

BMJ Open A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer

M G Keane,¹ L Horsfall,² G Rait,² S P Pereira¹

To cite: Keane MG, Horsfall L, Rait G, *et al.* A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer. *BMJ Open* 2014;**4**:e005720. doi:10.1136/bmjopen-2014-005720

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-005720>).

Received 21 May 2014
Revised 29 October 2014
Accepted 30 October 2014



CrossMark

¹Institute for Liver and Digestive Health, University College London, London, UK

²Research Department of Primary Care and Population Health, University College London, London, UK

Correspondence to

Dr SP Pereira;
stephen.pereira@ucl.ac.uk

ABSTRACT

Objectives: Pancreatic ductal adenocarcinoma (PDAC) and biliary tract cancers (BTC) are often diagnosed late and at an advanced stage. Population-based screening programmes do not exist and diagnosis is primarily dependent on symptom recognition. Recently symptom-based cancer decision support tools (CDSTs) have been introduced into primary care practices throughout the UK to support general practitioners (GPs) in identifying patients with suspected PDAC. However, future refinement of these tools to improve their diagnostic accuracy is likely to be necessary.

Setting: The Health Improvement Network (THIN) is a primary care database, which includes more than 11 million electronic patient records, from 562 GP practices in the UK.

Participants: All patients with a diagnosis of PDAC or BTC between 2000 and 2010 were included in the study along with six matched controls; 2773 patients with PDAC, 848 patients with BTC and 15 395 controls.

Primary and secondary outcome measures: The primary aim of this study was to determine the early symptom profiles of PDAC and BTC. Secondary aims included comparing early symptom trends between BTC and PDAC, defining symptom onset in PDAC and evaluating trends in routine blood tests nearest to the time of diagnosis.

Results: In the year prior to diagnosis, patients with PDAC visited their GP on a median of 18 (IQR 11–27) occasions. PDAC was associated with 11 alarm symptoms and BTC with 8. Back pain (OR 1.33 (95% CI 1.18 to 1.49) $p < 0.001$), lethargy (1.42 (95% CI 1.25 to 1.62) $p < 0.001$) and new onset diabetes (OR 2.46 (95% CI 2.16 to 2.80)) were identified as unique features of PDAC.

Conclusions: PDAC and BTC are associated with numerous early alarm symptoms. CDSTs are therefore likely to be useful in identifying these tumours at an early stage. Inclusion of unique symptoms, symptoms with an early onset and routinely performed blood tests is likely to further improve the sensitivity of these tools.

BACKGROUND

Pancreatic ductal adenocarcinoma (PDAC) and biliary tract cancers (BTC) are lethal tumours that are often diagnosed late when

Strengths and limitations of this study

- A significant strength of this study is that the Health Improvement Network (THIN) database includes routinely recorded primary care data that is not subject to recall bias.
- The database includes a large cohort of patients with pancreatic ductal adenocarcinoma and biliary tract cancers (BTC).
- The patient population within the database has been shown to be representative of the UK population.
- This is the first study to evaluate early symptoms of BTC in primary care.
- A patient's final diagnosis within the THIN database is dependent on correct coding by their general practitioner. However, this has generally been found to be accurate in patients with cancer, when electronic records have been compared to hospital correspondence or written medical records.
- The study was limited to Read code analysis alone. A potential concern is that additional information is contained in the 'free-text' area of the notes. Previous studies that have explored this and have found it to only occur rarely.
- Currently, information about histology, stage of disease at diagnosis or treatment received is not linked to presenting symptoms within the THIN database. It is therefore impossible to determine which symptoms are associated with the earliest stages of the disease and would therefore offer the greatest opportunity for early intervention and treatment. However, linkage of THIN data to Hospital Episode Statistics (HES) data is currently underway.
- Although conditional logistic regression was not utilised to take advantage of our matched data, the study's sample size was large enough to make losses in statistical power negligible.
- The primary aim of the figures in this study was to present a visual representation of general trends within the data.

the disease is at an advanced stage and no longer amenable to curative surgical resection.^{1 2} PDAC is the ninth most common cancer in the UK with more than 8000 cases

diagnosed each year compared with less than 2000 cases of BTC. Overall 5-year survival in both tumours is less than 4%.^{2 3} Despite advances in diagnostic technology and the identification of a number of promising biomarkers, impact on survival has been limited and novel diagnostic strategies are therefore urgently needed.⁴

Recently prediagnostic symptom profiles have been investigated as a method of enabling earlier diagnosis, in a number of common cancers including PDAC.^{5 6–8} The diagnosis of PDAC is heralded by the insidious onset of a heterogeneous collection of symptoms. Although symptom profiles are recognised to vary between patients with PDAC, certain symptoms appear to occur with sufficient frequency to be useful as early diagnostic markers of the disease. To date, prediagnostic symptom profiles for PDAC have largely been defined through postdiagnosis retrospective interview studies of secondary care patients⁸ and through the interrogation of large primary care databases with predefined symptom lists.^{6 7} Almost no studies have explored the symptom profile of BTC.

