Early Detection of Pancreatic Cancer

Stephen P Pereira¹, PhD; Lucy Oldfield², PhD; Alexander Ney¹, MD; Phil A Hart³, MD; Geri Keane¹, MD; Stephen J Pandol⁴, MD; Debiao Li⁵, PhD; William Greenhalf², PhD; Christie Y Jeon⁴, ScD; Eugene J Koay⁶, PhD; Christopher V Almario⁴, MD; Christopher Halloran², MD; Anne Marie Lennon⁷, PhD; Eithne Costello², PhD

¹Institute for Liver and Digestive Health, University College London, London UK
²Department of Molecular and Clinical Cancer Medicine, University of Liverpool, UK
³Division of Gastroenterology, Hepatology, and Nutrition, The Ohio State University Wexner Medical Center, Columbus, OH, USA
⁴Cedars-Sinai Medical Center, Los Angeles, CA, USA
⁵Biomedical Imaging Research Institute, Cedars-Sinai, Los Angeles, CA, USA
⁶The University of Texas MD Anderson Cancer Center, Houston, TX, USA
⁷Division of Gastroenterology and Hepatology, The Johns Hopkins University, Baltimore, MD, USA

Key words: Pancreatic cancer, screening in high-risk individuals, familial pancreatic cancer, cancer syndromes, biomarkers of early diagnosis

For Correspondence
Prof Eithne Costello, Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Ashton Street, Liverpool L69 3GE, United Kingdom. ecostell@liverpool.ac.uk

Abbreviations:
CPDPC; Consortium for the Study of Chronic Pancreatitis, Diabetes, and Pancreatic Cancer
EHR; electronic health record
HNPCC; Hereditary non-polyposis colorectal cancer
HRI; high-risk individuals
NOD; new-onset diabetes;
PDAC; pancreatic ductal adenocarcinoma (i.e., pancreatic cancer)
UK-EDI; United Kingdom-Early Detection Initiative
CDST; cancer decision support tools
HP; hereditary pancreatitis
Abstract
Pancreatic ductal adenocarcinoma (PDAC) is most frequently detected at an advanced stage. This limits treatment options and contributes to a dismal 5-year survival rate of 3 to 15%. PDAC is relatively uncommon and with current modalities, screening of the asymptomatic adult population is not feasible or recommended. However, screening of individuals in high-risk groups is undertaken. Here we review high-risk groups for PDAC, including individuals with inherited predisposition and patients with pancreatic cystic lesions. We discuss new studies aimed at finding ways of identifying PDAC in high-risk groups, such as individuals with new-onset diabetes mellitus and those attending primary and secondary care practices with suggestive symptoms. We review early detection biomarkers, explore the potential of exploiting social media for PDAC detection, appraise prediction models developed using electronic health records and research data, and examine the application of artificial intelligence to imaging for the purposes of early PDAC detection.

Introduction
Pancreatic ductal adenocarcinoma (PDAC) remains an intractable cancer and a leading cause of cancer deaths\(^1\). Despite advances in the diagnosis and treatment of other gastrointestinal malignancies, such as colorectal and gastric cancers\(^2\), mortality rates for PDAC only marginally surpass the number of newly diagnosed cases and the 5-year survival rates remain as low as 3-15\(^%\)\(^3\)\(^4\). These dismal figures are attributed to both late- and incurable-stage diagnosis as well as high tumour chemo-resistance\(^5\). The latter renders most treatment options ineffective.

The need for prompt diagnosis is recognised globally\(^6\) and a shift toward early detection has been recommended by several healthcare organisations\(^7\). A number of observations support the benefits of earlier PDAC detection. Patients diagnosed with stage I disease experience longer survival times than patients diagnosed with more advanced stages\(^8\). Similarly, incidentally diagnosed PDAC is associated with longer median survival compared to PDAC that is diagnosed when patients have symptoms\(^8\). Survival rates are also dramatically better for those patients diagnosed at an operable versus inoperable stage\(^9\). Unfortunately, up to 85\(\%\) of cases are not surgically resectable when identified and in the UK over half are diagnosed following a non-specific disease course leading to an emergency hospital admission\(^10\)-\(^13\).

Although identifying PDAC as early as possible is essential\(^14\)-\(^15\), there are several challenges. Despite a high mortality rate, PDAC is relatively uncommon, with an incidence of 8 to 12/100,000 and a 1.3\(\%\) lifetime risk of developing the disease\(^3\). The low disease incidence renders screening of the asymptomatic adult population currently unfeasible. Existing diagnostic methods would incur unacceptably high rates of false positive findings\(^16\), and in the context of PDAC, false positives carry significant ramifications. In some cases, a definitive diagnosis requires surgery, which carries the risk of significant morbidity and mortality.

Screening individuals in high-risk groups increases the rate of detection and reduces false-positive results.\(^8\) A number of groups at high risk of PDAC have been identified, including individuals in families with an inherited risk\(^17\)-\(^18\), people with cystic lesions\(^19\)-\(^21\), and those over 50 years of age with newly diagnosed diabetes mellitus (DM)\(^22\)-\(^23\). The following discussion...
focuses on the practicality of screening for PDAC and appraises current clinical studies aimed at facilitating early detection of PDAC.

**High-risk groups**

**Familial/inherited risk**

Two indications for an inherited risk for PDAC are established. A family history of the disease is a risk factor for development of PDAC and germline mutations in specific genes are associated with PDAC.

