head, neck, body and tail [5]. Approximately 80% of the pancreas is made up of exocrine pancreatic acini, which are pyramidal acinar cells arranged in circular groups with the apex directed towards a central lumen. These cells contain zymogen granules apically, and a nucleus and endoplasmic reticulum towards the cell base which aids the synthesis of digestive enzymes. The enzymes are stored in the Golgi complex of these cells until signalling pathways trigger their secretion into the central lumen. Acinar cells contain receptors for secretin, cholecystokinin and acetylcholine – neurotransmitters which regulate exocytosis of the digestive enzymes [5].

The pancreas also contains one to two million islets of Langerhans, measuring on average between 100-150µm in diameter. 1-2% of the total pancreatic mass is islet tissue, which are responsible for the production of glucagon, insulin, somatostatin, and pancreatic polypeptide [6]. The islets of the anterior head and tail of the pancreas are derived from the primordial dorsal bud. These are made up pre-dominantly of α-cells (15-20%), which produce glucagon to elevate blood glucose, and β-cells (3-10%) which produce insulin to lower blood glucose.

6.2 Pancreatic ductal adenocarcinoma

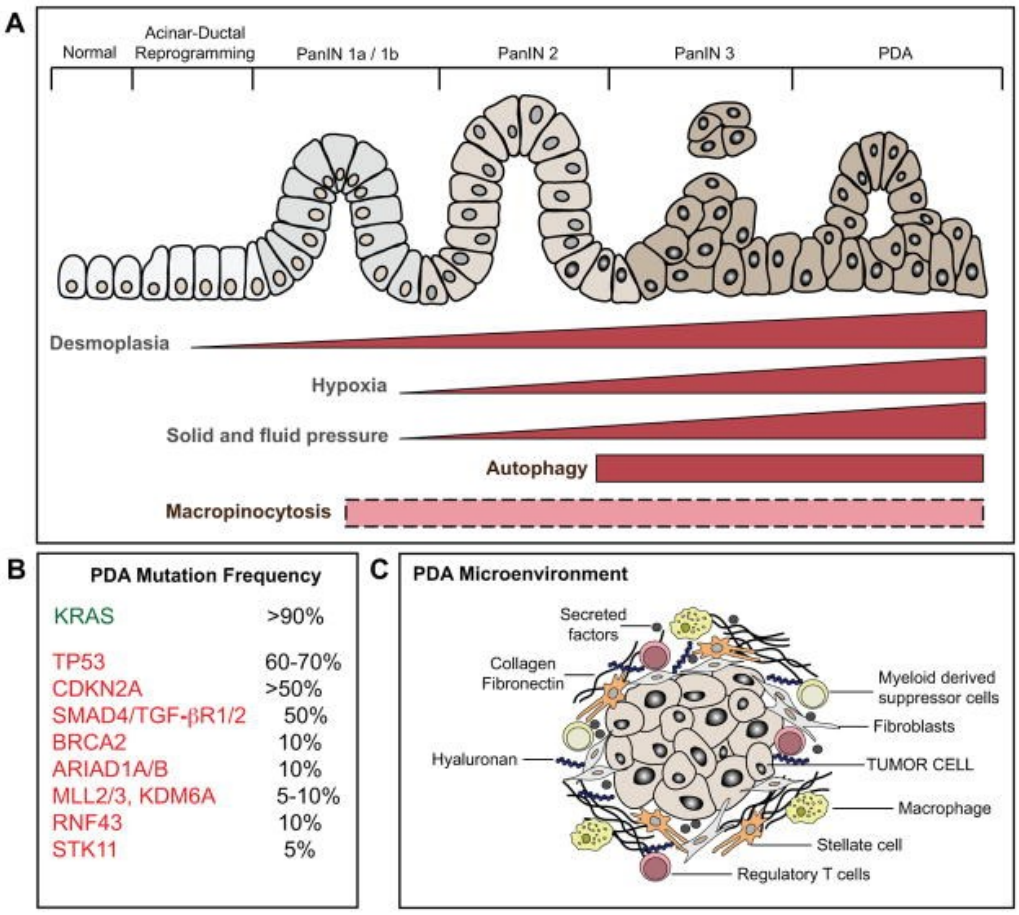


Figure 2. Schematic of the multi-stage progression of PDAC. (A) Some forms of PDAC may arise from multi-stage progression of precursor lesions known as pancreatic intraepithelial neoplasia (PanIN). (B) KRAS, TP53 and CDKN2A mutations are three most commonly identified mutations associated with disease progression. (C) The tumour microenvironment is altered with increased fibrosis, extracellular matrix deposition (desmosplasia), and recruitment of immune and inflammatory cells. Republished from Perera and Bardeesy, 2015;**5**(12) [7], with PMC Open Access Subset content copyright licence permissions.

Pancreatic cancers are subdivided by tumour origin into two groups: exocrine and endocrine. Endocrine pancreatic cancer is responsible for the production of various hormones including insulin. These occur in 2-4% of all PC cases and originate from the islet cells. Exocrine PCs account for about 95% of all pancreatic cancers and almost all exocrine PCs are pancreatic ductal adenocarcinomas. Other much rarer exocrine cancers include acinar cell carcinoma (1% of pancreatic cancers), solid pseudopapillary neoplasms (%), and pancreatoblastomas [8]. Of these, acinar cell carcinomas are more common in men and found in about 1% of pancreatic cancer cases. Solid papillary neoplasms are more common in younger women and are the most common pancreatic tumour in children. Pancreatoblastomas are also mostly

found in children and are extremely rare in adults [9]. Thus, in the context of sporadic pancreatic cancers that develop in later adult life, the likelihood of presenting with these rarer exocrine cancer diagnoses is extremely small.

Within the scientific literature, two definitions (PC and PDAC) have been observed to be used interchangeably. This is due to the disproportionately high incidence of PDAC diagnosis verses any other subtype being referred to when described more simply as pancreatic cancer. Likewise, it has been argued that the number of rarer subtypes included in studies where PC subtypes are not clearly defined are frequently assumed to be so small as to be statistically insignificant. For consistency, both PC and PDAC in referenced literature will be referred to as PDAC in this study to represent any pancreatic cancers of non-neuroendocrine origin.

PDAC follows an evolutionary pathway characterised by hypovascularization, desmoplasia (fibrotic overgrowth around the tumour), genomic complexity and metabolic reprogramming [10]. The altered metabolism of PDAC cells allows it to thrive in poorly perfused, hypovascular conditions by upregulating nutrient acquisition and utilization pathways simultaneously (Figure 2.) [7]. The overexpression of glucose transporters and glycolytic enzymes positively influences tumour growth regardless of the oxygen availability for aerobic metabolism in PDAC. This type of paradoxical metabolic reprogramming is known as ‘the Warburg effect’ that can be found in most cancers [11]. The cancer pathway for PDAC follows a typical pattern for many cancers, involving three critical periods of change: Time 1 – from tumour initiation and cell birth to the rise of a parental clone, Time 2 – from the parental clone to metastasis, and Time 3 – from cellular dissemination to patient death [12]. PDAC primaries contain a mixture of distinct sub-clones which exist many years before metastases becomes evident [12]. These have been calculated to take an average estimate of 11.7 years to progress from initiation to overt cancer, and a further 6.8 years on average to become metastatic [12]. This finding may offer potential windows of opportunity to detect PDAC lesions early, including precursor lesions whilst they are still in the curative stages of disease [12].

6.2.1 Pathogenesis of PDAC from a genetic perspective

Advances made with next generation sequencing techniques and the regular sequencing of tumours have led to the recognition of PDAC as a genetic disease. By identifying tumour genomes and analysing their pattern and frequency of expression, genetic studies continue to provide greater insight into its pathogenesis [13]. PDAC has around 60 genetic alterations per tumour [13] and a significant feature of the PDAC genome is that the combination of alterations that arise is unique to each case [14, 15]. The nature and function of this tumour's heterogeneity are increasingly being elucidated and emerging knowledge about the complex biology of this tumour offers up new approaches to targeting for early detection and subsequent treatment. From these extensive gene studies, four genes have been identified to be frequently mutated in PDAC: K-Ras, CDKN2A, p53, and SMAD4.

K-Ras oncogene mutations occur in low-grade PanIN lesions and in over 90% of early PDAC tumours [16]. Such a high frequency of this mutation suggests that tumour pathogenesis may revolve predominantly around the molecular pathways regulated by K-Ras and provides a strong argument for focussing research on K-Ras inhibitor therapies. K-Ras also inactivates the tumour suppressor gene p16/CDKN2A in early pancreatic intraepithelial neoplasms (PanIN), and the tumour suppressor gene p53 in 50-75% of PDAC cases, often in the late PanIN phase of disease (Figure 3.) [16]. Other oncogenes and tumour suppressor genes are also associated with PDAC but the prevalence and importance of K-Ras, p16 and p53 gene mutations specifically has led to their inclusion as key pancreatic juice analysis gene markers for PDAC in the European Registry of Hereditary Pancreatitis and Familial Pancreatic Cancer (EUROPAC) screening study. EUROPAC is gathering data with the aim of developing new screening programmes for the detection of early-stage PDAC in high risk individuals (HRI) and the inclusion of pancreatic juice gene markers marks another step in this complex process [17].

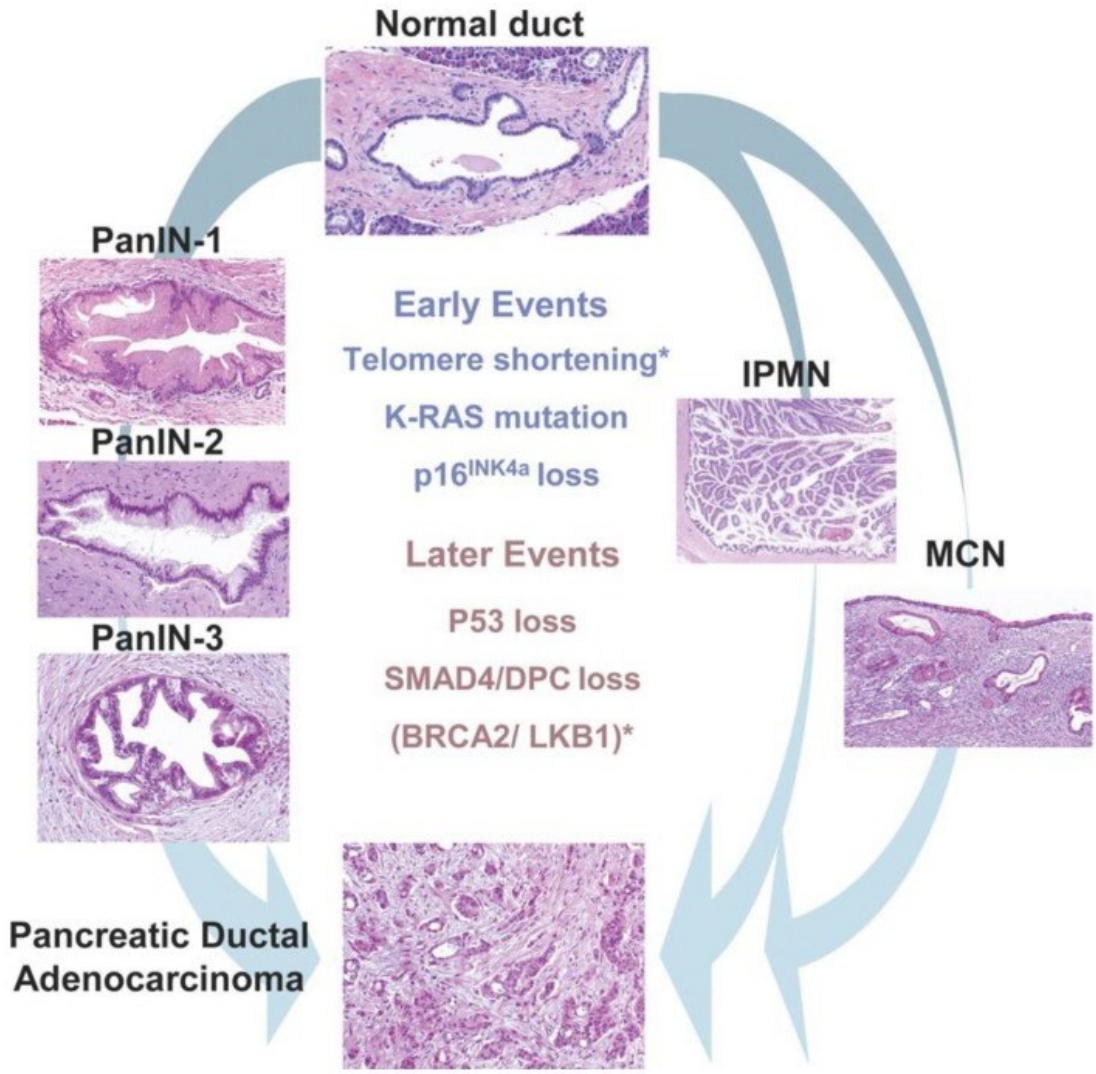


Figure 3. Pancreatic precursor lesions and genetic events involved in pancreatic adenocarcinoma progression. PanIN, IPMN, and MCN represents three known precursor lesions of PDAC. Early genetic alteratsion (K-ras mutations, P16) and late genetic alterations (p53 loss, SMAD4 loss), that occur in adenocarcinomas also occur in PanIN and to a lesser extent in IPMNs and MCNs. Asterisks indicate events that a re not common to all precursor lesions. Republished from Polireddy and Chen 2016;7(11), with permission from Ivyspring International Publisher

CDKN2A is involved in the regulation of the RB1 gene which has an important role in G1/S checkpoint inhibition of the cell cycle. Mutations in CDKN2A occur in early PDAC and are associated with a 20-fold higher risk for PDAC when identified in patients with the familial pancreatic cancer-melanoma syndrome compared to patients without these mutations [18]. Certain genes highly correlated with CDKN2A and PDAC progression include CDK4, CDK6, MYC, MDM2 and p53. CDK4 is a major regulator of insulin signalling [19] and its activity is shown to be critical for the proliferation of differentiated β cells in mouse islets to accommodate the steady-state requirements of insulin secretion [18]. CDK4 activation has a

close relationship with cancer development and is inhibited by CDKN2A. It is also involved in the induction of gluconeogenesis in the liver and has an important metabolic role in the development and progression of PDAC. CDK6, like CDK4, is involved in the regulation of G1 to S cell cycle progression, transcription, differentiation [18]. The combination of CDK6 and cyclin D3 (CD3) plays an important role in glucose metabolism and promotes cancer cell survival [20]. Pre-clinical and clinical trials have shown that these properties can be exploited with CDK4 and CDK6 inhibitors having a role in PDAC therapy. As such, CD3-CDK6 inhibitors may enhance the clinical efficacy of anti-CDK4/6 treatment and prognosis in patients with PDAC [20].

MYC has an important role in activating multiple genes involved in DNA replication, transcription, translation, chromatin modification and protein synthesis and degradation. MYC mutations induce errors in the normal cell cycle with amplification of the gene resulting in overexpression, uncontrolled cell growth and multiplication [21]. The overexpression of MYC is found in up to 42% of late PDAC diagnoses, and associated with poorer clinical outcomes, increased probability of recurrence, worsening disease and decreased survival [21]. MYC also protects PDAC cells from failure and inhibits their differentiation, highlighting the importance of developing treatments targeting this gene [22].

P53 plays a role in inducing cell cycle arrest at G1 or G2 checkpoints. Mutations in this gene are found in more than 50% of malignant cancers with considerable variation in frequency depending on the cancer type and are associated with a loss of tumour suppressor function [23]. Low expression of p53 is associated with an increased risk of PDAC with high oncogene MDM2 overexpression – another gene closely associated with CDKN2A [24]. MDM2, among other oncogenes, attenuates the function of p53, resulting in its mutation and thus contribution to the development of cancer [25]. Its expression is associated with poor prognosis for PDAC both in response to p53 mutation and as a p53-independent effector of tumour development and progression. As such, MDM2 provides researchers with another important target for future therapies in addition to p53.

TGFβ is a key mediator of immune evasion in PDAC, and the deletion of TGFβ target SMAD4 is identified in approximately 55% of PDAC tumours [25]. Although the exact contribution of SMAD4 to PDAC development and overall prognosis remains unclear, its inactivation leads to modification of the TGFβ response, resulting in the TGFβ signalling pathway changing from a tumour suppressor to a tumour promoter with disease progression. Some therapeutic benefits may exist with a SMAD4-targeted treatment approach. However, the existence of a

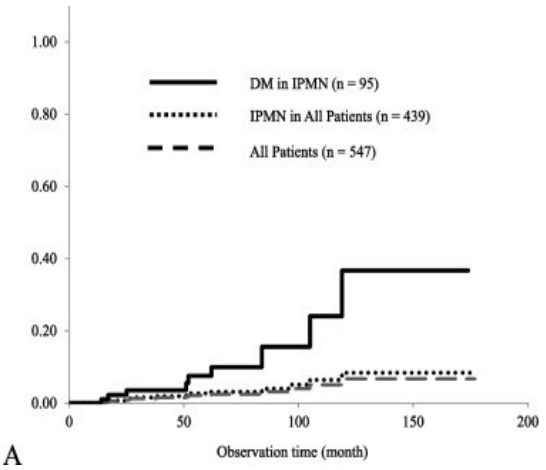
SMAD-independent TGFβ signalling pathway may bypass such strategies, thus limiting the overall effectiveness of this approach [26].

Other gene mutations that occur less frequently are as important in their roles in PDAC. Many of these mutations appear to converge on a handful of key pathways and processes [14]. These include NOTCH, Hedgehog, β-catenin, axon guidance, chromatin remodelling, and DNA repair pathways. For example, the BRCA2, PALB2, FANCC, FANCG gene mutations constituting less than 10% of the mutated genes in PDAC are all involved in the genome maintenance DNA repair pathway [13].

Tumours either deficient or showing mutant variation of these genes may provide openings for potential treatment pathways. The frequent and lesser gene mutations all point towards potentially effective targeted therapies depending on the unique genetic makeup of the lesion in each case of PDAC. Although genetic studies have primarily focussed on the search for therapeutic targets, their contribution to the understanding of the role genes play within the cell cycle and metabolic character of this cancer also provides vital insights into its etiopathogenesis and provide vital clues to effective early detection strategies.

6.2.2 Precursors and cancer development

The pathogenesis of exocrine PDACs is believed to involve at least two distinct pathways that includes the evolution of intraductal papillary mucinous neoplasms (IPMN) into its invasive form, and pancreatic intraepithelial neoplasms (PanIN) that develop into PDAC [27]. IPMN risk factors include long-standing diabetes (LSDM) treated with insulin, chronic pancreatitis, and family history of PDAC (Figure 4.) [28, 29], whereas PanINs are associated with obesity and pancreatic fatty infiltration [29, 30]. These risk associations support the link between neoplastic change and clinical disease states that may also aid the early detection of PDAC as part of future screening models.



B

	At 5 Years	At 10 Years
All Patients with Pancreatic Cyst		
Incident ratio	2.12%	6.7%
Expect ratio	1.81%	2.28%
SIR	1.17	2.93
IPMN in Pancreatic Cyst		
Incident ratio	2.65%	8.39%
Expect ratio	1.09%	1.59%
SIR	2.43	5.29
DM in IPMN		
Incident ratio	7.51%	36.7%
Expect ratio	2.54%	3.06%
SIR	2.96	11.99

Figure 4. (A) Cumulative carcinogenic curve in all patients with cyst, patients with intraductal papillary mucinous neoplasm (IPMN), and IPMN plus diabetes mellitus (DM). (B) Cumulative carcinogenic ratio (Incident ratio), Expected ratio in the general Japanese population (Expect ratio), and standardized incidence ratio (SIR) in each group. Reproduced from Yamaguchi et al., 2022, with PMC Open Access Subset content copyright licence permissions.

6.2.2.1 Pancreatic intraductal neoplasms

PanINs are precursor lesions that are typically microscopic (<5mm), flat or papillary, arising from the small intralobular pancreatic ducts [31]. These lesions acquire progressive mutations which are classified by their morphological features as low grade PanIN-1A (flat), PanIN1-1B (papillary), intermediate grade PanIN-2 (mild to moderate cytological and architectural atypia), and high grade PanIN-3 (showing severe cytological and architectural atypia) [32]. Not all PanIN lesions develop into PDAC and not all patients diagnosed with PDAC are found to have PanIN lesions. However, high grade PanINs and PDAC show similar immunohistochemical protein expression by markers for the apomucins MUC1, MUC3, MUC4, and MUC5AC, suggesting that PDAC stemming from PanINs may be identified through

shared gene expression patterns [32]. Furthermore, pancreatic cancer patients with a family history of pancreatic cancer and carriers of the CDKN2A gene have both been shown to frequently develop multifocal, microscopic PanINs around the primary tumour [33, 34]. The grade of PanIN can also be linked to specific alterations to gene expression, with CDKN2A inactivation being detectable in early PanIN, and p53 and SMAD4 inactivation being associated with later alterations in tumour progression [31].

6.2.2.2 Intraductal papillary mucinous neoplasms

Up to 15% of PDACS are thought to arise from mucinous pancreatic cysts, which include IPMNS and mucinous cystic neoplasms (MCNs) [13]. Morphologically, these can be categorised by the location and extent of involvement within the pancreas, with 15-21% of IPMNs originating from the main duct segment located within the pancreatic head. They typically result either in dilatation of the main duct of more than 6mm or in the formation of cystic lesions that appear well defined on radiological imaging. Approximately one third of patients with IPMN are associated with invasive carcinoma and some of the genetic changes seen are also found in PDAC, including K-ras, CDKN2A, SMAD4 and p53 [35]. PDAC derived from IPMN appear to be smaller, less invasive, and less extensive than ordinary PDAC, and maybe more prognostically favourable in terms of their biological behaviour [36].

PDAC patients presenting with concomitant PanIN lesions appear to demonstrate longer post-resection survival compared to PDAC patients who present without PanIN [37]. In a separate study, patients found with higher-grade PanIN lesions (PanIN2 and PanIN 3) have also been shown to have a better overall survival than those with lower-grade PanIN lesions [38]. The authors of both studies suggested that such findings may be explained by the existence of a subset of tumours which lack PanIN, that demonstrate more aggressive local invasion, and have a tendency to develop into more poorly differentiated cancers through an as-yet undefined pathway [37, 38]. These differences in clinicopathological presentation are known to exist elsewhere such as with colon cancers [39]. For example, it is well recognised that most sporadic colon cancers arise from the progression of adenoma to carcinoma. However, other less common colon cancer subtypes may arise independently of adenomatous formation, such as those demonstrating defects in DNA mismatch repair in patients with hereditary nonpolyposis colorectal cancer syndrome [38].

Early resection of IPMNs has also been shown to improve survival [40]. Clinically, IPMNs tend to be visible on conventional imaging, whereas PanINs are not easily detectable, thus limiting the effectiveness of radiological imaging as a standalone screening tool for early PDAC. To

aid the search for effective screening models for PDAC, more studies examining the significance of having PDAC with concomitant IPMN vs PDAC derived from IPMN and others are needed. The findings from such studies would help develop optimal methods of monitoring and early treatment of PanIN, IPMN and MCN precursor lesions. Vital insights into the prognostic value of screening for PanIN and IPMN as part of a wider protocol for early detection of sporadic PDAC are also needed from further research. Currently, no effective screening method for sporadic PDAC exists which emphasises the need to identify effective biomarkers for the early detection of PDAC and potentially its precursor lesions.

6.2.2.3 Pancreatic ductal adenocarcinoma

PDAC was originally divided into three subtypes based on expression of certain genes: Classical (high expression of adhesion-associated and epithelial genes), Quasimesenchymal (high expression of mesenchyme associated genes), and Exocrine-like (showing relatively high expression of tumour cell derived digestive enzyme genes) [41]. The International Cancer Genome Consortium study updated these definitions to Progenitor (similar to classical subtype), Squamous (similar to quasimesenchymal; associated with poor prognosis similar to basal subtype carcinomas of other organs), and aberrantly differentiated endocrine exocrine (ADEX; similar to exocrine-like subtype). A new immunogenic subtype has also been identified that partially overlaps with the previously described classical subtype lesions due to their marked enrichment of immune signature [41]. More recently, the heterogeneity of PDAC identified through RNA-based sequencing has revealed further insights into its pathogenesis.

The clonal nature of shared mutations among PanINs and advanced tumours supports a stepwise-progression model for pancreatic cancer. Multiple other alterations are believed to occur simultaneously, each having their own driver or passenger roles in pathogenesis [41]. Despite the intra-tumoral heterogeneity demonstrated by PDAC in situ, metastatic PDAC shows high uniformity of driver mutations when examined within the same patient [41]. This suggests that potential treatments modelled to target patient-specific driver gene mutations based on their individual mutant gene profile has the potential to significantly limit the malignant processes involved. Another important pathophysiological component of PDAC is the stroma that surrounds these neoplastic cells. Poor vascularity and intense desmoplastic stroma are predominant components of PDAC in which the neoplastic cells survive. Transcriptome analysis of these stromal cells has shown that desmoplastic stroma associated with the malignant changes of PDAC comprises two subtypes with different gene expression

characteristics: a normal subtype that resembles pancreatic stellate cells and an activated ‘inflammatory’ subtype [42]. The biphasic nature of this stroma may provide an additional target for detection of metastatic PDAC and enhance protocols for grading and staging as the interaction between this stroma and neoplastic cells may also have a role in tumour heterogeneity [42].

6.2.3 Incidence and mortality

In the United Kingdom, pancreatic cancer accounted for 3% of newly diagnosed all-cancer cases in 2014 with around 9000 new cases reported yearly. Incidence rates for pancreatic cancer have increased by 14% since the early 1990's and it is expected to rise by a further 6% to 21 cases per 100,000 people by 2035 [43]**.** PC is the 11th commonest cancer in the United Kingdom (UK), yet accounts for 5% of all cancer-related deaths and the fifth commonest cause of cancer mortality in the UK [43]. Mortality rates have remained unchanged in the UK over 40 years, with fewer than 5% surviving beyond 5 years, and less than 1% at 10 or more years. By comparison, breast cancer survival in the UK has doubled from 40% to 78% in the last 40 years [44].

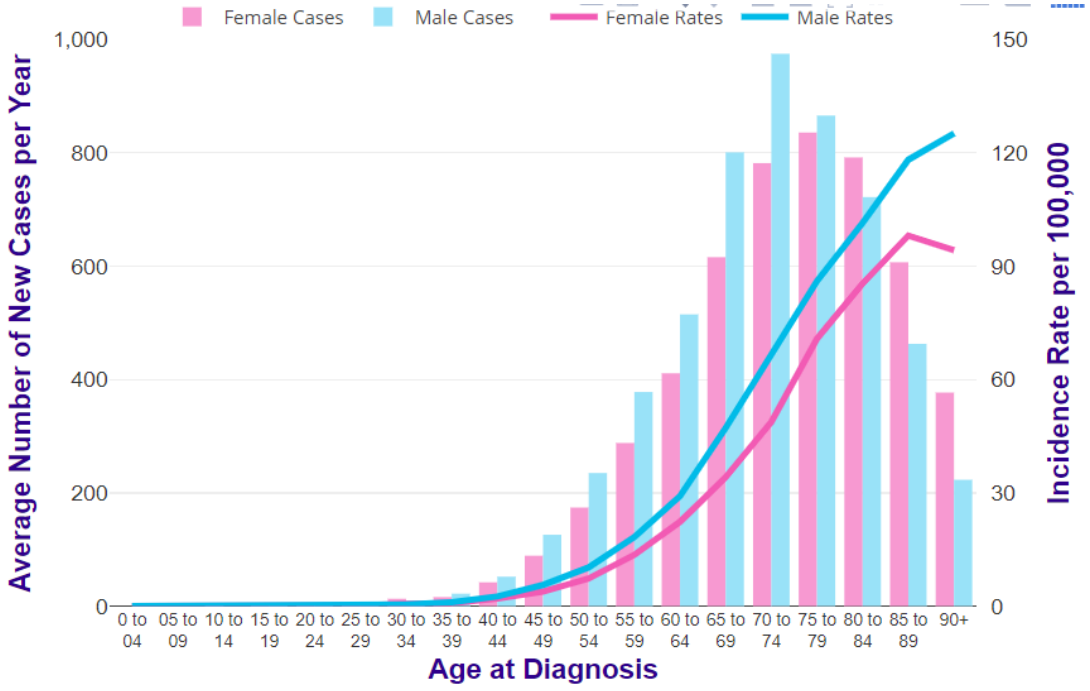


Figure 5. Pancreatic cancer (ICD10: C25), average number of new cases per year and age-specific incidence rates per 100,000 population, UK, 2016-2018. A steady rise in risk is seen from around the age of 35-39, with the greater rises seen after age 50 in both male (blue) and female (pink) cohorts. Bars represent total cases per year per age cohort, line represents incident rate. Reproduced with permissions from Cancer Research UK; Internet URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/incidence#heading-One; accessed September 2022.

PC accounts for approximately 4% of cancer-related deaths worldwide and expected to rise in incidence regardless of location, socio-economic status, age, and gender [45]. In the UK, overall mortality rates have remained unchanged but distribution by age and gender has shifted over time (Figure 5.). The overall mortality rate for men has decreased, with the

largest reduction observed in the aged 25-49 group and highest rises seen in the 50-54 age group. Women's mortality rates have increased across all age groups in over the same period of time.

The shift in incidence towards later life may be associated with higher living standards and changes in socio-economic circumstances and cultural or lifestyle choices. Ecological studies suggest that PC is associated with nations with high human development index (HDI) scores, which measures quality of health, knowledge, and standard of living (Figure 6.) [45-47]. The PC incidence and mortality in high HDI -scoring countries may be six- to sevenfold higher than lower HDI strata[45]. Individuals tend to live longer in higher HDI nations and therefore have a greater accumulated lifetime exposure risk for cancer, which may partially explain this relationship [46, 47]. However greater inequality is also associated with higher HDI scores [47], which may influence lifestyles and exposure to risk factors that have been shown to be directly linked with cancer including PC such as obesity, tobacco use, alcohol use, and diabetes [48]. All large ecological studies come with limitations. Higher HDI countries tend to have more accurate healthcare data records. Less developed nations will be prone to under-reporting, misrepresentation bias and attribution bias. Combined, the differences in strength of association may be less than what has been reported. However, the corroboration of the same findings amongst repeat studies and linkage to diseases associated with both cancer and improved HDI scores does provide support for the notion that HDI is linked to increased risk of PC, even though the reasons why are not clear.

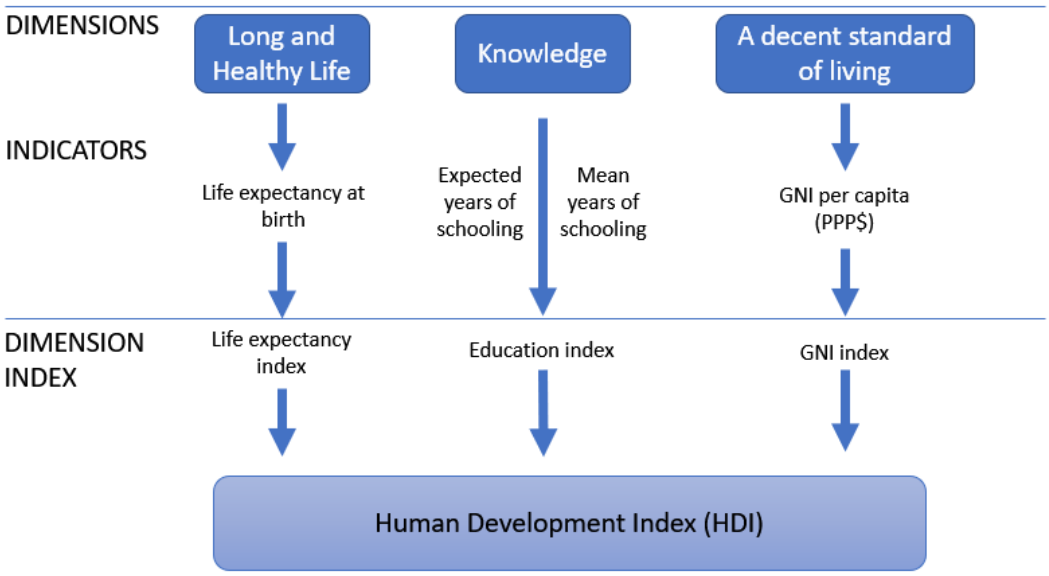


Figure 6. Human Development Index (HDI) as defined by its key assets. GNI = gross national income, PPP$ = purchasing power parity. Pancreatic cancers are associated with higher HDI scores that are

likely attributed to greater accumulated lifetime exposure risk of cancer [46, 47]. Diagram adapted from Training Material for Producing National Human Reports, March 2015; UNDP Human Development Report Office.

6.2.3.1 Early detection and clinical presentation

PDAC has a broad symptom signature consisting of multiple symptoms, with only a few considered as 'alarms' that are strongly suggestive of cancer [49]. 79% of new cases with recorded staging at diagnosis in England are diagnosed with stage 3 and 4 cancer, with only 21% found at the earlier disease stages (Figure 7.) [8].

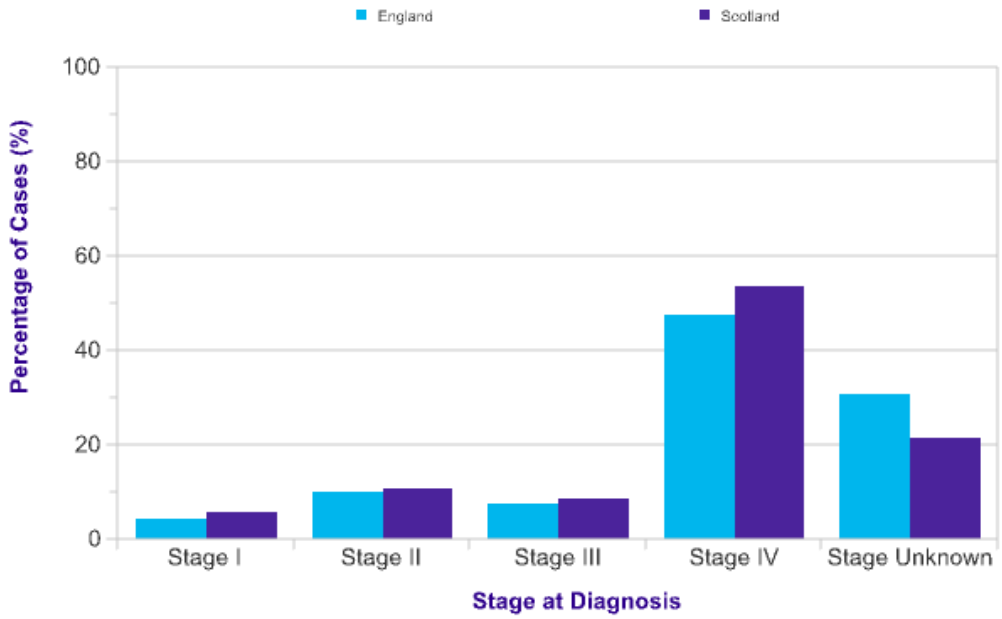


Figure 7. Overall diagnosis by stage. 69-79% of pancreatic cancer cases in England and Scotland have stage at diagnosis recorded. Of these, 68-69% of patients are diagnosed at stage IV with 79% of patients with a known stage diagnosed at stage III or IV. Reproduced with permissions from Cancer Research UK ; Internet URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/incidence#heading-Four; accessed September 2022.

One of the most profoundly impactful issues limiting effective management of PDAC is that early-stage disease commonly occurs without symptoms, or with relatively few symptoms that are non-pathognomonic for the disease at the time of their physical presentation to the individual [50]. Symptoms can also occur with an intermittent course that can be falsely reassuring to patients, leading to further delays to diagnosis [51]. The most common symptoms to present clinically are abdominal or back pain, jaundice, and weight loss [50-55].

Other less common symptoms across all stages can include nausea, vomiting, early satiety, altered stools/steatorrhea, bloating, discoloured urine, pruritis, fatigue, and gastroesophageal reflux [51, 54, 55]. Around 70% of cases present with symptoms of pain with or without jaundice or weight loss, and 10% are investigated for painless jaundice [55]. In Japan, a retrospective observational multicentre study performed over a 9-year period found that 25% (50/200) patients with early stage PDAC were detected by investigations due to abdominal pain (72%), back pain (26%), nausea (8%), diarrhoea (2%) and jaundice (2%) [55]. The remaining patients were detected through abnormalities found on routine medical check-ups and abnormalities during examination or follow-up for other disease (17% and 51% respectively). Kikuyama et al.'s study within the Japanese cohort highlighted abdominal pain as the most frequent symptom associated with subsequent follow-up investigations and diagnosis of early stage PDAC. In a prospective cohort study involving data across seven hospitals in England for individuals aged 40 years or older with suspicion of PDAC, the authors found that no initial symptoms were reported any more frequently by patients with cancer than by those with no cancer, indicating that early stage PDAC cannot be accurately detected by symptoms alone [50]. However, subsequent symptoms of jaundice, fatigue, change in bowel habit, weight loss and decreased appetite were all shown to be features more prevalent in the PDAC patient group verses those with no cancer ($p<0.0001$) [50]. The authors concluded that jaundice, weight loss, and diabetes may be key indicators signalling the progression of PDAC at certain stages of disease. They also suggested that Individuals presenting with these features atypically should be considered seriously for the possibility of an underlying PDAC diagnosis [50].

The presentation with abdominal pain in early-stage disease did not feature as a symptom associated with early stage PDAC in the English population study (48). A Thai study involving 100 patients by Kongkan et al. did report abdominal pain or discomfort as a common presenting symptom in 71% of cases identified in one hospital setting [56]. However, most of the patients were in advanced stages of the disease at the time of diagnosis, suggesting that these symptoms were not a feature significantly associated early-stage disease and comparable to the results from a multi-institutional series of 185 patients from Spain which found that most types of pain experienced were significantly associated with increasing stage [57]. It is unclear why there is a disparity between the Japanese study and other studies above. Reasons suggested for racial disparities in cancer presentation include differences in levels of access to diagnostic, screening, and treatment modalities [58]. Differences may also be observed as a result of changes in clinical epidemiology over time where routine screening

may result in overdiagnosis bias [58]. In these circumstances, routine screening may increase detection of early stage potentially lethal disease, but may also generate an increase in the detection of indolent cancers in the process. Although the overall effect is likely to be marginal, this could still give the impression that a higher proportion of cancers are detected at early stage.