Defining the early symptom profiles of PDAC has enabled the development of symptom-based cancer decision support tools (CDSTs). Recently these tools have been introduced into primary care practices across 15 cancer networks, throughout the UK.⁶ Their impact on referral practice is subject to an ongoing audit.⁹ CDSTs for PDAC have been validated independently within other primary care data sets. Initial results suggest that although they can effectively discriminate patients with PDAC, they may overestimate cancer risk in certain groups, in particular in older patients.¹⁰ Future modification of existing tools to improve their overall diagnostic accuracy is therefore likely to be required.

Patients with PDAC frequently encounter a number of delays during their route to diagnosis.¹¹ Although PDAC is no longer considered to be a symptomatically silent disease, debate exists about how long patients are symptomatic for and if symptoms occur simultaneously or sequentially. A recent qualitative interview study suggested very early symptoms might actually be intermittent and therefore reassuring to patients leading them to ignore them until they increase in severity or other symptoms arise.¹² Once symptomatic, large primary care database studies and patient surveys indicate that patients with PDAC visit their general practitioner (GP) frequently with alarm symptoms in the months and years prior to diagnosis.^{6 7 11 13} However, almost half of patients are still diagnosed as a result of an emergency presentation to hospital.¹¹ Reasons why the disease is not identified earlier are complex. The average GP will only see one new case of PDAC every 5 years and alarm symptoms overlap with a number of other more common benign and malignant conditions, as a result it is recognised as a very challenging disease to identify at an early stage.¹¹ Simple screening tests do not exist and very few GPs have access to cross-sectional imaging; therefore, diagnosis of PDAC in primary care is primarily

dependent on symptom recognition. How alarm symptoms for PDAC overlap with other conditions has rarely been evaluated, and it is unclear if there are certain symptoms or combinations of symptoms that are unique to PDAC. In addition, very little is known about what features of the disease prompt GPs to suspect cancer and initiate investigations and referrals. Current National Institute for Health and Care Excellence (NICE) referral guidelines for suspected cancer in primary care contain limited specific information on the best method for referring patients with suspected PDAC or BTC for further investigation, but new guidelines are underway.¹⁴

The primary aim of this study was therefore to determine the early symptom profiles of PDAC and BTC in a large primary care cohort. Secondary aims include comparing early symptom trends between BTC and PDAC, defining symptom onset in PDAC and evaluating trends in routine blood tests nearest to the time of diagnosis.

METHODS

Data source

In the UK most GPs record patient data electronically. A subset of GP practices have opted to provide anonymous electronic patient records for use in clinical and epidemiological research. The Health Improvement Network (THIN) is a primary care database, which includes more than 11 million electronic patient records, from 562 GP practices, covering around 6% of the UK population (<http://csdmruk.cegedim.com/>). The data are broadly representative of the UK general practice population in terms of demographics and consultation behaviour.^{15 16} Diagnoses, symptoms and referrals to secondary care are electronically recorded using the Read code system.¹⁷ Clinical diagnoses recorded by GPs electronically have recently been shown to be accurate compared with other reliable sources.^{16 18} All drug prescriptions and variables such as body mass index (BMI), blood pressure, smoking status, alcohol intake and laboratory results are also recorded.

Study design

A case–control design was used to compare ‘alarm’ symptoms and commonly performed blood test results in patients with a diagnosis of PDAC or BTC. Unaffected controls were matched for age, sex, practice and year of diagnosis.

Study population

All patients with a Read code diagnosis of PDAC or BTC between 1 January 2000 and 31 December 2010 were extracted from the database. Read code lists to identify diagnosed patients were developed using previously described methodology.¹⁹ The date of diagnosis was set as the index date and for control patients a random consultation date was selected to become the index date. All patients were required to have contributed 2 years of data prior to the index date. Two years was selected as

the time period of interest based on preliminary data suggesting alarm symptoms were uncommon beyond this time period (figure 1A, B). To help ensure the data analysed were of adequate quality, only patients from GP practices which had achieved both acceptable mortality recording²⁰ and an acceptable computer usage²¹ were included.