Although multiple genes have mutations associated with PDAC, the mutations are almost as frequent in sporadic disease as in individuals with a family history. Moreover, families with a strong history of PDAC mostly lack mutations in the genes known to be associated with PDAC. It is probable that for most people, the genetic risk is multigenic, but it is difficult to explain families with multiple cases affecting several generations without assuming a single mutation or tightly linked mutations at a single locus. Registries of such high-risk families indicate a roughly 50% lifetime risk, regardless of whether a causative mutation is known, consistent with a single highly penetrant germline mutation. When the causative mutation is known, it segregates with the disease. However, the lack of a family history in individuals with those same specific mutations indicates that these mutations are context specific.

One such context is a specific single gene mutation that predisposes to PDAC. This is known as Hereditary Pancreatic Cancer (HPC). Another context is predisposition to other cancer types; certain hereditary cancer syndromes, including Hereditary Breast Ovarian Cancer Syndrome (HBOC), Lynch II, Familial Atypical Multiple Mole Melanoma (FAMMM-PC), Peutz-Jegher’s and Li Fraumeni syndrome are associated with PDAC. As described, in some families the inheritance of an HPC mutation is not associated with PDAC. In some of these non-HPC families other forms of cancer are prevalent. For example, HBOC is characterised as an autosomal dominant predisposition for breast/ovarian cancer and many HBOC families have mutations in \textit{BRCA2}. \textit{BRCA2} mutations are also associated with PDAC. In some \textit{BRCA2} HBOC families there are cases of PDAC as well as breast/ovarian. These families are classified as both HBOC and HPC. A spectrum of different \textit{BRCA2} mutations are associated with HBOC. This spectrum does not seem to be significantly different in HBOC families with or without PDAC. Even in HBOC families with no \textit{BRCA2} (or \textit{BRCA1}) mutation the risk of PDAC is greatly elevated. Similarly, Lynch syndrome (Hereditary Non-Polyposis Colorectal Cancer) is associated with mismatch repair mutations that have been linked to PDAC. Not all families with Lynch syndrome have an established high-risk for PDAC. Lynch I syndrome families have elevated risk of colorectal cancer only, while Lynch II syndrome comprises other cancer types, including PDAC. Both Lynch I and Lynch II include families with no identified causative mutation. A third example of a syndrome that includes families with- and without- causative mutations associated with PDAC is FAMMM. Many FAMMM kindreds have mutations in the \textit{CDKN2A} gene (encoding the p16 tumour suppressor), some have cases of PDAC which defines a sub-syndrome called FAMMM-PC. FAMMM-PC can be clinically defined in families without a known causative mutation.
While HBOC, Lynch II, FAMMM-PC, Peutz-Jegher’s and Li Fraumeni syndrome are associated with PDAC, none of these syndromes is characterized by multiple cases of PDAC. In contrast Familial Pancreatic Cancer (FPC)\textsuperscript{36} is a syndrome that is defined as having multiple first degree relatives with PDAC and is exclusive of other cancer syndromes. FPC can be associated with \textit{BRCA2} mutations and possibly \textit{CDKN2A} or even \textit{PALB2} (a breast cancer susceptibility gene whose protein product enables \textit{BRCA2} anchorage to nuclear structures) mutations, but in most cases the causative mutation is unknown. It is only in families with autosomal dominant predisposition, with or without a known causative mutation, that screening can be justified. In the absence of autosomal dominance, risk for an individual may be very high but it is difficult or impossible to prospectively predict. Identifying autosomal dominance requires careful examination of family history, ideally with identification of germline mutations\textsuperscript{37}.

**Cystic lesions**

Pancreatic cysts are found in approximately 8\% of individuals aged over 70\textsuperscript{38}. They are of interest because two of the three precursors to PDAC, intraductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNs), are pancreatic cysts. IPMNs and MCNs are collectively referred to as mucinous cystic lesions (MCLs). In contrast to the third precursor lesion, pancreatic intraepithelial neoplasia (PanIN), which can only be identified on surgical histopathology, pancreatic cysts are easy to detect and are incidentally found in 3\% of individuals undergoing a CT scan. Identification of MCLs therefore offers the potential for early detection of PDAC. Two issues complicate this simple concept. First, not all pancreatic cystic lesions are IPMNs and MCNs. Many are cystic lesions with no risk of malignant transformation which do not require surveillance. It is estimated that a cyst seen incidentally on Magnetic Resonance Imaging (MRI) has a 10 in 100,000 chance of being a mucinous invasive malignancy and a 17 in 100,000 chance of being a ductal cancer\textsuperscript{21}. Second, most IPMNs and MCNs do not progress to PDAC. Currently available clinical tools are imperfect at differentiating benign cysts that are safe to discharge, from MCLs that harbor high-grade dysplasia or PDAC and require surgical resection, versus those MCLs which have low-grade dysplasia and are safe to watch. This is a significant problem as highlighted by the fact that 25\%\textsuperscript{39} of patients who have a presumed MCL removed are ultimately found to have a cyst with no malignant potential, while up to 78\%\textsuperscript{40} of branch duct IPMNs which undergo surgical resection do not have high-grade dysplasia or PDAC, and in hindsight did not require surgery. Progress has been made in overcoming these challenges. We now know that different types of pancreatic cysts have specific mutational profiles which can be used to identify the type of cysts with a high level of confidence. For example, serous cystadenomas, one of the commonest types of cysts with no malignant potential, are associated with mutations in \textit{VHL}. MCLs can have mutations in \textit{KRAS}, while \textit{GNAS} mutations are highly specific. Overall, almost two thirds of IPMNs and MCNs harbor a mutation in \textit{KRAS} or \textit{GNAS} in the cyst fluid\textsuperscript{41}. A novel recently-described approach is to combine the most specific clinical features with molecular markers. A recent study of 862 individuals undergoing surgery for pancreatic cysts, found that a combined molecular and clinical marker panel was more accurate than current clinical features alone, and use of these markers would have decreased the number of unnecessary operations by 60\%\textsuperscript{41}.