The experience of cancer-associated pain is a complex multidimensional construct and a patient's ethnicity may influence the experience of pain [59]**.** A systematic review of 11 studies examining the relationship between ethnicity and pain experience of cancer patients in the United States of America identified two general themes to pain: cultural differences in barriers to pain treatment, and the experienced severity of pain itself [59]. Their study identified that Asian Americans tended to normalise pain verses Western patients. Asian Americans reported more barriers to pain treatment than Western patients but not when compared with Afro-American or Spanish-speaking Latina patients. Evidence also suggested that Asian Americans report lower levels of moderate pain compared to other ethnicities. Caucasian reference populations were a common theme amongst the papers studied in this systematic review and it was suggested that pain severity appears to be higher generally in ethnic groups verses white cancer patients in specific circumstances. The systematic review acknowledged multiple limitations including publication bias and study heterogeneity. It was unable to differentiate between first- and second-generation immigration status of individuals within the studies, and the complexity of defining ethnicity may have contributed to type 1 error in the results of each study. The omission of non-English language papers would also have limited insight into the experience of pain by patients in different ethnic groups in different countries for comparison. This paper was however able to highlight the influence of cultural differences on the way in which pain is perceived and approached. Local socioeconomic factors such ease of access to high quality health care services, and the cultural influences specific to the population within Kikuyama's study may be part of a specific multi-dimensional construct of factors responsible for the increased presentation of pain in early-stage pancreatic cancer within the Japanese cohort not seen in studies performed elsewhere.

Steatorrhea may be associated with an underlying pancreatic cancer when associated with weight loss in individuals aged 60 years or over [60]. It is defined as an increase in fat excretion in the stools which can cause overlapping clinical symptoms of diarrhoea, due to increased frequency and looseness of bowel movements. Clinically, steatorrhea is the result of fat malabsorption which can arise from different conditions such as exocrine pancreatic

insufficiency, coeliac disease, and tropical sprue [61]. With pancreatic cancer, steatorrhea may be associated with pancreatic localisation of the tumour as this symptom and cholestatic symptoms (e.g. jaundice, hypocolia, choluria, and pruritis) are significantly more common in tumours affecting the pancreatic head [57]**.** However, steatorrhea secondary to the development of exocrine pancreatic insufficiency in pancreatic cancer may be more typically indicative of advanced disease. Studies have shown that steatorrhea presents across all stages of tumour. Kikuyama's group found that 2% of individuals visiting hospital with an eventually diagnosed stage 1 pancreatic cancer included diarrhoea as part of their initial presenting symptoms [55]**.** In the Thai study by Kongkam et al., steatorrhoea was a presenting symptom in 1% of cases (n= 100) although the disease stages associated with this symptom could not be ascertained due to limitations with the data. In the Spanish study by Porta et al., steatorrhea was a presenting sign in 25.5% (n=46) of all pancreatic cancer cases. For individuals with available data on clinical symptoms at presentation and associated tumour staging, 31.1% presented with stage I tumours and 25.3% of individuals presented with stage II-IV disease. The presentation with steatorrhea across all stages of disease may be due to a mixture of site and disease progression-related factors. As the study only analysed cases with confirmed diagnoses of exocrine pancreatic cancer, the authors were unable to assess the predictive value of signs and symptoms to determine if correlations existed between steatorrhea and detection of early-stage head of pancreas lesions. Despite the significant problems and limitations associated with accurate measurement of symptoms for studies attempting predictive analysis, findings from multiple authors demonstrate the importance of recognising steatorrhea and other cholestatic signs and symptoms as part of a strategy for identifying patients with pancreatic cancer at an earlier stage.

6.2.3.2 Clinical pathways as barriers to diagnosis and treatment

The type of symptom presentation for underlying pancreatic cancer may also be associated with a large variance in diagnostic testing and subsequent delays in treatment (Figure 8.) [51, 62]. Shorter intervals from symptom-onset to diagnosis, and shorter intervals from symptom-onset to specialist referral have both shown a positive association with overall survival [62, 63]. In the context of PDAC diagnosis, non-specific abdominal pain was the most common first symptom but correlated with significantly longer delays in seeking first medical contact [51]. It has therefore been suggested that patient symptom-to-assessment intervals of less than 30 days, and a diagnostic interval of less than 60 days for symptomatic PDAC are associated with a clinically meaningful improved probability of upfront surgical treatment [51]. The diagnosis of PDAC typically arises from three different initial processes: (i) patients

seeking clinical review for symptoms; (ii) on incidental findings through the detection of abnormal results from routine medical checks, and (iii) from investigations for other disorders [55]. Thus, identifying early disease is also largely dependent on the circumstances in which the earliest suspicions of pancreatic malignancy may first arise.

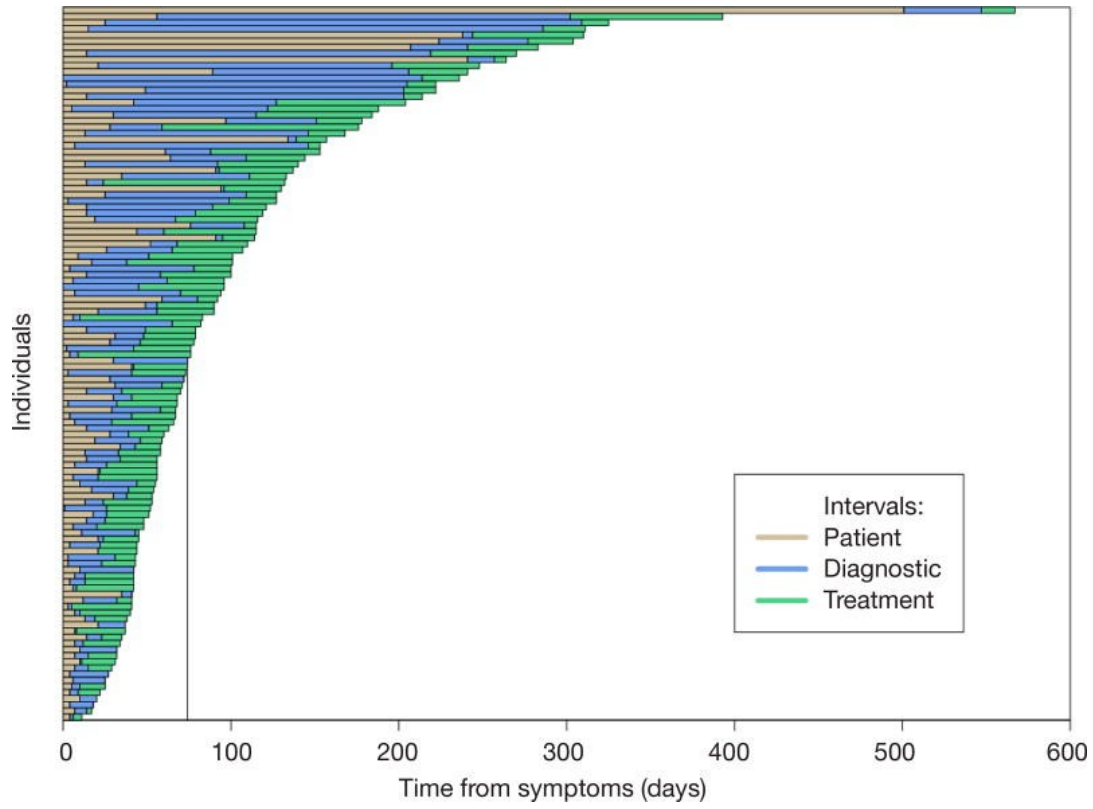


Figure 8. Patient, diagnostic, and treatment intervals for all 116 patients included in the analysis: Delays are typically highest between symptom onset and the zero-day time point for each patient represents the point at which patients first reported symptoms of pancreatic cancer. The median interval for all patients (including patient, diagnostic, and treatment intervals) is 74 days (black vertical line). Reproduced from Deshwar et al. 2018 [51], with PMC Open Access Subset content copyright licence permissions.

6.3 Diabetes as a risk factor for pancreatic cancer

PDAC has a complex inter-relationship with certain modifiable risk factors such as diabetes, smoking, diet and obesity [64-67] as well as a high-risk association with certain inheritable conditions [68]. There is general consensus amongst researchers that these associations are key to the development of effective methods to detect early stage PDAC and contribute to the development of future screening tools [12]. Long-standing diabetes, new onset diabetes, and alterations to glycaemic status still remains one of the few clinical features reliably associated with the development of PDAC.

A population-based cohort study by Rentsch et al. examining the risk of 16 cancers across the full glycaemic spectrum of HbA1c levels confirmed that a statistically significant association existed between HbA1c and pancreatic cancer (Figure 9.) [69]. The relationship persisted after adjustment for body mass index (BMI), physical activity, and underlying cardiovascular disease. Furthermore, Pancreatic cancer was the only type of cancer within the study to demonstrate a persistent risk-relationship with HbA1c after adjustments for other variables. Previous systematic reviews found no evidence of an association between HbA1c and any cancers including pancreatic cancer [70, 71]. However, as pancreatic cancer is a relatively rare cancer, low site-specific cancer events identified within smaller population cohorts of older studies and the systematic reviews based on these studies was likely to have limited their power. A more recent study involving 440,000 patients with diabetes and 26,887 detected cancer events did find a positive association between HbA1c and pancreatic cancer which was not found in other cancers [64]. However, the study cohort did not include non-diabetic individuals as a control group and were thus subject to the effects of confounding by exposure to glucose-lowering therapies and other unknown variables that could not be adjusted for [69, 72]. Compared to previous systematic reviews and other recent studies examining the risk association between HbA1c and cancer risk, the large sample size of the UKB dataset in Rentsch's study and inclusion of individuals with and without diagnosed diabetes with a long follow-up window and covariate data overcame some of the inherent weaknesses in previous work [69].

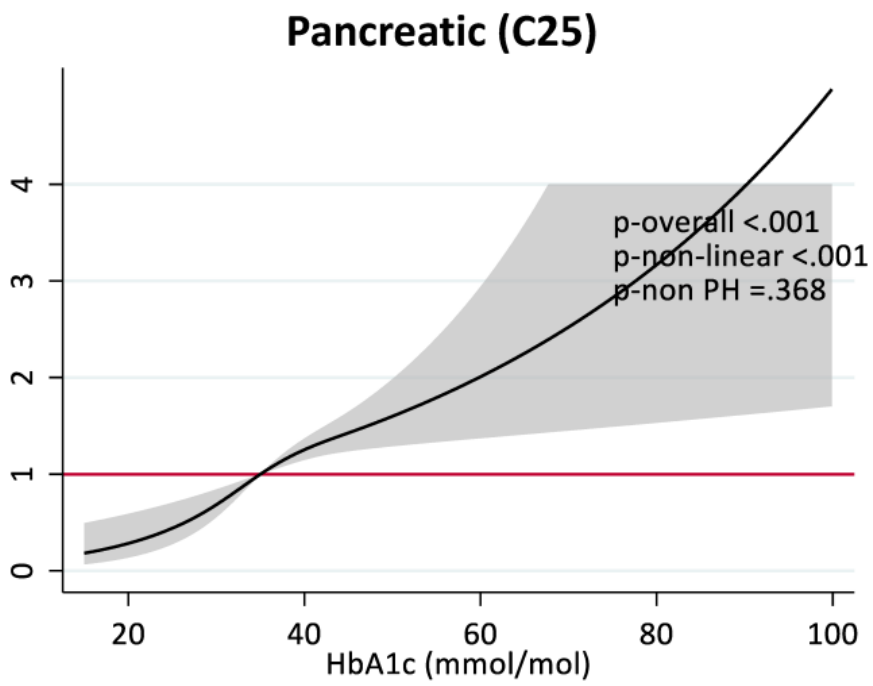


Figure 9. Associations between HbA1c and incidence of cancer outcomes, excluding participants who reported metformin exposure at baseline. For pancreatic cancer, there was relatively strong evidence that HbA1c concentration levels was associated with risk for pancreatic cancer. Low HbA1c was associated with lower cancer risk and high HbA1c was associated with elevated cancer risk. Reproduced from Rentsch et al. 2020 [69], with PMC Open Access Subset content copyright licence permissions.

In the same study by Rentch et al., comparisons of cancer risk in people with and without Type 2 Diabetes Mellitus (T2DM) at baseline showed that participants with a history of T2DM at baseline had an increased risk for pancreatic cancer (HR 1.59, 95% CI 1.21 to 2.09), uterine cancer (HR 1.52, 95% CI 1.13 to 2.04), bladder cancer (HR 1.60, 95% CI 1.21 to 2.11) and colorectal cancer (HR1.23, 95% CI 1.05 to 1.44) (Figure 10.). Together, their findings demonstrated that whilst T2DM is associated with increased risk for several different cancers, HbA1c appears to have an independent positive association with pancreatic cancer that isn't found with other cancers.

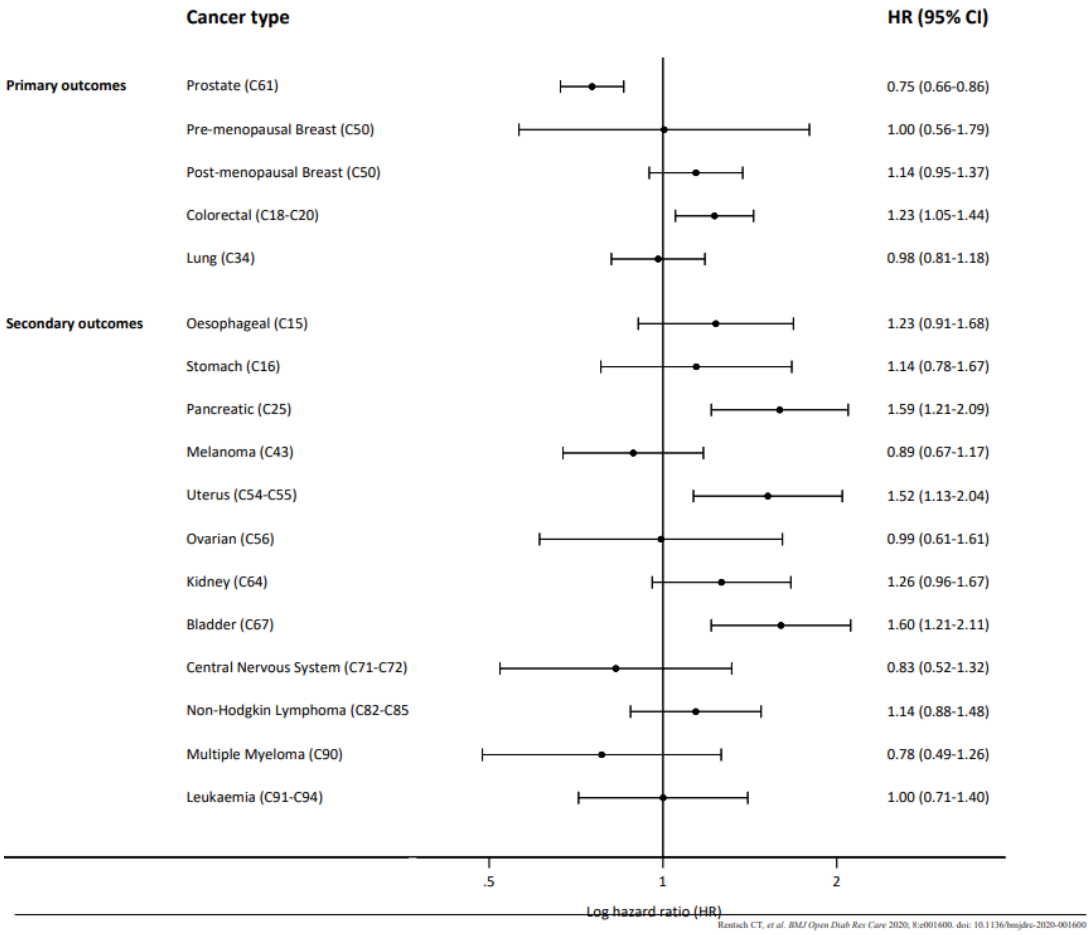


Figure 10. Associations between baseline diagnosis of type 2 diabetes (T2DM) and cancer incidence in the UK Biobank. Comparing cancer risk in people with and without T2DM at baseline showed an increased risk of colorectal, uterus, pancreatic and bladder cancers. Reproduced from Rentsch et al. 2020 [69], with PMC Open Access Subset content copyright licence permissions.

6.3.1 Type 2 diabetes: definition and diagnostic criteria

Diabetes represents a complex set of underlying metabolic and endocrine changes that occur within the body resulting in poor glucose homeostasis. Pancreatic β cells are responsible for the production of the hormone insulin which facilitates the uptake of glucose into the body and reducing blood sugar levels. In clinical practice the most described subtypes are type 1 diabetes mellitus (or insulin-dependent diabetes mellitus/T1DM/IDDM) and type 2 diabetes mellitus (Non- insulin-dependent diabetes mellitus/T2DM/NIDDM).

Type 2 diabetes is a chronic metabolic condition characterised by insulin resistance and relative insufficient pancreatic insulin production, resulting in high blood glucose levels [73]. It accounts for nearly 90% of diabetes in the UK and its development arises as a result of both

genetic susceptibility and exogenous influences over time [74]. T2DM is associated with obesity, physical inactivity, raised blood pressure, disturbed lipid levels, and median age of onset occurs in the 6th decade of life. It's association with thrombosis links diabetes with cardiovascular disease but it is also independently associated with increased risk of cardiovascular disease and related mortality [75].

The diagnosis of diabetes does not indicate the point at which impaired glucose regulation starts physiologically. More accurately, it defines a point at which the state of glucose control is considered poor enough to require regular monitoring with or without the addition of treatment interventions. The UK National Institute for Health and Care Excellence (NICE) defines T2DM based on recommendations provided by the World Health Organization (WHO). Diagnosis is confirmed by blood biomarkers such as HbA1c with a concentration level of 48 mmol/mol (6.5%) or higher, or with a fasting blood plasma glucose (FBG) level of 7.0 mmol/L or higher where using HbA1c may be inappropriate or unavailable [76]. Characteristic features of diabetes include thirst, polyuria, blurred vision, weight loss, recurrent infections, and tiredness. However, these may not be severe and may even be completely absent in some individuals. Such individuals may be unaware of their rising HbA1C or FBG levels, allowing further deterioration to occur before symptoms are investigated and a formal diagnosis is made. Certain individuals may be at high risk of developing diabetes or occasionally identified incidentally with asymptomatic early diabetes. This group of individuals falls into the 'Pre-diabetes' cohort with HbA1c concentration levels ranging from 42-47 mmol/mol (6.0-6.4%) or FBG levels between 5.5 and 6.9 mmol/L [76, 77]. In such individuals, lifestyle changes are recommended that include regimens to alter diet, reduce weight and increase exercise with further monitoring. Diagnosis of diabetes or initiation of anti-diabetic treatment is typically considered only after these attempts to modify lifestyle have failed, which can delay a formal diagnosis by months or years.

6.3.2 The relationship between diabetes and pancreatic cancer

Several large population studies and meta-analyses have demonstrated a significant link between diabetes and cancer-related deaths (Table 1) [72, 78-83]. In a recently pooled analysis of 19 Asian cohorts, more than 771,000 participants were followed for up to 21 years, showing that diabetes was associated with a 26% increase in risk of death from any cancer (Hazard ratio (HR) 1.25; 95% confidence interval (CI) 1.19 - 1.31) [84]. For individuals with diabetes associated with pancreatic cancer the association was consistently stronger if they were aged under 60 at enrolment, and after adjusting for BMI, smoking status and

alcohol consumption. The link between diabetes and pancreatic cancer is historically well recognised, with reporting of the prevalence of diabetes in pancreatic cancer patients within the range of 4-20% [85]. More recently, studies have reported this prevalence to be as high as 68% [86]. The majority of diabetes diagnoses in studies also occurs in the 2-3 years prior to pancreatic cancer diagnosis (new-onset diabetes: NODM) among cases verses controls, accounting for 52-74% of diabetes diagnoses within the PDAC cohort [85-87]. In addition, impaired fasting glucose falling below the diagnostic criteria for diabetes may be found in a further 38% of PDAC patients [88]. When individuals in this cohort are considered alongside those with clinical diabetes diagnoses, the combined prevalence may be as high as 80% [89]. With many diabetes diagnoses occurring shortly before PDAC diagnosis, evidence suggests that alongside those with a history of LSDM, a new-onset diabetes variant may arise in certain individuals as part of a pathophysiological consequence of underlying pancreatic cancer which would have otherwise remained undetected until further along the disease process.

To date, several meta-analyses of studies on the association between PDAC and diabetes have been performed [79-84, 90]. The first group to examine the temporal relationship between PDAC and diabetes in this manner reviewed 36 studies encompassing over 9,000 patients in total and concluded that a causal association between T2DM and PDAC existed. The collective findings added support to the hypothesis that T2DM is likely to be another independent modifiable risk factor alongside cigarette smoking and possibly obesity [80].

In 2015, a follow up meta-analysis was performed on 88 case-control and cohort studies, incorporating new studies with those analysed by Huxley et al. to produce a summary relative risk association that was very similar to the previous meta-analysis (Table 1) [82].

Publication	No. studies (case-control/cohort/total)	OR case-control studies (95% CI)	OR cohort studies (95% CI)	Overall RR (95% CI)
Everhart and Wright, 1995	11/9/20	1.8 (1.1-2.7)	2.1 (1.6-2.8)	-
Huxley et al., 2005	17/19/36	2.82 (1.6-1.99)	1.73 (1.59-1.88)	1.82 (1.66-1.89)
Ben et al., 2011	0/35/35	-	1.94 (1.66-2.27)	1.94 (1.66-2.27)
Batabyal et al., 2015	39/50/88	2.08 (1.87-2.32)	1.88 (1.71-2.07)	1.97 (1.78-2.18)
Zhang et al., 2018	26/0/26	3.69 (3.12-4.37)	-	3.69 (3.12-4.37)

Table 1: Aggregate table showing summary odds ratios (OR) and overall relative risk (RR) of association between diabetes and pancreatic cancer for four meta-analyses. Diabetes is shown repeatedly to be associated with a near 2-fold relative risk for PC. (Results adapted from Everhart and Wright, 1995; Huxley et al., 2005; Ben et al., 2011; and Batabyal et al., 2015.)

When diabetes duration prior to PDAC diagnosis was divided into time categories of ≤1, 1-4, 5-9 and ≥ 10 years, individual relative risk (RR) ranged from 6.69 at ≤1 year to 1.36 at 10 years (39). The negative risk association with increasing duration from diabetes diagnosis was consistently reported in each meta-analysis, with nearly doubled risk for PDAC in those with a diabetes diagnosis within 1-4 years falling to a RR of approximately >1.5 in studies beyond 10 years (Table 2) [80-82, 91].

(a)

Duration	**Huxley et al** (no. of studies; RR (95% CI)	**Ben et al** (no. of studies; RR (95% CI)	**Batabyal et al** (no. of studies; RR (95% CI)
<1 year	N/A	3; 5.38 (3.49-8.30)	3; 6.69 (3.80-11.78)
1-4 years	9; 2.05 (1.87-2.25)	5; 1.95 (1.65-2.31)	21; 1.86 (1.56-2.21)
5-9 years	9; 1.54 (1.31-1.81)	4; 1.49 (1.05-2.12)	23; 1.72 (1.47-2.00)
>10 years	7; 1.51 (1.16-1.96)	4; 1.47 (0.94-2.31)	28; 1.36 (1.19-1.55)

(b)

Duration	**Zhang et al (Studies; RR (95% CI)**
<2 years	7; 4.93 (4.18-5.82)
2-4 years	5; 1.86 (1.25-2.77)
5-10 years	4; 2.15 (1.49-3.10)
>10 years	4; 1.96 (0.71-5.44)

Table 2: Relative risk for pancreatic cancer decreases over time from point of diabetes diagnosis. (a) RR of association between DM diagnosis and PDAC with 95% confidence intervals (CI), as RR grouped by DM duration prior to PDAC diagnosis. (b) Meta-analysis of Chinese studies only, showing changes in relative risk over time that are consistent with other non-location/ethnicity-specific meta-analyses (Zhang et al., 2018)

Batabyal's group plotted diabetes duration against relative risk using multiple values, which identified a non-linear relationship existing between the two. A greater study-level risk that arose in individuals with the shortest history of diabetes likely reflected the cohort of individuals who developed NODM because of underlying pancreatic cancer pathology. Where risk dropped to a moderately increased level above the baseline for those with diabetes for 7.5 years or more, the pattern of association was attributed to the chronic effects of long-term T2DM on the pancreas that was likely to be inducing cancerous change (Figure 11.) [82].

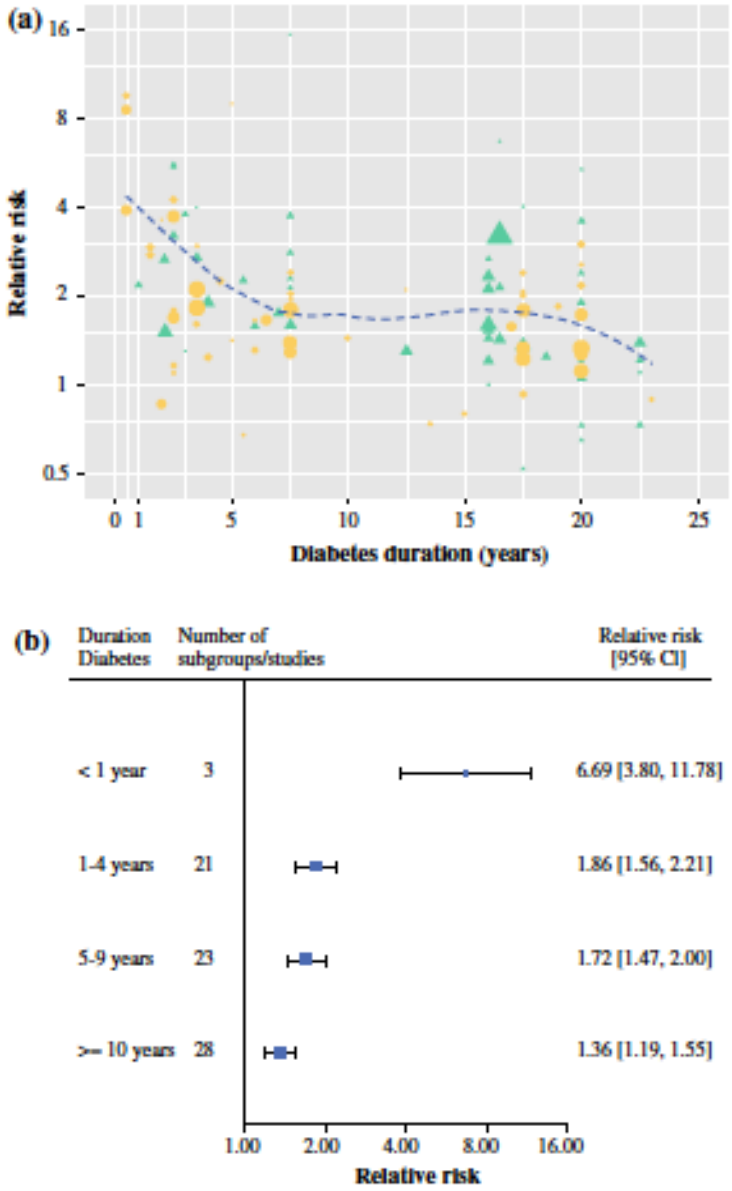


Figure 11. Study-level and population level representations of change in relative risk for pancreatic cancer by duration of diabetes. (a): Meta-regression of 122 estimates from 51 studies within the Batabyal group meta-analysis show that after an initially elevated risk for pancreatic cancer in newly diagnosed diabetic patients of less than 1 year, risk decreases before plateauing at 5 years onwards. The effect size for each study-level subgroup is plotted against the midpoint of each diabetes duration for the corresponding population. Size of symbols are proportional to the precision of estimated effect size. **Yellow Circles**: effect size from cohort studies; **Green Triangles**: effect size from case-control studies. **(b)** Individual-level RRs from 75 estimates across 31 studies showing relative risk of PDAC by duration of diabetes. Relative risk is as high as 6.69 (95%CI 3.80, 11.78), before a similar drop is seen with plateauing after the first year since diabetes diagnosis. (Reproduced from Batabyal et al., 2014. 21:7 , with PMC Open Access Subset content copyright licence permissions.)

A further meta-analysis by Song et al. also identified a moderate association between the presence of LSDM and PDAC compared to non-diabetic controls (RR = 1.64, CI 95% 1.52-1.78) (47). When the results of forty-four individual studies were pooled together by DM duration into those with diabetes for ≥5 years and ≥10 years, the RR (95% CI) dropped to 1.58 (1.42-1.75) and 1.50 (1.28-1.75) for ≥5 years and ≥10 years respectively [90].

Baseline RR for patients with long-standing DM may be lower than the RR for NODM, but remains significantly higher verses controls [80-82, 90]. This indicates that the strength of association between long-term diabetes and PDAC development is equal to or slightly weaker than that between new-onset diabetes and PDAC development [90].

The mildly elevated baseline risk for long-standing DM individuals for PDAC and the markedly elevated risk associated with NODM individuals has been demonstrated consistently amongst researchers. Such a distinct difference between these diabetes cohorts is strongly indicative of at least two distinctly different clinicopathological pathways leading to the clinical presentation of PDAC.

These meta-analyses prove useful in demonstrating the strength of association between DM and PDAC, however wide heterogeneity existed between the studies included within each meta-analysis making direct comparisons difficult. Some studies included in these meta-analyses did not identify PDAC from other rarer non-neuroendocrine subtypes. However, as PDAC represents 90% of all primary pancreatic cancer types, there is consensus that the inclusion of other non-PDAC exocrine tumours of the pancreas is unlikely to have a substantial statistical impact on the overall findings in each study. Study population sizes also varied significantly which may have influenced the strength of the relationships between diabetes and PDAC that were identified. Finally, different diabetes subtypes were not separated to exclude type 1 diabetes from type 2 diabetes in all association studies used in the meta-analyses. As with the prevalence of PDAC vs other non-NET pancreatic cancers, 80% to 90% of DM prevalence is expected to be Type 2 in nature, therefore the impact of T1DM individuals on large case-control or cohort studies was considered not to be significant by the authors.

6.3.3 Surgical Resection of pancreatic cancers and diabetes

The development of diabetes following surgical resection of the pancreas is a well-recognised complication in some patients. Patients undergoing pancreatectomy are more likely to develop NODM based on various factors that include age and pre-operative BMI [92]. Location and extent of resection seems to influence the expected risk, although studies have shown varying results regarding the extent of risk for diabetes. In a study by Kim et al, 10-24% of subjects developed DM following pancreaticodudenectomy, whilst 8-60% develop DM after distal pancreatectomy [93].

In another study, 3.6% of pancreatectomies resulted in NODM as a result of impaired function in central pancreatectomy [94]. This number increases to 8% in distal pancreatectomy. In the same study, tests showed that 23% (n=18) of patients developed diabetes immediately in relation to the surgical resection. Just over half this cohort (51%, n= 39) developed either glucose intolerance or frank diabetes following surgery during the study's follow up period. Increased percentage pancreatectomy also resulted in more chance of diabetes (P = 0.025) and BMI was also associated with the development of diabetes, with suggestion that it increases the risk for diabetes following pancreatic resection [94].

The association between pancreatectomy and development of diabetes is related to the location and extent of resection involved. Despite this, the pathology of post-pancreatectomy related diabetes may be more influenced by the preservation of insulin secretion than extent of resection [95]. In addition, NODM incidence increases with time progression [92], with an accumulative incidence of 8.9% at postop 6 months 14% at year 1, and 22.3% at year 3. This rises to 27.1% at year 5 and 35.5% at year 10.

The independent association between pancreatic resection and pancreatogenic diabetes is strong enough that a post-pancreatectomy diabetes index (PDI) has been developed and validated to predict the development of diabetes and pre-diabetes in patients as a post-operative outcome [96]. Peri-operative fasting and postprandial (OGTT, oral glucose tolerance test) plasma glucose, glycated haemoglobin, insulin and c-peptide measured in these patients identified five measures for predicting diabetes development (PDI) . These were HbA1c % at evaluation, BMI > 30kg/m2, age ≥ 6, and type of pancreatectomy. Although BMI and age were not significantly predictive in this model, the role of these variables in the development of diabetes and insulin resistance is well documented so they were kept [96]. To the knowledge of the authors, the PDI is the first index of its kind to predict development of postoperative diabetes in patients undergoing major pancreatectomy.

6.3.4 New onset diabetes prior to pancreatic cancer diagnosis

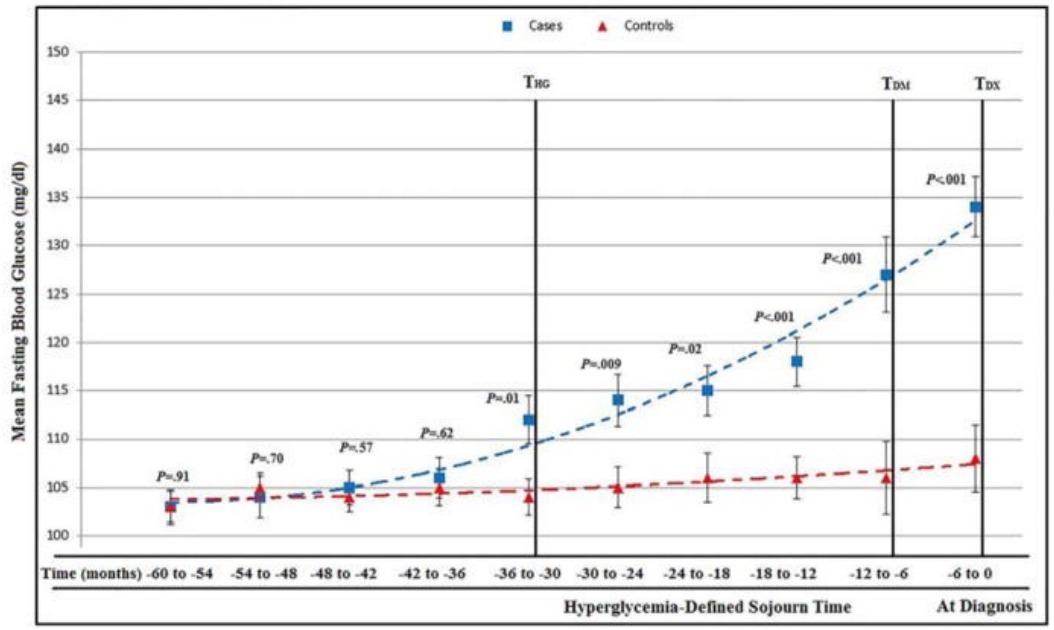
Because DM can occur within a short period of time prior to PDAC symptoms, 33-40% of patients identified as NODM with no history of diabetes prior to PDAC diagnosis may be missed clinically [97]. This could have a significant impact on its value as a time-dependent indicator of early PDAC, as a late diagnosis of diabetes may indicate a subsequently shorter interval to PDAC diagnosis. The median delay in diagnosis of diabetes in patients with NODM and without cancer-specific symptoms at DM onset is thought to be around 2.5 months, ranging from 0.25 to 14 months between cases [97]. In the same study, the median duration of NODM before the onset of cancer-specific symptoms was 6 months, with 25% (6/24) patients never receiving a DM diagnosis before PDAC symptoms emerged [97]. Because the sample size in this study was small, it was difficult to identify any pattern associated with the wide range of delays observed. Repeat studies with larger sample sizes would help understand the association between certain causes for delays and the average length of delays incurred by each cause.

Delays caused by clinician-based diagnoses alone become less relevant in cases diagnosed in later years of the study cohort, possibly due to more vigorous implementation of glycaemic-criteria diagnostic guidelines and more frequent blood glucose monitoring by clinicians to aid diagnosis. The high rate of failure associated with physician-diagnosed DM has therefore been indicated as a significant cause for the weaker associated risk between DM and PDAC reported in some older studies, where physician-based diagnosis was part of the inclusion criteria [98]. Epidemiological studies have commonly relied on the date of diabetes diagnosis to determine the temporal strength of association between the two diseases (reference). However, diabetes utilised as a categorical value in research does not reflect the time-dependent changes seen with gradual impairment of normal glucose metabolism that eventually surpasses the pre-determined disease threshold clinically defined as diabetes. This pre-diabetic period may occur over months or years with individuals clinically categorised as either ‘non-diabetic’ or ‘pre-diabetic’, depending on multiple environmental and individual factors. As PDAC is believed to take as long as 10 years to evolve from overt cancer to metastatic disease, the diabetogenic changes induced by PDAC-related effects on the pancreas and surrounding tissue may also be gradual.

Detectable signs of impaired glucose metabolism may arise before diabetes is subsequently diagnosed. In this context the presence and worsening of pre-diabetic status in an individual is also likely to be suggestive of an underlying PDAC diagnosis. Pancreatic cancer incidence

risk has been recognised to have a linear dose-response relationship with FBG across both diabetes and pre-diabetes states [99]. In a meta-analysis of 9 studies, the rate of pancreatic cancer per-unit increase in fasting blood glucose within each study was noted to be similar, showing a roughly 14% increase in rate of pancreatic cancer diagnosis associated with every 0.56 mmol/K increase in FBG [99].

This relationship has also been observed with HbA1c. A near-linear association between HbA1c levels and pancreatic cancer has been found to be statistically significant in both diabetic (chronic hyperglycaemic) and non-diabetic participants [100]. The pre-diabetic relationship between FBG and PDAC was explored further in a recent retrospective case-control study that collected individual FBG profiles at 6-monthly intervals up to 60 months before PDAC diagnosis [101]. Mean FBG for patients and controls was similar up from 60 months up until 36 months before the reference PDAC diagnosis dates (Figure 12.) [101]. By comparison, relative hyperglycaemia for those with eventual PDAC diagnosis rose from 36 months to index date. A progressive increase in FBG was observed at each time interval nearing PDAC diagnosis, with levels peaking above diabetes level criteria at 6 months before index date [101]. This study concluded that the hyperglycaemia-defined duration of pre-diagnostic progression of PDAC was estimated to be 3 to 36 months [101]. Resected tumour volumes were compared to contemporaneous FBH levels in the same study, showing that hyperglycaemia was also correlated positively with tumour size. With these findings, PDAC-induced hyperglycaemia could aid early detection of PDAC months before diabetes is typically identified.



Time (months)	-60 to –54	-54 to –48	-48 to –42	-42 to –36	-36 to –30	-30 to –24	-24 to –18	-18 to –12	-12 to –6	-6 to 0
Cases	81	72	60	61	68	71	69	90	88	159
Controls	106	122	99	125	116	128	86	123	106	123

cases: $y = 0.3561x^2 - 0.6803x + 103.83$; $R^2 = 0.9743$ controls: $y = 0.0341x^2 + 0.025x + 103.75$; $R^2 = 0.7849$

Figure 12. Temporal glycaemic profile of a population-based controls and pancreatic cancer after excluding subjects with diabetes at baseline (60 to 54 months). Patients (blue) with hyperglycaemia show a statistically different pattern of FBG from controls (red) for a mean period of 36-30 months before PDAC diagnosis. The table shows participant numbers in the case and control cohorts for each time interval. Reproduced from Sharma et al., 2018 155:2, reproduced with PMC Open Access Subset content copyright licence permissions.)