The control sample contained randomly selected patients without a diagnosis of PDAC or BTC. Stratified sampling within the same GP practices from where patients with a cancer diagnosis were identified was used to ensure control patients had similar characteristics to those with a cancer diagnosis in terms of age, sex, practice and equivalent year of consultation (control group) to year of diagnosis (cancer group). Up to six control patients were selected per patient with a cancer diagnosis.

Outcomes

Alarm symptoms and laboratory tests were selected based on clinical knowledge and the existing literature.^{6 7 22–31} To ensure that no symptoms had been missed by the literature review, Read codes for 10% of patients with PDAC (n=296) were reviewed in their entirety to identify any additional common or biologically plausible symptoms (table 1). For each individual symptom, frequency, median onset and average number of presentations were recorded. Symptoms were grouped according to pathological aetiology and onset (greater or less than 6 months prior to diagnosis). All symptoms with a frequency of greater than 5% were identified as potential alarm symptoms and included in the subsequent case–control study (table 1).

Laboratory tests were restricted to routinely performed tests to ensure adequate numbers were recorded for the control population and included haemoglobin and liver function tests: serum bilirubin, alkaline phosphatase (ALP), alanine aminotransferase (ALT).

Covariates

Age, gender, time period and Townsend score, smoking status and BMI were selected as potential confounders. Where multiple measures of BMI and smoking status were recorded, the earliest record in the 2-year time frame from the index date was selected. Deprivation was examined using quintiles of Townsend score from 'one' (least deprived) to 'five' (most deprived). The Townsend score is a combined measure of owner occupation, car ownership, overcrowding and unemployment based on a patient's postcode and linkage to population census data for 2001 for approximately 150 households in that postal area.

Statistical analyses

Multivariable logistic regression was used to estimate the ORs for symptoms in the 2 years prior to PDAC or BTC diagnosis versus the 2 years prior to the index date in patients with and without cancer. Linear regression was used to estimate adjusted mean differences in clinical

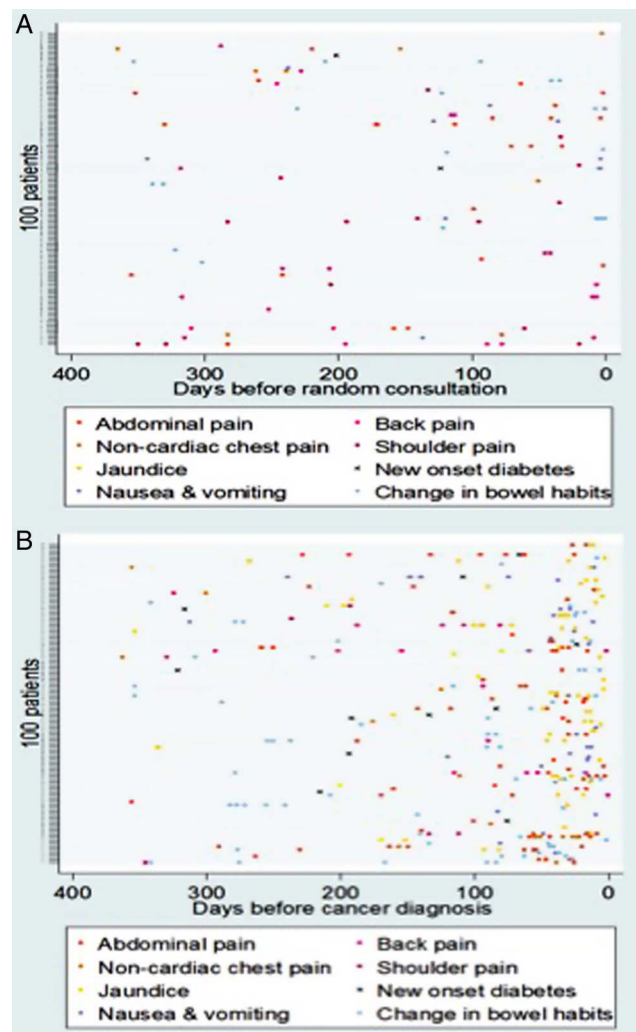


Figure 1 (A) Incidence and timing of common presenting symptoms in 100 matched controls 1 year before a random consultation. (B) Incidence and timing of common presenting symptoms in 100 randomly selected patients with pancreatic cancer in the year prior to diagnosis.

measures between patients with and without cancer. Although many of these laboratory values are slightly skewed, we decided to analyse the data without transformation—the large sample size means the statistical analyses should be robust to deviation from normality and transformation can make results difficult to interpret. To account for data clustering within GP practice, we used a multilevel regression model with the practice identifier entered as a random effect. All p values were two-tailed and a value of less than 5% (≤ 0.05) was considered statistically significant. All analyses were done using Stata V.12.1.