Progress towards identifying those MCLs that harbor high-grade dysplasia or early invasive PDAC has been made. There is increasing knowledge about the progression of IPMNs from a
genetic perspective which suggests that early lesions are heterogeneous, while those with high-grade dysplasia have a smaller number of homogeneous driver genes\textsuperscript{42}. Preliminary studies evaluating the presence of mutations in \textit{PIK3CA}, \textit{SMAD4} and \textit{TP53} in cyst fluid are promising, identifying almost 80\% of IPMNs with high-grade dysplasia or cancer\textsuperscript{43}. Advances have also occurred in endoscopy. Here, confocal endoscopes have been developed which can be passed into the cyst, providing \textit{in vivo} real time imaging of the cyst lining. Studies have shown good sensitivity with excellent specificity for differentiating benign cysts requiring no follow-up from IPMNs\textsuperscript{44-46}. Further work is needed to identify the small number of individuals with MCLs who are at highest risk of progression to PDAC and require intensive surveillance, from the vast majority with IPMNs and MCNs which will never progress.

**New-onset diabetes mellitus**

The relationship between PDAC and DM is multifaceted. People with long-standing type 2 DM (>5 years) have a 1-1.5 fold increased risk of PDAC. However in those with DM of <1 year duration, the relative risk of PDAC increases to 5.4 fold\textsuperscript{47} with substantial research pointing towards PDAC-induced hyperglycaemia and DM\textsuperscript{48}. When diagnosed with PDAC, around 80\% of patients have abnormal fasting glucose or glucose intolerance\textsuperscript{49,50}. This is not the case for other common cancers where the prevalence of DM is similar to that of non-cancer controls\textsuperscript{51}. The DM experienced by most PDAC patients is of recent onset, diagnosed less than 24–36 months before PDAC diagnosis, and is improved by surgical resection of the tumour\textsuperscript{48}. Thus new-onset DM (NOD) can be considered an early warning sign of PDAC\textsuperscript{22}, and individuals with NOD are the highest risk group for sporadic PDAC.

Approximately 0.8-1\% of individuals aged >50 years with new-onset DM have DM secondary to PDAC. However, distinguishing PDAC-associated DM from the more prevalent type 2 DM is challenging, and strategies that enrich for those with PDAC in the NOD group are urgently required. The opportunities that NOD presents for early detection have been recently reviewed elsewhere\textsuperscript{52}.

**Practicalities of screening**

**Who should be screened?**

The US Preventive Services Task Force and other guidelines recommended against screening for PDAC in asymptomatic average risk adults\textsuperscript{6,53,54}. Screening of the general population carries risk and there is currently no evidence that it reduces mortality or is cost effective\textsuperscript{54,55}. However, international guidelines and a white paper on the early detection of PDAC recommend targeted screening of individuals at a >5\% lifetime risk of developing PDAC (Figure 1)\textsuperscript{6,54}. This includes individuals with at least two first-degree-relative with PDAC, patients with hereditary pancreatitis (HP) or certain genetic syndromes (e.g. Peutz-Jeghers syndrome, p16, BRCA2, PALB2 and Lynch syndrome)\textsuperscript{6,54} along with those with MCLs of the pancreas\textsuperscript{19-21}.

More recently, other high-risk groups have been identified, such as individuals with NOD, discussed above, and those with modifiable (e.g. smoking, obesity) or non-modifiable (e.g. age) risk factors, certain comorbidities (e.g. obesity, chronic pancreatitis) or early symptoms
(some of which are present for >1 year prior to diagnosis) \(^{10,12,13,56}\). With the exception of jaundice, weight loss and NOD, most individual risk factors or symptoms only provide a modest (1.5-fold to 3-fold) increased risk of developing PDAC \(^{10,54}\). However, by combining these factors through cancer decision support tools (CDST) the 5% risk threshold can be reached in a proportion of individuals; defining a further group suitable for screening \(^{10,12,13}\).

How should individuals be screened?
Successful early detection should enable treatments that improve patient survival and well-being, without solely increasing the time between diagnosis and death, or lead time. All PDAC screening programmes aim to detect and treat T1N0M0 cancers or high-grade dysplastic premalignant lesions (i.e. PanIN3 or mucinous cystic tumours with high-grade dysplasia) \(^{6,54}\).

Established screening programmes exist e.g. the pan-European EUROPAC registry (http://www.europac-org.eu)\(^{36}\), North American National Familial Pancreatic Tumor Registry \(^{57}\) and the German National Case Collection for familial pancreatic Cancer (FaPaCa)\(^{58}\). The protocols of these prospective registry studies vary between each programme but generally include cross-sectional imaging and blood tests (with tumour markers) at registration and then annual non-radiating forms of imaging by MRI or endoscopic ultrasound (EUS). If a suspicious lesion is identified further investigations and treatment are arranged as clinically necessary.