6.3.5 Diagnostic Markers for Diabetes

The recommended criteria for diagnosing diabetes have changed over the last few decades, in which time many studies on pancreatic cancer and diabetes association have taken place. Current UK guidelines recommend the use of HbA1c and FBG in diagnosing Diabetes and pre-diabetes , yet historically, oral glucose tolerance tests (OGTT) and fasting FBG were the gold standards for confirming a diagnosis [76].

In the UK, HbA1c is the preferred biomarker for confirming DM diagnosis although FBG is recommended where HbA1c is contra-indicated other disease factors in certain groups of individuals that may influence haemoglobin glycation may be present (e.g. people with acute pancreatic damage including pancreatic surgery, end-stage renal disease, people on medication that may cause hyperglycaemia such as long-term corticosteroid treatment, or people with HIV infection) [76]. Testing for impaired HbA1c is part of the risk assessment strategy recommended by the NICE guidelines. It identifies individuals who may be diabetic based on other clinical features or considered as high-risk and recommended for tests. Retrospective studies are dependent on past medical records and are therefore restricted by how information was collected at the time. A change in clinical criteria testing for DM from FBG and OGTT to HbA1c would therefore have a significant impact on the variability within studies examining data collected retrospectively during this period of transition.

Each of the three test modalities have their own advantages and disadvantages and may subject studies to bias [102]. OGTT and FBG identify a greater proportion of individuals with prediabetes or diabetes than HbA1c, but the practicality and technical challenges associated with achieving accurate and consistent repeat results puts these older tests at a disadvantage to HbA1c in the clinical setting [103]. Many epidemiological studies between PDAC and DM used FBG values either as definition criteria for DM, or in measuring DM status change over time. The main disadvantage of FBG is the requirement for subjects to fast prior to having their blood test which can be difficult to achieve objectively, making FBG difficult to reproduce on repeat tests, although this appears to be more pronounced in children verses adults [104, 105]. Studies using FBG are therefore prone to over- and under- estimating the true diabetic status of individuals.

6.3.6 HbA1c

HbA1c is a measure of the amount of glucose that binds to haemoglobin over the 120-day lifespan of the red blood cell and interprets as an average glucose level in the blood over a 90-120 day period [106]. As a result, HbA1c is better suited to detecting signs of chronic

hyperglycaemia than FBG or OGTT. HbA1c also has a stronger correlation with chronic complications of diabetes than FBG making it a better representative indicator of risk for these secondary problems such as retinopathy and cardiovascular disease [102]. Despite its clinical strengths as a monitoring tool of the body's ability to maintain normal glucose homeostasis, HbA1c also carries certain disadvantages which can limit its usefulness in terms of research. HbA1c is reliable indicator of chronic hyperglycaemia but provides limited information on the underlying pathophysiological abnormalities causing diabetes [102]. The postprandial state of the individual when glucose levels are highest, or after fasting when they are lowest, are important periods to test in the evaluation of ß-cell capacity and ß-cell function respectively [102]. The acute values detected by FBG and 2-hour OGTT glucose tests are able to indicate where the defect lies in glucose regulation which HbA1c is unable to do [102]. Because of this limitation, studies on PDAC-induced NODM that use HbA1c as the sole criteria for DM would struggle to determine how the pathophysiological changes arise in terms of pancreatic ß-cell dysfunction.

HbA1c is better at distinguishing between prediabetes and diabetes than healthy verses prediabetic individuals who may be asymptomatic [103, 106, 107]. In a population study in the USA that used HbA1c ≥ 6.5% (48mmol/mol) as the reference value for diabetes, only 30-40% of previously undiagnosed patients as diabetic were identified, compared to approximately 50% by FBG and 90% using 2-hour OGTT [103]. A pooled analysis of 96 population-based studies further confirmed that HbA1c alone did not identify a large portion of previously undiagnosed individuals compared to surveys using glucose-based testing (FBG or OGTT) [108].

Variation in haemoglobin glycation may also be influenced by the various factors associated with race, including genetics, social, and environmental factors [109]. In a Chinese study examining the accuracy of HbA1c for diagnosing DM compared to OGTT, an HbA1c reference of ≥ 6.5% gave a sensitivity and specificity of 33.2% and 93.5% respectively [110]. The optimal cut-off threshold was suggested to be 6.3% for identification of DM in this particular cohort with higher specificity and sensitivity of 56.3% and 85.5% achieved respectively [110]. Malaysia implements the use of HbA1c with a threshold of 6.3%, whereas Japan uses a cut-off of 6.5% and New Zealand's threshold is higher at 6.7% [111]. The use of HbA1c to detect early diabetes for the purpose of screening for other diseases therefore requires that the limitations of this modality be recognised and accounted for. The 'rule-in, rule-out approach' is one proposed method that considers the HbA1c variability between distinct population groups and the poor sensitivity associated with detection of early diabetes [112]. In this

model, HbA1c between 5.6% and 6.9% may be interpreted as impaired, with values of at least 5.5% considered as normal, and ≥ 7.0% as confirmed diabetes without need for secondary testing. Individuals falling within the impaired HbA1c range are then tested by FBG or OGTT to aid diagnosis, thus reducing the frequency of classification error [112].

HbA1c and plasma glucose tests currently represent the best tools available for detection of changes in glucose metabolism and both have merits in research if implemented properly. Prospective studies examining the glycaemic changes associated with PDAC progression should therefore consider the inclusion of multiple complementary glucose blood biochemistry tests for optimal detection and monitoring of hyperglycaemic changes within the pre-diabetic cohort.

6.3.7 NODM as a 'first-sieve' indicator of underlying pancreatic cancer

The possibility of identifying NODM as a risk factor for PDAC within a screening model was suggested by Chari et al. based on their findings that approximately 1% of diabetics aged over 50 years old would develop PDAC within 3 years [113]. Unfortunately, DM is relatively common in the general population and isolating pancreatic cancer-induced NODM from other forms of diabetes requires tools to refine this group of high-risk individuals from those who develop DM without pancreatic cancer [113-115].

In addition to T2DM, obesity, dyslipidaemia, and insulin resistance are part of a cluster of metabolic conditions that are linked to an increased risk of PDAC [116]. According to WHO definitions, insulin resistance plus the presence of any two of: obesity (body mass index; BMI >30 kg m^{-2}), hypertension (BP≥140/90mmHg), dyslipidaemia (Triglycerides >150 mg dL^{-1}; High density lipoprotein (HDL) cholesterol <35 mg dL^{-1} (male) or <39 mg dL^{-1} (female)) are key risk factors that defines metabolic syndrome (MetS) [117].

A meta-analysis of 9 studies found an association between metabolic syndrome and PDAC for women (RR 1.58; 95% CI 1.35-1.84), which was less significant for men (RR 1.20; 95%CI 0.80-1.80). The mechanisms linking MetS to PDAC are not fully understood, however it may be a surrogate marker for other risk factors such as decreased physical activity, consumption of high-calorie dense foods, high dietary fat intake, low fibre intake, and oxidative stress [118].

Fatty infiltration in the pancreas is positively correlated with high BMI or obesity, and prevalence of diabetes mellitus, which are well-known risk factors of pancreatic cancer [119]. In a study by Hori et al., non-alcoholic fatty pancreas disease (NAFPD) in hamsters with fatty

infiltration as a feature of the pancreas demonstrated a significantly higher risk association between this feature and risk of developing PDAC (OR of 6.1; P<0.001) after adjusting for BMI and DM prevalence [119]. Despite this early finding, the true relationship of NAFPD with all the criterion for MetS, including obesity and diabetes in both animal models and humans remains to be fully elucidated. Further studies are also needed on the prevalence of fatty pancreas in the general population and within populations who develop incidental PDAC will be required to aid clarification on the causal relationship between fatty pancreas and PDAC [120].

Obesity is recognised as a major independent risk factor for PDAC [121]. A meta-analysis of 16 studies across 3 different continents (America, Europe, and Asia), identified a potential non-linear association between BMI and risk of PDAC [121]. Although the exact mechanisms are unclear, obesity-associated increased adipokine production, inflammation, insulin resistance and altered intestinal microbiota may play an important role in the hastening of PDAC onset [122]. New-onset type 2 DM (DM duration > 36 months) is typically associated with obesity and weight gain, however evidence also suggests that both PDAC-associated NODM and long-standing T2DM are associated with weight loss or decreased BMI prior to PDAC diagnosis [88, 123-125].

PDAC-NODM patients (DM duration ≤ 36 months) undergo more pronounced weight loss (>10% of original weight) verses LSDM patients [88, 126] and appear to require escalation of their antidiabetic therapy more frequently [127]. In one key study, 59% of PDAC-DM subjects were identified with paradoxical weight loss up to between 12 and 15 months before diabetes onset [128]. The weight loss couldn't otherwise be linked to other signs of cachexia such as anorexia, fatigue, or other cancer-specific symptoms of abdominal pain, back pain and jaundice. BMI and family history of DM have been shown previously to be poor differentiators of PDAC-NODM and T2DM [85] and PDAC symptoms are poorly correlated temporally to PDAC stage at presentation, suggesting that weight loss preceding DM diagnosis may be an important clue in the detection and pathogenesis of PDAC-NODM [128].

The moderate association between T2DM and PDAC found in previous studies and meta-analyses may actually be confounded. Evidence from a mendelian randomization (MR) study that found increases in BMI and fasting insulin are causally associated with increased risk of PDAC [67]. No causal relationship was observed for T2DM in this study, although the authors suggested that the hyperinsulinemia found supports the existence of reverse causality with diabetes manifesting as a result of pancreatic cancer and that the hyperinsulinemia induced

by PDAC was being mistaken for early T2DM [67]. The study assumed a linear relationship between genetic instruments and the risk factor of interest, as well as a log-linear association between risk factors and PDAC risk and acknowledged this limitation. No causal association was found between dyslipidaemia and PDAC in the same MR study, however an earlier population study identified the risk of PDAC to be 40% higher than the control population [116]. When combined with NODM of <1yr, the relative risk elevation was even greater (RR 2.512; 95% CI 1.169-5.398) [116]. Such differing outcomes may be due to the potential violation of assumptions of linearity and pleiotropy (one gene influencing two or more phenotypical traits that are seemingly unrelated) [67], but more studies are required to confirm the relationship one way or the other. There is some evidence to suggest that dietary cholesterol may be associated with risk of pancreatic cancer [125]. However, a meta-analysis of these studies concluded that more results are required to strengthen current evidence(76). Furthermore, the meta-analysis on dietary cholesterol examined serum total cholesterol but not triglyceride, HDL or LDL levels which might link dietary cholesterol with specific types of dyslipidaemia profiles associated with other metabolic syndrome risk factors and PDAC.

NODM in pancreatic cancer has been given various definitions and taken to represent diabetes diagnoses made anywhere within 1, 2 or 3 years prior to PDAC diagnosis. Despite a lack of consensus between investigators, the pooled RRs for studies incorporated into the PDAC-DM association meta-analyses have repeatedly shown a marked increase in RR from approximately 1.5 (baseline for Long-standing diabetes (LSDM) RR) to 2 at 1-4 years duration of DM to index that does not occur at any point earlier in the association timelines [80-82]. At <1 year to PDAC diagnosis, it is difficult to rule out other confounders and there is likely to be significant detection bias for diabetes due to investigations for other symptoms at this stage. For example, a recent case-control study of patients with PDAC has shown that a marked relative rise in FBG levels from month 36 to index (PDAC diagnosis date) was associated with time to diagnosis, tumour volume and grade of PDAC [101]. At 36 months to index, paradoxical weight loss has also been observed in 59% individuals with NODM diagnosis for between 12 and 15 months prior [128]. Based on these and other studies, it appears that that the 3-year lead-in to PDAC diagnosis may represent a window of opportunity to detect hyperglycaemic changes associated with PDAC-DM.

6.3.8 T3cDM – pancreatogenic diabetes

Reverse causality may explain why the prevalence of diabetes is so high in the cohorts of individuals with NODM who develop pancreatic cancer. Diabetes in some individuals may in fact be due to a combination of both direct and indirect effects of the development of pancreatic cancer, rather than a contributing factor in the development of this malignancy. The development of NODM in pancreatic cancer is thus likely to follow a different pathological pathway to that of either Type-1 or Type-2 Diabetes.

In the last decade there has been an increased focus on the pathophysiology of secondary diabetes of the exocrine pancreas, also referred to as Type 3c Diabetes (T3cDM). Historically, diseases of the pancreas resulting in diabetes have been described within the literature as pancreatogenic or pancreatogenous diabetes mellitus, and T3cDM [127]. In order to standardise reporting of this diabetes subtype, the American Diabetes Association (ANA) has recently suggested that Diabetes of the Exocrine Pancreas should be the agreed nomenclature to unify the various definitions under a single term [129]. However, T3cDM is still a term commonly used in scientific literature to describe this subtype and also used by the British Association of Diabetes and Pancreatic Cancer UK. For consistency, we will continue to refer to this subtype as T3cDM.

T3cDM can arise from any process that diffusely injures the pancreas such as inflammation, infection, trauma, neoplasia, or surgical resection. Clinical conditions associated with T3cDM include acute, relapsing and chronic pancreatitis, pancreatic cancer, cystic fibrosis and haemochromatosis [130]. These processes disrupt the global architecture or physiology of the pancreas, often resulting in both exocrine and endocrine dysfunction [127, 131, 132].

6.3.8.1 Misdiagnosis of T3DM – a hidden classification

Systematically identifying cohorts of diabetic individuals in primary care with preceding pancreatic disease remains a novel exercise and is not routinely addressed in primary care. As a result, the comparative incidence, and precise clinical characteristics of T3cDM remains unclear. In 2013 Ewald *et al.* highlighted the importance of missing T3cDM diagnoses clinically and found that there was sparse data on T3cDM at the time [131]. Based on their study, approximately 9% of all diabetes cases are thought to be type 3c in nature, and nearly 80% of cases were linked to a history of chronic pancreatitis [131]. Ewald's team also found that most T3cDM cases were frequently mis-classified initially as T2DM due to a lack of awareness of this subtype and the challenges of differentiating between the two [131].

Few studies have attempted to examine the true proportion of T3cDM in the population. In one study using retrospectively collected primary care data from the UK Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) between 2005 and 2016, 559 individuals with NODM were identified in a cohort of 31,789 adults with a history of pancreatic disease (approx. 1.8%) [132]. Within this cohort, adult-onset type 2 diabetes had the highest incidence of 142 per 100,000 person years. After diabetes re-classification, adults with preceding pancreatic disease were identified as T3cDM (1.8%) - a higher proportion than that for T1DM (1.1%) [132]. Although the proportion of T3cDM cases was still relatively small, only 2.7% of these cases were correctly labelled by clinicians with 87.8% of T3cDM cases misdiagnosed as T2DM, and 7.7% as T1DM [132].

In Western diabetic populations, T3cDM may account for 5-10% of all individuals with diabetes, of which 73% of diagnoses are due to chronic pancreatitis [133]. For individuals with diseases of the exocrine pancreas generally (acute or chronic pancreatitis, or pancreatic cancer) the prevalence of diabetes is approximately 0.11% [134]. Patients with inflamed and fibrous pancreas also have a high risk of developing PDAC and after chronic pancreatitis, PDAC is the second most common cause of T3cDM, responsible for 8% of all cases [131].

6.3.8.2 Pathophysiology of T3cDM

PDAC is one of the strongest and most consistent diabetogenic forces known to humans and destabilises glucose homeostasis in nearly all patients in whom it occurs [135]. As PDAC is strongly correlated with the development of T3cDM, the correct diagnosis of this diabetes subtype would allow appropriate investigation and treatment for these patients [136]. On diagnosis of NODM, distinguishing between T3cDM and T2DM is important as both are risk factors for PDAC, with T3cDM possibly representing an even higher at-risk group of individuals within the NODM cohort.

The general pathophysiology of T3cDM is involves pancreatic inflammation and further irreversible fibrosis of islet cells that progresses to islet cell loss [137]. Damage to β-cell mass and polypeptide secreting cells occurs in the early phase of the disease [137]. Altered glucose metabolism due to these early signs of endocrine insufficiency begins as asymptomatic or mild hyperglycaemia. Over time, this leads to higher detectable HbA1c concentration levels and shorter interval to insulin dependence by patients, compared to those with T2DM [127].

6.3.9 Antidiabetic medication and pancreatic cancer

Medication used in the treatment of diabetes can affect factors that mediate the association between diabetes and pancreatic cancer. Both insulin and oral antidiabetic medications (OAM) have been suggested to directly affect key factors mediating the association between T2DM and PDAC and may influence the development, progression and outcome of PDAC as a result [138].

From a research standpoint, deciphering associations of antidiabetic medications with cancer can clarify mechanisms underlying the relationship between diabetes and cancer, which are well established but not yet totally understood [139]. Revealing the relationship between antidiabetic medications and cancer can also help distinguish the roles of hyperglycaemia and hyperinsulinaemia in the association between diabetes and cancer [139]. Diabetes and cancer have metabolic profiles that are relevant to both diseases, including obesity, insulin resistance, hyperinsulinaemia, oxidative stress, chronic inflammation, and alterations in sex hormones [140]. Certain risk factors related to lifestyle as well as genetic susceptibilities are also associated with both diseases. This raises the possibility that OAM treatment may influence the development progression of certain cancers in addition to diabetes. Studying associations between OAM and cancer risk is subject to the same challenges and biases that arise when investigating associations between diabetes mellitus and cancer risk [139]. These include relatively long latency periods of diseases, competing comorbidities, detection bias, and the possibility of reverse causality. The different pathophysiology of various cancer types also means associations should be measured in the context of each cancer subtype and not as a collective entity.

The way diabetes medications are used for diabetes management also makes it difficult to separate associations with cancer relating to diabetes from those relating to the medications. With pancreatic cancer, the relatively short period of time between presentation of symptoms for the disease and confirmation of cancer diagnosis is at odds with the relatively long period of time taken for T2DM to develop and progress. The presence of various types of diabetes in the population adds further complexity to the understanding of underlying pathology when observed through changes made to medication treatment. In certain individuals, NODM emerging just prior or concomitant with the diagnosis of pancreatic cancer and rapid deterioration in diabetes control in others further supports the notion that a bi-directional relationship exists between diabetes and pancreatic cancer. In addition to biochemical diagnostic measures confirming diabetes diagnosis categorically, the

pattern of changes made to antidiabetic therapy in the form of continuous time data may play an important role in the detection of underlying malignancy.

Untangling the associations between diabetes, OAMs and pancreatic cancer requires quantitative assessment and monitoring of drug exposure including dose, duration, and continuity. This depth of information required to comprehensively analyse drug usage is often challenging to obtain and not always performed. The time-varying nature of progressively worsening hyperglycaemia and initiation of treatments is also one example of some of the challenges facing the investigation of associations between antidiabetic medications and cancer risk.

Time-lag bias and confounding by indication arises in the study of associations between antidiabetic medication and pancreatic cancer where the severity of diabetes itself is the confounder. Diet-controlled diabetes and diet-controlled pre-diabetes may influence the severity and progression of underlying pathophysiological processes responsible, depending on how successful individuals are in controlling the disease through non-medication treatment alone. Likewise, the timing of introducing or removing different antidiabetic medications in response to poorly controlled diabetes may influence the progression of an undiagnosed malignancy such as pancreatic cancer [139]. With diabetes, different medications are introduced based on the stage of diabetes and response to previous treatment. Time-lag bias fails to consider the stage of diabetes associated with patterns in medication use over time as well as the impact prior and current treatments have on underlying disease. To control for this, studies examining the relationship between diabetes medication and pancreatic cancer should aim to only compare patients with the same stage of diabetes with one another [141].

Immortal time bias describes when individuals stratified into cohorts by the type of medication they used have failed to be recognised as being on a different medication or starting a different medication at an earlier point in time. In one example identified in a study by Suissa and Azoulay [141], a patient initiated and continued treatment with a sulfonylurea and subsequently switched to or added metformin. However, they were classified as a metformin user during the entire follow-up. The time between entry into the cohort and the first metformin prescription became immortal because the subject had to survive to receive this first metformin prescription and was misclassified as exposed to metformin when in fact they were exposed to sulfonylurea, leading to immortal time bias.

Confounding by indication might occur where individuals with diabetes may be started on specific antidiabetic medication due to diabetes that is more difficult to control. These same individuals may also be more susceptible to complications of diabetes or adverse events such as developing cancer.

The timing of introducing various antidiabetic medications would also influence the strength of their association with an outcome measure such as pancreatic cancer. Studies combining insulin with OAMs as a singular category may fail to consider the stepwise pattern of treatment that occurs in clinical practice. Type 2 diabetics may typically start OAMs before insulin therapy, whereas Type 1 diabetics would start on insulin therapy and consider the addition of OAMs as an adjunct where indicated [142].

Reverse causality in the study of associations between diabetes and pancreatic cancer suggests that different patterns of antidiabetic medication use might also help to differentiate the cause for different disorders of glycaemia and identify specific drug-response patterns associated with a possible underlying pancreatic cancer. In some cases, an adult newly diagnosed with diabetes may atypically transition from OAM to Insulin therapy over a short period of time, thus suggesting that an underlying pathophysiological process is more malignant than first indicated. Dose-response relationships are another cause for bias in observational studies that attempt to explore the influence of medication on disease, either as a confounder or an independent variable. Some studies will stratify people according to cumulative dose or cumulative time under treatment. This form of bias arises because the stratification occurs while these variables are highly associated with disease severity and disease association. To mitigate all these potential biases, a time-dependent statistical approach that both adjusts for disease duration and accounts for the timing of initiation of new medication may help to overcome some of the biases that arise in either context [143].

6.3.9.1 Insulin

A link between hyperinsulinemia has been found across multiple cancers including breast cancer, but it may be particularly strong with pancreatic cancer [144]. Both experimental and epidemiological studies indicate that insulin use is associated with an increased risk of PDAC. There are at least several hypothetical mechanisms linking diabetes with an increased risk of pancreatic cancer [145]. The role of hyperinsulinemia appears to be important to the progression of cancer and has been documented in experimental studies as well as in epidemiological observations [145]. As far back as the 1970s, insulin has been demonstrated to increase DNA synthesis in breast cancer cell cultures [146]. The administration of insulin

to rats with breast tumours has also been observed to increase tumour mass at a faster rate compared to controls, with the greatest acceleration seen when insulin and glucose administration were combined [147]. More recently, studies have demonstrated that risk of cancer incidence and mortality are associated with high levels of endogenous insulin. In an experimental study by Zhang et al., diet-induced increase in circulating insulin was found to play a causal role in PDAC initiation, independent of high BMI [148]. Gene manipulated mice with a reduced ability to produce endogenous insulin without impairing glucose homeostasis had significantly fewer high grade PanINs over 1 year of monitoring verses controls, despite being fed with a high-fat diet which is known to promote hyperinsulinemia. A recent prospective cohort study observing the relationship between hyperinsulinaemia and cancer diagnoses found that individuals with hyperinsulinaemia and with no history of diabetes were at significantly higher risk of developing cancer vs the controls (adjusted HR 2.04, 95% CI 1.24-3.34, p = 0.005), independent of BMI [149].

A pooled analysis on 8305 cases of PDAC across 15 case-control studies within the Pancreatic Cancer Case-Control Consortium (PanC4) examined the relationship between diabetic individuals on insulin therapy and development of PC [150]. Insulin use was associated with a significantly higher risk for PC (OR 2.66) vs diabetics not using insulin. The odds ratio was even higher for shorter insulin use (OR 5.60 for less than 5 years). In this study, the increased pancreatic risk seen in the cohorts with duration of exposure to insulin therapy of more than 5 years since diagnosis of diabetes strongly supports a causal role of diabetes on pancreatic cancer.

In 2017, a Mendelian randomization study also found strong evidence that genetically increased fasting insulin levels were causally associated with an increased risk of pancreatic cancer [67]. In this study using genome-wide data from a cohort of 7110 pancreatic cancer patients and 7264 control subjects identified within the PanC4 and Pancreatic Cancer Cohort Consortium (PanScan) datasets, the odds ratio for risk of developing pancreatic cancer was 1.66, 95% CI = 1.05 to 2.63, per SD [44.4 pmol/L]) [67]. However, the effect of fasting insulin also differed significantly by sex, with men showing an odds ratio estimate of 2.59 (95% CI 1.39 to 4.80) verses 0.94 for women (95% CI 0.48 to 1.85).

6.3.9.2 Biguanides

Biguanides, and in particular Metformin hydrochloride are the recommended initial choice of antidiabetic medication treatment for patients with T2DM in the UK [151]. Metformin is also recommended as part of dual- or multi-drug combination therapy where tolerated

[151]. Metformin is one of the most commonly used forms of Biguanide drugs in clinical practice, and its relationship with cancer has been widely studied [152]. The blood glucose-lowering actions of metformin occurs via: (a) a hepatic effect: improving hepatic insulin resistance and reducing gluconeogenesis, therefore reducing hepatic glucose output; (b) a muscle effect: increased adenosine monophosphate (AMP) kinase activity to increase glucose uptake in skeletal muscle and fat; and (c) an intestinal effect: by increasing the circulating concentration of glucagon-like peptide 1 (GLP1) which in turn increases insulin secretion and glucose uptake, slows gastric emptying, and modulates appetite [153].

The hypoglycaemic effect of metformin may play a role in reducing pancreatic carcinogenesis. Preclinical animal studies have demonstrated that metformin may have a role in preventing or slowing down the rate of carcinogenesis brought on by metabolic stress induced by high-fat diets. For example, Metformin in drinking water has been shown to prevent pancreatic carcinogenesis in hamsters on a high-fat diet [154]. Metformin may also have a role in the suppression of PC growth in diet-induced obese/pre-diabetic mice [155]. It is thought to do this by blunting tumoral activation of mammalian target of rapamycin (mTOR), which is involved in the regulation of protein translocation [155].

A study by Chen et al. on genetically engineered mice models with oncogenic Kras mutations (KC mice model) found that intake of metformin delayed pancreatic tumorigenesis represented by decreased percentage of early lesions (acinar-to-ductal metaplasia and mouse PanIN (mPanIN) 1, and late mouse PanIN lesions (mPanIN2 and mPanIN3) [156]. Metformin was also found to diminish chronic pancreatitis-mediated development of early lesions within the same study.

The majority of studies on the effect of Metformin on risk for PC have shown that subjects have a survival benefit on this medication [157] [158]. However, the outcomes of different meta-analyses of the published literature are mixed and the role of Metformin in PC development remains unclear. A meta-analysis by Wang et al. showed that Metformin is protective with a reduction in incidence of PC by up to 37% (n=13 studies, RR 0.63, 95% CI 0.46-0.86, p=0.003) [159]. Another meta-analysis performed by Zhang et al. demonstrated a 46% reduction in risk based on seven studies which was non-statistically significant [91]. After restricting the inclusion criteria to studies with only diabetes patients (n=4 studies), Zhang et al. only found a neutral role for Metformin, leading them to speculate that metformin cannot decrease the incidence of PC [91]. Similar conclusions were drawn by an

older meta-analysis by Singh et al. who found no significant association between Metformin use and risk of developing PC (n=4 studies, OR 0.76, 95% CI 0.57-1.03, P=0.073) [160].

In a more recently published systematic review, a meta-analysis of 24 articles from different countries including more than 2 million subjects found that compared with no use of Metformin, the use of Metformin could reduce the risk of PC in patients with T2DM (OR = 0.82, 95% CI (0.69, 0.98) [152]. Higher study participant numbers and the high quality of their study enhanced the statistical power of their data analysis and provided more reliable estimates than previous studies. However, the remarkable heterogeneity found in previous meta-analyses, which is a well-recognised limitation of these types of studies, was still a significant limitation highlighted in this study. Hu's group also attempted to examine the relationship between duration of diabetes and occurrence of PC in more detail by identifying whether study subjects were newly diagnosed with diabetes. Unfortunately, this investigation yielded inconsistent results and it was recommended that future studies should aim to include longitudinal data for detailed exploration of the dose-response relationships.

6.3.9.3 Sulfonylureas

Sulfonylureas (SUs), alongside thiazolidinediones are some of the most commonly used antidiabetic agents after biguanides. However, concerns remain over their safety due to the results of some studies indicating that their use is associated with an increase in risk for certain cancers. A historical systematic review found that SUs were associated with a 70% increase in the odds of PC (n=8 studies; adjusted OR 1.70 95% CI 1.27-2.28, P<0.001) [160]. Despite the apparent strength of these findings, the authors highlighted that considerable heterogeneity was likely related to confounding by indication and reverse causality [160]. Other studies around the same period identified similar risk outcomes with use of SUs [159, 161].

A more recent Korean population-based study of the effects of antidiabetic medications on the risk of pancreatic cancer concluded that subjects on sulfonylurea showed an increased risk for pancreatic cancer compared to subjects with no drug exposure (HR 1.73 95% CI 1.57 * 1.91) ([162]. In another study, the administration of pioglitazone, insulin and its analogues to the diabetic patients were found to be associated with increased risk of pancreatic cancer by 45% [163]. A separate retrospective cohort study within a municipal district of China found that T2DM patients who were new users of sulfonylurea were found not to be associated with a greater risk of pancreatic cancer verses those using metformin only (HR 0.94; 95% CI 0.46-1.96) [164]. However, the case numbers (T2DM + sulfonylurea treatment

+ PC) were limited in this study (n=18; Incidence = 33/100,000 person years) and the authors noted that extrapolating their findings to other populations should be made with caution.

6.3.9.4 Incretins

Incretin-based therapies like glucagon-like peptide-1 receptor agonists (GLP-1a) and dipeptidyl peptidase-4 inhibitors (DPP-4i) help maintain glycaemic control in patients with T2DM with additional systemic benefits and little risk of hypoglycaemia [165]. However, overstimulation of the GLP-1 receptor by incretin-based therapies is also speculated to increase risk for the development of pancreatic cancer either directly or indirectly from inflammation of the pancreas.

Results regarding the association between DPP-4i treatment and pancreatic cancer are mixed. Earlier meta-analyses found DPP-4i to be linked with pancreatitis but there was no evidence of a statistically significant increased risk for pancreatic cancer [166, 167]. One meta-analysis by Azoulay et al. examining 157 trials which reported pancreatic cancer found no associated increased risk with DPP-4i use across different types of DPP-4i molecules, suggesting that this is a drug class effect [168]. These meta-analyses of studies published between 2005-2017 acknowledged several limitations including short follow-up durations, reporting bias and small numbers of PC cases. A more recent study published after these meta-analyses examined the relationship between 10,218 new users of DPP-4i and pancreatic cancer within a cohort of 33,208 subjects who were newly diagnosed with T2DM and started on antidiabetic drug treatments [169]. In this Korean nationwide population-based cohort study, DPP-4i significantly increased the risks of pancreatitis (adjusted HR 1.24, 95% CI 1.01-1.52; P = 0.037) and pancreatic cancer (adjusted HR 1.81, 95% CI 1.16-2.82; P = 0.0009). The risk of pancreatic cancer was generally consistent in the first 12 months after the initial prescription of antidiabetic medication, without showing an increased trend according to exposure duration. Incidentally, the authors recognised that subjects were selected for inclusion not only for being newly diagnosed with T2DM but also because they were taking antihyperglycemic drugs due to poorly controlled hyperglycaemia. In this context the results of this study were likely to be a result of reverse causality with an underlying pancreatic cancer driving deterioration in glycaemic control. Such changes were detected as changes to clinical status in the form of attempts to manage worsening diabetes via the introduction of antidiabetic medication or additional OAMs to pre-existing treatment.

In the Korean population-based cohort study by Lee et al., DPP-4i exposure was associated with a decreased risk of future PC (HR = 0.57, 95% CI 0.51-0.64) [162]. Furthermore, subjects

with dual exposure to metformin plus PDD-4i were at lower risk of PC compared with metformin-only treated subjects. The synergistic effect appears to reflect the results of a pre-clinical study suggesting that the anti-tumour effect of DPP-4i is due to downregulation of autophagy, increased apoptosis, and cell cycle arrest [170]. However, DPP-4is appear to have heterogeneous effects on cancer cells depending on tumour types, stages, microenvironment, and host condition [165]. This means they can potentially impact cancer-bearing T2DM patients either favourably or unfavourably [171]. Further studies on the synergistic effect of metformin + DPP-4i are required to understand the potential benefits.

6.3.9.5 Thiazolidinediones

The thiazolidinediones (TZD) are considered to preserve β-cell function indirectly by acting as insulin sensitizers and inducers of insulin secretion [72]. Some studies suggest that Thiazolidinediones inhibit the proliferation and metastasis of human PC cells [172], while it has also been suggested by both older and more recent meta-analyses that TZDs do not have a protective or harmful effect on overall incidence of pancreatic cancer [160, 173]. In the most recent of these, TZD vs non-use of TZDs was associated with an OR of 1.13 (95% CI 0.73-1.75) with heterogeneity observed across the seven included studies ($p < 0.01$). A few observational studies have also concluded that pioglitazone may increase the risk of pancreatic cancer[174, 175]. In a separate study exploring the direct targets of pioglitazone and associated genes with this drug, two of these genes - TGFB1 and RELA – were shown to be strongly amplified in pancreatic cancer, suggesting that pioglitazone may be a causal risk factor. There is still much debate about the effects of TZDs on risk for cancer and further investigation is required if the molecular mechanisms linking diabetes, TZDs and PC are to be clarified.

6.3.9.6 Antidiabetic medication, clinical application and research

T2DM can increase the risk of PC and certain antidiabetic medications can modify this risk [176]. On the other hand, NODM signalling the progression of underlying PC may be observed clinically as a sudden need for escalation in antidiabetic treatment for the treatment of resistant dysglycaemia. Metformin as a first-line OAM is frequently introduced to treatment regimens for T2DM at an earlier, milder stage. It also appears to have a beneficial effect on risk for PC although the significance of this effect continues to be debated. By contrast, the risk of PC is elevated in T2DM insulin users. However, insulin tends to be prescribed in patients with more advanced T2DM and the elevated risk for shorter-term insulin use seen in these individuals may in fact be explained by reverse causation [150]

[88]. Other OAMs are typically introduced as second line therapies, either as an alternative to Metformin or as part of multidrug regimens. Large population datasets are required to achieve adequate case-control numbers and specific details about drug regimens and variation in dosing and time can be absent or limited. As such, the proportionally fewer numbers of individuals on multiple OAMs within the T2DM and wider population adds to the challenges faced by researchers performing retrospective observational studies.

Understanding the relationship between antidiabetic medication use and PC is currently limited. However, further exploration of these relationships may facilitate our understanding of PC development when masked by clinical diabetes. There are many limitations typically associated with drug association studies and assumptions about associations between antidiabetic medications and PC should thus currently be approached with caution. To mitigate this problem, future research may benefit from the current expansion of population-wide electronic health data record keeping at both the primary and tertiary care level. As future retrospective and prospective observational studies gain access to increasingly detailed electronic health records data, it is likely that more longitudinal information about the dose-response relationships between antidiabetic medications and risk for PC will become available for researchers. Furthermore, increased access to full dose- and time-sensitive information may significantly reduce some of the sources of bias that typically limit the strength of findings within drug-association studies published to date.

6.4 Screening for Pancreatic Cancer

In the UK, no effective screening programme for sporadic pancreatic cancer exists for the general population. Identifying individuals from clinical features alone is not a reliable option as symptoms are commonly non-specific. They also tend to occur in the latter stages of disease, when treatment to achieve cure is no longer possible. A bi-directional relationship between the pancreatic cancer and diabetes appears to exist, suggesting that some individuals develop diabetic complications secondary to pancreatic cancer.

NODM diagnosis in certain individuals appears to indicate an increased risk of underlying PDAC which is especially marked within the first 3 years following DM diagnosis. This pattern of timing is consistent with findings that significant differences in glucose regulation becomes detectable around the same length of time prior to PDAC diagnosis and correlates with changes in tumour volumes [98]. Changes from tumour development to metastatic disease occurs over an approximate 7- to 10-year period of time [12]. If these hypotheses are accurate, further studies examining the temporal relationships of the two diseases in detail will be required to add validity to these findings.

6.4.1 NICE criteria for possible pancreatic cancer

Current clinical criteria for diagnosing diabetes are not optimised to identify diabetes secondary to cancer. However in 2021 the NICE criteria for possible PDAC was revised to include a combination of diagnosis of NODM with weight loss and age ≥60 as high risk for PDAC that warrants further investigation [3]. A recent study showed that NODM was associated with a 3-fold rise in risk of PC which increased to 6-fold or 10-fold higher once weight loss incidence within a two year period was factored in [177]. NODM with concomitant weight loss of 1 to 8lb was associated with an HR of 3.61 (95% CI 2.14-6.10), whilst a weight loss of more than 8lb was associated with an HR of 6.75 (95% CI 4.55-10.00). The elevated risk association with this combined exposure cohort puts it on par with the increased levels of risk associated with individuals with familial pancreatic cancer or hereditary gene mutations such as BRCA2 and CDKN2A [178].

6.4.2 Metabolomics

Metabolomics is an emerging discipline that enables examination of changes in endogenous and exogenous metabolites in cells, tissues and biofluids depending on the metabolic state of the individual. Performing this technique on human blood serum provides researchers with a tool to map out distinct phenotypical patterns according to different stressors or

disease statuses by comparing these mapped phenotypes against healthy individuals. A study using this technique has led to the discovery of specific metabolite signatures for PDAC [179, 180], and appears to demonstrate effective stratification of diabetic patients and their complications which could be useful in the early distinction of NODM due to PDAC verses T2DM.

Metabolomics introduces an alternative approach to testing for risk of PDAC by observing changes in blood serum chemistry for those at risk of both diabetes and PDAC. For ranges of HbA1c measurements, clinical thresholds for defining pre-diabetes and diabetes are set differently depending on where the guidelines originate. As such, performing risk association studies using these guidelines must take these localised variations into account.