RESULTS

Read code analysis of a subgroup 296 patients with PDAC

The Read codes of 10% of randomly selected patients with PDAC (296 cases) were reviewed in their entirety.

Table 1 Frequency and onset of common and biologically plausible symptoms in a 10% cohort of patients with pancreatic ductal adenocarcinoma

Symptom*	N=296 (%)	Median symptom onset—days prior to diagnosis [range]	Mean number of presentations with symptom prior to diagnosis [range]
Pain			
Abdominal pain	130 (44)	106 [1–1092]	1.20 [0–12]
Back pain	90 (30)	483 [7–1092]	0.54 [0–6]
Non cardiac chest pain	39 (13)	159 [10–1093]	0.18 [0–5]
Shoulder pain	21 (7)	671 [48–1095]	0.13 [0–5]
Upper GI symptoms			
Dyspepsia/reflux	77 (26)	136 [12–1095]	0.52 [0–15]
Nausea and vomiting	58 (20)	73 [1–1073]	0.30 [0–13]
Abdominal mass	11 (4)	49 [0–255]	0.04 [0–1]
Bloating	9 (3)	87 [27–607]	0.00 [0–1]
Upper gastrointestinal bleeding	8 (3)	49 [2–689]	0.00 [0–1]
Dysphagia	7 (2)	227 [0–603]	0.03 [0–1]
Hepatomegaly	3 (1)	10 [0–68]	0.01 [0–1]
Bile duct obstruction			
Jaundice	104 (35)	31 [0–648]	0.35 [0–1]
Pruritus	23 (8)	114 [13–1059]	0.10 [0–5]
Lower GI symptoms			
Change in bowel habit	104 (35)	188 [0–1078]	0.73 [0–14]
Pancreatic dysfunction			
Pancreatitis	11 (4)	108 [21–922]	0.03 [0–1]
Steatorrhoea	4 (1)	62 [0–593]	0.02 [0–1]
Other constitutional symptoms			
Weight loss	29 (10)	144 [14–937]	0.09 [0–1]
Lethargy	23 (8)	219 [8–988]	0.10 [0–3]
Anorexia	14 (5)	46 [2–337]	0.00 [0–1]
DVT/PE	11 (4)	24 [0–1084]	0.04 [0–1]
Insomnia	6 (2)	45 [34–74]	0.00 [0–1]
Fracture	4 (1)	78 [0–242]	0.00 [0–1]
Change in taste/smell	2 (0.7)	40 [38–42]	0.00 [0–1]

*Each symptom was assessed independently.

[■]=Symptom onset >6 months prior to diagnosis. []=Symptom onset <6 months prior to diagnosis.

GI, gastrointestinal.

In this group, symptoms were common (table 1); 91% (268/296) had relevant symptoms in the 2 years prior to diagnosis. Patients attended their GP on a median of 3 occasions with alarm symptoms (range 0–22) during this period but visits did cluster nearest to the time of diagnosis (figure 1A, B). In those who were symptomatic, 51% (136/268) reattended with the same symptom during the 2-year time period and 75% (202/268) reattended with an alternative alarm symptom. Common alarm symptoms that prompted reattendance included abdominal, back, chest or shoulder pain, dyspepsia and change in bowel habit. 11% (32/296) of patients had previously been diagnosed with another cancer. The length of time a symptom had been present for was measured from first presentation to time of diagnosis. All prediagnosis serum measurements of bilirubin (345 tests), glucose (188 tests) and haemoglobin (335 tests) were also obtained for this cohort. A rising trend in glucose and bilirubin nearest to the time of diagnosis was observed (figure 2).

Very few biologically plausible symptoms were reported more than 1 year prior to diagnosis (table 1 and figure 1A, B). The limit of the study period was therefore set at 2 years, for the subsequent case-control study.

Case-control study

In total, 2773 patients with PDAC, 848 patients with BTC and 15 395 controls were included in this study (table 2). In the year prior to diagnosis, patients with PDAC visited their GP on a median of 18 (IQR 11–27) occasions and patients with BTC visited their GP on a median of 22 (IQR 12–22) occasions. This is compared with control patients who visited their GP on a median of 14 occasions (IQR 8–21). In PDAC a median of three of these visits were with alarm symptoms and a quarter visited their GP on more than four occasions with alarm symptoms in the year prior to diagnosis.