Approximately 8% of all PDACs are believed to arise from premalignant mucinous cystic lesions of the pancreas\(^{59}\). Guidelines recommend immediate surgical resection for any lesions with high-risk stigmata and regular surveillance with interval imaging (MRI or EUS) for all mucinous or indeterminate lesions\(^{21,60}\). Surveillance protocols vary between the different guidelines, but typically recommend 6-monthly cross-sectional imaging for the first year, and then annually thereafter if no changes are detected. More frequent surveillance is advocated for larger lesions or those with worrisome features\(^{21,60}\).

For the newly established high-risk groups such as NOD there are no established guidelines or screening programmes. However, studies underway in the US\(^{61}\) and UK\(^{62}\) are actively investigating how best to detect PDAC in individuals with NOD. In the UK, CDSTs have recently been introduced into primary care practices\(^{12}\). Due to the low incidence of PDAC, general practitioners (GPs) may see only one case of PDAC every 5 years. Despite repeated primary care consultations with symptoms which may be attributable to PDAC prior to diagnosis \(^{12,13,63}\), patients often encounter delays in the workup process\(^{64}\). To address this, GPs in the UK are gaining greater access to diagnostic investigations e.g. computerized tomography (CT) or rapid referral clinics such as the newly developed Multidisciplinary Diagnostic Centres (MDC)\(^{65}\) which enable the rapid assessment of patients with concerning but non-specific symptoms. Referrals can be triggered by CDST-defined risk or GP’s clinical assessment. The utility and outcomes of the current programme are subject to ongoing assessment via the Accelerate, Coordinate, Evaluate (ACE) Programme, an initiative supported by Cancer Research UK and Macmillan Cancer Support\(^{65}\).

Role for biomarkers in early detection of PDAC

Challenges and opportunities
Biomarkers could play a vital role in early detection of PDAC by enriching for individuals in high-risk groups with the highest chances of a cancer diagnosis, thus helping clinicians prioritise individuals for screening (Figure 2). However, despite thousands of published papers, no single candidate biomarker has translated to clinical use for the early detection of PDAC. Indeed, biomarker development for this disease faces unique challenges. The low incidence of PDAC means that acquiring the quantity of samples necessary for biomarker development is not easy, and large national and international collaborations are required. Pancreatic tumours are highly heterogeneous, both within and between individuals. Consequently, single biomarkers will most likely lack high-sensitivity for PDAC detection, and robust panels of biomarkers will be required.

Understanding and accounting for potential confounding factors is an important component of PDAC biomarker development. Most major studies now incorporate samples from disease controls, such as chronic pancreatitis. Moreover, awareness that the presence of obstructive jaundice can lead to false-positive biomarker findings is increasing. Currently, the knowledge that a high proportion of PDAC patients has DM is not well accounted for in biomarker studies, and emerging biomarkers risk having an association with DM rather than with PDAC. Finally, a very significant deficit in early detection studies for PDAC has been the lack of bespoke pre-diagnostic cohorts. Although large population-based cohorts that have recruited healthy individuals, such as UKCTOCS, EPIC and others contain patient samples that predate PDAC diagnosis, crucial information, such as family history and co-morbidity data, including the presence of chronic pancreatitis and/or DM, are not consistently available.

When calculating the cost versus benefit of a biomarker-assisted screening programme, the costs of both the initial screening and subsequent tests required to confirm the diagnosis should be taken into account. As both true positive and false positive tests require further investigation, high specificity biomarkers are required. Ghatnekar et al. developed a framework for modelling cost and quality adjusted life-years of serum biomarker-mediated early detection of PDAC in individuals with NOD. This study, conducted within the Swedish healthcare system, concluded that biomarker-mediated screening is highly desirable.

**Current promising biomarkers for the early detection of PDAC**

Using samples from patients already diagnosed with PDAC in biomarker development may result in biomarkers indicative of symptomatic disease. An alternative strategy is to utilise samples from mouse models, early-stage PDAC (stage I/II), and precursor lesions (PanIN, IPMN, MCN) in biomarker development. Research into the molecular changes occurring during the sequential progression from lesion through early-stage PDAC and finally to advanced disease is generating a greater understanding of PDAC development and progression and providing a basis for designing rational early detection biomarker development strategies for PDAC. Markers generated through this process represent aberrant changes at the genetic, transcriptomic, metabolomic and proteomic level that correlate with the earliest histological stages of PDAC development (Table 1). Finally, the use of high-risk groups as controls and combining molecular markers with clinical features will be an important consideration when identifying biomarkers capable of selecting populations at highest risk of PDAC development for screening.
Currently there are no biomarkers validated for early detection of PDAC. However, a number of published biomarkers demonstrate potential for future evaluation. Elevation of Carbohydrate antigen 19-9 (CA19-9), the only biomarker routinely used in the management of PDAC, has been shown in samples taken prior to diagnosis\(^76\). However, it is not recommended for screening. CA19-9 is also not expressed in individuals with a Lewis-negative genotype, only 65% of patients with resectable PDAC have elevated serum levels\(^77\) and it is elevated in other benign and malignant diseases. The diagnostic value of CA19-9 may, in the future, be improved by measuring the antigen on individual proteins or using it in combination with additional markers. In the distinction of PDAC from benign disease and healthy controls, Lee et al.\(^78\) reported an improved AUC using a combination of CA19-9 and CEMIP (cell migration-inducing hyaluronan binding protein), a protein involved in the degradation of hyaluronan. compared to CA19-9 alone (0.94 vs 0.89). Importantly, CEMIP showed a diagnostic yield of 86.1% (68/79) in CA19-9 Le-negative PDAC.