6.4.3 Targeted methylation analysis and circulating cell-free DNA

Targeted methylation analysis on circulating cell-free DNA in blood plasma (cfDNA) is a technique that may provide an alternative to targeted mutation panels and whole genome sequencing as a tool for cancer detection [2]. In a recent study by Liu et al., cfDNA sequencing detected a broad range of cancer types at all stages with specificity and sensitivity performance approaching the goal for population-level screening. For pancreatic cancers, stage 1 disease sensitivity was 63% (95% CI 35% to 97%), stage 2 disease sensitivity was 83% (95% CI 36% to 100%), stage 3 sensitivity was 75% (95% CI 35%-97%) and stage 4 sensitivity was 100% (95% CI 80%-100%). The results of this study support the feasibility of employing targeted methylation analysis of cfDNA for early cancer detection and clinical validation in intended use populations is current ongoing [2]. One of these validation studies is a randomised control trial of the GRAIL Galleri^{TM} multi-cancer screening test which is being planned for the National Health Service in England [181]. The trial aims to have 140,000 healthy participants aged 50-79 (70,000 exposed to screening and 70,000 unexposed), tested for 50 different cancers responsible for a third of all-cause mortality. A 25% relative risk reduction in all-cancer mortality is proposed with cfDNA testing although the true reduction will likely be smaller. The expected discrepancy is due to various aspects of the model design that may increase risk of statistical errors. For example, the efficacy estimates are based on the test's ability to detect end-stage cancers which, as a surrogate measure, can be misleading [182]. The ability to detect early-stage cancers using this technique is recognised to be low. Nevertheless, some later-stage cancers are expected to be detected sooner than they would without the test, and this is where cfDNA testing has been proposed to be of most benefit. As for its potential use in early detection of sporadic PDAC, pancreatic cancer

is believed to be part of a group of cancers that shed more cfDNA into the circulation than other cancers [183]. Thus its use as a secondary or tertiary screening tool for detecting early-stage PDAC could therefore be highly relevant to the cohort of adults aged over 60 who develop NODM.

6.4.4 Screening with old and new methods

Once a new diabetes or pre-diabetes diagnosis is identified in older adults, practical measures for monitoring PDAC risk could include repeat HbA1c testing at agreed time intervals to track changes. Co-existing and unintended weight loss should also be closely monitored following diagnosis of pre-diabetes or new diabetes to aid recognition of a possible connection between the two exposures and a possible underlying PDAC diagnosis. As previously highlighted, clinical signs associated with sporadic PDAC are non-specific at presentation. Screening for early-stage disease is limited by the difficulty in detecting early unintended weight loss with concomitant early stages of glucose dysregulation. These limitations are difficult to overcome unless regular health checks or personal monitoring routines are already established. Time is also lost from symptom presentation to first clinical presentation and specialist referral for further investigations. As such, models for screening of early-stage PDAC require the addition of second-line tests to refine the at-risk cohort further and reduce the time taken to confirm a diagnosis of PDAC.

PDAC also follows a complex genetic and epigenetic pathway of pathogenesis that appears to arise from multiple prior disease states which contribute to different gene mutation patterns unique to each individual. With the insights gained through genetic and epigenetic research, relying on identifying a new adult diagnosis of diabetes may permit this presentation to serve only as a first clinical indicator of possible underlying PDAC or high-risk precursor lesion. After symptoms, non-modifiable and modifiable risk factors are accounted for, the new diabetes diagnosis may warrant follow-up with targeted gene screening and metabolomic profile mapping to rapidly determine if the new diabetes diagnosis is due to T2DM or NODM-PDAC. Combining clinical and genetic risk modelling could refine the cohort of those most at risk of PDAC for further investigation. This may increase early-stage PDAC detection rates and also identify those without PDAC but who are at high-risk and require close monitoring.

6.5 Large population prospective studies in the United Kingdom

The low prevalence of PDAC makes it difficult to study large cohorts of individuals with this disease easily. In the last two decades, the digitalisation of medical records and record keeping has increased in both scale and ease of access. The ability to quickly gather very specific information across large populations with greater control over the quality and consistency of this data has given researchers an opportunity to deepen our understanding of disease using 'big data'. Studies relying on the examination of clinical information retrospectively are subject to sources of error due to confounding and bias. However prospective studies have fewer potential sources of bias and confounding by comparison but are easiest to perform when the outcome of interest is common. Unfortunately, because pancreatic cancer is relatively uncommon in the general population, the size of the population cohort required to observe enough cases to be statistically significant would need to be very large, and therefore require a significant number of resources to achieve logistically.

In response to the limitations of previous large population epidemiology studies, Two UK based prospective cohort studies have been designed with the potential to inform on future strategies on screening individuals with NODM for pancreatic cancer. The United Kingdom Early Detection Initiative (UK-EDI) for pancreatic cancer aims to recruit 2500 individuals with NODM aged 50 years and over, with follow-up every 6 months over a 3-year period [184]. For study eligibility, diabetes diagnosis will be considered according to HbA1c clinical measurements of ≥48 mmol/mol. Detailed clinical information and biospecimens will be collected at baseline and importantly at each follow-up to support the development of molecular, epidemiological and demographic biomarkers for earlier detection of pancreatic cancer in the high-risk NODM cohort [184].

The UK Biobank database is a large population database which currently represents one of the best opportunities to examine the development of pancreatic cancer within a UK population cohort. The database has been set up to prospectively collect medical information from 500,000 volunteers aged between 40-69 from 2006 to 2010 with ongoing follow-up thereafter. These individuals have continued to be monitored with repeat investigations and linked national record updates to build the database and record changes to baseline characteristics. Genotyping, biochemical tests and imaging have also been performed as part of the database collection for the purpose of facilitating research.

One recent study using the UK Biobank dataset examined the integration of polygenic risk scores (PRS) with modifiable risk factors to demonstrate that combining cancer-specific PRS measures with family history and modifiable risk factors improves prediction accuracy for 16 cancers examined, although the magnitude of improvement varied depending on the cancer type [1]. PRS for pancreatic cancer reached significant risk stratification and was the primary determinant of risk stratification verses modifiable risk factors when they were examined together and separately as nested cohorts. The authors developed a risk model for pancreatic cancer consisting of four classes of risk factors based on their findings: (i) demographic factors (age and sex); (ii) family history of cancer in first-degree relatives or history of Lynch syndrome, Familial atypical multiple mole melanoma syndrome (FAMMM), Peutz-Jeghers syndrome, and Familial adenomatous polyposis (FAP); (iii) modifiable risk factors; and (iv) genetic susceptibility, represented by the PRS. As our intent is to explore the risk associations for PDAC in individuals with no history of familial pancreatic cancer or gene-linked cancers, predictive modelling for risk of sporadic PDAC is likely to depend more on modifiable risk factors and PRS with GWAS-identified risk variants specific to this sub-cohort of individuals.

6.6 Conclusion and Future Work

Diabetes and cancers are some of the most challenging diseases in terms of diagnosis due to their heterogenous and complex nature [185]. The risk of pancreatic cancer in individuals with T2DM is well recognised but the development of NODM can also be a manifestation of pancreatic cancer. Growing evidence suggests that NODM and diabetes with rising HbA1c are both independent risk factors for pancreatic cancer [186]. When determining the best approaches to studying HbA1c as a measure of risk for PDAC, there is value in categorising HbA1c for ease of application. However, the relationship between HbA1c and PDAC may be better demonstrated by HbA1c measurements when treated as numerical values within a set range for statistical analysis. This would also act to minimise loss of power and risk of a Type 1 statistical error [187].

This literature review focussed on presenting the current knowledge about the relationship between sporadic pancreatic cancer and diabetes. It also highlighted the numerous issues associated with studying these two complex diseases. The limitations of diagnostic criteria, variation in clinical definitions, limited availability of biochemical tests, and lack of longitudinal data in retrospective datasets are some examples of study design factors that contribute to the difficulties encountered when studying relationships between diabetes,

pancreatic cancer, and other potential covariates. In addition to NODM as an early indicator of PDAC, we discussed potential etiopathologies associated with the development of PDAC.

Potential risk factors associated with PDAC include non-modifiable (e.g., age and sex) and modifiable risk factors (e.g., weight/BMI, smoking, alcohol, diet), which may contribute to an optimised clinical model for screening. Although diabetes appears to be the main risk factor for sporadic PDAC, other clinical features may still have an important role as confounders or covariates.

Optimal risk models for screening may be based on a collection of clinical features composited with risk models using novel genetic analysis techniques such as PRS for risk prediction and cfDNA +/- radiological imaging for early-stage tumour detection. Further investigation of the interplay between diabetes, antidiabetic medication and pancreatic cancer is also needed to improve our understanding of how medication influences the development of pancreatic cancer. As such, the availability of antidiabetic intervention data (i.e., lifestyle, antidiabetic medications, and insulin adjuncts) within the UK Biobank dataset will also be assessed as part of the preliminary work up for possible inclusion in our study.

6.7 Study Goals

In individuals with NODM who are subsequently diagnosed with PDAC, higher HbA1c levels found prior to cancer diagnosis may be directly associated with malignancy progression. The time-to-event interval from NODM diagnosis to PDAC diagnosis may also differ to the time intervals associated with HbA1c changes seen in longstanding T2DM patients who subsequently develop PDAC. Evidence showing that HbA1c rises before clinical detection of PDAC demonstrates the potential value of HbA1c as a predictive marker [188, 189]. However, HbA1c measurements as a predictor of PDAC, or re-interpreted categorically according to range thresholds for normoglycaemia, pre-diabetes and diabetes has not been widely examined.

Hypothesis: HbA1c as a single value or grouped as ranges of values in subjects with NODM is a useful predictive marker of PDAC.

6.7.1 Aim & Objective

The aim of this study was to investigate the correlation between HbA1c levels and occurrence of PDAC in individuals diagnosed with NODM. Specifically, our objective was to quantify the risk association between elevated HbA1c levels measured at a single time-point and the incidence of PDAC as the primary outcome variable. Furthermore, we sought to

evaluate the effectiveness of HbA1c when assessed both as discrete numerical values and when categorised according to clinical thresholds for pre-diabetes and diabetes. This analysis aimed to ascertain whether the categorisation of HbA1c values offers any advantages in the context of interpreting HbA1c as a potential marker for PDAC.

6.8 Study Outline

HbA1c is the best currently available diagnostic measure for diabetes and glycaemic status and development of diabetes is associated with 60-80% of sporadic PDAC. It is therefore the most appropriate marker to use as a target exposure in our study as both a continuous variable and categorical variable using UK cut-off concentration level values for normoglycaemic (less than 42 mmol/mol), pre-diabetes (42 to 47.9 mmol/mol), and diabetes (48 mmol/mol or higher).

In this study using UK Biobank data, we first planned to compare the risk association and bi-directional relationship between diabetes and pancreatic cancer in this dataset with previous study outcomes. With PDAC as the key outcome, participants were divided into non-diabetic, NODM and LSDM sub-groups based on HbA1c concentration levels taken at first assessment and information available on prior diabetes status. These groups underwent separate cox regression analysis to compare hazard ratios between these groups for risk of PDAC. The overall diagnostic performance of HbA1c as a predictive marker for PDAC was examined for each sub-group using receiver operated characteristic (ROC) curve graph plots. To demonstrate the predictive value of HbA1c as a predictive test for PDAC when used alone, these plots were used to generate area under curve values to demonstrate the predictive value of HbA1c. Optimal cut-off values for determining the presence or absence of PDAC and the sensitivity and specificity of cut-off values associated with pre-diabetes diagnostic thresholds and diabetes diagnostic thresholds will also be examined.

For our study to be scientifically valid, the quality of each data item required from the UK Biobank dataset was evaluated prior to inclusion for further analysis. This was to examine the strengths and limitations of the UK Biobank dataset items which could influence the study design and our findings. We also expect to interpret any findings with consideration of the bias and confounding that pre-exists within the UK Biobank dataset and also introduced within our own study methodology.

7 Study Design and Methodology

To generate the nested case-control cohorts for this study, the population dataset provided by UK Biobank first required optimisation. To do this, each variable representing our key exposure and outcomes was carefully examined. Inclusion and exclusion criteria are described below. Other potential risk factors included in this study were also optimised prior to inclusion as either matched covariates or those to be adjusted for during analysis of the key exposure.

7.1 Data source: UK biobank

The entirety of our study data originated from the UK Biobank (UKB) dataset, which contains prospectively collected healthcare records on approximately 500,000 participants first recruited into the UKB study between 2006 and 2010. The participants were recruited from a wide variety of demographics to provide socioeconomic and ethnic heterogeneity, and urban-rural mix. This was to ensure broad distribution across all exposures and allow reliable detection of generalizable associations between baseline characteristics and health outcomes [190]. Each participant attended one of twenty-two assessment centres across England, Scotland, and Wales, where they completed a touchscreen questionnaire followed by a brief computer-assisted interview. Physical and functional measures, a 24-hour recall diet questionnaire, exercise tests, and fluid samples (blood, urine, and saliva) were collected to establish baseline measures. Linked data for death, hospital inpatient, cancer, and primary care (GP) records were also collected by the UKB in retrospect and prospectively with researcher access to these updates made available on a regular basis.

7.2 Data extraction: UK biobank

Permission to access the dataset was granted by UK Biobank to our research group upon review of the research proposal and signed material transfer agreements (MTA), which were provided by the University of Southampton and University Hospital Southampton NHS Foundation Trust (UHSNHSFT). The UK Biobank Project had approval from the North-West Multi-centre Research Ethics Committee as a Research Tissue Bank (RTB) approval that covered the proposed use of the UK Biobank dataset in most cases where researchers have requested data. The RTB was granted initially in 2011 and renews every 5 years on an ongoing basis. UK Biobank also possesses a Human Tissue Authority (HTA) licence to allow researchers access to participants' samples under certain obligations without the need to obtain a separate HTA licence [191]. For our own study no separate ethics approval process

was required by us through our affiliated bodies (University of Southampton and UHSNHSFT) for the use of UK Biobank data resource in our work. In addition, no tissue samples were required for our research, and this was stated as part of the MTA agreement submitted to UK Biobank.

Encrypted data files were downloaded from the UK Biobank (UKB) dataset in September 2021, following a UKB Data Showcase Update in August 2021. Full details explaining how the encrypted data supplied by UK Biobank was downloaded and decoded for research purposes can be accessed via the UK Biobank website [192]. Different variables within the UKB dataset were organised by data fields (DF) and examined to determine what inclusion and exclusion criteria were required to optimise each variable prior to further data analysis. A key of the data fields used in our study is included in the Appendices section. Additional custom variables were also created from the DFs available where representative sub-cohorts were not designed specifically within the original UKB dataset for our needs.

7.3 Dataset optimisation

The large population dataset provided by UKB allowed us to examine comprehensive healthcare information on all participants collected from the start of the study. Additional prospectively added health data via linked national records database updates were also provided by UKB in the form of periodic scheduled updates. Longitudinal outcome data within the dataset enabled us to perform our study using models for time-to-event data. As the aim of our study was to examine the effect of HbA1c concentration levels on risk for PDAC within the UK Biobank dataset, we embedded a matched nested case-control study within the larger prospective cohort to investigate the causal nature of this relationship. For the matched nested case-control design to be implemented correctly, several elements of the design needed to be properly defined [193].

7.4 Case definition: PDAC

To create the case cohort, the outcome measure was first defined as a pancreatic cancer diagnosis that occurred after entry into the study. The UK Biobank participants were considered to have nonendocrine pancreatic cancer (PDAC) if they had a diagnosis within linked death registry and cancer registry records using ICD10 coding (ICD10 codes: C25.1, C25.2, C25.3, C25.7, C25.8, C25,9). Cases were identified where PDAC was described as a primary cause of death, a contributary (secondary) cause of death, type of cancer was pancreatic using ICD10 codes, or if a distinct diagnosis was recorded across any hospital

inpatient records in either the primary or secondary position. The corresponding date of each diagnosis first recorded across all episodes was also provided by a separate linked time variable.

For this study, PDAC cases were included if PDAC was the first of any incident cancer diagnosis diagnosed on or after the participant's first UKB assessment date. Cases were excluded if PDAC diagnoses occurred prior to the date of first assessment, if there was a previous history of other cancer, or the pancreatic cancer subtype was labelled as neuroendocrine (ICD10 code C25.4). Histological classification was confirmed where possible using Cancer Register-linked comparison data. If multiple episodes of different ICD10 codes for pancreatic cancer were associated with a single individual, the date of the first episode and subtype was used as the single representative outcome. PDAC cases could only be included as a case once, and subsequent episodes of non-endocrine pancreatic cancer in the same individual were not counted again as additional cases. Once a participant had been identified as a case, they could not be reused in any future risk sets for other cases as a 'control' individual.

7.5 Controls (Risk set) definition: Diabetes and HbA1c

One of the challenges encountered at the beginning of the model-building process was the identification of an appropriate risk set from which study controls could be sampled. The control set needed to incorporate the time dependent log data of our key exposure with adjustment for known or potential confounders. The current consensus is that diabetes as our key exposure, is both causally associated with pancreatic cancer and associated with pancreatic cancer consequentially by reverse causality. In either case, the clinical manifestation of diabetes in association with pancreatic cancer causes disruption to normal glycaemic control, and we chose to focus on this altered glucose control as our main measure and key exposure for this study. By investigating the link between diabetes and PDAC by the measures used to detect changes in glycaemic control, we hoped to understand more about how changes to HbA1c concentrations attenuates risk of a PDAC diagnosis. It was therefore essential that our study included information on time-to-event data for diabetes and PDAC. Our study needed to include HbA1c concentration levels taken before PDAC diagnosis and ensure that diagnosis of diabetes with dates of diagnosis at any time before or after the study start date were also available. This would enable us to divide participants between those with a prior history of diabetes and those with no history of diabetes on entry into the study for further sub-cohort analysis.

To identify those with a diabetes diagnosis, several data-fields were available to indicate each participant's diabetes status at different points in time based on different criteria. Participants were considered to have diabetes if there was a confirmation of their diagnosis from questionnaire and verbal interview data, linked ICD10 records, or HbA1c taken at Instance 0 was 48mmol/mol or greater.

For the variable 'diabetes diagnosed by a doctor', participants were asked "Have you ever been diagnosed with diabetes by a doctor" on the initial assessment visit and answered with the options: 'Yes', 'No', 'Do not know', and 'Prefer not to answer'. ICD10 records of diabetes diagnosis were identified in two formats. 'First occurrence' data-fields were both variables created by the UK Biobank by mapping information from Primary care data, Hospital inpatient data, Death Register records, and self-reported medical conditions reported at the baseline or subsequent UK Biobank assessment centre visits. These provided dates of first occurrence of any code mapped to 3-character ICD10 E10 (insulin-dependent diabetes mellitus (IDDM)) and E11 (non-insulin-dependent diabetes mellitus (NIDDM)) respectively. Event dates given that were apparently before birth were omitted when constructing this data field before it was made available for research purposes. ICD10 records data summarised distinct diagnosis codes of each participant as recorded across all their hospital inpatient records in either the primary or secondary position. 3-character ICD10 codes mapped to IDDM (E10) and NIDDM (E11) were identified from this cohort. Although these data fields already incorporated information from Hospital Inpatient data, we also performed our own manual check of the UKB-provided ICD10 records data to ensure that other diabetes diagnoses linked to dates were not omitted from our study.

We identified 4 key data-fields that provided age and date information linked to the categorical variables above: Age diabetes diagnosed; Date of first in-patient diagnosis – ICD10; Date E10 first reported insulin-dependent diabetes mellitus; and Date E11 first reported non-insulin-dependent diabetes mellitus. 'Age diabetes diagnosed' required participants answering "yes" to the question of if they had ever been diagnosed with diabetes by a doctor to follow up with an answer to the question "What was your age when the diabetes was first diagnosed?". Answers were given as age in years, with options for "Do not know", and "Prefer not to answer" also available. 'Date of first inpatient diagnosis' contained date information in string format data for all 3-character ICD10 codes found in the ICD10 diagnoses data field. String format data was also provided for the data fields for dates of 'E10 – IDDM' and 'E11 – NIDDM' first reported.

To define the earliest date of diabetes diagnosis, linked diagnosis dates and partial time data from the data fields described were cross-examined. Partial dates and dates coded as erroneous were removed. The earliest full date available across these data fields for a diabetes diagnosis were then taken as the date of diabetes diagnosis for the study. The values from data fields that identified diabetes diagnoses were merged into a new categorical variable to represent any participants with a diabetes diagnosis at any point in time and coded 0 for no diabetes diagnosis, and 1 for a diabetes diagnosis. An additional variable was also generated for participants with a diabetes diagnosis at any point in time with available linked dates of diagnosis.

HbA1c values taken from all participants at date of first attendance were divided into categories defined by clinical boundaries: normal = <42 mmol/mol; pre-diabetes = 42 to 47.9 mmol/mol ; diabetes = 48 mmol/mol or greater and examined as continuous data. Paired time data was available from when the blood sample was taken for HbA1c concentration levels Date of attending assessment centre; string format data). As HbA1c was the key exposure in this study, participants with missing Hba1c concentration level records from initial assessment were excluded from our optimised dataset and from further analysis.

By comparing dates of diabetes diagnosis to the date of the blood test at first assessment, participants with no history of diabetes prior to entering the study could be separated into further cohorts by HbA1c concentration levels based on the clinical ranges for normoglycaemia (<42mmol/mol), pre-diabetes (42 to 47.9mmol/mol), and diabetes (≥48mmol/mol). This variable represented all individuals with no history of diabetes diagnosis prior to entry into the UK Biobank, divided into range categories for HbA1c concentration levels: 0 = normoglycaemic, 1 = new prediabetes, and 2 = new diabetes.

7.6 Selection of additional factors

Additional data was extracted on the following risk factors in all individuals within our nested cohort. The covariate data were assessed based on records available at the start of the study at index date in each cohort.

7.6.1 Non-modifiable risk factors

‘Age when attended a UK Biobank assessment centre’ was a derived variable based on date of birth and date of attending assessment centre. Age at initial assessment visit was treated as age at index date. Sex (Male and Female) was acquired from central (NHS) registry at

recruitment and in some cases updated by the participant during the initial assessment period.

7.6.2 Anthropometric measures

Other covariates were included in the study to adjust for potential influences on the key relationship between HbA1c concentration levels and risk of PDAC diagnosis. Body mass index (BMI) and waist circumference are anthropometric measures frequently used to examine obesity-related health risks. BMI values were constructed from height and weight measured during the initial assessment centre visit. Categorical variables were created out of the continuous data for BMI using clinically defined boundaries set according to the National Institute for Health and Care Excellence (NICE) definitions [194]. BMI was subdivided into different weight classes: Underweight = less than 18.5 kg/m^2; Healthy weight = 18.5-24.9 kg/m^2; Overweight = 25-29.9 kg/m^2; and Obese = 30 kg/m^2 or higher. Because of the known association of BMI with diabetes and pancreatic cancer, BMI was considered an important covariate and the optimised dataset had further participants excluded if no BMI record was recorded at initial assessment. Waist circumference was measured during initial assessment centre visit in centimetres and analysed as a continuous variable.

7.6.3 Socio-economic status

Townsend deprivation index (TDI) is a combined measure of owner occupation, car ownership, overcrowding and linkage to population census data within a postal area [53]. TDI at recruitment was calculated for each participant immediately prior to joining UK Biobank by assigning a score corresponding to the output area in which their postcode was located. Continuous data for TDI was converted into subcategories ranging from 0 to 19 for this study. A score of 6 to 7 covered an original TDI range of -0.99999 to 0.99999, with 0 representing the area with the overall mean values based on preceding national census output areas. Scores of less than 6 on the modified scale leaned towards higher socioeconomic status, with a higher score indicating poorer socioeconomic status. Deprivation was also examined using quintiles of Townsend scores from ‘1’ (least deprived) to ‘5’ (most deprived).

7.6.4 Smoking

Smoking is a well-established risk factor for pancreatic cancer [195, 196] and associated with a reduction in survival among patients with pancreatic cancer [197]. In this study, ‘smoking status at study entry’ summarised current and past smoking status of participants. The data

field containing information on the number of cigarettes currently smoked daily amongst current cigarette smokers was used to explore the categorical and dose-response relationship of smoking with diabetes and pancreatic cancer. The quantitative values for this field were subdivided into groups for categorical analysis (0 = non-smoker, 1 = 1 to 5 cigarettes per day, 2 = 6-10 cigarettes per day, 3 = 11 or more cigarettes per day).

7.6.5 Alcohol

For alcohol profiles, alcohol drinker status data was obtained via questionnaire response options of 'Never', 'Previous and 'Current'. Alcohol frequency values were obtained by asking participants how often they drank alcohol on average, with the participants choosing from the options: 'Never', 'Special occasions only', 'One to three times a month', 'Once or twice a week', 'Three or four times a week', and 'Daily or almost daily'.

7.7 Cohort time axis

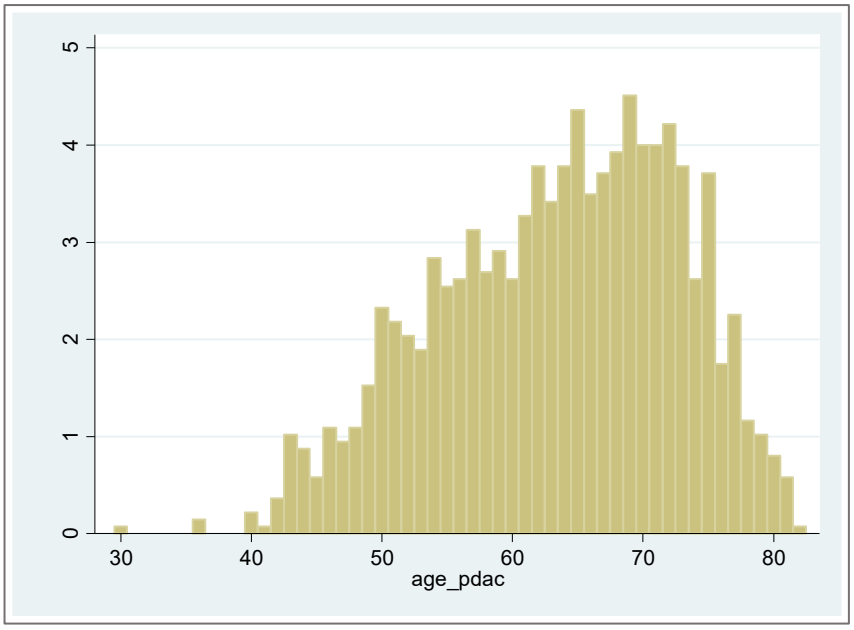
For the nested case-control analysis, date of first attendance to a UK Biobank centre for initial assessment were used to determine the age of each individual at this time point. Age in months at initial assessment was then defined time zero (i.e., time of entry into the survival study). We chose to use age as our measure of time. Because most of the time data available was provided in date format, calculations of the age of participants at the time were performed based on the date of birth information provided by the UK Biobank. This was available by 'year' and 'month' only, and no 'day' information was provided for confidentiality reasons. All measures in the study with date information such as our risk set (diabetes and HbA1c) were also converted to month and year format (mmyyyy) to the last complete month. This would align them with the date of birth information, thus standardising the date format for ease of use when performing calculations. Participants were followed up by time in months until their exit time, which was defined by the first occurrence of PDAC (primary event failure), or right censoring for participants leaving the study before event occurrence. Right censored individuals included those who were lost to follow-up, or to death, and study end date for everyone else (March 2021).

7.8 Selection of matching factors

Matching of case-control studies is a commonly implemented technique used in the field of public health and medical literature [198]. The purpose of matching cases to controls was to improve the efficiency of studying the large population dataset and to a lesser degree, help eliminate confounding by matched variables. Age and sex are both non-modifiable risk

factors that are known to impact the likelihood of developing pancreatic cancer. Older age is known to be associated with increased incidence of pancreatic cancer, which the UK Biobank cohort also demonstrated (Figure 13). By using Log-rank test of equality of survival functions we also identified from the pre-optimised dataset that risk for pancreatic cancer was significantly different between men and women ($p<0.05$) (Figure 14.).

(a)



(b)

Descriptor	Coefficient	Standard error	P-value	95% confidence interval (lower to upper)
Age at first attendance	0.0118	0.00359	<0.01	0.00472 - 0.0188

Figure 13. Older age is associated with increased incidence of PDAC. (a) Histogram showing x-axis: Age of participants at the time they develop PDAC within the UKB cohort by years, & y-axis: percentage of participants in each age group who develop PDAC relative to the number of PDAC cases in total. (b) Cox proportional hazard regression analysis with 'age at first assessment' as the independent variable shows a significant positive correlation (coefficient value of >0; P <0.01) between increasing age and risk for PDAC.

Sex	Numbers observed	Numbers expected
Female	649	720.31
Male	681	609.69
Total	1330	1330.00

Figure 14. Female and Male participants have statistically significantly different risks for PDAC. Log-rank test for equality of survivor functions show a significantly higher risk of PDAC in women vs men (P-value <0.01).

By matching cases to controls by sex, we were able to control for their effects on our survival analysis model and reduce the distortion effect they might have on the associations we sought with HbA1c. In our study we chose not to differentiate between female and male participants when examining overall risk for pancreatic cancer. This was to ensure that the number of pancreatic cancer events in our dataset was sufficient to perform regression analysis without adding bias to regression coefficients. Sufficient numbers were also essential to maintaining predictive accuracy [199].

On matching of cases to controls by age, matched controls achieved greater adequacy the closer they were in age to the case at time zero. In addition, case-control ratio sampling of controls had to consider the total number of controls available for age-matching from within the optimised dataset. Ideally, controls would be matched to the exact age of each case by month and year. After matching by sex, the study model also had to ensure that enough controls in each cohort were then available for matching by age in sufficient numbers to prevent the need to recycle any controls for matching again with other cases. To achieve this compromise, the age range for matching controls to each case allowed matching of controls up to +0.9999 years older than the case age to ensure most cases were successfully matched to the recommended number of controls required.

7.9 Matching of cases to controls: Power and sample size estimates

The UKB dataset involved over 500,000 participants and performing sample size estimations on this cohort ensured that the number of participant controls required was enough to avoid a type II error by providing it with sufficient power (i.e., the ability for tests to detect an effect that truly exists). An additional benefit of reducing the cohort to a 1:n ratio (where n = no. of controls matched to each case) was that it would improve the efficiency of the study by not including any more controls than was required to detect a true effect with confidence. Power and sample size calculations were conducted to determine the smallest sample size needed to detect the effect of diabetes (as the key exposure) on the likelihood of PDAC (as the primary outcome) at the desired level of significance. The alpha value (Type I error, i.e. probability of rejecting the null hypothesis when it is in fact correct) was initially set at 0.05, with power set at 95%. The probability of exposure among controls (p0) was determined by looking at the prevalence of the diabetes variables within the UKB dataset. Crude odds ratios (OR) for the risk of developing PDAC in association with a diabetes diagnosis was used to determine the minimal detectable OR required. The details of how each input for the power and sample size calculation were derived are explained in the following paragraphs.

7.9.1 p(0) – the proportion of diabetes in the control population

In the UK, diabetes prevalence across all ages was approximately 6% in 2019 [200]. The UKB dataset cohort had an age range at study start from 36 to 74 years and the proportion of participants with any diagnosis of diabetes was 10.6%. This proportion reduced to 9.3% (mean age at diagnosis = 60 years) when the cohort was refined to only include cases with linked dates of diagnosis. Incidentally, Health Survey for England (HSE) statistical data from 2018-2019 found that the proportion of doctor-diagnosed diabetes in those aged 45-64 years of age was close to this UKB value, lying between 8.5-9.1% from 2018 and 2019 [201]. Sample size calculations were performed using both a p0 value of 6% and at 9% to compare the number of controls needed for the sample sizes to be adequate in both cases. Comparing the calculations between p0 inputs also helped inform the decision on an appropriate case-control ratio for the remainder of the study.

7.9.2 OR – the minimum detectable odds ratio

To perform sample size estimates, the minimum detectable odds ratio needed to be stated. Previous meta-analysis of studies examining the odds ratios (OR) for diabetes and risk of pancreatic cancer ranged from OR=1.36 (95%CI 1.16-1.96) [82] to OR=1.51 (1.19-1.55) [80]

at 10 years from diabetes diagnosis when the association appeared weakest. Based on these findings, we could expect that a minimum detectable OR between 1.36-1.51 would be a reasonable estimate of the true OR. In our study, the crude odds ratios calculated for the individual UKB data fields with linked dates ranged from 2.08 (95%CI 1.74-2.48) to 5.09 (95%CI 3.90-6.55) as shown in Table 3. The crude odds ratios for HbA1c in the pre-diabetes range and diabetes range were 1.81 (95%CI 1.45-2.25) and 2.63 (95%CI 2.4-3.22) respectively. The lowest 95% confidence interval amongst these calculations was taken from the HbA1c pre-DM range OR results, with a value 1.45. As this figure satisfied our target range for possible values to represent the minimum detectable odds ratio, 1.45 was used as the reference odds ratio in the sample size estimates.

Variable	Odds Ratio	95% CI (upper – lower)	P-value
Diabetes diagnosed by a doctor	2.08	1.74-2.48	< 0.05
Date E10 (IDDM) first reported	5.09	3.90-6.55	< 0.05
Date E11 (NIDDM(first reported	4.14	3.66-4.68	< 0.05
HbA1c concentration levels			
42 – 47.9 mmol/mol (pre-DM range)	1.81	1.45-2.25	< 0.05
≥48 mmol/mol (DM range)	2.63	2.14-3.22	< 0.05

Table 3. Table of figures showing the odds ratios for risk of PDAC calculated from variables provided by the UK Biobank. With HbA1c as the key exposure, a minimum detectable odds ratio of 1.45 based on the lowest end of the 95% confidence interval calculated for HbA1c across the different models for defining diabetes and pre-diabetes ranges. This gave us confidence that our ratio of matched controls would help us detect an odds ratio or hazards ratio equivalent to this value in our study.

7.9.3 Sample size calculations

The crude total number of PDAC identified from the UKB dataset based on our calculations was n=1329. With probability of exposure (p0) set to 0.06, a ratio of 1:6 cases was deemed sufficient to detect an odds ratio of 1.45 with 90% power and 95% significance. If the power was increased to 95% whilst keeping the exposure estimate at 0.06, the number of cases within the study was insufficient for an adequate number of controls to be matched for

detection of the same minimal odds ratio with the same level of significance. Both scenarios are displayed in Table 4.

alpha	power	odds ratio	p0	M	N
0.05	0.90	1.45	0.06	1	2284
0.05	0.90	1.45	0.06	5	1342
0.05	0.90	1.45	0.06	6	1303
|	|	|	|	|	|
0.05	0.95	1.45	0.06	10	1538
0.05	0.95	1.45	0.06	100	1411
0.05	0.95	1.45	0.06	500	1400
0.05	0.95	1.45	0.06	2000	1398
0.05	0.95	1.45	0.06	5000	1398
0.05	0.95	1.45	0.06	10000	1397

Table 4. Sample size estimates for establishing the number of controls needed for the matched case-control study. Key: alpha = 0.05 (5% risk of Type I error occurring), **power** = 0.90 (90% likelihood of a true effect being detected), **odds ratio** = minimum detectable odds ratio (1.45), **p0** = 0.06 (the prevalence of the independent variable – diabetes - in the population, e.g., 6%), **M** = number of controls needed for each case, **N** = number of cases required in the study (n=1329). With power set to 0.9 (90%), a ratio of 1:6 cases to controls could be achieved with the case numbers in our dataset. Increasing the power to 0.95(95%) to improve the chance of detecting a true difference with all other parameters staying the same, results in numbers of controls needed to match cases being unattainable.

In addition to the issues described above, the PDAC frequency after factoring in exclusion criteria in our optimised dataset was n=1208. This reduction in cases would require a respective increase in the total number of controls to mitigate the issue. With power kept at 95% and p0 value increased to 0.09, a ratio of 1:5 cases was sufficient to detect the minimal odds ratio at 95% significance, with all other input values left unchanged. As we intended to investigate specific covariates at a later stage, it was decided that each case would be matched to nine controls. The p0 value was also increased to consider the variance in diabetes exposure observed within the dataset (i.e., between 6-11%) and the likelihood that the true p0 for our study age cohort was also closer to 9% than 6% as discussed earlier. Another purpose of increasing the control numbers was to account for the expected loss of participant numbers due to exclusion criteria used to generate sub-groups of individuals within the HbA1c cohort. Any significant drop in participant numbers would reduce the

statistical power of the study and increase the risk of type II errors. Finally, as we planned to perform multivariate analyses to adjust for other potential risk factors, we anticipated that case and control numbers would drop further still, as more covariates were included in each risk model, which would again increase the likelihood of a Type II error. Thus, retaining a larger control population for the analysis would aim to limit these negative effects on our potential findings. The distribution of sample size calculations with successful matching by a true N value of 1208 by 1:5 cases to controls is displayed in Table 5.

alpha	power	odds ratio	p0	M	N
0.05	0.95	1.45	0.09	1	1967
0.05	0.95	1.45	0.09	2	1470
0.05	0.95	1.45	0.09	3	1305
0.05	0.95	1.45	0.09	4	1222
0.05	**0.95**	**1.45**	**0.09**	**5**	**1173**
0.05	0.95	1.45	0.09	6	1140
0.05	0.95	1.45	0.09	7	1116
0.05	0.95	1.45	0.09	8	1098
0.05	0.95	1.45	0.09	9	1085
0.05	0.95	1.45	0.09	10	1074

Table 5. Sample size estimates for establishing the number of controls needed for the matched case-control study. Key: alpha = 0.05 (5% risk of Type I error occurring), **power** = 0.95 (95% likelihood of a true effect being detected), odds ratio = minimum detectable odds ratio (1.45), **p0** = 0.09 (the prevalence of the independent variable - diabetes - in the population, i.e., 9%), **M** = number of controls needed for each case, **N** = number of cases required in the study (n=1208). With power set to 0.95 (95%), we can achieve a ratio of 1:5 cases to controls with the case numbers in our dataset.

7.10 Statistical analysis

After optimisation of the dataset nine controls were matched to each case to generate the two cohorts required for the nested case-control model. Baseline characteristics of the participants within the UK Biobank dataset were compared with the optimised dataset for any differences. This allowed us to observe and account for any changes to the distribution of individuals caused by our optimisation process. The baseline characteristics of PDAC cases and 1-to-9 matched control cohort were then also compared. As the control cohort was now matched by age and sex, we were able to begin observing key numerical differences between the two groups free from significant interference by either of these factors.

On the basic characteristics table frequency (with percentages) were presented for binary and categorical variables, and means (with standard deviations), and median (interquartile ranges, with total range) for continuous variables. Chi-squared goodness of fit modelling was performed on discrete distributions of data to examine for significant differences between the study cohorts for each categorical variable, and Student's t-test was performed on variables with continuous data to show if a statistically significant difference between the means of two independent study cohorts was present.