While a comprehensive review of candidate biomarkers for early detection is beyond the scope of this review, we have highlighted several novel biomarkers resulting from studies designed with early detection in mind, summarised in Table 2. Note, genetic biomarkers were discussed in more detail in an earlier section of this review\(^41\).

<table>
<thead>
<tr>
<th>Source</th>
<th>Biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteomic</td>
<td>CA19-9, CEA, CEMIP, TSP-1, TSP-2, VNN1 (downstream markers), MUC1, MUC2</td>
</tr>
<tr>
<td>Metabolomic</td>
<td>M2-pyruvate kinase (M2-PK), palmitic acid, inositol, proline, ceramide, phosphatidyl choline, Isocitrate</td>
</tr>
<tr>
<td>Genetic</td>
<td>KRAS, GNAS, SMAD4, TP53</td>
</tr>
<tr>
<td>Transcriptomic</td>
<td>miR-486.5p, -16, -24, -27a, -30a.5p, -323.3p, -20a, -25, -29c, -483.5p</td>
</tr>
</tbody>
</table>

Table 1 Biomarkers representative of early PDAC development organized by source

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Marker Type</th>
<th>Sample</th>
<th>Technique</th>
<th>AUC</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSP-2 + CA19-9</td>
<td>Protein</td>
<td>Blood - Plasma</td>
<td>ELISA / MS</td>
<td>0.96 (PDAC all stages vs HC)</td>
<td>79</td>
</tr>
<tr>
<td>TSP-1 + CA19-9</td>
<td>Protein</td>
<td>Blood - Serum</td>
<td>MS</td>
<td>0.86 (Pre-diagnostic PDAC vs HC)</td>
<td>80</td>
</tr>
<tr>
<td>MUC1 + MUC2</td>
<td>Protein</td>
<td>Pancreatic juice</td>
<td>IHC</td>
<td>0.85 (malignant vs benign IPMN)</td>
<td>81</td>
</tr>
<tr>
<td>LYVE-1 + REG1A + TFF1</td>
<td>Protein</td>
<td>Urine</td>
<td>ELISA</td>
<td>0.93 (stage I, II PDAC vs HC)</td>
<td>82</td>
</tr>
<tr>
<td>miR-16, -24, -27a, -30a.5p, -323.3p, -20a, -25, -29c, -483.5p + CA19-9</td>
<td>miRNA / Protein</td>
<td>Blood - Serum</td>
<td>microRNA array</td>
<td>0.93 (stage I, II PDAC vs HC)</td>
<td>83</td>
</tr>
<tr>
<td>Cysteamine / GSH / PPAR-γ (VNN1)</td>
<td>Protein</td>
<td>Blood - Serum</td>
<td>HPLC/ELISA</td>
<td>0.84, 0.86, 0.82 (PDAC vs HC)</td>
<td>84</td>
</tr>
</tbody>
</table>
Future validation of these candidate biomarkers and others must demonstrate sufficient evidence to support their application in early detection through the use of pre-diagnostic human samples and appropriate controls (e.g. from established high-risk groups and benign disease). Ultimately, the most promising early detection strategy will likely come from a discrete panel of biomarkers used in combination with clinical features.  

### Synthetic biomarkers

An emerging area of research and development involves the engineering of probes that are activated by tumour cells or the stromal cells within the tumour microenvironment. The activated probes can be detected in any number of ways, including in the blood, urine, or imaging methods. These synthetic biomarkers have advantages over endogenous biomarkers (e.g., CA19-9) because signals from normal tissues are reduced or eliminated, the signals can be engineered to be highly specific for cancer, and the method can be tailored for highly sensitive detection tools. Proof-of-concept has been demonstrated in several settings, including early detection of ovarian cancer in mouse models through use of cleavable substrates by upregulated proteases in the tumour microenvironment\(^9\), sequential activation by tumour acidity and hypoxia for ultrasensitive imaging detection of tumours as small as 1 mm in mice\(^9\), and identification of residual cancer after surgery in humans through a protease-activated fluorescent probe\(^9\). Development of synthetic biomarkers for PDAC may significantly impact early detection efforts and other clinical applications in the future.

### Ongoing trials and studies for early PDAC detection

**Chronic Pancreatitis, Diabetes and Pancreatic Cancer (CPDPC) Consortium and the New Onset Hyperglycaemia and Diabetes Cohort**

The lack of tools for early diagnosis of pancreatic disease (both chronic pancreatitis and PDAC), triggered a number of initiatives from the National Institutes of Health in the USA aiming for improvements in outcome for patients with these diseases. These initiatives were reinforced by the Recalcitrant Cancer Act passed by the US Congress in 2012 and signed in 2013. These actions were greatly facilitated by several partners, including the National Pancreas Foundation, the Kenner Family Research Fund, and the Pancreatic Cancer Action Network.
A key outcome of these efforts was the creation and funding of the CPDPC Consortium. The main goals of the CPDPC are to establish large prospective cohorts of carefully phenotyped patients (both paediatric and adult) with collection of radiologic data and biospecimens to be used for diagnosing and monitoring of disease status. The NOD cohort of CPDPC is designed to recruit, phenotype and collect biospecimens on 10,000 patients with new-onset of diabetes after age 50. The age of onset of diabetes after age 50 was chosen as the best clinical marker available for early detection of PDAC although the incidence of PDAC in this group is predicted to be only 1-2% over a 3-year period. The collected biospecimens and clinical data will be available for validation of promising biomarkers for the early detection of PDAC.