With survival study end points defined and study outcomes controlled for age and sex, initial univariate analyses were performed using Cox proportional hazards regression analyses to estimate hazard ratios and 95% confidence intervals (95%CI) for risk of PDAC with each study variable and covariate. Kaplan-Meyer survival curves were plotted for all categorical variables to test the assumption that risk for PDAC was the same in exposed and unexposed participants. Plotting these graphs provided insight into the shape of the survival function for each group and gave an idea of whether the groups were proportional by observing if the curves were roughly parallel to one another. Log-rank test of equality provided a statistical comparison between the same exposed and unexposed categorical covariates in the form of a p-value degree of significance for the categorical values. The test for equality with continuous variables was performed by univariate Cox proportional hazard regression. These tests are commonly used as part of forward or backward stepwise regression modelling for multivariate analyses. However, performing these calculations at this stage gave us the opportunity to make early observations about the strength of each variable's one-to-one association with PDAC and flag them up as potentially significant predictors for inclusion in multivariate models we planned to build beyond the focus of this study. The covariates in this study were included because of their likely or possible effect on the outcome as either a

confounder or independent variable. Figures showing the Kaplan-Meier plots, log-rank calculations, and univariate cox proportional hazard regression calculations were included in the Appendices, under section *13.1.1 Univariate analyses: Testing the predictive value of variables.*

The models for HbA1c (continuous and categorical) were further adjusted for age, sex, and potential confounders including BMI, socioeconomic status (TDI), smoking status and alcohol status as these were described in section *6.4.6 Selection of additional factors*, and section *6.4.6 Selection of additional factors*. To evaluate the impact of HbA1c concentration levels on the time-to-event relationship with PDAC diagnoses, the mean, median, skewness as a measure of degree of lop-sidedness in the frequency distribution, and the coefficient values from cox proportional hazard regression were all examined for the HbA1c concentration levels at study start for participants with no prior history of diabetes at study start, that were within the clinical ranges of Normoglycaemia, Pre-diabetes, and Diabetes.

Finally, ROC curve graphs were plotted to assess the overall diagnostic performance of HbA1c as a predictive marker for PDAC for each of the subgroups. Theoretical optimal cut-off values were also calculated to evaluate the usefulness of HbA1c as a standalone predictive marker for PDAC. All analyses were performed using STATA v16 Data Analysis and Statistical Software.

8 Results 1: Basic Characteristics

Of 502,459 patients in the UK Biobank cohort, we identified 1215 incident PDAC diagnoses occurring within the study period. After exclusion of individuals due to missing HbA1c and BMI data, 463,346 control participants were available within the optimised cohort for matching. Matching cases to controls by age and sex, and to a ratio of 1:9 yielded a study cohort with 1208 cases to 10,872 controls. 7 cases (0.6% of total PDAC) failed to achieve adequate matching to 9 unique controls without the need to reuse controls and were therefore omitted from our final study cohort. The flow diagram of the progress of all participants through our protocols to create the nested case-control cohort for our study is shown in Figure 15.

8.1 Study population

At baseline (UKB Dataset) the average age of entry (±SD) into the study was 56.7(±8.16) years of age. More women than men (54.4% vs 45.6%) enrolled within the UKB study and more participants within the UKB dataset were of white ethnicity (94%) compared to the UK census findings (89%) for England and Wales in 2010. After recategorization of TDI for this study, we observed that the UKB population were on average marginally better off than the census average in the UK (5.21 for UKB vs 6 for UK average; a score of <6 was associated with higher socioeconomic status). Mean BMI was 27.4 with most of the UK Biobank cohort lying within the overweight category.

Following optimisation of the UKB data for this study, most variables showed statistically significant differences in distribution of characteristic outcomes between this cohort and the original UKB dataset ($p<0.05$). The full table of basic characteristics showing distribution of subjects within the UKB, and the optimised study dataset are displayed in Table 6.

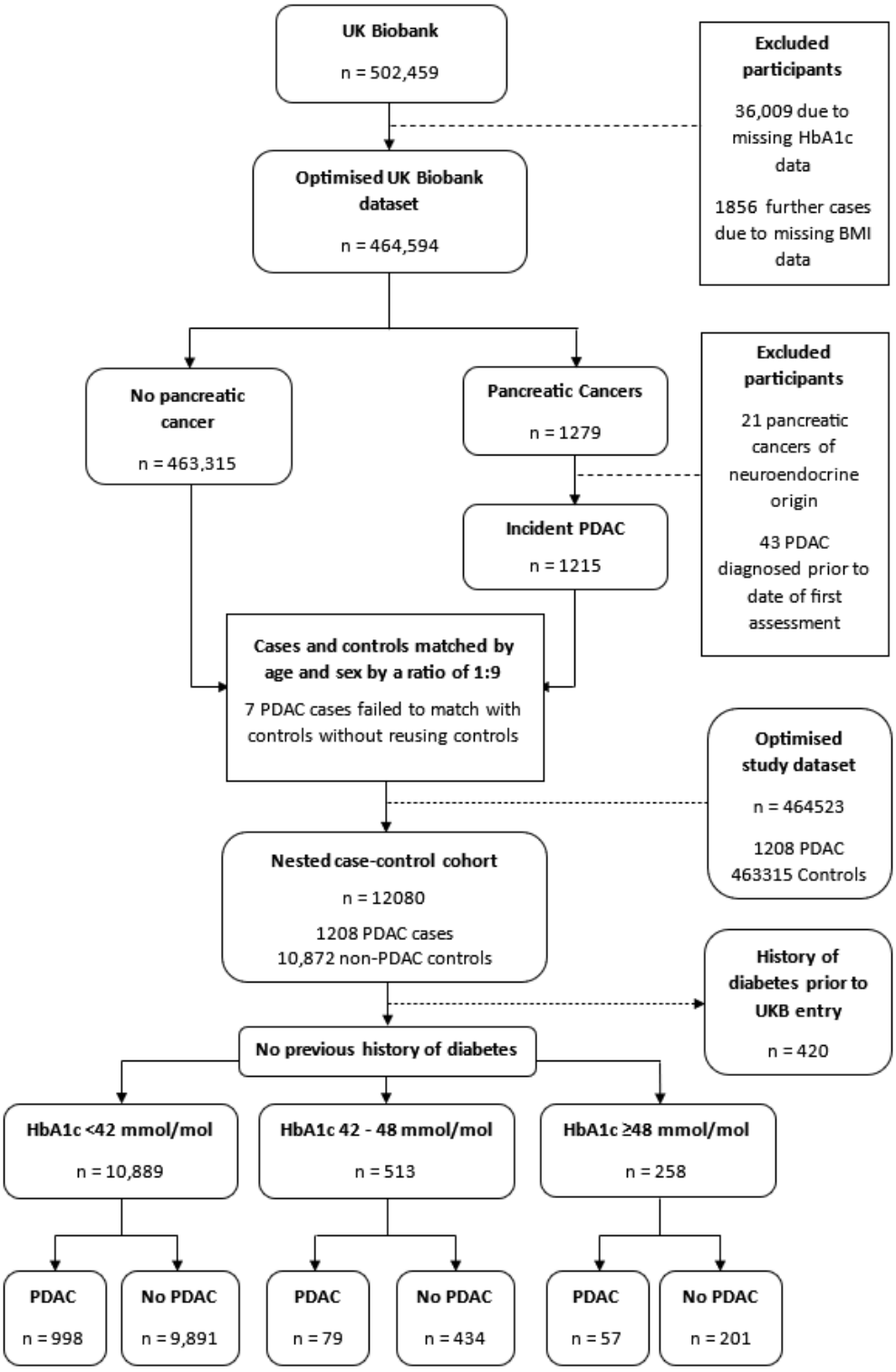


Figure 15. Flow diagram of the study design. From 502,459 participants within the UK Biobank dataset, we identified 1208 individuals with a diagnosis of PDAC which were matched to controls by a ratio of 1:9. 7 cases were excluded for failing the matching process. After excluding 420 participants for a history of diabetes before first UKB assessment, the remaining subjects (n=11660) were categorised at study entry as having no diabetes, new pre-diabetes, and new diabetes according to HbA1c values recorded at their first assessment.

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence. p<0.05 indicate significant differences between groups compared.

Variable	UKB Dataset	Optimised UKB dataset	Test of independence (p-value)	Optimised dataset cases vs controls		Test of independence (p-value)
Demographic information	**Frequency (proportion,%)**	**Total (proportion,%)**		**PDAC (proportion,%)**	**No PDAC (proportion,%)**	
Total no.	502,459	464,523		1,208 (0.26)	463,315 (99.74)	
Age at study start, n mean (IQR) (range), y	502,459 56.7 ± 8.16 (50-63) (36-74)	464,523 56.6 ± 8.15 (50-63) (37-74)	0.72	1,208 56.7 ± 8.35 (50-63) (38-72)	463,315 56.6 ± 8.16 (50-63) (37-74)	0.72
Sex, n Female Male	502,459 273,353 (54.4) 229,106 (45.6)	443,523 251,964 (54.2) 212,559 (45.8)	<0.05	1,208 582 (48.2) 626 (51.8)	463,315 251,382 (54.3) 211,933 (45.7)	<0.05
Ethnicity, n White Mixed background Asian Black Chinese Other	499,683 472,656 (94.6) 2,956 (0.6) 9,880 (2.0) 8,060 (1.6) 1,573 (0.3) 4,558 (0.9)	462,404 439,473 (95.0) 2,634 (0.6) 8,744 (1.9) 6,176 (1.3) 1,442 (0.3) 3,935 (0.8)	<0.05	 1,156 (96.2) 10 (0.8) 10 (0.8) 16 (1.3) 3 (0.3) 7 (0.6)	 438,317 (95.0) 2,624 (0.6) 8,734 (1.9) 6,160 (1.3) 1,439 (0.3) 3,928 (0.9)	<0.05
Pancreatic cancer, n No Yes	502,481 501,083 (99.7) 1398 (0.3)	464,523 463,315 (99.7) 1,208 (0.3)	<0.05	-	-	-

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence (continued)

Age at PDAC diagnosis, n mean (IQR) (range), in years	1375 63.4 ± 9.2 (57-71) (30-82)	1,208 63.9 ± 9.0 (57-71) (40-82)	0.29			
Location of lesion, n	1,398	1,208				
Head	491 (40.7)	491 (40.7)				
Body	128 (10.6)	128 (10.6)				
Tail	143 (11.8)	143 (11.8)				
Duct	22(1.8)	22 (1.8)				
Other	24 (2.0)	24 (2.0)				
Overlapping	10 (0.8)	10 (0.8)				
Unspecified	390 (32.3)	390 (32.3)	1.00			
Diabetes diagnosed by a doctor, n	501,530	464,062		1,206	462,856	
Prefer not to answer	402 (0.1)	308 (0.1)		1 (0.1)	307 (0.1)	
Do not know	1,248 (0.3)	1,119 (0.2)		6 (0.5)	1,113 (0.2)	
No	471,745 (94.1)	436,984 (94.2)		1,067 (88.5)	435,917 (94.2)	
Yes	28,135 (5.6)	25,651 (5.5)	<0.05	132 (10.5)	24,519 (5.5)	<0.05
ICD10 Diagnoses (Main or Secondary)						
E10 Insulin-dependent diabetes mellitus (NIDDM), n	502,459	464,523		1,208	463,315	
No	497,382 (99.0)	459,950 (99.0)		1,150 (95.2)	458,800 (99.0)	
Yes	5,077 (1.0)	4,573 (1.0)	<0.05	58 (4.8)	4,515 (1.0)	<0.05

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence (continued)

E11 Non-insulin dependent diabetes mellitus (NIDDM), n	502,459	464,523		1,208	463,315	
No	460,383 (91.6)	426,220 (91.8)		879 (72.8)	425,314 (91.8)	
Yes	42,076 (8.4)	38,303 (8.2)	<0.05	329 (27.2)	37,974 (8.2)	<0.05
Any diabetes, n	460,467	453,545		1,176	452,369	
No	411,608 (89.4)	408,950 (90.2)		819 (69.6)	408,131 (90.2)	
Yes	48,859 (10.6)	44,595 (9.8)	<0.05	357 (30.4)	44,238 (9.8)	<0.05
Diabetes with date of diagnosis, n	59,963	42,787		369	42,523	
No	13,047 (21.8)	816 (1.9)		15 (4.1)	801 (1.9)	
Yes	46,916 (78.2)	42,076 (98.1)		354 (95.9)	41,722 (98.1)	<0.05
HbA1c, n	466,450	464,523	<0.05	1,208	463,315	
Continuous						
mean ± SD (IQR) (range), mmol/mol	36.1 ± 6.8 (32.8-37.9) (15-515.2)	36.1 ± 6. 8 (32.8-37.9) (15-515.2)	0.52	38.5 ± 8.5 (34.1-39.8) (20-97.1)	36.1 ± 6.7 (32.8-37.9) (15-515.2)	<0.05
Categorical						
< 42 mmol/mol	427,541 (91.7)	425,954 (91.7)		1,006 (83.3)	424,948 (91.7)	
42 to 47.9 mmol/mol	21,301 (4.6)	21,151 (4.6)		92 (7.6)	21,059 (4.6)	
>48 mmol/mol	17,608 (3.8)	17,418 (3.8)	<0.05	110 (9.1)	17,308 (3.7)	<0.05

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence (continued)

HbA1c range in non-diabetics before study, n	452,737	450,989		1,134	449,855	
Normoglycaemic (<42mmol/mol)	424,965 (93.9)	423,407 (93.9)		998 (88.0)	422,409 (93.9)	
New pre-diabetes (42 to 47mmol/mol)	18,642 (4.1)	18,522 (4.1)		79 (7.0)	18,443 (4.1)	
New diabetes (≥48mmol/mol)	9,130 (2.0)	9,060 (2.0)	<0.05	57 (5.0)	9,003 (2.0)	<0.05
Townsend Deprivation Index, n mean ± SD (IQR) (range) *(<6 = better, >6 worse)*	501,836 5.21 ± 3.12 (3-7) (0-18)	463,957 5.17 ± 3.24 (3-7) (0-18)	<0.05	1,207 5.31 ± 3.10 (3-7) (0-16)	462,750 5.17 ± 3.09 (3-14) (0-18)	0.12

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence (continued)

Body mass index, n mean ± SD (IQR) (range), kg/m^2	499,355 27.4 ± 4.8 (24.1-29.9) (12.1-74. 7)	464,523 27.4 ± 4.8 (24.1-29.9) (12.1-74.7)	0.10	1,208 28.4 ± 5.1 (25.0-30.9) (17.4-52.7)	463,315 27.4 ± 4.8 (24.1-29.9) (12.1-74.7)	<0.05
Categorical: <18.5 18.5 – 24.9 25.0 – 29.9 ≥30.0	 2,497 (0.50) 159,967 (31.8) 213,326 (42.5) 126,669 (25.2)	 2,298 (0.5) 149,055 (32.1) 198,835 (42.8) 114,335 (24.6)	 <0.05	 3 (0.2) 296 (24.7) 532 (44.3) 377 (30.8)	 2,295 (0.5) 148,759 (31.3) 198,303 (43.7) 113,958 (24.5)	 <0.05
Waist circumference, n Mean ± SD (IQR) (range), cm	500,299 90.3 ± 13.5 (80-99) (20-197)	464,429 90.3 ± 13.5 (80-99) (20-197)	0.16	1,208 94.1 ± 14.0 (84-103)(62-147)	463,221 90.3 ± 13.5 (80-99)(20-197)	<0.05
Smoking status, n Prefer not to answer Never Previous Current	501,568 2,057 (0.4)± 273,496 (54.5) 173,044 (34.5) 52,971 (10.6)	464,069 1,852 (0.4) 252,716 (54.5) 160,744 (34.6) 48,757 (10.5)	 <0.05	11,207 6 (0.5) 548 (45.4) 465 (38.5) 188 (15.6)	462,862 1,846 (0.4) 252,168 (54.5) 160,279 (34.6) 48,569 (10.5)	 <0.05
Cigarettes smoked per day, n Non-smoker 1 to 10 11 to 20 21 or 30 31 to 40 41 or more	502,459 466,308 (92.8) 12,939 (2.6) 17,984 (3.6) 4,152 (0.8) 846 (0.2) 230 (0.0)	464,523 431,273 (92.8) 11,880 (2.6) 16,572 (3.6) 3,817 (0.8) 770 (0.2) 211 (0.0)	 <0.05	1,208 1,067 (88.3) 37 (3.1) 79 (6.5) 22 (1.8) 3 (0.3) 0 (0.0)	463,315 430,206 (92.8) 11,843 (2.6) 16,493 (3.6) 3,795 (0.8) 767 (0.2) 211 (0.0)	 <0.05

Table 6. Basic Characteristics of the UK Biobank and optimised datasets with tests of independence (continued)

Alcohol status, n	501,562	464,066		1207	462,859	
Prefer not to answer	755 (0.1)	646 (0.1)		2 (0.2)	644 (0.1)	
Never	22,384 (4.5)	20,167 (4.4)		49 (4.1)	20,118 (4.4)	
Previous	18,099 (3.6)	16,562 (3.6)		50 (4.1)	16,512 (3.6)	
Current	460,324 (91.8)	426,691 (91.9)	<0.05	1,106(91.6)	425,585 (91.9)	<0.05
Alcohol frequency, n	500,959	463,555		1206	462,349	
Never	40,635 (8.1)	36,864 (8.0)		100 (8.3)	36,764 (7.9)	
Special occasions only	58,001 (11.5)	52,990 (11.4)		139 (11.5)	52,851 (11.4)	
1-3 times a month	55,848 (11.1)	51,643 (11.1)		112 (9.3)	51,531 (11.1)	
1-2 times a week	129,282 (25.7)	119,993 (25.8)		296 (24.5)	119,697 (25.8)	
3-4 times a week	115,428 (23.0)	107,434 (23.1)		271 (22.3)	107,163 (23.1)	
Daily or almost daily	101,765 (20.3)	94,631 (20.4)		288 (23.8)	94,343 (20.4)	
Missing	1,500 (0.3)	968 (0.2)	<0.05	2 (0.2)	966 (0.2)	<0.05

8.2 Pancreatic cancer

A total of 1375 classifications of pancreatic cancer were identified from within the UKB dataset with dates of diagnosis ranging from 26th April 1998 to 2nd April 2021. After implementing exclusion criteria and removal of duplicates, 1215 participants with PDAC were eligible for inclusion in our nested case-control study. After study optimisation the mean age at diagnosis of PDAC was 63.9 years ± 9.0, with an interquartile range (IQR) of 57 (25%) to 71(75%). Age at diagnosis ranged from 40 to 82 years with a median age of 64 years. Most cases in this study were located to the pancreatic head (40.7%), with 10.6% and 11.8% in the body and tail respectively. 32.3% of cases did not have a specified location.

8.3 Diabetes

48,859 cases of diabetes were recognised within the UKB dataset, of which 46,916 (96%) were linked with corresponding dates of diagnosis. Within the optimised study dataset, a significantly higher proportion of individuals developed diabetes within the PDAC cohort (30.4%) compared to the control population (9.8%; $p<0.05$). A higher proportion of participants with HbA1c concentration levels within the clinical ranges of pre-diabetes and diabetes was also found in the PDAC cohort vs controls ($p <0.05$). Examination of the distribution of diabetes diagnoses within the optimised study dataset revealed that a higher number of people were found to have diabetes or new-onset diabetes after enrolling in the UKB study, in contrast to the percentage of individuals with pre-existing diabetes (6.2% compared to 2.9%). 1134 Individuals with no history of diabetes prior to entering the study developed pancreatic cancer during our study's time-window. Based on HbA1c concentration levels taken at time of entry, 998 (88%) were classified as normoglycaemic, 79 (7%) were classified as newly pre-diabetic, and 57 (5%) were classified as newly diabetic. Table 7 presents additional information, displaying how participants were distributed within the 1:9 matched nested case-control cohort. Furthermore, this table breaks down the cohort based on categories derived from HbA1c concentration levels, including normoglycaemic, new pre-diabetes, and new diabetes.

8.4 Basic Characteristics – Comparison of cases to controls

After matching of controls to cases by age and sex, certain differences in distribution of individuals between these two groups became apparent. Chi-square tests of differences with P-values of <0.05 performed on categorical data demonstrated a statically significant difference between the outcome distribution for PDAC and expected distribution for its

respective controls. A p-value of <0.05 for Student's t-tests of continuous cohorts also indicated that the difference in means between the groups being compared were statistically different.

A higher proportion of participants with a PDAC diagnosis were men (Men: 51.8% vs. Women: 45.7%). PDAC was also associated with slightly lower socioeconomic status (p <0.05). The distribution of clinically overweight participants between cases and controls in the study was 44.3% vs 43.7%, and 30.8% vs 24.5% respectively in the obese category. The overall differences in distribution of weight between cases and controls was statistically significant (p<0.05). This was observed when differences in BMI distribution was interpreted categorically (p<0.05) and as a continuous variable (p<0.05). Similarly, waist circumference showed a significant difference in distribution of size between the PDAC and control cohorts with a higher mean circumference in the PDAC cohort (94.1cm ± 14.0cm) vs controls (90.3cm ± 13.5cm). A difference in distribution of individuals was noted in the study variable for current smoking status with a higher proportion of participants in the PDAC case cohort reporting themselves as previous smokers (38.5% for cases vs 34.6% for controls) or current smokers (15.6% vs 10.5% respectively). Smokers in the case cohort also smoked more when the two groups were compared by the number of cigarettes smoked per day (p<0.05). Alcohol status showed differences between the case and control cohort populations (p<0.05). A slightly higher proportion of individuals in the control cohort recorded themselves as 'Never' drinkers compared to those in the case cohort (4.4% vs 4.1% respectively). Although more individuals in the case cohort described themselves as 'Previous' drinkers, more of the control cohort reported themselves as 'Current' drinkers (91.9% vs 91.6% for cases). Frequency of alcohol intake revealed that proportionally more individuals who developed PDAC and drank alcohol were likely to report drinking alcohol daily or almost daily (23.8%) whereas control individuals tended to drink less frequently (20.4%). The table of basic characteristics including frequency distribution for these variables within the nested case-control cohort is shown in Table 7.

Results 1: Basic Characteristics

Table 7. Basic Characteristics Table for the matched nested case-control cohort. 1:9 case control cohort includes participants with diabetes diagnoses prior to study entry. Participants without a previous diabetes diagnosis before study entry were subdivided into groups (normoglycaemia, new pre-diabetes, and new diabetes) according to HbA1c measurements taken at study entry. '-' indicates that no participants were found by this distribution criteria.

	1:9 case control cohort (n=12080)	**1:9 case control cohort without previous diabetes diagnosis (n=11660)**					
		Normoglyaemic (HbA1c = <42 mmol/mol)		**New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)**		**New diabetes (HbA1c = ≥ 48 mmol/mol)**	
Demographic information	**Frequency (proportion, %)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**
Total no.	12,080	998	9,891	79	434	57	201
Age at study entry, n mean (IQR) (range), y	12,080 56.7 ± 8.3 (50-63) (38-72)	998 56.8 ± 8.3 (50-64) (38-72)	9,891 56.7 ± 8.3 (50-63)(38-72)	79 56.3 ± 8.3 (51-63) (38-70)	434 57.3 ± 8.5 (51-65) (38-72)	57 56.0 ± 9.3 (47-64) (40-68)	201 56.2 ± 7.9 (50-62) (38-70)
Sex, n Female Male	12,080 5,759 (47.7) 6,321 (52.3)	998 495 (49.6) 503 (50.4)	9,891 4,805 (48.6) 5,086 (51.4)	79 39 (49.4) 40 (50.6)	434 193 (44.5) 241 (55.5)	57 26 (45.6) 31 (54.4)	201 68 (33.8) 133 (66.2)
Ethnicity, n White Mixed (any) Asian Black Chinese Other	12,030 11,455 (95.2) 72 (0.6) 226 (1.9) 154 (1.3) 32 (0.3) 91 (0.8)	993 964 (97.1) 7 (0.7) 6 (0.6) 9 (0.9) 2 (0.2) 5 (0.5)	9,445 9,448 (95.9) 58 (0.6) 155 (1.6) 103 (1.1) 24 (0.2) 69 (0.7)	78 73 (93.6) 0 (0.0) 2 (2.6) 2 (2.6) 1 (1.3) 0 (0.0)	432 379 (87.7) 3 (0.7) 23 (5.3) 15 (3.5) 3 (0.7) 9 (2.1)	57 50 (87.7) 0 (0.0) 2 (1.8) 3 (5.3) 0 (0.0) 1 (1.8)	199 173 (86.9) 0 (0.0) 14 (7.0) 8 (4.0) 1 (0.5) 3 (1.5)

Table 7. Basic Characteristics Table for the matched nested case-control cohort. (continued)

	1:9 case control cohort	Normoglyaemic (HbA1c = <42 mmol/mol)		New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)		New diabetes (HbA1c = ≥ 48 mmol/mol)	
Demographic information	**Frequency (proportion, %)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**
Pancreatic cancer, n No Yes	12,080 10,872 (90.0) 1,208 (10)	10,889 9,891 (90.8) 998 (9.2)	-	513 434 (84.6) 79 (15.4)	-	258 201 (84.6) 57 (15.4)	-
Age at PDAC diagnosis mean (IQR) (range), in years	63.4 ± 9.2 (57-71) (30-82)	64.2 ± 9.0 (58-71) (40-82)		62.7 ± 9.8 (51-65) (38-72)		61.8 ± 9.5 (55-69)(42-79)	
Location of lesion, n	469 (40.7)	404 (40.5)		36 (45.6)		20 (35.1)	
Head	128 (10.6)	103 (10.3)		10 (12.7)		7 (12.3)	
Body	143 (11.8)	126 (12.6)		5 (6.3)		6 (10.5)	
Tail	22 (1.8)	20 (2.0)		1 (1.3)		1 (1.8)	
Duct	24 (2.0)	22 (2.2)		1 (1.3)		0 (0.0)	
Other	10 (0.8)	9 (0.9)		0 (0.0)		0 (0.0)	
Missing	390 (32.3)	314 (31.5)		26 (32.9)		23(40.4)	

Table 7. Basic Characteristics Table for the matched nested case-control cohort. (continued)

	1:9 case control cohort	Normoglyaemic (HbA1c = <42 mmol/mol)		New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)		New diabetes (HbA1c = ≥ 48 mmol/mol)	
Demographic information	**Frequency (proportion, %)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**
Diabetes diagnosed by a doctor, n	12,071	996	9,886	79	434	57	200
Prefer not to answer	6 (0.0)	1 (0.1)	5 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Do not know	32 (0.3)	3 (0.3)	17 (0.2)	1 (1.3)	7 (1.6)	2 (3.5)	1 (0.5)
No	11,262 (93.3)	982 (98.6)	9,760 (98.7)	64 (81.0)	350 (80.7)	19 (33.3)	73 (36.5)
Yes	771 (6.4)	10 (1.0)	104 (1.1)	14 (17.7)	77 (17.7)	36 (63.2)	126 (63.0)
ICD10 Diagnoses (Main or Secondary)							
E10 - Insulin dependent diabetes mellitus (NIDDM), n	12,080	998	9,891	79	434	57	201
No	11,891 (98.4)	974 (97.6)	9,881 (99.9)	75 (94.9)	426 (98.2)	45 (79.0)	183 (91.0)
Yes	189 (1.6)	24 (2.4)	10 (0.1)	4 (5.1)	8 (1.8)	12 (21.0)	18 (9.0)
E11 - Non-insulin dependent diabetes mellitus (NIDDM), n	12,080	998	9,891	79	434	57	201
No	10,817 (89.5)	835 (83.7)	9,593 (97.0)	32 (40.5)	280 (64.5)	11 (19.3)	51 (25.4)
Yes	1,263 (10.5)	163 (16.3)	298 (3.0)	47 (59.5)	154 (35.5)	46 (80.7)	150 (74.6)
Any diabetes, n	11,779	996	9,877	79	434	57	201
No	10,336 (87.8)	819 (82.2)	9,515 (96.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Yes	1,443 (12.3)	177 (17.8)	362 (3.7)	49 (62.0)	179 (41.2)	57 (100.0)	201 (100.0)
Missing	-	-	-	30 (38.0)	255 (58.8)	-	-

Table 7. Basic Characteristics Table for the matched nested case-control cohort. (continued)

Diabetes with date of diagnosis, n							
No	1,396	187	321	51	166	57	201
Yes	23 (1.7)	12 (6.4)	9 (2.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
missing	1,373 (98.4)	175 (93.6)	312 (97.2)	48 (94.1)	162 (97.6)	57 (100.0)	201 (100.0)
HbA1c, n	-	-	-	3 (5.9)	4 (2.4)	-	-
Continuous	12,080	998	9,891	79	434	57	201
mean ± SD (IQR) (range), mmol/mol	36.4 ± 7.4 (32.8-38.1) (15.6-266.3)	35.7 ± 3.3 (33.6-41.6) (20-41.9)	34.7 ± 3.4 (32.4-41.5) (15.6-41.9)	44.4 ± 1.7 (43-45.9) (42.1-47.6)	44.1 ± 1.7 (42.7-45.4) (42-47.9)	57.3 ± 9.2 (50.7-97.1) (48-97.1)	60.2 ± 20.1 (50.2-126.1) (48-266.3)
Categorical							
< 42 mmol/mol	10,949 (90.6)		9,891 (100.0)	-	-	-	-
42 to 47.9 mmol/mol	599 (5.0)	998 (100.0)	-	79 (100.0)	434 (100.0)	-	-
>48 mmol/mol	532 (4.4)	-	-	-	-	57 (100.0)	201 (100.0)
		-					

Table 7. Basic Characteristics Table for the matched nested case-control cohort. (continued)

	1:9 case control cohort	Normoglyaemic (HbA1c = <42 mmol/mol)		New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)		New diabetes (HbA1c = ≥ 48 mmol/mol)	
Demographic information	**Frequency (proportion, %)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**
Townsend Deprivation Index, n Mean ± SD (IQR) (range), modified scale *(<6 = better, >6 worse)*	12,069 5.2 ± 3.1 (3-7) (0-17)	997 5.2 ± 3.1 (3-7) (0-16)	9,881 5.1 ± 3.0 (3-7) (0-17)	79 5.3 ± 3.4 (3-7) (1-15)	434 5.9 ± 3.4 (3-8) (0-15)	57 7.2 ± 3.8 (4-10) (1-15)	201 5.9 ± 3.5 (3-8) (0-15)
Body mass index, n mean ± SD (IQR) (range), kg/m^2 BMI clinical ranges: Underweight <18.5 Healthy (18.5 –24.9) 25.0 – 29.9 ≥30.0	12,080 27.6 ± 4.8 (24.3-30) (15.2-60.6) 53 (0.4) 3,680 (30.5) 5,254 (43.5) 3,093 (25.6)	998 27.8 ± 4.7 (24.7-30.2) (17.4-52.7) 2 (0.2) 269 (30.0) 462 (46.3) 265 (26.6)	9,891 27.1 ± 4.5 (24-29.5) (15.2-68.4) 49 (0.5) 3,271 (33.1) 4,359 (44.1) 2,212 (22.4)	79 30.0 ± 5.6 (26.3-33.4) (18.9-44.3) 0 (0.0) 12 (15.2) 34 (43.0) 33 (41.8)	434 30.5 ± 5.9 (26.4-33.1) (17.3-61.5) 1 (0.2) 55 (12.7) 167 (38.5) 211 (48.6)	57 31.9 ± 5.6 (28.5-34.2) (21-49.3) 0 (0.0) 4 (7.0) 15 (26.3) 38 (66.7)	201 30.8 ± 5.3 (27.4-33.4) (18.5-47.4) 0 (0.0) 23 (11.4) 72 (35.8) 106 (52.7)
Waist circumference, n Mean ± SD (IQR) (range), cm	12,077 91.3 ± 13.4 (82-100) (55-171)	998 92.4 ± 13.1 (83-101) (62-144)	9,890 89.9 ± 12.8 (81-98) (57-171)	79 98.4 ± 14.6 (90-106) (65-130)	434 99.4 ± 13.7 (90.5-107) (55-156)	57 105.4 ± 13.3 (97-114) (72-135)	201 101.8 ± 13.2 (93-110) (69-141)

Table 7. Basic Characteristics Table for the matched nested case-control cohort. (continued)

	1:9 case control cohort	Normoglyaemic (HbA1c = <42 mmol/mol)		New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)		New diabetes (HbA1c = ≥ 48 mmol/mol)	
Demographic information	**Frequency (proportion, %)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**	**PDAC Frequency (%)**	**No PDAC Frequency (%)**
Smoking status, n	12,025	994	9,853	79	430	57	198
Never	6,374 (53.0)	474 (47.7)	5,384 (54.6)	26 (32.9)	207 (48.1)	25 (43.9)	93 (47.0)
Previous	4,302 (35.8)	366 (36.8)	3,460 (35.1)	37 (46.8)	150 (34.9)	27 (47.4)	77 (38.9)
Current	1,349 (11.2)	154 (15.5)	1,009 (10.2)	16 (20.3)	73 (17.0)	5 (8.8)	28 (14.1)
Cigarettes smoked per day, n	12,080	998	9,891	79	434	57	201
Non-smoker	11,146 (92.3)	882 (88.4)	9,206 (93.1)	66 (83.5)	383 (88.3)	53 (93.0)	179 (89.0)
1 to 10	308 (2.5)	31 (3.1)	242 (2.4)	2 (2.5)	17 (3.9)	1 (1.7)	2 (1.0)
11 to 20	479 (4.0)	64 (6.4)	347 (3.5)	8 (10.1)	22 (5.1)	3 (5.6)	13 (6.5)
21 or 30	119 (1.0)	18 (1.8)	77 (0.8)	3 (3.8)	11 (2.5)	0 (0.0)	4 (2.0)
31 to 40	15 (0.1)	3 (0.3)	10 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.0)
41 or more	13 (0.1)	0 (0.0)	9 (0.1)	0 (0.0)	1 (0.2)	0 (0.0)	1 (0.5)
Alcohol status, n	12,060	995	9,877	79	434	57	200
Never	497 (4.1)	37 (3.7)	370 (3.7)	5 (6.3)	38 (8.8)	5 (8.8)	17 (8.5)
Previous	433 (3.6)	35 (3.5)	323 (3.3)	2 (2.5)	24 (5.5)	5 (8.8)	11 (5.5)
Current	11,130 (92.3)	923 (92.8)	9,184 (93)	72 (91.1)	372 (85.7)	47 (82.4)	172 (86.0)
Alcohol frequency, n	12,063	996	9,879	79	434	57	200
Never	933 (7.7)	73 (7.3)	695 (7.0)	7 (8.9)	62 (14.3)	10 (17.5)	28 (14.0)
Special occasions only	2,534 (21.0)	253 (25.4)	2,105 (21.3)	20 (25.3)	64 (14.8)	7 (12.3)	31 (15.5)
1-3 times a month	2,871 (23.8)	236 (23.7)	2,441 (24.7)	14 (17.7)	80 (18.4)	9 (15.8)	34 (17.0)
1-2 times a week	3,090 (25.6)	247 (24.8)	2,565 (26.0)	18 (22.8)	96 (22.1)	15 (26.3)	50 (25.0)
3-4 times a week	1,274 (10.6)	85 (8.5)	1,038 (10.5)	12 (15.2)	58 (13.4)	6 (10.5)	22 (11.0)
Daily or almost daily	1,361 (11.3)	102 (10.2)	1,035 (10.5)	8 (10.1)	74 (17.1)	10 (17.5)	24 (17.5)

9 Results 2: Survival analysis

9.1 Diabetes and PDAC

Univariate analysis performed on participants with a history of diabetes within the whole matched case-control cohort (n = 12080) showed a consistently strong link between having a diabetes diagnosis and risk for pancreatic cancer. Having diabetes diagnosed by a doctor was associated with an increased risk for PDAC (HR 1.86 95%CI 1.55-2.23; p <0.01) within the nested case-control cohort (Table 8.). Definitions of diabetes incorporating IDC10 data from national records demonstrated a risk association with PDAC. Risk for PDAC with an ICD10 diagnosis of IDDM was 3.91 (95% CI 3.00 – 5.09, p <0.05) and risk for PDAC with an ICD10 diagnosis of NIDDM was 3.67 (95% CI 3.23 – 4.16, p <0.05). The strength of this association remained when the criteria for diabetes diagnosis were combined into variables for diabetes diagnosis that were created specifically for our study ('Record of any diabetes': HR 3.55 95% CI 3.14-4.02, p <0.05; 'Diabetes with date of diagnosis': HR 3.84 95% CI 3.40 – 4.34, p <0.05).

HbA1c as a continuous variable showed significant association with risk for PDAC. For each increase in HbA1c by 1mmol/mol, the risk of PDAC rose by approximately 2% respectively (HR 1.02 95% CI 1.01-1.02, p <0.05). For the purpose of categorising HbA1c measurements obtained at study entry, participants were divided based on specific clinically descriptive ranges corresponding to normoglycaemia, pre-diabetes, and diabetes, depending on where their HbA1c measurements fell. Using HbA1c <42 mmol/mol as the reference value, we identified increased risks for PDAC which were statistically significant among participants falling within the pre-diabetes category (HR 1.80 95% CI 1.46 – 2.23, p <0.05), as well as the diabetes category (HR 2.60 95% CI 2.13-3.16, p <0.05) of measurements. These findings strongly indicated that pre-diabetes range HbA1c as a single measurement may be associated with incident PDAC, independent of diabetes status.

9.2 Risk stratified by HbA1c measurement in participants without prior diabetes diagnosis at study entry

Upon exclusion of participants with a diabetes diagnosis prior to study entry, hazard ratios were calculated to determine risk for PDAC in the remaining individuals after being grouped by the HbA1c range into which their study entry HbA1c measurements fell within.

For participants defined using the 'diabetes diagnosis by a doctor' criteria, calculation of risk for PDAC could not be performed reliably. This is because the diabetes diagnoses identified within this definition were mostly taken from retrospective data collected before UKB

biometric assessments took place. A large proportion of these participants with key exposure outcomes were eliminated as part of the exclusion criteria for this step of analysis, thus limiting the usefulness of this data beyond our previous observations.

In participants who developed a diabetes diagnosis after entry into the study, the risk for PDAC was significantly raised ($p < 0.05$) in those with HbA1c levels taken at study entry within the normoglycaemic range. Hazard ratios ranged from 4.72 to 12.0 when they were grouped according to any of the various diabetes definitions we examined. Participants with an HbA1c measurement within the pre-diabetes range of values, who subsequently developed NIDDM within the study time period demonstrated an increased risk for PDAC (HR 2.40, 95% CI 1.53 - 3.77, $p < 0.05$). Increased risk for PDAC was also observed in the subgroup representing all diabetes diagnoses with linked dates of diagnosis (HR 2.61, 95% CI 1.65 - 4.14, $p < 0.05$). This association did not reach statistical significance for those in the IDDM subgroup (HR 2.50, 95% CI 0.92 - 6.86, $p = 0.07$).

Conversely, participants with no diabetes diagnosis before study entry who entered with an HbA1c measurement within the diabetes range and had an eventual diagnosis of IDDM or NIDDM were associated with a statistically significant risk association with PDAC if their eventual diagnosis was IDDM (HR 2.24, 95% CI 1.19 – 4.25, $p < 0.05$). This association was not seen in the NIDDM subgroup (HR 1.29, 95% CI 1.19 – 4.25, $p < 0.05$).