**UK Early Detection Initiative for Pancreatic Cancer**

Similarly, in the United Kingdom, Cancer Research UK are funding the UK Early Detection Initiative (UK-EDI) to recruit 2,500 individuals aged >50 years who were diagnosed with NOD in the previous six months (UK-NOD). The UK-NOD cohort is designed to recruit from both primary and secondary care centres, and to collect questionnaire and clinical data, alongside longitudinal biosamples over three years. As with the NOD cohort of the CPDPC, data and biospecimens will be made available for research on early detection of PDAC, including validation of existing biomarkers that have shown promise for early detection as well as supporting new discovery programs. Working across both US and UK cohorts, the sensitivity/specificity for the detection of PDAC by combined biomarker and epidemiological/clinical feature analysis will be evaluated. Finally, the cost-effectiveness of diagnosing PDAC earlier in the setting of NOD will be assessed.

With appropriate EU funding directed towards early PDAC detection, studies such as the NOD cohort of the CPDPC and UK-EDI could be undertaken more widely in Europe. This would ensure that future biomarker-driven PDAC screening is relevant to other countries and healthcare systems within Europe.

**ADEPTS**

ADEPTS (Accelerated Diagnosis of neuroendocrine and Pancreatic TumourS) is a multi-centre diagnostic accuracy study funded by Pancreatic Cancer UK (PCUK), which aims to improve diagnostic pathways in pancreatic cancer. The overall objective of the study is to develop a diagnostic tool (which combines the use of refined CDSTs and a minimally-invasive blood test for circulating biomarkers) that can be used for screening in selected higher risk patient cohorts. The health economics of implementing such a tool that will allow prioritisation of urgent investigations in patients with a raised combined risk score, are also being studied. Under this study, a large, multicentre, prospective sample collection of liquid and tissue biopsies from healthy and symptomatic individuals, as well as from those known to have a genetic association or high-risk cystic lesions of the pancreas, is underway.

**CPDPC DETECT**

A small pilot study suggested that PDAC-DM may be distinguished from the more prevalent type 2 DM by a blunted pancreatic polypeptide response to a mixed meal. An ongoing study supported by the CPDPC Consortium called DETECT (Evaluation of a Mixed Meal Test for Diagnosis and Characterization of Pancreatogenic Diabetes Secondary to Pancreatic Cancer and Chronic Pancreatitis) seeks to validate this observation. Additionally, this study will comprehensively examine differences in glucose homeostasis between these subtypes of...
diabetes, including insulin secretion, beta cell function, glucagon response, and incretin hormone response. These results will provide the opportunity to further refine our approach to early detection of PDAC in adults with NOD.

The European Registry of Hereditary Pancreatitis and Familial Pancreatic Cancer (EUROPAC)

The European Registry of Hereditary Pancreatitis and Familial Pancreatic Cancer (EUROPAC) was established in Liverpool, UK as a registry of patients at high-risk of PDAC. Such a registry is invaluable in developing secondary screening techniques for the identification of early cancer. EUROPAC quantified the risk of PDAC in HP as approximately 40% lifetime risk. They further characterized this risk in terms of different genotypes. EUROPAC has also shown that the age of onset of PDAC becomes progressively earlier with each generation in FPC and that DM predisposes for cancer in HP. On this basis EUROPAC has proceeded with the development of a screening programme based on blood tests and imaging.

Current advances in artificial intelligence and deep learning methodologies in relation to the early detection of PDAC

Employing Social Media for early PDAC detection

Social media has forever changed how society communicates. There are 3 billion social media users worldwide and this ‘big data’ resource presents a unique opportunity to learn about the lives of non-experimental patients outside the walls of healthcare facilities. As patients with PDAC first experience subtle, vague symptoms that can precede the diagnosis by years, it becomes important for medical researchers to learn how to harness the power of social media for identifying online signals of early PDAC.

Using social netnography techniques - a type of ethnography that analyzes perceptions and behaviours of individuals online, it may be possible to develop social media and online behavior pattern recognition algorithms for diagnosing PDAC at an early stage. To achieve this, researchers can partner with social media mining services to identify individuals online with a self-reported PDAC diagnosis and then collect all of their prior, de-identified and publicly-available online posts, many of which will have been posted before their cancer diagnosis. By conducting topic modeling and thematic analysis on the corpus of posts, one can identify themes and online behavior signals that may be predictive of early PDAC.