Table 8. Univariate analysis using cox proportional hazards regression. Hazard ratios (HR) for the various definitions of diabetes (key exposure) are presented with 95% confidence intervals and p-values of significance. From left to right: The first column represents HR calculated for the whole optimised cohort. The second (Normoglycaemic), third (New pre-diabetes), and fourth (New diabetes) columns show HR for these subgroups.

	1:9 case control cohort		**Normoglycaemic (HbA1c = <42 mmol/mol)**		**New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)**		**New diabetes (HbA1c = ≥ 48 mmol/mol)**	
Demographic information	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**
Diabetes diagnosed by a doctor								
No	1.00 (ref)							
Yes	1.86 (1.55-2.23)	<0.05	0.83 (0.52-1.32	00.44	1.06 (0.61-1.78)	0.87	0.80 (0.48-1.35)	0.41
ICD10 Diagnoses (Main or Secondary)								
E10 - Insulin dependent diabetes mellitus								
No	1.00 (ref)							
Yes	3.91 (3.00-5.09)	<0.05	12.0 (8.02-18.0)	<0.05	2.50 (0.92-6.86)	0.07	2.24 (1.19-4.25)	<0.05
E11 - Non-insulin dependent diabetes mellitus								
No	1.00 (ref)							
Yes	3.67 (3.23-4.16)	<0.05	5.11 (4.32-6.04)	<0.05	2.40 (1.53-3.77)	<0.05	1.29 (0.67-2.49)	0.45
Record of any diabetes								

Table 8. Univariate analysis using cox proportional hazards regression. (continued)

No	1.00 (ref)							
Yes	3.55 (3.14-4.02)	<0.05	4.72 (4.01-5.55)	<0.05	-	-	-	-
Diabetes with date of diagnosis								
No	1.00 (ref)							
Yes	3.84 (3.40-4.34)	<0.05	5.47 (4.66-6.41)	<0.05	2.61 (1.65-4.14)	<0.05	-	-
HbA1c								
Continuous (mmol/mol)	1.02 (1.01-1.02)	<0.05	1.09 (1.07-1.11)	<0.05	1.09 (0.96-1.24)	0.16	0.99 (0.97-1.01)	0.27
Categorical								
< 42 mmol/mol	1.00 (ref)							
42 to 47.9 mmol/mol	1.80 (1.46-2.23)	<0.05	1.81 (1.44-2.28)	<0.05	-		-	
>48 mmol/mol	2.60 (2.13-3.16)	<0.05	2.80 (2.15-3.66)	<0.05				

9.3 HbA1c as an independent risk factor

To validate our findings, univariate analysis was repeated after excluding all individuals with a history of diabetes prior to study entry. The likelihood of developing PDAC with HbA1c measurements within the pre-diabetes and diabetes clinical ranges was still increased in both groups. Newly detected pre-diabetes was associated with increased risk for PDAC (HR = 1.81, 95% CI 1.44 - 2.28, p<0.01). When multivariate analysis was performed with adjustment for age, sex, BMI, socioeconomic status, smoking status and alcohol status, the risk continued to be significant (HR = 1.62, 95% CI 1.28-2.05; p<0.01). The results of the univariate analysis calculations for covariates within the adjusted multivariate model are available within the Appendices section. Within the same model shown in Table 9., newly detected diabetes was also associated with a higher risk for PDAC (HR = 2.80, 95% CI 2.15-3.66; p<0.01) versus the new pre-diabetes cohort, which persisted after multivariate analysis with adjustments for the same covariates (HR = 2.56, 95% CI 1.95-3.37; p<0.01).

	Nested case-control cohort (crude)*		Nested case-control cohort (adjusted)**	
	HR 95% CI (lower/upper)	p-value	HR 95% CI (lower/upper)	p-value
HbA1c concentration range (n)				
Continuous	1.02 (1.01 – 1.02)	< 0.01	1.02 (1.01 – 1.02)	< 0.01
<42 mmol/mol (10,889)	1 (ref)		1 (ref)	
New prediabetes (513)	1.81 (1.44 – 2.28)	< 0.01	1.62 (1.28 – 2.05)	< 0.01
New diabetes (258)	2.81 (2.15 – 3,67)	< 0.01	2.51 (1.91 – 3.30)	< 0.01

* matched by age and sex
** matched by age and sex, and adjusted for age, sex, BMI, socioeconomic status, smoking status and alcohol status

Table 9. Table showing hazard ratios (HR) for non-diabetic participants with HbA1c concentration levels within the pre-diabetes and diabetes range of values. The association between newly elevated HbA1c concentration levels in the non-diabetic population into the UK Biobank study and risk for PDAC after adjusting for age, sex, BMI, socioeconomic status, smoking status and alcohol status is significant in the newly pre-diabetic cohort as well as the newly diabetic cohort.

A Kaplan-Meyer graph of failure estimates was plotted to better show the cumulative incidence of the 1208 PDAC cases for the new pre-diabetes and new diabetes cohorts over time, displayed in months (Figure 16.). Those found to be newly pre-diabetic (dark red line) and newly diabetic (dark green line) at study start had a consistently higher probability of developing PDAC up to the study end date when compared with those found to be

normoglycaemic (dark blue line) at the beginning of the study. When 95% confidence intervals were factored in (pale green, pale red, and pale blue areas), a clear separation between the normoglycaemic and pre-diabetic confidence intervals was observed from as early as 12 months beyond the study start date. The graph demonstrated how higher single HbA1c measurements are associated with increased risk of incident PDAC over time, independent of BMI and other potential risk factors for PDAC.

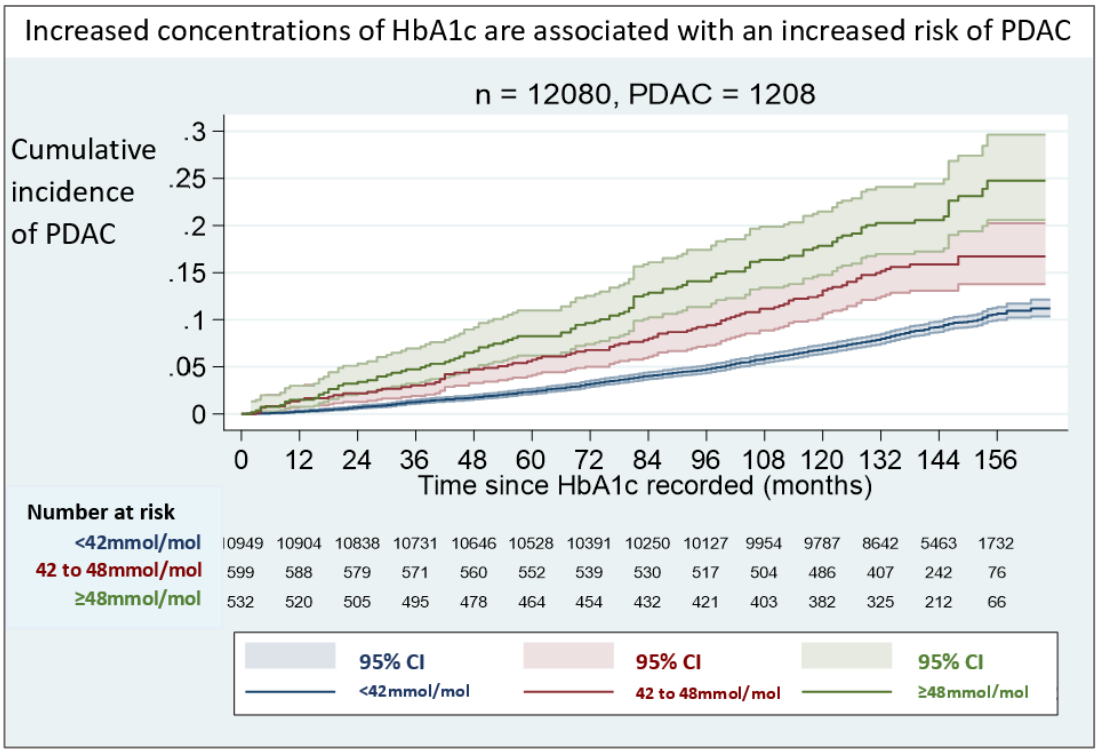


Figure 16. Kaplan-Meyer graph of failure estimates showing cumulative incidence of PDAC diagnoses over time in months for non-diabetic participants on entry into the UK Biobank with normoglycaemic, pre-diabetes range, and diabetes range HbA1c levels on first assessment. Key: <42mmol/mol = normoglycaemia, 42 to 48mmol/mol = new pre-diabetes, ≥48mmol/mol = new diabetes. Table below the graph shows the number of participants remaining at risk at each 12-month interval after numbers are removed for a diagnosis of PDAC, loss to follow-up, and deaths over time.

9.4 The effect of higher measurements of HbA1c on time-to-event PDAC outcomes

To compare the risk for PDAC between groups, log-rank test for equality of survivor functions was performed. These calculations comparing normoglycaemic, newly pre-diabetic and newly diabetic subgroups generated a p-value of <0.05 (See Appendices) allowing us to reject the null hypothesis and determine that differences between the three groups were found to be statistically significant.

Additional log-rank tests were subsequently performed to identify if the differences existed previously were still present when direct 1-to-1 comparisons between each HbA1c category were made (i.e., normoglycaemia vs new prediabetes; normoglycaemia vs new diabetes; new pre-diabetes vs diabetes), with all comparisons returning p-values of <0.05. This showed that a statistically significant difference in probability of developing a PDAC diagnosis that persisted over time existed between all three different clinical classifications based on the HbA1c measurement taken at study start. Furthermore, when we reviewed the mean time-to-event data, we noted that the mean time to PDAC diagnosis for each of these decreased as the HbA1c concentration levels rose, dropping from 7.38 years (±3.33) to 6.35 years (±3.33) to 5.82 years (±3.83) as we moved from the normoglycaemic > new pre-diabetes > new diabetes cohorts respectively. The other attributes of time-to-event data for these categories are shown with their histogram plots in Figure 17. (normoglycaemia), Figure 18. (New pre-diabetes), and Figure 19. (New diabetes). Right skewness was also noted to increase with each elevation in HbA1c concentration range (Skewness = -0.295 > - 0.245 > 0.187), although variance also increased.

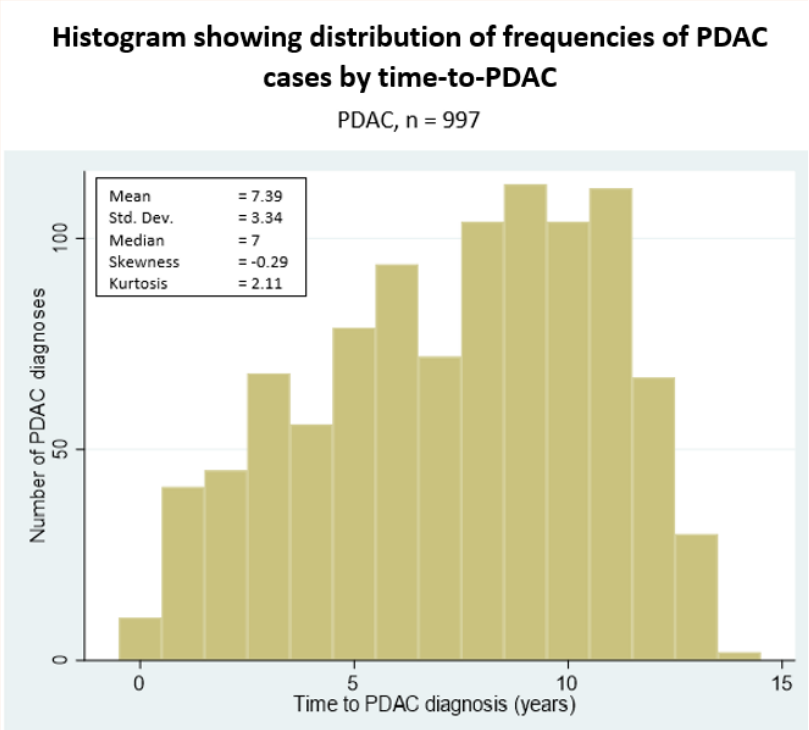


Figure 17. Histogram of time to PDAC diagnosis in participants with no previous history of diabetes and a normoglycaemic HbA1c concentration level at initial assessment. X-axis: time to PDAC diagnosis in years, Y-axis: frequency of PDAC diagnoses in each time interval

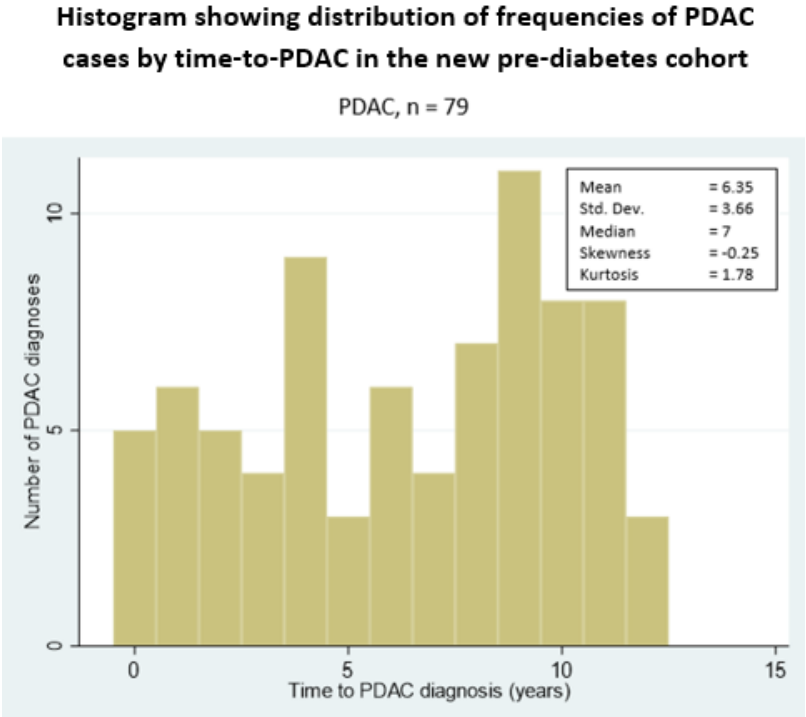


Figure 18. Histogram of time to PDAC diagnosis in participants with no previous history of diabetes and a new pre-diabetes HbA1c concentration level at initial assessment. X-axis: time to PDAC diagnosis in years, Y-axis: frequency of PDAC diagnoses in each time interval

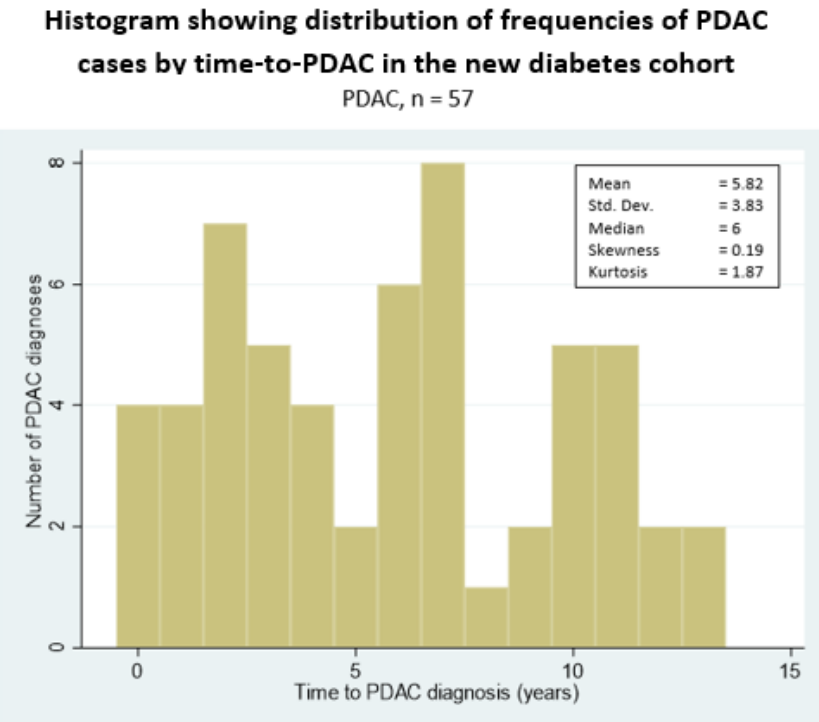


Figure 19. Histogram of time to PDAC diagnosis in participants with no previous history of diabetes and a new diabetes HbA1c concentration level at initial assessment. X-axis: time to PDAC diagnosis in years, Y-axis: frequency of PDAC diagnoses in each time interval

Coefficients were generated using Cox regression analysis of time-to-event association for normoglycaemic, pre-diabetes, and diabetes cohorts of HbA1c concentration levels who develop PDAC. Participants without a previous diagnosis of diabetes at study entry with normoglycaemic HbA1c levels taken at study start had an HR-associated coefficient of -2.40 (95% CI -2.52 to -2.28, p <0.05). Participants with single measurements of HbA1c at study entry within the new pre-diabetes range showed coefficients of -2.48 (95% CI -2.99 to -1.97, p <0.05), participants with HbA1c measurements within the new diabetes range showed coefficients of -2.56 (95% CI -3.24 to -1.88, p <0.05). The more negatively trending coefficients seen with each elevation in HbA1c concentration range indicated that rises in HbA1c concentrations are not only proportional to the risk for developing PDAC but are also associated with shorter time-to-event intervals to developing a PDAC diagnosis.

10 Results 3: Testing the predictive performance of HbA1c

10.1 Predictive value of HbA1c for PDAC in participants with new pre-diabetes and new diabetes diagnosis at study entry.

For individuals with no prior history of diabetes and a new pre-diabetes or new diabetes diagnosis on entry into the study, HbA1c failed as a standalone predictive marker for PDAC. ROC curve graph plots for both groups gave area under curve (AUC) values of 0.593 after adjusting for age, sex, BMI, socioeconomic status, smoking status and alcohol status (Figure 21.).

Dichotomising the predictive value of HbA1c according to optimal cut-point thresholds failed to generate a diagnostic threshold that was useful for predicting risk of PDAC in this participant cohort. Youden index modelling for participants with a new pre-diabetes diagnosis generated an empirical optimal cut-point of HbA1c = 36.45 mmol/mol (Youden index = 0.121; Sensitivity = 0.45; Specificity = 0.66), which was within the normoglycemic range. Youden index calculations for the new diabetes diagnosis group yielded the same optimal cut-point of HbA1c = 36.45 mmol/mol (Youden index = 0.137; Sensitivity = 0.67; Specificity = 0.57).

10.2 Predictive value of HbA1c for PDAC in individuals with known diabetes and elevated HbA1c on study entry

HbA1c also failed as a predictive marker for PDAC when used alone in the groups of participants with known history of diabetes who would subsequently develop PDAC. ROC curve plots with AUC calculations for these participants were 0.581 for HbA1c in the pre-diabetes range, and AUC = 0.587 for those with an HbA1c in the diabetes range (Figure 22.).

Diabetic participants with HbA1c levels in the pre-diabetes range had an optimal cut-point of HbA1c = 35.05 mmol/mol; Youden index = 0.104; Sensitivity = 0.57; Specificity = 0.53). The optimal cut-off point for diabetic participants with HbA1c concentration levels in the diabetes range at study entry was 36.45 (Youden index = 0.116; Sensitivity = 0.43; Specificity = 0.68). As with all the other cohorts in this study, these cut-points fell well within the normoglycaemic range, demonstrating that basing prediction models for PDAC on categorical thresholds for HbA1c as a standalone test is not a viable strategy.

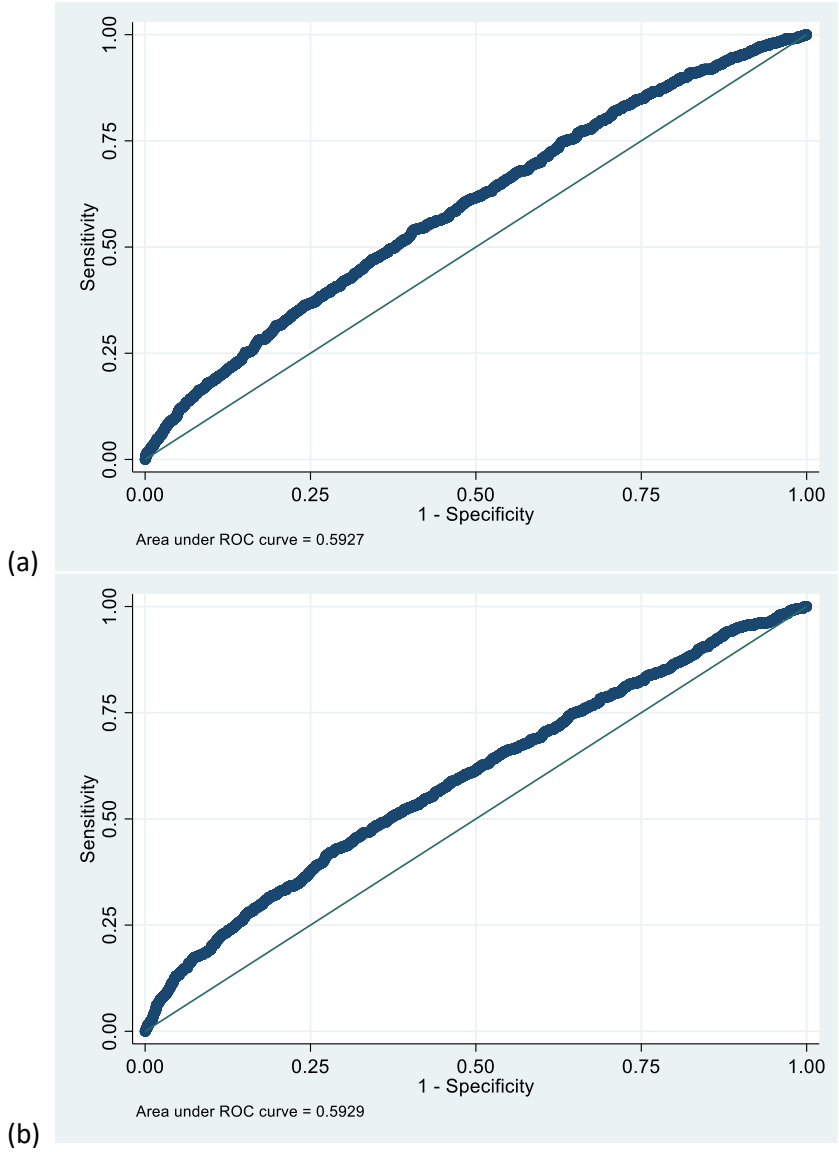
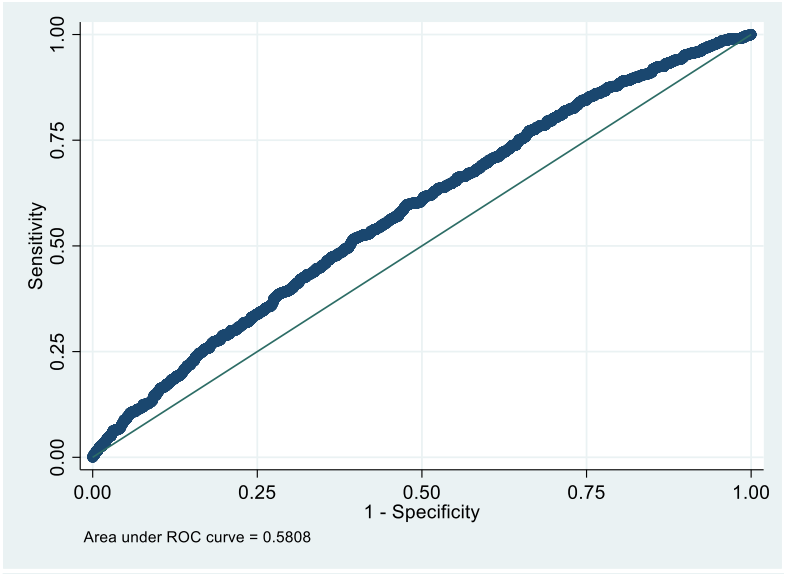


Figure 21. ROC curve analysis with and AUROC curve values representing the predictive performance of HbA1c for PDAC in (a) individuals with new pre-diabetes diagnosis (AUC = 0.593) and (b) individuals with a new diabetes diagnosis (AUC = 0.593). The sensitivity (y-axis) and 1-specificity (x-axis) of HbA1c for PDAC at each HbA1c concentration level recorded at study entry for included participants are plotted (blue curved line).

(a)



(b)

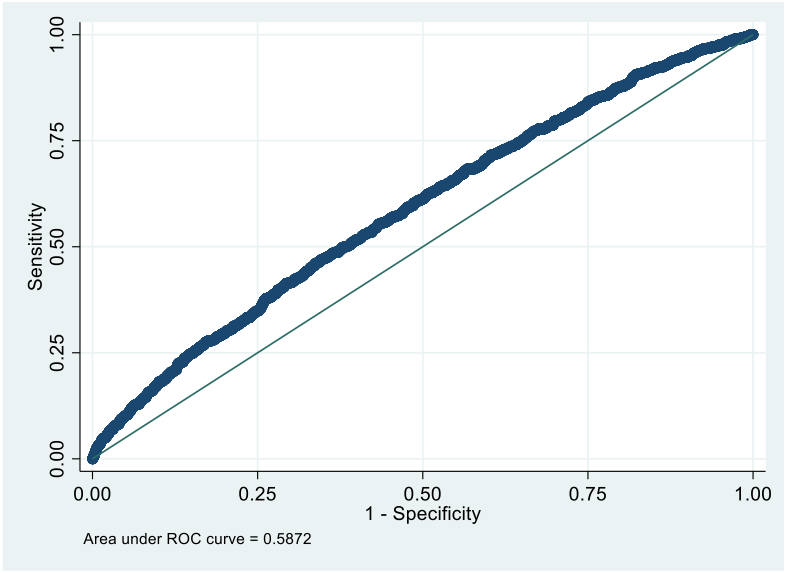


Figure 22. ROC curve graph plots and AUROC curve value demonstrating predictive performance of HbA1c for PDAC in participants with pre-existing diabetes whom either (a) have HbA1c concentration levels in the pre-diabetes range at study entry (AUC = 0.581), and (b) have HbA1c concentration levels in the diabetes range (AUC = 0.587). The sensitivity (y-axis) and 1-specificity (x-axis) of HbA1c for PDAC associated with each HbA1c value recorded at study entry for included participants are plotted (blue plotted dots forming the curved line).

The sensitivity and specificity at cut-off points for the pre-diabetes diagnostic threshold of 42mmol/mol and the diabetes threshold ≥48 mmol/mol in both the NODM and LSDM cohorts were calculated. In all groups, HbA1c showed low sensitivity and high specificity test performance outcomes. This indicated that using HbA1c as a standalone predictive test for PDAC would carry a higher risk of missing PDAC diagnosis where one exists (low sensitivity),

despite demonstrating a strong ability to ensure positive tests detected are truly positive (high specificity) (Table 10.).

Diabetes status	HbA1c Cut-off point	Sensitivity (%)	Specificity (%)
NODM	42 mmol/mol 48 mmol/mol	19.94 10.86	91.15 95.98
LSDM	42 mmol/mol 48 mmol/mol	10.30 5.81	94.48 97.83

Table 10. HbA1c is unable to act as a predictive marker for PDAC when used alone. The low sensitivity (%) and high specificity (%) for cut-off points on the ROC curve of 42 mmol/mol (pre-diabetes diagnostic threshold) and 48 mmol/mol (diabetes diagnostic threshold) makes HbA1c a test that is likely to miss many underlying diagnoses. Setting categorical cut-off points for HbA1c is of no benefit to its usefulness as an early risk marker for PDAC.

11 Discussion

In this study, we explored the value of HbA1c as a predictive marker of risk for PDAC. We demonstrated that for participants with DM, a rise in HbA1c concentration levels by any magnitude raises the overall risk for PDAC in a proportional manner without violating the proportional hazards assumption. We were also able to demonstrate that for individuals with no prior history of diabetes, higher HbA1c levels at study entry were associated with a higher likelihood of having a shorter time-interval until diagnosis of PDAC. This was despite the absence of repeat HbA1c test results to explore this relationship in greater detail, which was a significant limitation of the UK Biobank dataset. ROC curve analysis confirmed that HbA1c fails as a predictive marker for PDAC with NODM and LSDM individuals when used as a standalone test. Finally, optimal cut-off points derived from the ROC curves showed that categorising HbA1c does not make it a useful predictive marker of risk for PDAC when used as a single value measure. Generally, the findings of our study were in keeping with the work of other authors. Validation of our findings and a review of the limitations, bias and confounding identified in our study will be discussed further in this section.

11.1 Validation of HbA1c as a marker of diabetes

To validate HbA1c as a surrogate marker of diabetes within our dataset, we compared adjusted HRs for HbA1c as a continuous and categorical marker internally against other UKB dataset definitions for diabetes and found them to be in concordance. The other definitions of diabetes we examined within the UKB dataset also yielded HRs that were in keeping with the increased risk found in previous studies. For example meta-analyses and pooled analyses suggested that in diabetic cohorts, pancreatic cancer risk increased by 1.5-2.0-fold [138, 202] and one study found that pancreatic cancer risk increased by 1.77-fold in individuals with pre-existing diabetes [203]. A large-scale study in Korea also demonstrated that diagnosis of diabetes prior to pancreatic cancer diagnosis was a significantly independent risk factor on multivariate analysis (HR = 1.48; 95% CI, 1.43-1.53; $p < 0.0001$) [204]. Although our HR yields were higher than those identified in the external studies this could be due to many factors such as demographic differences in ethnicity, and criteria for diabetes definition [204].

11.2 Elevation in HbA1c and PDAC

Higher HbA1c concentration levels measured at study entry were associated with significant rises in adjusted HR for PDAC when HbA1c was treated as a continuous variable and when grouped into clinically defined diagnostic thresholds for pre-diabetes and diabetes. In the

sub-cohort of participants with no prior history of diabetes, we showed that elevated HbA1c concentrations were independently associated with increased risk of developing PDAC after multivariate analysis with adjustment for age, sex, BMI, socioeconomic status, smoking status and alcohol status, which was in keeping with previous studies. This risk association was significant for those with HbA1c measurements within the pre-diabetes and diabetes ranges. We also demonstrated that when treated as separate categorical cohorts of individuals, the differences in risk for PDAC associated with newly pre-diabetic individuals compared with risk for PDAC in newly diabetic individuals was statistically significant. This may advocate the incorporation of HbA1c in either continuous or categorical form into future strategies for early detection. For example, as a continuous variable, routine monitoring for incremental changes over time may provide the earliest clues to an underlying and as-yet undetectable lesion, prompting closer monitoring or further investigation. Such a protocol might overlap with protocols where rising HbA1c levels reaching thresholds for pre-diabetes or diabetes might trigger further investigation and follow-up.

11.3 HbA1c analysed at single time points

Categorisation of continuous risk factors is common practice in observational epidemiological studies. Good practice recommendations include appropriate choice of groupings and explanation for grouping, clear analysis strategies and presentation of findings, and drawing valid conclusions from categorical analyses, avoiding injudicious use of multiple alternative analyses [205]. However, the simplicity gained from categorisation can also lead to loss of statistical power in studies and incomplete correction for confounding factors [206]. In this study, the direct correlation between higher single time-point HbA1c concentration levels at study entry and increased risk for PDAC supports the use of individual HbA1c values in statistical analysis. However, we also demonstrated that using single HbA1c values as a standalone marker to predict risk of PDAC based on optimal cut-off values is not a useful approach. Based on our findings we recommend that future studies utilise HbA1c single values alongside interval HbA1c repeat measurements for the purpose of risk modelling as it provides the most accurate measure of any manifestation of changes which may be associated with an underlying PDAC diagnosis over time.

11.4 Risk of PDAC in newly diabetic vs non-diabetic population cohorts

In our study, the average length of time from the measurement of HbA1c at study entry to PDAC diagnosis was 1.5 years less for participants with measurements within the new diabetes range compared to those with normoglycaemia at study entry. Although the range

of time-to-events overlapped between sub-cohorts, the shorter mean time to PDAC diagnosis in the diabetes sub-cohort could be interpreted as a greater need to prioritise the NODM cohort over the normoglycaemic non-diabetic cohort with regards to referral for further investigations.

Although our study suffered a significant limitation from the absence of HbA1c repeat values taken at regular intervals, our ability to demonstrate that higher HbA1c concentration levels are also associated with a higher likelihood of quicker PDAC diagnosis is supported by the results of several other studies that had repeat HbA1c measurements over time to examine this hypothesis in greater detail. For example, hyperglycaemia based on FBG profiles has been seen to arise up to 36 months before PDAC diagnosis in those with or without diabetes at baseline, and in those with or without resection at diagnosis [101]. FBG levels were also shown to increase with tumour volume, although activity was not associated with tumour grade. More recently Lemanska et al. published the first population-based study using large nationally representative database from the UK to demonstrate the temporal associations between HbA1c and pancreatic cancer [207]. In this matched case-control study, data from 590 primary care practices in England identified 8,777 patients with a diagnosis of pancreatic cancer between January 2007 and August 2020, which were matched by age, gender and diabetes to 34,979 controls. Log regression models were calculated for HbA1c data for a cohort of individuals with a history of diabetes and for a cohort without diabetes at six points in time, starting from five years before the diagnosis of PDAC. Adjustments were made for BMI, other HbA1c measurements within the same year, ethnicity, index of multiple deprivation as quintiles, smoking and alcohol consumption. Longitudinal plots for HbA1c demonstrated an increase before diagnosis of pancreatic cancer in both the non-diabetic and diabetic subgroups, with changes starting about 2 years earlier in the diabetic subgroup compared to cases without diabetes (-3 years verses -1year). The degree of increase was also larger in the diabetic group as compared to the non-diabetic group. Interestingly, average HbA1c for the subsample without diabetes was 43.1 (95%CI 42.6 to 43.6) at baseline which is consistent with our findings that non-diabetic hyperglycaemia still carries a significant increase in risk for PDAC. The pattern of changes in HbA1c levels associated with diagnosis of pancreatic cancer differed between the diabetes and non-diabetes subgroups. A higher risk of pancreatic cancer was associated with an increase in HbA1c for people without diabetes than for people with diabetes. For a 10 mmol/mol increase in HbA1c, non-diabetic individuals were 2.1 (95% CI 1.9 to 2.4, $p < 0.01$) times more likely to be diagnosed with pancreatic cancer than people whose HbA1c did not increase. For the diabetic subgroup, the

increased likelihood was significant but to a lesser degree, with the likelihood increased by 1.4 times (95% CI 1.9 to 2.4, p<0.001). The length of time over which HbA1c change occurred also had a different effect on the two subgroups. Increases in HbA1c by 1mmol/mol between cases and controls in the year before pancreatic cancer was associated with an adjusted OR of 1.04 (95% CI 1.03 to 1.04, p <0.001) for people with diabetes and 1.09 (95% CI 1.07 to 1.11, p < 0.001) for people without diabetes.

In a different study, Wu et al sought to apply criteria other than those traditionally used in the definition of diabetes to improve sensitivity for early detection of patients with pancreatic cancer [208] . They concluded that screening with pancreatic cancer based solely on elevation in HbA1c was not an effective way, and that factors such as racial/ethnic differences in cancer risk should be taken into consideration while examining the effect of glycaemic parameters on risk of cancer [208, 209].

In our study of individuals without a prior diagnosis of diabetes before study entry who subsequently developed DM within the study time period, we identified risk for PDAC which appeared to decrease in association with higher HbA1c levels measured at study entry. This pattern of risk appeared to be in direct contradiction to our predictions that higher HbA1c measurements would be associated with proportionally increased risk for PDAC. However, on further examination of the data, we recognised that a significant number of individuals entering our study with newly identified dysglycaemia developed PDAC diagnoses within a very short period early into the study (i.e., within the first 12 months) which provided a clue as to the cause of our abnormal findings. Following the generation of our outcomes within the limitations of the study, a colleague within our research group re-examined the dataset, taking an alternative approach in determining the association between HbA1c and PDAC [210]. In addition to the positively skewed number of PDAC cases identified towards the start of the study period, the strength of association between the HbA1c parameters and incident PDAC attenuated over time. This revealed the non-linear relationship between exposure and outcome, which was subsequently acknowledged to cause a violation of the proportional hazards assumption. This violation raised the possibility that the non-linear relationship between HbA1c and PDAC identified in our participant cohort was likely due to the presence of reverse causality [210]. Such a violation of the proportional hazards assumption could also explain why examination of the data within the limits of our own approach resulted in seemingly conflicting findings. Within this extended study, the alternative approach applied new time-varying coefficient interaction terms which were created to link HbA1c at the time of enrolment to censor date and DM range (categorically, those with HbA1c within the pre-

diabetes and diabetes range of values) for participants with no prior diagnosis of diabetes. The time-varying coefficient interactions for this group were then stratified into 12-monthly time intervals. The hazard ratios generated thus took the varying interactions over time into account. In this study by McDonnell et al., for those without prior diagnosis of DM, an inverse association between HR and duration of follow-up was identified. Hazard ratios for newly dysglycaemic individuals were 2.10 (95% CI 1.31 – 3.37, p <0.002) and 8.55 (95% CI 4.58-15.99, p = 0.002) after multivariate adjustments [210], confirming our original findings but showing an even stronger association between single HbA1c measurements within these dysglycaemic range cohorts and PDAC once the violation of proportional hazards assumption was accounted for.

11.5 Limitations of the UK Biobank dataset

All epidemiological studies are subject to multiple sources of chance, bias and confounding with measures taken to mitigate these as much as possible. Although the results of our study may reflect the true effect of elevated Hba1c concentration levels on risk for pancreatic cancer, there is always the possibility that the findings may be due to an alternative explanation. Findings because of chance, bias, or confounding may have provided a singular or combined effect of generating spurious results, giving the impression that a valid statistical association existed where one did not actually exist. Conversely, such results could have led to the conclusion that no association exists when one was present.

For example, diabetes is a complex disease that can be diagnosed through a combination of presenting symptoms and clinical investigations. In this study, DF variables for diabetes-related measures ranged from self-completed questionnaire data to hospital coded ICD10 records. Questionnaire and interview answers rely heavily on memory and are a significant source of recall bias. Although these were overcome largely by the inclusion of data from medical records, our study required a further level of detail in the form of dates of diagnosis. The primary source of dates of diagnosis in our study came from ICD10 records. These dates were provided within the context of the clinical circumstances in which the diagnoses were given. For example, Hospital Record level data may have provided a date of diagnosis of diabetes based on coding as a secondary diagnosis in relation to a primary condition that led to the individual's admission to hospital. The diabetes diagnosis may have been present before, or undiagnosed until hospital admission. Variability in presentation of diabetes and patient-led management through diet control and lifestyle factors also influences the rate of disease progression. In the primary care setting, disease presentation, management, and

effectiveness of monitoring by both the patient and clinician would all have an influence on when a date of diagnosis was determined.

All data within our study was derived from the UKB database for exploration of the risk association relationship between HbA1c and PDAC. For any inferences drawn from the study to be warranted, the effects of potential bias, chance, and confounding had to be accounted for and minimized where possible. UKB performed its own set of internal validation steps prior to releasing the data it had collected to researchers. The details of which regarding initial study protocols and strategy for obtaining healthcare data can be found in the original study rationale [211]. This meant that pre-existing selection and information bias arising from the UK Biobank dataset became a bias in our own study, in addition to any further bias arising from our own methodology. One of the major forms of selection bias encountered on recruitment of participants for the UKB study was identified as the "healthy volunteer effect" [212]. On average, individuals volunteering for research studies tend to be more health-conscious than non-volunteers and the voluntary basis on which participants enrolled into the UKB made this cohort vulnerable to the same effects. To understand more about the value of UKB data as a representative sampling of the general population, a study by Fry et al in 2017 examined if the differences between the UK Biobank cohort and different sampling frames (e.g. Health Survey for England, 2008; HSE 2008) existed with regard to a range of characteristics due to this effect [212]. When compared with the same variables within the HSE census data for England, Scotland, and Wales, participants within the UKB study cohort were on average, taller, leaner and had a smaller waist circumference than the general population. Mean body mass index in UK Biobank men and women was 27.9 and 27.3 respectively, compared with 28.5 and 28.0 in the general population based on data from the HSE 2008. UK Biobank men and women were also less likely to be obese (BMI >=30) across all age groups examined in comparison with the general population. The percentage of men and women categorised as clinically obese and aged 55-64 years in the UK Biobank was 26.3% and 23.0% respectively, compared with 35.6% and 32.2% of the UK population from HSE 2008 data.

11.6 Other sources of bias, and study limitations

In this study the information used to define prevalent diabetes for each participant was reliant on the quality and detail of UK Biobank baseline assessment data, and on our ability to identify dates of diabetes diagnosis occurring by this date.

Self-reported data from questionnaire-derived data fields 'Diabetes diagnosed by a doctor' and data 'Age diabetes diagnosed' were subject to recall bias, a known and major limitation with regards to data quality [213, 214]. For example, the prevalence of self-reported diabetes within the UKB was lower than the general population, consistent with previous comparisons of UKB with data obtained from HSE 2009 data [215]. In the study by Fry et al., 4.5% of men within the UKB cohort aged 45-54 were identified as having diabetes verses 8.1% in HSE 2009. This pattern persisted with older men (Age 55-64 years, UKB = 7.8% vs. HSE = 10.5%), and women in the lower and upper age cohorts (Age 45-54 years, 2.4% and 3.5% respectively; Age 55-64 years, 6.3% and 8.0% respectively). For further comparison, Public Health England released a diabetes prevalence report in 2016 based on HSE data from 2012,2013 and 2014, showing that the expected diabetes prevalence for England in 2015 for all adults aged 45-54 years was 9.0% (Public Health England, Diabetes Prevalence Model. September 2016 publication).

Incorporating data from linked ICD-10 Hospital records enabled diagnoses of diabetes to be obtained directly from clinical records databases. However, using secondary care data yielded its own forms of bias as secondary care admission data with coding for diabetes in most cases is not the primary or secondary cause for hospital admissions (Eastwood et al).

A study by Eastwood et al. examining the quality of UKB data aimed to develop algorithms to define prevalent and incident diabetes for UK Biobank with the further aim of having these implemented by the UKB for results to be made available to researchers on request [216]. The results of this study confirmed that just under half of those identified with diabetes in primary care also had a record of this diagnosis on hospital admission data. As can be expected, the cohort of individuals were typically older, more likely to be on insulin, and more likely to have diabetes complications than those without secondary care evidence of diabetes. For individuals identified with diabetes from secondary care data, a third had no evidence of diabetes in primary care. Where a proportion of individuals were identified as having type 2 diabetes from secondary care records alone, there was less evidence of antidiabetic medication use, hyperglycaemia, and microvascular complications in their primary care records compared to those identified with the Read code for T2DM in primary care. From the information provided by Eastwood et al. these significant gaps are indicative of the various risk factors, presentation of complications and diabetes management seen in clinical practice and the challenges faced in deriving accurate diagnostic data from individuals across the full spectrum of clinical diabetes diagnoses.

A more optimal method for deriving diabetes diagnosis dates from linked healthcare records comes from having sufficient primary care and secondary care data for all participants within the dataset. A mid-point between the dates of last primary care consultation or hospital admission and the date of first diabetes code could then be taken as the event date and be closer to the true onset of diabetes than the date of first diabetes code alone.

Having incomplete primary care record data for the full UKB cohort was a significant limitation for our study, as most of our pancreatic cancer case participants were not included within the primary care data that was available at the time. As of 2016, UKB was only able to obtain GP primary care data for half of the volunteer cohort. For the remaining participants, UKB had to request permissions from each Primary Care group/practice, which was then given on a case-by-case basis. This process, combined with varied individual concerns about handling of confidential patient data added complications that hindered the UKG group aims of building a complete dataset of primary care items. Without primary care data for the full cohort, we were unable to implement a more refined method to define date of diabetes diagnosis for each participant case and were reliant on the earliest date of diagnosis data from ICD10 records provided by UK Biobank.

11.7 Confounding

Confounding variables are often defined as variables that correlate positively or negatively with both the dependent variable and the independent variable [217]. The presence of these variables affects the variables being studied so that the results do not reflect the actual relationship between the variables under study [217].

In this study, the nested-cohort design enabled us to examine those with and without the key exposure (Hba1c at different clinically diagnostic thresholds, and as a continuous variable), and see how they differed with respect to the outcome. This type of study design requires the elimination of confounding and to approximate a randomizer control trial as best possible. As such, matching, weighting, and regression-based methods were appropriate.

A major strength of our study was the implementation of an experimental design that sought to minimize the effects of Type I error. For example, we described in detail the use of power and sample size calculations to generate an adequate number of controls to be matched with cases in our case control study. Working with the large UK Biobank base-cohort, calculations for sample size allowed us to achieve greater power despite the relatively small case

outcome numbers available. The large control population made matching by 1 case to 9 controls possible, which was more than the required minimum to achieve the power and sensitivity demanded of the study. This also helped avoid type II errors by adjusting sample size to account for the power demanded of the study to detect a result at the level of significance required.

The process of matching was also restricted by selection of participants within the full biobank population based on inclusion and exclusion criteria to eliminate potential confounding by age, sex, and other potential confounders such as a prior history of non-melanoma cancer. In the statistical analysis phase of the study, cox regression and log regression analyses were all performed with adjustments for covariates to also control for confounders.

12 Conclusion

Our study confirmed that single time-point HbA1c concentration levels are not helpful as a standalone predictive marker for PDAC. However, a standalone HbA1c result combined with knowledge of previous diabetes status is capable of estimating levels of risk for underlying PDAC in individuals found to be newly pre-diabetic or newly diabetic. The findings of our study and follow-up study by McDonnell et al. also highlighted the importance of recognizing the potential for violation of the proportional hazards assumption when reverse causality becomes a factor in the risk relationship between HbA1c and PDAC. Careful examination of the data for violation of the proportionality assumption and the adjustment for this with time-varying coefficient interaction terms may be an important step to refining our approach to HbA1c as grouped values with the aim of integrating this biomarker into future composite risk model building strategies.

In conclusion, this study demonstrated that for non-diabetic individuals aged over 50 in the UK, newly elevated HbA1c concentration levels measured within the pre-diabetes or diabetes ranges of dysglycaemia without a prior history of diabetes had an increased likelihood of underlying PDAC being the primary cause. The current NICE criteria advise a combination of NODM diagnosis, age over 60 and evidence of weight loss to be identified in an individual for their risk to be sufficient for urgent referral for further investigations. Our study has shown that future thresholds for referring individuals for specialist investigations might benefit from inclusion of individuals aged 50 and above, with new pre-DM or NODM diagnosis as part of the clinical evidence for further investigations. Repeated HbA1c testing over time to generate longitudinal data may play an important and vital role in monitoring for changes to glycemic status associated with increased risk for PDAC. Performing more prospective large cohort studies with time-dependent follow-up datasets and repeat HbA1c values will enable researchers to study variations in sequential HbA1c values and determine how the magnitude of change in HbA1c over time may modulate the risk for PDAC. It is also our recommendation that future studies should consider combining HbA1c testing with novel methods such as polygenic risk scoring. This approach may enhance front-line clinical risk detection models that are needed to quickly identify individuals most at risk of sporadic PDAC sooner, with the overall aim of achieving better outcomes by improving opportunities for curative treatment and reducing overall mortality.

13 Appendices

13.1 Key of UK Biobank data fields used in this study

Data Field (DF)	Description
31	Sex
48	Waist circumference (cm)
56	Date of attendance to a UK Biobank assessment centre
189	Townsend deprivation index (TDI)
1558	Alcohol intake frequency
2443	Diabetes diagnosed by a doctor
2976	Age diabetes diagnosed (linked to 2443)
3456	Number of cigarettes currently smoked amongst current cigarette smokers
21001	Body mass index (BMI)
21003	Age when attended a UK Biobank assessment centre
20116	Smoking status
20117	Alcohol drinker status
30750	HbA1c concentration level measurements (mmol/mol)
40001	Primary cause of death
40002	Contributory (secondary) cause of death
40006	Type of cancer: ICD10
40011	Histology of cancer tumour -Record of histological subtype from records
41270	Diagnoses - ICD10; Record of primary or secondary diagnoses from hospital patients records
41280	Date of first inpatient Diagnosis - ICD10 (linked to 41270 - ICD10 record of diagnosis)
130706	Date E10 first reported – Insulin-dependent diabetes mellitus
130708	Date E11 first reported – Non-insulin-dependent diabetes mellitus

13.2 Additional Results for Chapter 7: Statistical analyses

13.2.1.1 Age at study start

```
. stcox age_attendukb, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -11150.636
Iteration 1:   log likelihood = -11150.575
Iteration 2:   log likelihood = -11150.575
Refining estimates:
Iteration 0:   log likelihood = -11150.575

Cox regression -- Breslow method for ties

No. of subjects =       12,080                 Number of obs    =       12,080
No. of failures =        1,208
Time at risk    =      1643786
                                               LR chi2(1)       =         0.12
Log likelihood  =   -11150.575                 Prob > chi2      =       0.7266
```

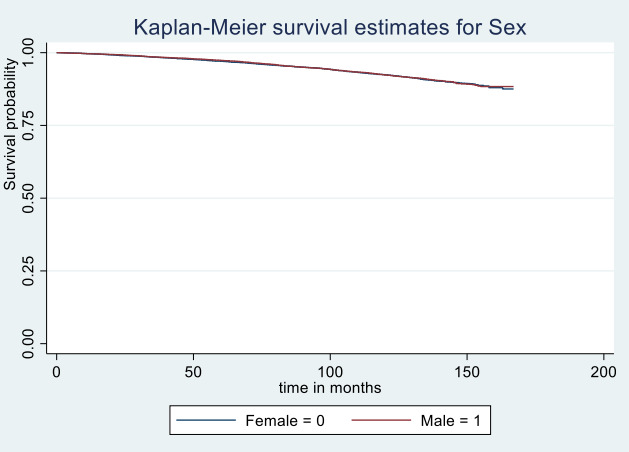
_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age_attendukb	.0012089	.0034597	0.35	0.727	-.005572	.0079897

13.2.1.2 Sex

```
. sts test sex, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions
```

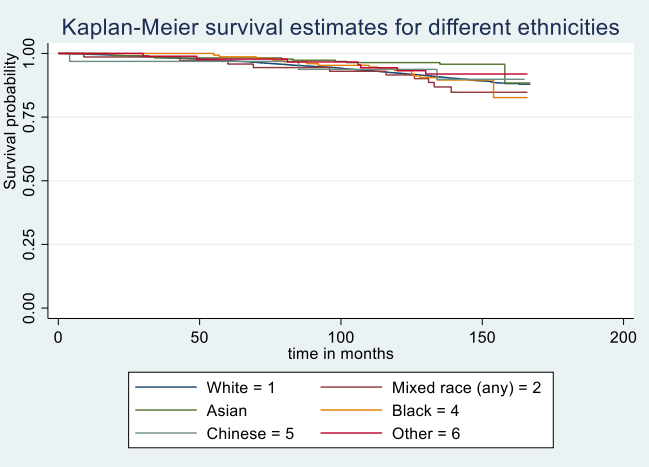
sex	Events observed	Events expected
0	582	580.24
1	626	627.76
Total	1208	1208.00

```
          chi2(1) =       0.01
          Pr>chi2 =     0.9191
```



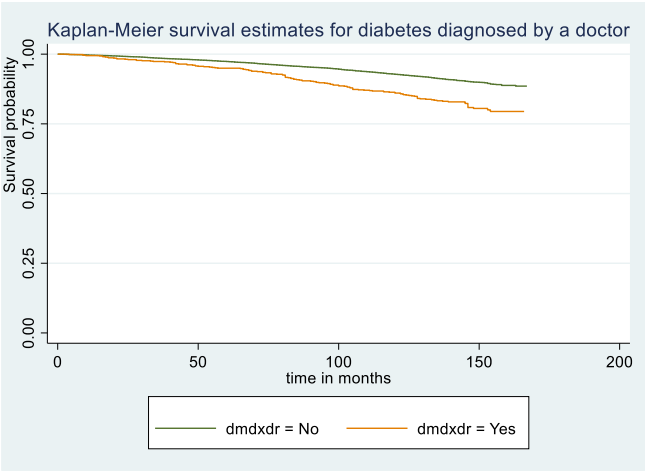
13.2.1.3 Ethnicity

```
. sts test ethnicity, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
ethnicity    |  observed     expected
-------------+------------------------
1            |      1156      1145.71
2            |        10         7.15
3            |        10        22.30
4            |        16        14.68
5            |         3         3.27
6            |         7         8.89
-------------+------------------------
Total        |      1202      1202.00

                chi2(5) =       8.56
                Pr>chi2 =     0.1277
```



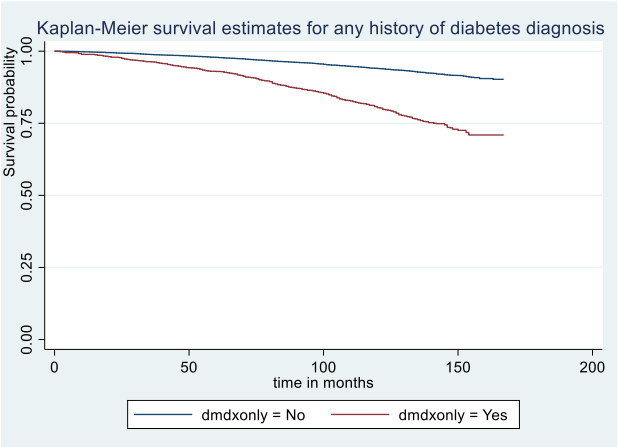
13.2.1.4 Diabetes diagnosed by a doctor (dmdxdr)

```
sts test dmdxdr, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
dmdxdr       |  observed     expected
-------------+------------------------
-3           |         1         0.52
-1           |         6         2.87
0            |      1067      1132.27
1            |       132        70.35
-------------+------------------------
Total        |      1206      1206.00

                chi2(3) =      61.73
                Pr>chi2 =     0.0000
```



13.2.1.5 Any diabetes diagnosis (dmdxonly)

```
sts test dmdxonly, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
dmdxonly     |  observed     expected
-------------+------------------------
0            |       819      1047.34
1            |       357       128.66
-------------+------------------------
Total        |      1176      1176.00

                chi2(1) =     455.73
                Pr>chi2 =     0.0000
```



13.2.1.6 Age at diabetes diagnosis

```
. stcox mage_dm, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -2571.6898
Iteration 1:   log likelihood = -2570.0984
Iteration 2:   log likelihood = -2570.0977
Refining estimates:
Iteration 0:   log likelihood = -2570.0977

Cox regression -- Breslow method for ties

No. of subjects =        1,403                 Number of obs    =       1,403
No. of failures =          369
Time at risk    =       173974
                                               LR chi2(1)       =        3.18
Log likelihood  =   -2570.0977                 Prob > chi2      =      0.0744
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mage_dm	.0007323	.0004133	1.77	0.076	-.0000778	.0015424

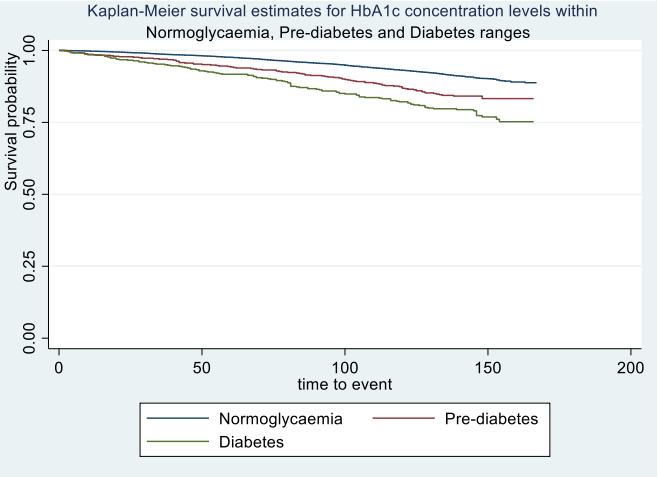
13.2.1.7 HbA1c (continuous variable)

```
. stcox hba1c0_cont, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -11150.636
Iteration 1:   log likelihood =  -11129.12
Iteration 2:   log likelihood = -11118.011
Iteration 3:   log likelihood = -11116.364
Iteration 4:   log likelihood =   -11116.3
Iteration 5:   log likelihood =   -11116.3
Refining estimates:
Iteration 0:   log likelihood =   -11116.3

Cox regression -- Breslow method for ties

No. of subjects =       12,080                 Number of obs    =      12,080
No. of failures =        1,208
Time at risk    =      1643786
                                               LR chi2(1)       =       68.67
Log likelihood  =     -11116.3                 Prob > chi2      =      0.0000
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
hba1c0_cont	.017706	.0014616	12.11	0.000	.0148414	.0205706

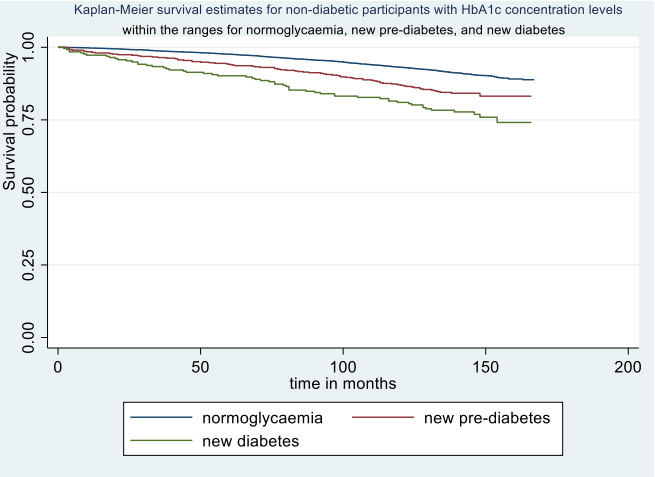
13.2.1.8 HbA1c (clinical categorical variable)

```
. sts test hba1c0_grp, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
hba1c0_grp   |observed     expected
-------------+------------------------
0            |    1006      1105.40
1            |      92        56.07
2            |     110        46.53
-------------+------------------------
Total        |    1208      1208.00

                chi2(2) =     118.68
                Pr>chi2 =     0.0000
```



13.2.1.9 New HbA1c concentration in Pre-diabetes or Diabetes range

```
sts test new_pdm_dm, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
new_pdm_dm   |observed     expected
-------------+------------------------
0            |     998      1065.71
1            |      79        46.58
2            |      57        21.71
-------------+------------------------
Total        |    1134      1134.00

                chi2(2) =      84.29
                Pr>chi2 =     0.0000
```



13.2.1.10 Townsend Deprivation Index (socioeconomic status)

```
stcox tdindxukb, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
 enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -11140.443
Iteration 1:   log likelihood = -11139.341
Iteration 2:   log likelihood = -11139.341
Refining estimates:
Iteration 0:   log likelihood = -11139.341

Cox regression -- Breslow method for ties

No. of subjects =       12,069                 Number of obs    =      12,069
No. of failures =        1,207
Time at risk    =      1642227
                                               LR chi2(1)       =       2.20
Log likelihood  =   -11139.341                 Prob > chi2      =     0.1377
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tdindxukb	.0136421	.0091323	1.49	0.135	-.0042569	.0315411

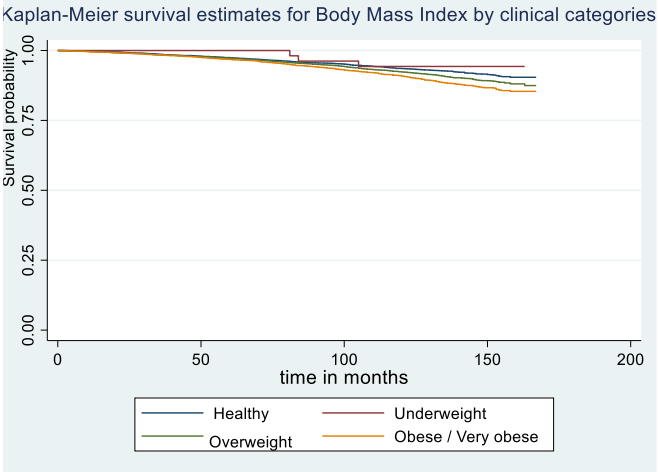
13.2.1.11 Body Mass Index (continuous)

```
. stcox bmi, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
 enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -11150.636
Iteration 1:   log likelihood = -11132.263
Iteration 2:   log likelihood = -11132.026
Iteration 3:   log likelihood = -11132.026
Refining estimates:
Iteration 0:   log likelihood = -11132.026

Cox regression -- Breslow method for ties

No. of subjects =       12,080                 Number of obs    =      12,080
No. of failures =        1,208
Time at risk    =      1643786
                                               LR chi2(1)       =       37.22
Log likelihood  =   -11132.026                 Prob > chi2      =     0.0000
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
bmi	.0345589	.0054567	6.33	0.000	.0238639	.0452539

13.2.1.12 Body mass index (clinical categories variable)

```
. sts test bmicat, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

bmicat | Events observed | Events expected
-------+-----------------+----------------
0      |             296 |          371.42
1      |               3 |            5.59
2      |             532 |          528.78
3      |             370 |          295.21
-------+-----------------+----------------
Total  |            1201 |         1201.00

           chi2(3) =   35.51
           Pr>chi2 =  0.0000
```



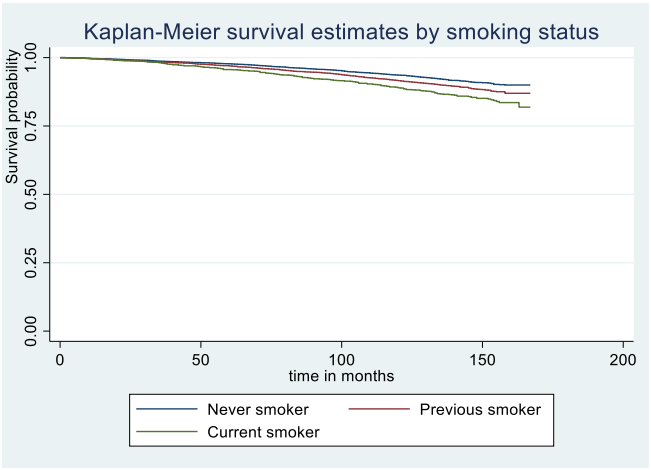
13.2.1.13 Waist circumference

```
stcox wcirc, nohr

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Iteration 0:   log likelihood = -11150.286
Iteration 1:   log likelihood = -11118.268
Iteration 2:   log likelihood = -11118.195
Iteration 3:   log likelihood = -11118.195
Refining estimates:
Iteration 0:   log likelihood = -11118.195

Cox regression -- Breslow method for ties

No. of subjects =       12,077                 Number of obs    =      12,077
No. of failures =        1,208
Time at risk    =      1643330
                                               LR chi2(1)       =      64.18
Log likelihood  =   -11118.195                 Prob > chi2      =     0.0000
```

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wcirc	.016776	.00206	8.14	0.000	.0127384	.0208136

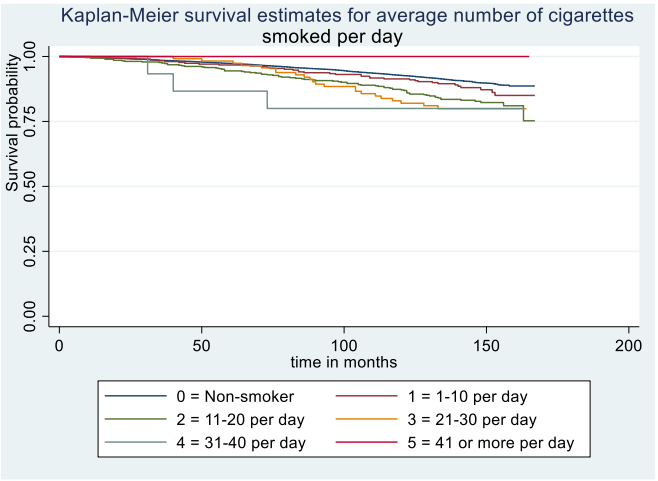
13.2.1.14 Smoking status

```
. sts test smoke_status, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
smoke_status |  observed     expected
-------------+------------------------
0            |       548       646.87
1            |       465       424.57
2            |       188       129.55
-------------+------------------------
Total        |      1201      1201.00

                chi2(2) =     45.38
                Pr>chi2 =     0.0000
```



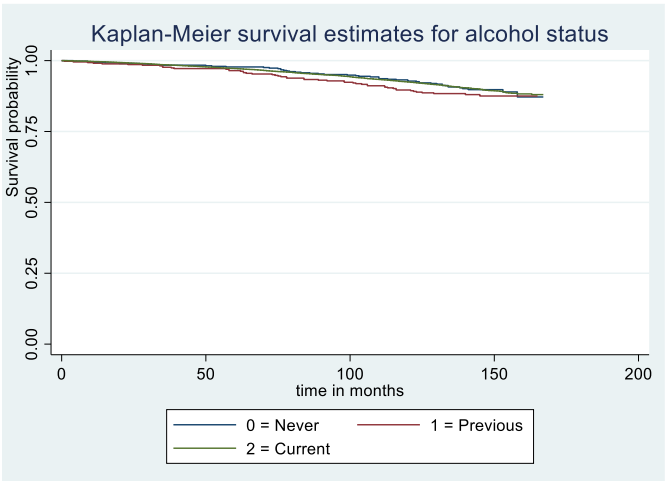
13.2.1.15 Cigarettes smoked per day (average)

```
sts test cigsmoke_d

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
cigsmoke_d   |  observed     expected
-------------+------------------------
0            |      1067      1119.18
1            |        37        30.38
2            |        79        44.91
3            |        22        10.80
4            |         3         1.41
5            |         0         1.33
-------------+------------------------
Total        |      1208      1208.00

                chi2(5) =     44.52
                Pr>chi2 =     0.0000
```



13.2.1.16 Alcohol status

```
sts test alcohol_status, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
  enter on or after:  time mage_attendukb

Log-rank test for equality of survivor functions

             |  Events       Events
alcohol_st~s |  observed     expected
-------------+------------------------
0            |        49        49.54
1            |        50        41.46
2            |      1106      1114.00
-------------+------------------------
Total        |      1205      1205.00

                chi2(2) =      1.82
                Pr>chi2 =     0.4019

.
```



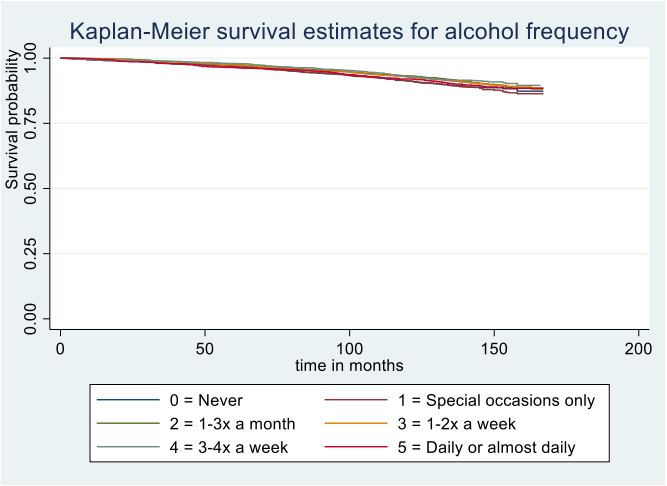
13.2.1.17 Alcohol frequency

```
. sts test alc_freq, logrank

        failure _d:  pdacukb == 1
  analysis time _t:  (mage_stend-origin)
            origin:  time mage_attendukb
 enter on or after:  time mage_attendukb
```

Log-rank test for equality of survivor functions

alc_freq	Events observed	Events expected
0	100	91.33
1	288	250.24
2	271	289.03
3	296	311.97
4	112	129.16
5	139	134.26
Total	1206	1206.00

chi2(5) = 10.92
Pr>chi2 = 0.0530



13.3 Additional results for Chapter 9: Survival analysis

Table 8. (Continued) Univariate analysis using cox proportional hazards regression. Hazard ratios (HR) for the covariates are presented with 95% confidence intervals and p-values of significance. From left to right: The first column represents HR calculated for the whole optimised cohort. The second (Normoglycaemic), third (New pre-diabetes), and fourth (New diabetes) columns show HR for these subgroups.

	1:9 case control cohort		**Normoglycaemic (HbA1c = <42 mmol/mol)**		**New pre-diabetes (HbA1c = 42 to 47.9 mmol/mol)**		**New diabetes (HbA1c = ≥ 48 mmol/mol)**	
Demographic information	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**	**HR (95%CI lower-upper)**	**p-value**
Age at study entry (continuous)	1.00 (0.99 – 1.01)	0.73	1.00 (1.00 - 1.01)	0.46	0.99 (0.96 – 1.01)	0.36	1.00 (0.97-1.03)	0.98
Sex								
Female	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-
Male	0.99 (0.89-1.11)	0.92	0.97 (0.86-1.10)	0.67	0.89 (0.57-1.38)	0.59	0.65 (0.38-1.09)	0.10
Ethnicity								
White	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-
Mixed (any)	1.39 (0.74-2.58)	0.30	1.15 (0.55-2.43)	0.71	-	-	14.8 (3.5-62.5)	< 0.01
Asian	0.44 (0.24-0.83)	0.01	0.40 (0.18-0.90)	0.03	0.48 (0.12-1.97)	0.31	0.26 (0.36-1.90)	0.19
Black	1.08 (0.66-1.77)	0.76	0.92 (0.48-1.77)	0.79	0.69 (0.17-2.80)	0.60	1.29 (0.40-4.13)	0.67
Chinese	0.91 (0.29-2.82)	0.87	0.81 (0.20-3.24)	0.77	1.78 (0.25-12.83)	0.57	-	-
Other	0.78 (0.37-1.64)	0.51	0.75 (0.31-1.81)	0.52	-	-	1.31 (0.18-9.53)	0.79
Pancreatic cancer Location								
Head	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-	1.00 (ref)	-
Body	1.01 (0.83-1.23)	0.93	0.95 (0.77-1.19)	0.67	1.61 (0.79-3.28)	0.19	1.35 (0.57-3.22)	0.50

Table 8. Univariate analysis using cox proportional hazards regression. (continued)

Tail	0.85 (0.70-1.02)	0.08	0.91 (0.75-1.11)	0.37	0.68 (0.26-1.79)	0.44	0.23 (0.08-0.65)	0.01
Duct	0.75 (0.49-1.16)	0.20	0.74 (0.47-1.16)	0.19	2.16 (0.29-16.2)	0.45	1.29 (0.17-9.74)	0.81
Other	0.99 (0.66-1.49)	0.97	1.06 (0.69-1.62)	0.80	1.38 (0.19-10.2)	0.75	-	-
Overlapping	1.22 (0.65-2.28)	0.53	1.31 (0.68-2.54)	0.43	-	-	-	-
Unspecified	0.96 (0.84-1.10)	0.57	0.98 (0.85-1.14)	0.82	1.25 (0.74-2.09)	0.41	0.60 (0.32-1.12)	0.11
Townsend Deprivation Index	1.01 (1.00-1.03)	0.14	1.01 (0.99-1.03)	0.40	0.95 (0.89-1.02)	0.18	1.10 (1.02-1.18)	0.01
Body mass index (continuous)	1.04 (1.02-1.05)	<0.01	1.03 (1.02-1.04)	<0.01	0.99 (0.95-1.03)	0.47	1.03 (0.98-1.08)	0.19
BMI clinical ranges:	1.00 (ref)		1.00 (ref)		1.00 (ref)		1.00 (ref)	
Healthy (18.5 –24.9)	0.67 (0.22-2.10)	0.50	0.49 (0.12-1.98)	0.32	-		-	
Underweight <18.5	1.26 (1.10-1.46)	<0.01	1.26 (1.09-1.47)	<0.01	0.94 (0.49-1.82)	0.86	1.16 (0.38-3.50)	0.79
25.0 – 29.9	1.56 (1.34-1.82)	<0.01	1.43 (1.21-1.70)	<0.01	0.75 (0.39-1.45)	0.39	1.86 (0.66-5.20)	0.24
≥30.0								
Waist circumference	1.02 (1.01-1.02)	<0.01	1.01 (1.01-1.02)	<0.01	1.00 (0.98-1.01)	0.64	1.02 (1.00-1.04)	0.07
Smoking status								
Never	1.00 (ref)		1.00 (ref)		1.00 (ref)		1.00 (ref)	
Previous	1.29 (1.14-1.46)	<0.01	1.20 (1.05-1.38)	<0.01	1.96 (1.19-3.24)	<0.01	1.33 (0.77-2.29)	0.30
Current	1.71 (1.45-2.02)	<0.01	1.72 (1.43-2.06)	<0.01	1.70 (0.91-3.17)	<0.01	0.82 (0.32-2.15)	0.69
Cigarettes smoked per day								
Non-smoker	1.00 (ref)		1.00 (ref)		1.00 (ref)		1.00 (ref)	
1 to 10	1.28 (0.92-1.77)	0.14	1.33 (0.93-1.91)	0.12	0.65 (0.16-2.64)	0.54	1.42 (0.20-10.3)	0.73
11 to 20	1.85 (1.47-2.32)	<0.01	1.90 (1.47-2.45)	<0.01	1.88 (0.90-3.92)	0.09	0.95 (0.30-3.06)	0.94
21 or 30	2.13 (1.40-3.26)	<0.01	2.33 (1.46-3.72)	<0.01	1.63 (0.51-5.19)	0.41	-	-
31 to 40	2.23 (0.72-6.93)	0.17	2.95 (0.95-9.18)	0.06	-	-	-	-
41 or more	-	-	-	-	-	-	-	-
Alcohol status								
Never	1.00 (ref)		1.00 (ref)		1.00 (ref)		1.00 (ref)	

Table 8. Univariate analysis using cox proportional hazards regression. (continued)

Previous	1.22 (0.82-1.81)	0.32	1.12 (0.70-1.77)	0.64	0.64 (0.12-3.31)	0.60	1.58 (0.46-5.48)	0.47
Current	1.00 (0.75-1.34)	0.98	1.00 (0.72-1.39)	0.99	1.50 (0.61-3.71)	0.38	0.97 (0.39-2.44)	0.95
Alcohol frequency							-	
Never	1.00 (ref)	0.67	1.00 (ref)		1.00 (ref)			
Special occasions only	1.05 (0.84-1.32)	0.19	1.12 (0.87-1.46)	0.38	2.80 (1.18-6.63)	0.02		
1-3 times a month	0.86 (0.68-1.08)	0.22	0.91 (0.70-1.18)	0.46	1.59 (0.64-3.94)	0.32		
1-2 times a week	0.87 (0.69-1.09)	0.09	0.90 (0.70-1.17)	0.43	1.66 (0.69-3.97)	0.26		
3-4 times a week	0.79 (0.60-1.04)	0.67	0.77 (0.56-1.05)	0.10	1.80 (0.71-4.58)	0.22		
Daily or almost daily	0.95 (0.73-1.22)	-	0.93 (0.69-1.27)	0.66	1.00 (0.36-2.77)	0.99		

13.4 Log-rank test for equality of survivor functions for new pre-diabetes and new diabetes subgroups

Log-rank testing of the subgroups within this cohort confirmed that statistically significant risks for PDAC were significantly different to one another.

HbA1c measurement group	**Events observed**	**Events Expected**
Normoglycaemia	998	1065.71
New pre-diabetes	79	46.58
New diabetes	57	21.71
TOTAL	1134	1134.00
P-value	<0.001	

14 References

1. Kachuri, L., et al., *Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction.* Nature Communications, 2020. **11**(1): p. 6084.
2. Liu, M.C., et al., *Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA.* Ann Oncol, 2020. **31**(6): p. 745-759.
3. NICE. *Suspected cancer: recognition and referral*. 2021 15 December 2021 [cited 2022 September]; NICE guideline [NG12]:[Available from: https://www.nice.org.uk/guidance/ng12/chapter/Recommendations-organised-by-symptom-and-findings-of-primary-care-investigations#primary-care-investigations.
4. Talathi SS, Z.R., Young M. *Anatomy, Abdomen and Pelvis, Pancreas*. 2022 Jul 25 [cited 2022 November]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK532912/figure/article-26567.image.f1/.
5. Talathi SS, Z.R. *Anatomy, Abdomen and Pelvis, Pancreas*. 2022 Jul 25; Available from: https://www.ncbi.nlm.nih.gov/books/NBK532912/.
6. Wynne, K., B. Devereaux, and A. Dornhorst, *Diabetes of the exocrine pancreas.* J Gastroenterol Hepatol, 2019. **34**(2): p. 346-354.
7. Perera, R.M. and N. Bardeesy, *Pancreatic Cancer Metabolism: Breaking It Down to Build It Back Up.* Cancer Discov, 2015. **5**(12): p. 1247-61.
8. UK, C.R. *Pancreatic cancer diagnosis and treatment statistics*. 2018 [cited 2022 Dec]; Available from: https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/survival.
9. UK, C.R. *Types of pancreatic cancer*. 2023 [cited 2023 June]; Available from: https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/stages-types-grades/types.
10. Sousa, C.M. and A.C. Kimmelman, *The complex landscape of pancreatic cancer metabolism.* Carcinogenesis, 2014. **35**(7): p. 1441-50.
11. Cameron, M.E., A. Yakovenko, and J.G. Trevino, *Glucose and Lactate Transport in Pancreatic Cancer: Glycolytic Metabolism Revisited.* J Oncol, 2018. **2018**: p. 6214838.
12. Yachida, S., et al., *Distant metastasis occurs late during the genetic evolution of pancreatic cancer.* Nature, 2010. **467**(7319): p. 1114-7.
13. Gupta, N. and R. Yelamanchi, *Pancreatic adenocarcinoma: A review of recent paradigms and advances in epidemiology, clinical diagnosis and management.* World J Gastroenterol, 2021. **27**(23): p. 3158-3181.
14. Ying, H., et al., *Genetics and biology of pancreatic ductal adenocarcinoma.* Genes Dev, 2016. **30**(4): p. 355-85.
15. Yao, W., A. Maitra, and H. Ying, *Recent insights into the biology of pancreatic cancer.* EBioMedicine, 2020. **53**: p. 102655.
16. Polireddy, K. and Q. Chen, *Cancer of the Pancreas: Molecular Pathways and Current Advancement in Treatment.* J Cancer, 2016. **7**(11): p. 1497-514.
17. UK, C.R. *Pancreatic cancer > Getting Diagnosed > Screening*. 2019 [cited 2020 June].
18. Wu, C., et al., *Is there a CDKN2A-centric network in pancreatic ductal adenocarcinoma?* Onco Targets Ther, 2020. **13**: p. 2551-2562.
19. Lagarrigue, S., et al., *CDK4 is an essential insulin effector in adipocytes.* J Clin Invest, 2016. **126**(1): p. 335-48.
20. Wang, H., et al., *The metabolic function of cyclin D3-CDK6 kinase in cancer cell survival.* Nature, 2017. **546**(7658): p. 426-430.