Scientists can also employ prospective methods that track users who opt-in across social and other online platforms. For example, researchers can use programmatic display ads and other digital channels to find online users that precisely match the attributes of a previously identified PDAC segment. Once the opt-in, PDAC ‘lookalike’ user group is built, vast and highly granular online insights can be prospectively collected and analyzed in real-time, covering every data category such as demographic, behavioral, contextual, and PDAC diagnosis, among many others. This data may then lead to robust pattern recognition algorithms informed by their social media and online behavior and commercial data that may signal early PDAC and facilitate earlier diagnosis and treatment.
Electronic health records-based models
Electronic health records (EHR) offer opportunities to utilize longitudinal and cumulative healthcare data to build prediction models. Parametric models that inform high-risk for PDAC among individuals with NOD or new-onset pre-diabetes have been developed, with varying predictive performances for PDAC. A model of NOD patients residing in Minnesota, USA incorporated age at onset of DM, and changes in weight and blood glucose to identify persons at 4.5% 3-year predicted risk of PDAC with 78% sensitivity and 80% specificity (AUC = 0.87). Another EHR model, developed using a UK population with NOD, incorporated age, BMI change, smoking, DM medications, proton-pump-inhibitors, changes in hemoglobin A1c, total cholesterol, creatinine and alkaline phosphatase. Amongst individuals with NOD, this model identifies a population with 5% 3-year predicted risk of PDAC with 11% sensitivity and 99.7% specificity (AUC = 0.82). The same model applied to individuals with pre-diabetes showed lower accuracy (AUC = 0.71) but consistent direction of association with PDAC as previously. A Medicare claims-based model among individuals with NOD incorporating multiple health indicators including pancreatitis, dyspepsia, depression, abdominal pain, weight jaundice, nausea/vomiting has been applied to an elderly population with NOD. The model identifies persons at 3.5% predicted risk of PDAC over 1-year (AUC = 0.73). Insurance claims-based models with diagnoses, such as pancreatitis, dyspepsia, abdominal pain, weight loss, and jaundice, incorporating parametric or machine-learning methods have also been developed with limited diagnostic performance (AUC = 0.73). Reliance on coded diagnoses without laboratory parameters limited the performance of these models. Data emanating from routine health examination in Korea have been used to develop a time-to-event prediction model incorporating age, sex, height, BMI, smoking, alcohol consumption, blood and urine glucose. This model performed as well (AUC = 0.81) as the UK model, but had low positive predictive value (0.3-0.4% risk over 10 years) because it was evaluated in the general unselected population aged ≥40 years. A consistent performance characteristic of EHR-based models is high specificity and low sensitivity, leading to high false negative rates. Thus, their value as the first sieve to identify persons who need further work-up for PDAC risk may be limited. Beyond aforementioned studies, development of models incorporating both parametric and machine-learning methods on selected patients are underway at large U.S. health systems, including Kaiser Permanente Southern California and the VA Health System (PRedictiOn Algorithms for the DeTECTion of Early Stage Pancreatic Cancer (PRO-TECT study).

Research data-based models
Data specifically collected for research purposes or as part of population-based surveys have been valuable for identifying populations at high-risk for PDAC. Data from the U.S. National Health Interview Survey and the Prostate, Lung, Colon and Ovarian Trial were used to develop an artificial neural network-based prediction model for PDAC. This model reached an AUC of 0.85 and incorporated data on age, sex, race/ethnicity, DM, emphysema, asthma, stroke, cardiovascular diseases, ulcers, other cancer, hypertension, smoking status, physical activity, alcohol consumption, and family history of PDAC. Other research interview-based models have not reached similar levels of performance, but consistently demonstrate age, sex, smoking, DM, family history of PDAC, abdominal symptoms and blood group as factors that point to increased risk of PDAC. Because information on smoking or drinking intensity and blood type are not readily available in EHR, implementation of models based on research data poses some challenge for the general population.
Deep learning methodologies applied to abdominal imaging

The role of imaging for early detection of PDAC has been reviewed recently. Here, we focus on the emergence of Artificial Intelligence (AI) toward this goal in terms of its application to imaging. Notably, AI may play an important role in early detection of PDAC by identifying not only the physical location of the primary tumor but also its secondary effects on the body. Toward identification of the primary tumor, Fishman and colleagues have described a radiomics-based machine learning algorithm to differentiate PDAC from benign situations (i.e., normal pancreas and pancreatitis) with high specificity and sensitivity. Other applications of radiomics and quantitative imaging approaches have shown that the enhancement and morphology of the primary tumors have biological underpinnings and clinical relevance, suggesting that quantitative imaging and further application of AI to these imaging features can provide non-invasive insight into the disease. This insight may have relevance to early detection through better stratification and personalized approaches to screening in high-risk individuals.

In terms of secondary effects of pancreatic cancer on the body, weight loss has been validated as one of three key factors to predict early stage disease in patients with NOD, and exocrine insufficiency appears to be a contributing cause. Indeed, it has been known that sarcopenia and fat loss are part of a wasting syndrome of PDAC and may have prognostic implications. The non-invasive measurement of different body compartments has recently been fully automated and applied to PDAC through AI methodologies, and this may be easily incorporated into the evaluation of routine diagnostic imaging. For patients undergoing screening or imaging for reasons other than the pancreas, the measurement of body compartments, especially over time, could provide important metrics in the earlier detection of pancreatic cancer.

AI assisted CT methods and novel MRI methods

A significant concern in the field is that current imaging methods may not be adequate to identify tumours in the pancreas when treatment would be optimal. A retrospective review of CT scans performed for other indications prior to PDAC diagnosis showed no evidence of a pancreas mass in the majority of patients 6 months or earlier before the diagnosis of PDAC. Thus, there is a need for advanced external body imaging techniques to improve detection of PDAC at an earlier stage than is presently possible. The advances in methods for external body imaging should be accompanied by endoscopic advancements in imaging in order to obtain tissue for validating the diagnosis.

Medical imaging such as CT and MRI plays an essential role in PDAC diagnosis by allowing a comprehensive evaluation of the morphological and biological changes in parenchyma and duct of the pancreas. The ability to detect pre-cancerous tissue changes in the pancreas using medical imaging among high-risk individuals has also been demonstrated. However, visual assessment by imaging physicians are qualitative, subjective, and prone to errors and intra- and inter-observer variabilities. More importantly, many distinguishing image features are hidden from human observers.