21. Witkiewicz, A.K., et al., *Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets.* Nat Commun, 2015. **6**: p. 6744.
22. Korc, M., *Beyond Kras: MYC Rules in Pancreatic Cancer.* Cell Mol Gastroenterol Hepatol, 2018. **6**(2): p. 223-224.
23. Perri, F., S. Pisconti, and G. Della Vittoria Scarpati, *P53 mutations and cancer: a tight linkage.* Ann Transl Med, 2016. **4**(24): p. 522.
24. Feng, Y., et al., *Potential functional variants in SMC2 and TP53 in the AURORA pathway genes and risk of pancreatic cancer.* Carcinogenesis, 2019. **40**(4): p. 521-528.
25. Grochola, L.F., et al., *Elevated transcript levels from the MDM2 P1 promoter and low p53 transcript levels are associated with poor prognosis in human pancreatic ductal adenocarcinoma.* Pancreas, 2011. **40**(2): p. 265-70.
26. Dardare, J., et al., *SMAD4 and the TGFβ Pathway in Patients with Pancreatic Ductal Adenocarcinoma.* Int J Mol Sci, 2020. **21**(10).
27. Farrell, J.J., *Intraductal papillary mucinous neoplasm to pancreas ductal adenocarcinoma sequence and pancreas cancer screening.* Endosc Ultrasound, 2018. **7**(5): p. 314-318.
28. Morales-Oyarvide, V., et al., *Diabetes mellitus in intraductal papillary mucinous neoplasm of the pancreas is associated with high-grade dysplasia and invasive carcinoma.* Pancreatology, 2017. **17**(6): p. 920-926.
29. Atsushi Yamaguchi, S.T., Yuzuru Tamaru et al., *Long-standing Diabetes Mellitus Increases Concomitant Pancreatic Cancer Risk in Intraductal Papillary Mucinous Neoplasms*. 2022.
30. Fric, P., et al., *Precursors of pancreatic cancer.* European Journal of Gastroenterology & Hepatology, 2017. **29**(3).
31. Hruban, R.H., et al., *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms.* Am J Surg Pathol, 2004. **28**(8): p. 977-87.
32. Distler, M., et al., *Precursor Lesions for Sporadic Pancreatic Cancer: PanIN, IPMN, and MCN.* BioMed Research International, 2014. **2014**: p. 474905.
33. Shi, C., et al., *Increased Prevalence of Precursor Lesions in Familial Pancreatic Cancer Patients.* Clin Cancer Res, 2009. **15**(24): p. 7737-7743.
34. Potjer, T.P., et al., *Variation in precursor lesions of pancreatic cancer among high-risk groups.* Clin Cancer Res, 2013. **19**(2): p. 442-9.
35. Distler, M., et al., *Pathohistological subtype predicts survival in patients with intraductal papillary mucinous neoplasm (IPMN) of the pancreas.* Ann Surg, 2013. **258**(2): p. 324-30.
36. Yamaguchi, K., et al., *Pancreatic ductal adenocarcinoma derived from IPMN and pancreatic ductal adenocarcinoma concomitant with IPMN.* Pancreas, 2011. **40**(4): p. 571-80.
37. Hassid, B.G., et al., *Absence of pancreatic intraepithelial neoplasia predicts poor survival after resection of pancreatic cancer.* Pancreas, 2014. **43**(7): p. 1073-7.
38. Yu, D.Y., et al., *Clinical significance of pancreatic intraepithelial neoplasia in resectable pancreatic cancer on survivals.* Ann Surg Treat Res, 2018. **94**(5): p. 247-253.
39. Lotfollahzadeh S, R.-B.A., Cagir B. *Colon Cancer*. 2023 2022 Dec 3 [cited 2023 June]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK470380/.
40. Aronsson, L., et al., *Intraductal papillary mucinous carcinoma versus pancreatic ductal adenocarcinoma: A systematic review and meta-analysis.* Int J Surg, 2019. **71**: p. 91-99.

41. Collisson, E.A., et al., *Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.* Nat Med, 2011. **17**(4): p. 500-3.
42. Moffitt, R.A., et al., *Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma.* Nat Genet, 2015. **47**(10): p. 1168-78.
43. UK, C.R. *Pancreatic cancer incidence statistics* 2019; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/incidence.
44. UK, C.R. *Breast cancer survival statistics*. 2020 [cited 2020 June]; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival.
45. Are, C., et al., *Predictive global trends in the incidence and mortality of pancreatic cancer based on geographic location, socio-economic status, and demographic shift.* J Surg Oncol, 2016. **114**(6): p. 736-742.
46. Wong, M.C.S., et al., *Global temporal patterns of pancreatic cancer and association with socioeconomic development.* Sci Rep, 2017. **7**(1): p. 3165.
47. Veisani, Y., et al., *Global incidence and mortality rates in pancreatic cancer and the association with the Human Development Index: decomposition approach.* Public Health, 2018. **156**: p. 87-91.
48. Fidler, M.M., F. Bray, and I. Soerjomataram, *The global cancer burden and human development: A review.* Scand J Public Health, 2018. **46**(1): p. 27-36.
49. Koo, M.M., et al., *Symptom Signatures and Diagnostic Timeliness in Cancer Patients: A Review of Current Evidence.* Neoplasia, 2018. **20**(2): p. 165-174.
50. Walter, F.M., et al., *Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study.* Lancet Gastroenterol Hepatol, 2016. **1**(4): p. 298-306.
51. Deshwar, A.B., et al., *Diagnostic intervals and pancreatic ductal adenocarcinoma (PDAC) resectability: a single-center retrospective analysis.* Ann Pancreat Cancer, 2018. **1**.
52. Stapley, S., et al., *The risk of pancreatic cancer in symptomatic patients in primary care: a large case-control study using electronic records.* Br J Cancer, 2012. **106**(12): p. 1940-4.
53. Keane, M.G., et al., *A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer.* BMJ Open, 2014. **4**(11): p. e005720.
54. Kanno, A., et al., *Multicenter study of early pancreatic cancer in Japan.* Pancreatology, 2018. **18**(1): p. 61-67.
55. Kikuyama, M., et al., *Early Diagnosis to Improve the Poor Prognosis of Pancreatic Cancer.* Cancers (Basel), 2018. **10**(2).
56. Kongkam, P., et al., *Pancreatic cancer in an Asian population.* Endosc Ultrasound, 2015. **4**(1): p. 56-62.
57. Porta, M., et al., *Exocrine pancreatic cancer: symptoms at presentation and their relation to tumour site and stage.* Clin Transl Oncol, 2005. **7**(5): p. 189-97.
58. Marcadis, A.R., et al., *Racial Disparities in Cancer Presentation and Outcomes: The Contribution of Overdiagnosis.* JNCI Cancer Spectr, 2020. **4**(2): p. pkaa001.
59. Kwok, W. and T. Bhuvanakrishna, *The relationship between ethnicity and the pain experience of cancer patients: a systematic review.* Indian J Palliat Care, 2014. **20**(3): p. 194-200.
60. CKS, N. *Sympoms suggestive of gastrointestinal tract (upper) cancers*. 2021 [cited 2023 June]; Available from: https://cks.nice.org.uk/topics/gastrointestinal-tract-upper-cancers-recognition-referral/diagnosis/symptoms-suggestive-of-gastrointestinal-tract-upper-cancers/.

61. Azer SA, S.S. *Steatorrhea*. 2023 2023 May 16 [cited 2023 June]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK541055/.
62. Gobbi, P.G., et al., *The prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic cancer.* Cancer Epidemiol, 2013. **37**(2): p. 186-90.
63. Raptis, D.A., et al., *Clinical presentation and waiting time targets do not affect prognosis in patients with pancreatic cancer.* Surgeon, 2010. **8**(5): p. 239-46.
64. Bao, B., et al., *The complexities of obesity and diabetes with the development and progression of pancreatic cancer.* Biochim Biophys Acta, 2011. **1815**(2): p. 135-46.
65. De Souza, A., et al., *Diabetes Type 2 and Pancreatic Cancer: A History Unfolding.* Jop, 2016. **17**(2): p. 144-148.
66. Korc, M., et al., *Tobacco and alcohol as risk factors for pancreatic cancer.* Best Pract Res Clin Gastroenterol, 2017. **31**(5): p. 529-536.
67. Carreras-Torres, R., et al., *The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study.* J Natl Cancer Inst, 2017. **109**(9).
68. Matsubayashi, H., et al., *Familial pancreatic cancer: Concept, management and issues.* World J Gastroenterol, 2017. **23**(6): p. 935-948.
69. Rentsch, C.T., et al., *Risk of 16 cancers across the full glycemic spectrum: a population-based cohort study using the UK Biobank.* BMJ Open Diabetes Res Care, 2020. **8**(1).
70. Hope, C., et al., *Relationship between HbA1c and cancer in people with or without diabetes: a systematic review.* Diabet Med, 2016. **33**(8): p. 1013-25.
71. Goto, A., et al., *High hemoglobin A1c levels within the non-diabetic range are associated with the risk of all cancers.* Int J Cancer, 2016. **138**(7): p. 1741-53.
72. Dankner, R., et al., *A historical cohort study on glycemic-control and cancer-risk among patients with diabetes.* Cancer Epidemiol, 2018. **57**: p. 104-109.
73. NICE. *Type 2 diabetes in adults: management* NICE Guidance and guidelines 2017 [cited 2019 December]; Available from: https://www.nice.org.uk/guidance/ng28/chapter/Introduction.
74. UK, D. *What is Type 2 diabetes?* . 2018; Available from: https://www.diabetes.org.uk/diabetes-the-basics/what-is-type-2-diabetes.
75. Fox, C.S., et al., *Update on Prevention of Cardiovascular Disease in Adults With Type 2 Diabetes Mellitus in Light of Recent Evidence: A Scientific Statement From the American Heart Association and the American Diabetes Association.* Diabetes Care, 2015. **38**(9): p. 1777-803.
76. CKS, N. *Diabetes - type 2*. 2017 [cited 2018 October]; Available from: Cks.nice.org.uk.
77. Bruenderman, E.H. and R.C. Martin, 2nd, *High-risk population in sporadic pancreatic adenocarcinoma: guidelines for screening.* J Surg Res, 2015. **194**(1): p. 212-9.
78. He, X., L. Shi, and J. Wu, *Retrospective database analysis of cancer risk in patients with type 2 diabetes mellitus in China.* Curr Med Res Opin, 2018. **34**(6): p. 1089-1098.
79. Everhart, J. and D. Wright, *Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis.* Jama, 1995. **273**(20): p. 1605-9.
80. Huxley, R., et al., *Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies.* Br J Cancer, 2005. **92**(11): p. 2076-83.
81. Ben, Q., et al., *Diabetes mellitus and risk of pancreatic cancer: A meta-analysis of cohort studies.* Eur J Cancer, 2011. **47**(13): p. 1928-37.

82. Batabyal, P., et al., *Association of diabetes mellitus and pancreatic adenocarcinoma: a meta-analysis of 88 studies.* Ann Surg Oncol, 2014. **21**(7): p. 2453-62.
83. Zhang, J.J., et al., *Diabetes mellitus and risk of pancreatic cancer in China: A meta-analysis based on 26 case-control studies.* Prim Care Diabetes, 2019. **13**(3): p. 276-282.
84. Chen, Y., et al., *Association between type 2 diabetes and risk of cancer mortality: a pooled analysis of over 771,000 individuals in the Asia Cohort Consortium.* Diabetologia, 2017. **60**(6): p. 1022-1032.
85. Pannala, R., et al., *Prevalence and clinical profile of pancreatic cancer-associated diabetes mellitus.* Gastroenterology, 2008. **134**(4): p. 981-987.
86. Aggarwal, G., P. Kamada, and S.T. Chari, *Prevalence of diabetes mellitus in pancreatic cancer compared to common cancers.* Pancreas, 2013. **42**(2): p. 198-201.
87. Chari, S.T., et al., *Pancreatic cancer-associated diabetes mellitus: prevalence and temporal association with diagnosis of cancer.* Gastroenterology, 2008. **134**(1): p. 95-101.
88. Pannala, R., et al., *Temporal association of changes in fasting blood glucose and body mass index with diagnosis of pancreatic cancer.* Am J Gastroenterol, 2009. **104**(9): p. 2318-25.
89. Wang, F., et al., *The relationship between diabetes and pancreatic cancer.* Mol Cancer, 2003. **2**: p. 4.
90. Song, S., et al., *Long-Term Diabetes Mellitus Is Associated with an Increased Risk of Pancreatic Cancer: A Meta-Analysis.* PLoS One, 2015. **10**(7): p. e0134321.
91. Zhang, P., et al., *Association of metformin use with cancer incidence and mortality: a meta-analysis.* Cancer Epidemiol, 2013. **37**(3): p. 207-18.
92. Dai, M., et al., *Risk factors for new-onset diabetes mellitus after distal pancreatectomy.* BMJ Open Diabetes Research & Care, 2020. **8**(2): p. e001778.
93. Kim, K.-J., et al., *Pancreatic Diabetes after Distal Pancreatectomy: Incidence Rate and Risk Factors.* Korean journal of hepato-biliary-pancreatic surgery, 2011. **15**(2): p. 123-127.
94. Hutchins, R.R., et al., *Long-term results of distal pancreatectomy for chronic pancreatitis in 90 patients.* Annals of surgery, 2002. **236**(5): p. 612-618.
95. Lee, B.W., et al., *Insulin secretory defect plays a major role in the development of diabetes in patients with distal pancreatectomy.* Metabolism, 2006. **55**(1): p. 135-41.
96. Maxwell, D.W., et al., *Post-Pancreatectomy Diabetes Index: A Validated Score Predicting Diabetes Development after Major Pancreatectomy.* Journal of the American College of Surgeons, 2020. **230**(4): p. 393-402.e3.
97. Aggarwal, G., et al., *New-onset diabetes in pancreatic cancer: a study in the primary care setting.* Pancreatology, 2012. **12**(2): p. 156-61.
98. Sharma, A. and S.T. Chari, *Pancreatic Cancer and Diabetes Mellitus.* Curr Treat Options Gastroenterol, 2018. **16**(4): p. 466-478.
99. Liao, W.C., et al., *Blood glucose concentration and risk of pancreatic cancer: systematic review and dose-response meta-analysis.* Bmj, 2015. **350**: p. g7371.
100. de Beer, J.C. and L. Liebenberg, *Does cancer risk increase with HbA1c, independent of diabetes?* Br J Cancer, 2014. **110**(9): p. 2361-8.
101. Sharma, A., et al., *Fasting Blood Glucose Levels Provide Estimate of Duration and Progression of Pancreatic Cancer Before Diagnosis.* Gastroenterology, 2018. **155**(2): p. 490-500.e2.

102. Bonora, E. and J. Tuomilehto, *The pros and cons of diagnosing diabetes with A1C.* Diabetes Care, 2011. **34 Suppl 2**(Suppl 2): p. S184-90.
103. Cowie, C.C., et al., *Prevalence of diabetes and high risk for diabetes using A1C criteria in the U.S. population in 1988-2006.* Diabetes Care, 2010. **33**(3): p. 562-8.
104. Ding, L., et al., *Hemoglobin A1c and diagnosis of diabetes.* J Diabetes, 2018. **10**(5): p. 365-372.
105. Guo, F., D.R. Moellering, and W.T. Garvey, *Use of HbA1c for diagnoses of diabetes and prediabetes: comparison with diagnoses based on fasting and 2-hr glucose values and effects of gender, race, and age.* Metab Syndr Relat Disord, 2014. **12**(5): p. 258-68.
106. Little, R.R. and C.L. Rohlfing, *The long and winding road to optimal HbA1c measurement.* Clin Chim Acta, 2013. **418**: p. 63-71.
107. Karakaya, J., et al., *The performance of hemoglobin A1c against fasting plasma glucose and oral glucose tolerance test in detecting prediabetes and diabetes.* J Res Med Sci, 2014. **19**(11): p. 1051-7.
108. *Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: a pooled analysis of 96 population-based studies with 331,288 participants.* Lancet Diabetes Endocrinol, 2015. **3**(8): p. 624-37.
109. Herman, W.H. and R.M. Cohen, *Racial and ethnic differences in the relationship between HbA1c and blood glucose: implications for the diagnosis of diabetes.* J Clin Endocrinol Metab, 2012. **97**(4): p. 1067-72.
110. Yu, E.Y., et al., *Can HbA1c replace OGTT for the diagnosis of diabetes mellitus among Chinese patients with impaired fasting glucose?* Fam Pract, 2015. **32**(6): p. 631-8.
111. Lim, W.Y., et al., *Screening for diabetes with HbA1c: Test performance of HbA1c compared to fasting plasma glucose among Chinese, Malay and Indian community residents in Singapore.* Sci Rep, 2018. **8**(1): p. 12419.
112. Cohen, R.M., S. Haggerty, and W.H. Herman, *HbA1c for the diagnosis of diabetes and prediabetes: is it time for a mid-course correction?* J Clin Endocrinol Metab, 2010. **95**(12): p. 5203-6.
113. Chari, S.T., et al., *Probability of pancreatic cancer following diabetes: a population-based study.* Gastroenterology, 2005. **129**(2): p. 504-11.
114. Ilic, M., B. Milicic, and I. Ilic, *Association between oral contraceptive use and pancreatic cancer risk: A systematic review and meta-analysis.* World J Gastroenterol, 2021. **27**(20): p. 2643-2656.
115. Eibl, G., et al., *Diabetes Mellitus and Obesity as Risk Factors for Pancreatic Cancer.* J Acad Nutr Diet, 2018. **118**(4): p. 555-567.
116. Tseng, C.H., *New-onset diabetes with a history of dyslipidemia predicts pancreatic cancer.* Pancreas, 2013. **42**(1): p. 42-8.
117. O'Neill, S. and L. O'Driscoll, *Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies.* Obes Rev, 2015. **16**(1): p. 1-12.
118. Esposito, K., et al., *Metabolic syndrome and risk of cancer: a systematic review and meta-analysis.* Diabetes Care, 2012. **35**(11): p. 2402-11.
119. Hori, M., et al., *Association of pancreatic Fatty infiltration with pancreatic ductal adenocarcinoma.* Clin Transl Gastroenterol, 2014. **5**(3): p. e53.
120. Takahashi, M., et al., *Fatty pancreas: A possible risk factor for pancreatic cancer in animals and humans.* Cancer Sci, 2018. **109**(10): p. 3013-3023.
121. Aune, D., et al., *Body mass index, abdominal fatness and pancreatic cancer risk: a systematic review and non-linear dose-response meta-analysis of prospective studies.* Ann Oncol, 2012. **23**(4): p. 843-52.

122. Xu, M., et al., *Obesity and Pancreatic Cancer: Overview of Epidemiology and Potential Prevention by Weight Loss.* Pancreas, 2018. **47**(2): p. 158-162.
123. Hart, A.R., H. Kennedy, and I. Harvey, *Pancreatic cancer: a review of the evidence on causation.* Clin Gastroenterol Hepatol, 2008. **6**(3): p. 275-82.
124. Mizuno, S., et al., *Risk factors and early signs of pancreatic cancer in diabetes: screening strategy based on diabetes onset age.* J Gastroenterol, 2013. **48**(2): p. 238-46.
125. Wang, J., et al., *Association of cholesterol with risk of pancreatic cancer: a meta-analysis.* World J Gastroenterol, 2015. **21**(12): p. 3711-9.
126. Olson, S.H., et al., *Weight Loss, Diabetes, Fatigue, and Depression Preceding Pancreatic Cancer.* Pancreas, 2016. **45**(7): p. 986-91.
127. Hart, P.A., et al., *Type 3c (pancreatogenic) diabetes mellitus secondary to chronic pancreatitis and pancreatic cancer.* Lancet Gastroenterol Hepatol, 2016. **1**(3): p. 226-237.
128. Hart, P.A., et al., *Weight loss precedes cancer-specific symptoms in pancreatic cancer-associated diabetes mellitus.* Pancreas, 2011. **40**(5): p. 768-72.
129. Petrov, M.S., *Diabetes of the exocrine pancreas: American Diabetes Association-compliant lexicon.* Pancreatology, 2017. **17**(4): p. 523-526.
130. *Diagnosis and classification of diabetes mellitus.* Diabetes Care, 2014. **37 Suppl 1**: p. S81-90.
131. Ewald, N., et al., *Prevalence of diabetes mellitus secondary to pancreatic diseases (type 3c).* Diabetes Metab Res Rev, 2012. **28**(4): p. 338-42.
132. Woodmansey, C., et al., *Incidence, Demographics, and Clinical Characteristics of Diabetes of the Exocrine Pancreas (Type 3c): A Retrospective Cohort Study.* Diabetes Care, 2017. **40**(11): p. 1486-1493.
133. Cui, Y. and D.K. Andersen, *Pancreatogenic diabetes: special considerations for management.* Pancreatology, 2011. **11**(3): p. 279-94.
134. Pendharkar, S.A., J. Mathew, and M.S. Petrov, *Age- and sex-specific prevalence of diabetes associated with diseases of the exocrine pancreas: A population-based study.* Dig Liver Dis, 2017. **49**(5): p. 540-544.
135. Singhi, A.D., et al., *Early Detection of Pancreatic Cancer: Opportunities and Challenges.* Gastroenterology, 2019. **156**(7): p. 2024-2040.
136. Jimenez-Luna, C., et al., *Novel Biomarkers to Distinguish between Type 3c and Type 2 Diabetes Mellitus by Untargeted Metabolomics.* Metabolites, 2020. **10**(11).
137. Bhattamisra, S., et al., *Type-3c Diabetes Mellitus, Diabetes of Exocrine Pancreas – An Update.* Current Diabetes Reviews, 2019. **15**.
138. Andersen, D.K., et al., *Diabetes, Pancreatogenic Diabetes, and Pancreatic Cancer.* Diabetes, 2017. **66**(5): p. 1103-1110.
139. Dankner, R. and J. Roth, *More recent, better designed studies have weakened links between antidiabetes medications and cancer risk.* Diabet Med, 2020. **37**(2): p. 194-202.
140. Zhang, Z.J., et al., *The effect of metformin on biomarkers and survivals for breast cancer- a systematic review and meta-analysis of randomized clinical trials.* Pharmacol Res, 2019. **141**: p. 551-555.
141. Suissa, S. and L. Azoulay, *Metformin and the risk of cancer: time-related biases in observational studies.* Diabetes Care, 2012. **35**(12): p. 2665-73.
142. NICE. *Type 1 diabetes in adults: diagnosis and management*. 2015 26 August 2015; Available from: https://www.nice.org.uk/guidance/ng17/chapter/Recommendations#diagnosis-and-early-care-plan.

143. Farmer, R.E., et al., *Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation.* Int J Epidemiol, 2017. **46**(2): p. 728-744.
144. Zhang, A.M.Y., et al., *Hyperinsulinemia in Obesity, Inflammation, and Cancer.* Diabetes Metab J, 2021. **45**(3): p. 285-311.
145. Dąbrowski, M., *Diabetes, Antidiabetic Medications and Cancer Risk in Type 2 Diabetes: Focus on SGLT-2 Inhibitors.* Int J Mol Sci, 2021. **22**(4).
146. Takizawa, S., J.J. Furth, and J. Furth, *Biological and technical aspects of nucleic acid synthesis in cultures of mammary tumors.* Cancer Res, 1970. **30**(1): p. 211-20.
147. Heuson, J.C., N. Legros, and R. Heimann, *Influence of insulin administration on growth of the 7,12-dimethylbenz(a)anthracene-induced mammary carcinoma in intact, oophorectomized, and hypophysectomized rats.* Cancer Res, 1972. **32**(2): p. 233-8.
148. Zhang, A.M.Y., et al., *Endogenous Hyperinsulinemia Contributes to Pancreatic Cancer Development.* Cell Metab, 2019. **30**(3): p. 403-404.
149. Tsujimoto, T., H. Kajio, and T. Sugiyama, *Association between hyperinsulinemia and increased risk of cancer death in nonobese and obese people: A population-based observational study.* Int J Cancer, 2017. **141**(1): p. 102-111.
150. Bosetti, C., et al., *Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium.* Ann Oncol, 2014. **25**(10): p. 2065-2072.
151. (BNF), B.N.F. *Type 2 diabetes*. 2023.
152. Hu, J., et al., *The relationship between the use of metformin and the risk of pancreatic cancer in patients with diabetes: a systematic review and meta-analysis.* BMC Gastroenterol, 2023. **23**(1): p. 50.
153. Scarpello, J.H. and H.C. Howlett, *Metformin therapy and clinical uses.* Diab Vasc Dis Res, 2008. **5**(3): p. 157-67.
154. Schneider, M.B., et al., *Prevention of pancreatic cancer induction in hamsters by metformin.* Gastroenterology, 2001. **120**(5): p. 1263-70.
155. Cifarelli, V., et al., *Metformin and Rapamycin Reduce Pancreatic Cancer Growth in Obese Prediabetic Mice by Distinct MicroRNA-Regulated Mechanisms.* Diabetes, 2015. **64**(5): p. 1632-42.
156. Chen, K., et al., *Metformin suppresses cancer initiation and progression in genetic mouse models of pancreatic cancer.* Mol Cancer, 2017. **16**(1): p. 131.
157. Shi, Y.Q., et al., *Relationships are between metformin use and survival in pancreatic cancer patients concurrent with diabetes: A systematic review and meta-analysis.* Medicine (Baltimore), 2020. **99**(37): p. e21687.
158. Zhou, P.T., et al., *Metformin is associated with survival benefit in pancreatic cancer patients with diabetes: a systematic review and meta-analysis.* Oncotarget, 2017. **8**(15): p. 25242-25250.
159. Wang, Z., et al., *Metformin is associated with reduced risk of pancreatic cancer in patients with type 2 diabetes mellitus: a systematic review and meta-analysis.* Diabetes Res Clin Pract, 2014. **106**(1): p. 19-26.
160. Singh, S., et al., *Anti-diabetic medications and risk of pancreatic cancer in patients with diabetes mellitus: a systematic review and meta-analysis.* Am J Gastroenterol, 2013. **108**(4): p. 510-9; quiz 520.
161. Bodmer, M., et al., *Use of antidiabetic agents and the risk of pancreatic cancer: a case-control analysis.* Am J Gastroenterol, 2012. **107**(4): p. 620-6.
162. Lee, D.Y., et al., *The influence of diabetes and antidiabetic medications on the risk of pancreatic cancer: a nationwide population-based study in Korea.* Sci Rep, 2018. **8**(1): p. 9719.

163. Liu, M., et al., *Activation of AMPK by metformin promotes renal cancer cell proliferation under glucose deprivation through its interaction with PKM2.* Int J Biol Sci, 2019. **15**(3): p. 617-627.
164. Zhao, H., et al., *Sulfonylurea and Cancer Risk Among Patients With Type 2 Diabetes: A Population-Based Cohort Study.* Frontiers in Endocrinology, 2022. **13**.
165. Suryadevara, V., et al., *Incretin based therapy and pancreatic cancer: Realising the reality.* World J Gastroenterol, 2022. **28**(25): p. 2881-2889.
166. Chen, H., et al., *Incretin-Based Therapy and Risk of Pancreatic Cancer in Patients with Type 2 Diabetes Mellitus: A Meta-analysis of Randomized Controlled Trials.* Diabetes Ther, 2016. **7**(4): p. 725-742.
167. Overbeek, J.A., et al., *Risk of dipeptidyl peptidase-4 (DPP-4) inhibitors on site-specific cancer: A systematic review and meta-analysis.* Diabetes Metab Res Rev, 2018. **34**(5): p. e3004.
168. Azoulay, L., et al., *Incretin based drugs and the risk of pancreatic cancer: international multicentre cohort study.* Bmj, 2016. **352**: p. i581.
169. Lee, M., et al., *Nationwide Trends in Pancreatitis and Pancreatic Cancer Risk Among Patients With Newly Diagnosed Type 2 Diabetes Receiving Dipeptidyl Peptidase 4 Inhibitors.* Diabetes Care, 2019. **42**(11): p. 2057-2064.
170. Jang, J.H., et al., *Suppression of lung metastases by the CD26/DPP4 inhibitor Vildagliptin in mice.* Clin Exp Metastasis, 2015. **32**(7): p. 677-87.
171. Kawakita, E., D. Koya, and K. Kanasaki, *CD26/DPP-4: Type 2 Diabetes Drug Target with Potential Influence on Cancer Biology.* Cancers (Basel), 2021. **13**(9).
172. Ninomiya, I., et al., *Pioglitazone inhibits the proliferation and metastasis of human pancreatic cancer cells.* Oncol Lett, 2014. **8**(6): p. 2709-2714.
173. Zhao, Z., X. He, and Y. Sun, *Hypoglycemic agents and incidence of pancreatic cancer in diabetic patients: a meta-analysis.* Frontiers in Pharmacology, 2023. **14**.
174. Lewis, J.D., et al., *Pioglitazone Use and Risk of Bladder Cancer and Other Common Cancers in Persons With Diabetes.* Jama, 2015. **314**(3): p. 265-77.
175. Qu, H., et al., *Global and Regional Effects of Bladder Cancer Risk Associated with Pioglitazone Therapy in Patients with Diabetes.* Sci Rep, 2017. **7**(1): p. 15804.
176. Roy, A., et al., *Diabetes and pancreatic cancer: Exploring the two-way traffic.* World J Gastroenterol, 2021. **27**(30): p. 4939-4962.
177. Yuan, C., et al., *Diabetes, Weight Change, and Pancreatic Cancer Risk.* JAMA Oncol, 2020. **6**(10): p. e202948.
178. Hu, C., et al., *Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer.* Jama, 2018. **319**(23): p. 2401-2409.
179. Ríos Peces, S., et al., *Untargeted LC-HRMS-Based Metabolomics for Searching New Biomarkers of Pancreatic Ductal Adenocarcinoma: A Pilot Study.* SLAS Discov, 2017. **22**(4): p. 348-359.
180. Ahonen, L., et al., *Targeted Clinical Metabolite Profiling Platform for the Stratification of Diabetic Patients.* Metabolites, 2019. **9**(9).
181. Carr, D., D.M. Kent, and H.G. Welch, *All-cause mortality as the primary endpoint for the GRAIL/National Health Service England multi-cancer screening trial.* J Med Screen, 2022. **29**(1): p. 3-6.
182. Prasad, V., et al., *The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses.* JAMA Intern Med, 2015. **175**(8): p. 1389-98.
183. Klein, E.A., et al., *Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set.* Ann Oncol, 2021. **32**(9): p. 1167-1177.

184. Oldfield, L., et al., *United Kingdom Early Detection Initiative (UK-EDI): protocol for establishing a national multicentre cohort of individuals with new-onset diabetes for early detection of pancreatic cancer.* BMJ Open, 2022. **12**(10): p. e068010.
185. Rahman, I., M.T. Athar, and M. Islam, *Type 2 Diabetes, Obesity, and Cancer Share Some Common and Critical Pathways.* Frontiers in oncology, 2021. **10**: p. 600824-600824.
186. Lu, Y., et al., *New-onset type 2 diabetes, elevated HbA1c, anti-diabetic medications, and risk of pancreatic cancer.* British Journal of Cancer, 2015. **113**(11): p. 1607-1614.
187. Altman, D.G. and P. Royston, *The cost of dichotomising continuous variables.* Bmj, 2006. **332**(7549): p. 1080.
188. Sadr-Azodi, O., S. Gudbjörnsdottir, and R. Ljung, *Pattern of increasing HbA1c levels in patients with diabetes mellitus before clinical detection of pancreatic cancer - a population-based nationwide case-control study.* Acta Oncol, 2015. **54**(7): p. 986-92.
189. Huang, B.Z., et al., *New-Onset Diabetes, Longitudinal Trends in Metabolic Markers, and Risk of Pancreatic Cancer in a Heterogeneous Population.* Clin Gastroenterol Hepatol, 2020. **18**(8): p. 1812-1821.e7.
190. Sudlow, C., 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 medicine, 2015. **12**(3): p. e1001779-e1001779.
191. UKBiobank. *UK Biobank research ethics approval*. 2021 16th March 2022]; Available from: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics.
192. UKBiobank. *Accessing UK Biobank Data*. 2020 October 2020 16th March 2022]; v2.3:[This document details the means by which data supplied by UK Biobank can be obtained

and manipulated once access has been approved.]. Available from: https://biobank.ndph.ox.ac.uk/~bbdatan/Accessing_UKB_data_v2.3.pdf.
193. Partlett, C., et al., *Application of the matched nested case-control design to the secondary analysis of trial data.* BMC Medical Research Methodology, 2020. **20**(1): p. 117.
194. NICE. *How should I confirm if a person is overweight or obese?* Clinical Knowledge Summaries 2021 December 2017 [cited 2021 November]; Available from: https://cks.nice.org.uk/topics/obesity/diagnosis/identification-classification/#:~:text=A%20waist%20circumference%20of%2094,of%20obesity%2Drelated%20health%20problems.
195. Bosetti, C., et al., *Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (Panc4).* Ann Oncol, 2012. **23**(7): p. 1880-8.
196. Vrieling, A., et al., *Cigarette smoking, environmental tobacco smoke exposure and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition.* Int J Cancer, 2010. **126**(10): p. 2394-403.
197. Yuan, C., et al., *Cigarette Smoking and Pancreatic Cancer Survival.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2017. **35**(16): p. 1822-1828.
198. Rose, S. and M.J.v.d. Laan, *Why match? Investigating matched case-control study designs with causal effect estimation.* The international journal of biostatistics, 2009. **5**(1): p. 1-1.

199. Ogundimu, E.O., D.G. Altman, and G.S. Collins, *Adequate sample size for developing prediction models is not simply related to events per variable.* Journal of clinical epidemiology, 2016. **76**: p. 175-182.
200. Diabetes.co.uk. *Diabetes Prevalence*. 2019 15th January 2019; Available from: https://www.diabetes.co.uk/diabetes-prevalence.html.
201. NHS. *Health Survey for England 2019*. 2019; Available from: http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/diabetes.aspx#:~:text=The%20increasing%20prevalence%20of%20doctor,17%25%20over%20the%20same%20period.
202. Maisonneuve, P. and A.B. Lowenfels, *Risk factors for pancreatic cancer: a summary review of meta-analytical studies.* Int J Epidemiol, 2015. **44**(1): p. 186-98.
203. Liao, K.F., et al., *Diabetes mellitus correlates with increased risk of pancreatic cancer: a population-based cohort study in Taiwan.* J Gastroenterol Hepatol, 2012. **27**(4): p. 709-13.
204. Park, B.K., et al., *Lifestyle, body mass index, diabetes, and the risk of pancreatic cancer in a nationwide population-based cohort study with 7.4 million Korean subjects.* Br J Cancer, 2022. **127**(3): p. 549-557.
205. Turner, E.L., J.E. Dobson, and S.J. Pocock, *Categorisation of continuous risk factors in epidemiological publications: a survey of current practice.* Epidemiol Perspect Innov, 2010. **7**: p. 9.
206. Naggara, O., et al., *Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms.* AJNR Am J Neuroradiol, 2011. **32**(3): p. 437-40.
207. Lemanska, A., et al., *BMI and HbA1c are metabolic markers for pancreatic cancer: Matched case-control study using a UK primary care database.* PLOS ONE, 2022. **17**(10): p. e0275369.
208. Wu, B.U., et al., *Association of Glycated Hemoglobin Levels With Risk of Pancreatic Cancer.* JAMA Network Open, 2020. **3**(6): p. e204945-e204945.
209. Kawada, T., *Diabetes Mellitus and Pancreatic Cancer: A Risk Assessment.* Clinical Gastroenterology and Hepatology, 2021. **19**(6): p. 1304.
210. McDonnell, D., et al., *Elevated Glycated Haemoglobin (HbA1c) Is Associated with an Increased Risk of Pancreatic Ductal Adenocarcinoma: A UK Biobank Cohort Study.* Cancers, 2023. **15**(16): p. 4078.
211. UKBiobank. *Protocol for a large-scale prospective epidemiological resource*. 2006 21 March 2007 [cited 2021; (AMENDMENT ONE FINAL):[Available from: https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf.
212. Fry, A., et al., *Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population.* American journal of epidemiology, 2017. **186**(9): p. 1026-1034.
213. Goto, A., et al., *Validity of diabetes self-reports in the Saku diabetes study.* J Epidemiol, 2013. **23**(4): p. 295-300.
214. Jackson, J.M., et al., *Validity of diabetes self-reports in the Women's Health Initiative.* Menopause, 2014. **21**(8): p. 861-8.
215. Fry, A., et al., *Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population.* Am J Epidemiol, 2017. **186**(9): p. 1026-1034.
216. Eastwood, S.V., et al., *Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank.* PLoS One, 2016. **11**(9): p. e0162388.
217. Pourhoseingholi, M.A., A.R. Baghestani, and M. Vahedi, *How to control confounding effects by statistical analysis.* Gastroenterol Hepatol Bed Bench, 2012. **5**(2): p. 79-83.

References

End of Thesis