AI is a powerful analysis tool based on human brain's neural structure. Roffman et al. used AI to predict non-melanoma skin cancer by using personal health data (e.g., gender,
race, Hispanic ethnicity, hypertension, heart disease, exercise habits, history of stroke, etc.) commonly available in electronic medical record (EMR) systems. Muhammad et al. used artificial intelligence to analyze available personal health data to calculate risk for PDAC in the general population and to identify high-risk individuals. Imaging represents more sensitive and specific information for parenchyma and duct of the pancreas than personal health data. Radiomic analysis of medical images using deep learning may allow identification of unique image features in pre-diagnostic images to allow accurate prediction of PDAC in the near future.

Summary
A recent study of 3.9 million cancer patients in seven countries (Australia, Canada, Demark, Ireland, New Zealand, Norway and the UK) examining seven sites of cancer (oesophagus, stomach, colon, rectum, pancreas, lung, and ovary) found PDAC to have the lowest 5-year survival rates (ranging from 7.9% in the UK to 14.6% in Australia). Early diagnosis will undoubtedly play an important role in improving these figures, and as our review points out, progress has been made. The establishment of new tailor-made cohorts (of individuals with NOD or with symptoms) provides unique pre-diagnostic resources for biological and epidemiological marker discovery and validation. Careful and ethical use of existing data, whether through social media or EHR has the power to facilitate prediction models, while AI applied to imaging offers the possibility of detecting earlier lesions. With respect to mucinous cysts, identifying the small number of individuals with MCLs at the highest risk of progression to PDAC is still a key knowledge gap. Much work remains to be done to improve the early detection of PDAC, including the ongoing studies reviewed here. The cohort studies underway are vital in many respects, not least because they serve to increase awareness amongst healthcare providers and patients alike, of the symptoms of PDAC and its link with NOD. Advances in early detection will go hand-in-hand with improvements in treatment to extend survival times for people with PDAC.

Figure Legends
Figure 1: Overview of Screening Programmes for PDAC
Figure 2: PDAC early detection strategies including identifying high-risk groups, creating resources, biomarker development (biospecimens and deep learning applied to diagnostic imaging), prediction model construction from medical health records and research databases.

Search Strategy and Selection Criteria References for this Review were identified through searches of PubMed for papers published in English with the search terms “early detection”, “pancreatic cancer”, “high-risk”, “screening”, “artificial intelligence” and “biomarker” from 1993 until September, 2019. Articles were also identified through searches of the authors’ own files. The final reference list was generated on the basis of relevance to the scope of this review, with a focus on the most recently published papers.

Contributors section
SPP is supported by the UCLH/UCL Comprehensive Biomedical Centre, which receives a proportion of funding from the Department of Health’s National Institute for Health Research
(NIHR) Biomedical Research Centres funding scheme. PH is funded by the National Cancer Institute and National Institute of Diabetes And Digestive and Kidney Diseases (NIDDK) U01DK108327. CYJ supported by National Institutes of Health (USA) R21CA220073, R01CA230442, U01DK108314. SP supported by National Institutes of Health (USA) U01 DK108314. AML was supported by the Lustgarten Foundation for Pancreatic Cancer Research, The Sol Goldman Center for Pancreatic Cancer Research, The Benjamin Baker Scholarship, and the National Institutes of Health Grants P50-CA062924. EC, LO, CH and WG are funded by Cancer Research UK grants, C7690/A26881 and C52547/A28210. EC is funded by North West Cancer Research, Pancreatic Cancer Action, Pancreatic Cancer Research Fund. CH is supported by National Institute for Health Research (Research for patient benefit), Pancreatic Cancer UK and the Royal College of Surgeons.

The content is solely the responsibility of the authors and does not necessarily represent the official views of funders.

Conflicts of interest section

Prof. Pereira has nothing to disclose.

Prof. Halloran reports grants from Cancer Research UK, The National Institute for Health Research (Research for patient benefit), Pancreatic Cancer UK and the Royal College of Surgeons.

Dr. Lennon reports grants from National Institutes of Health Grants P50-CA062924., grants from Lustgarten Foundation for Pancreatic Cancer Research, other from The Benjamin Baker Scholarship, other from The Sol Goldman Center for Pancreatic Cancer Research, during the conduct of the study; In addition, Dr. Lennon has a patent CancerSEEK pending.

Dr. Keane has nothing to disclose.

Dr. Hart has nothing to disclose.

Dr. Koay reports grants from Philips Healthcare, grants from Project Purple, grants from National Institutes of Health, during the conduct of the study; other from Taylor and Francis, LLC, outside the submitted work.

Dr. Pandol reports grants from National Institutes of Health, USA, during the conduct of the study;

Dr. Greenhalf has nothing to disclose.

Dr. Almario has nothing to disclose.

Dr. Costello has nothing to disclose.

Dr. Ney has nothing to disclose.

Dr. Jeon has nothing to disclose.

Dr. Oldfield has nothing to disclose.

Dr. Li has nothing to disclose.

Figure 2: Early detection efforts for pancreatic cancer (from top left, going counterclockwise): Efforts include biospecimens from multiple locations, including the mouth, blood, electronic medical record, urine, stool, tissue, and deep learning applied to diagnostic imaging. Interpretation and integration of these sources of data, and other emerging areas such as synthetic biomarkers, will be key areas of research in the future.

References


62. UK Early Detection Initiative for Pancreatic Cancer. [www.uk-edi.co.uk](http://www.uk-edi.co.uk) (accessed Sept 17 2019).
20