In the 2 years prior to diagnosis, alarm symptoms were more common in patients with PDAC or BTC compared with controls (table 3). For example, 43.9% of patients

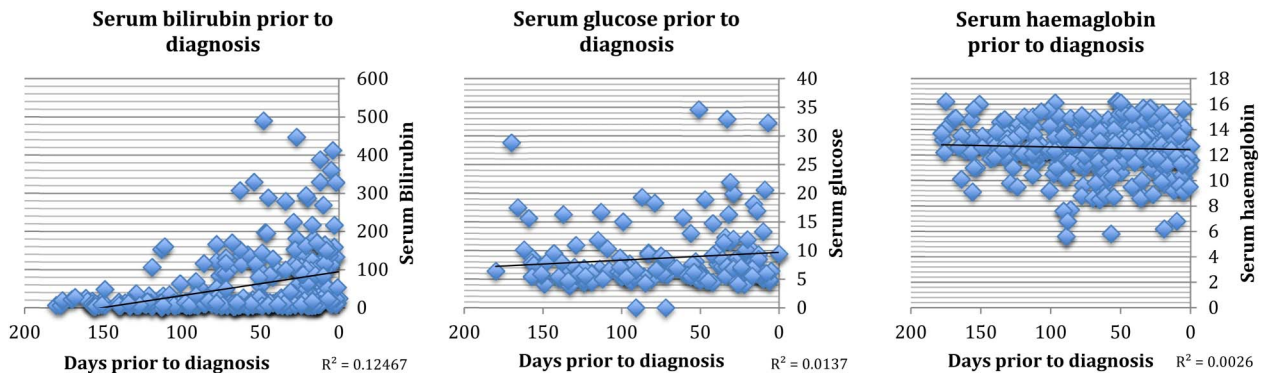


Figure 2 Trends in commonly performed blood tests in the year prior to diagnosis (bilirubin N=345 blood tests from 192 patients, glucose N=188 blood tests from 118 patients, haemoglobin N=335 blood tests from 195 patients).

with PDAC (OR=6.38 (95% CI 5.81 to 7.02)) and 37.3% of patients with BTC (OR=4.68 (95% CI 4.01 to 5.47)) consulted their GP with abdominal pain compared with just 11.3% of the control population. The incidence of dysphagia and pain other than abdominal or back was similar across all patient groups.

When comparing the symptom profiles of PDAC and BTC, some symptoms were only a feature of PDAC such as back pain, lethargy or new onset diabetes. However, other alarm symptoms did overlap between both cancers but generally were a more common feature of one cancer than the other. For example, abdominal pain (OR=1.35 (95% CI 1.15 to 1.59)) weight loss (OR=2.00 (95% CI 1.49 to 2.68)) and dyspepsia (OR=1.51 (95% CI 1.23 to 1.89)) were more frequently associated with PDAC and jaundice (OR=0.44 (95% CI 0.34 to 0.59)), and pruritus (OR=0.57 (95% CI 0.48 to 0.66)) was more frequently associated with BTC (table 3). Symptoms of unexplained weight loss were around twice as common in patients with PDAC compared with patients with BTC.

Mean liver biochemical tests including serum bilirubin, ALP and ALT closest to the date of diagnosis were substantially higher in patients with PDAC and BTC compared with controls ($p<0.001$; table 4). Mean serum bilirubin levels in BTC (26.1 $\mu\text{mol/L}$) and PDAC (20.7 $\mu\text{mol/L}$) were higher than in controls (10.2 $\mu\text{mol/L}$) but not at clinically detectable levels. The mean levels of bilirubin and ALP in patients with BTC were around double those of the control patients. With the exception of ALP, which was significantly higher in BTC compared with PDAC ($p<0.001$), there was no significant difference in routinely performed blood tests between the two cancer types (table 4).

BMI was significantly lower in patients with PDAC compared with patients with BTC or control patients. However, adjustments for BMI and smoking status had no meaningful effect on any of the relationships reported.

DISCUSSION

This study further defines the early symptom profile of PDAC and for the first time defines early alarm

symptoms that are associated with BTC in a primary care population. Prior to this study, very little was known about early alarm symptoms in BTC^{23 24} and how they overlapped with PDAC, although the clinical presentation of the two cancers was recognised to be similar. Associated alarm symptoms in BTC and PDAC in this study by and large reflected the underlying pathology and correlated with disease progression (tables 1 and 3).

With the exception of jaundice, the individual ORs of many of the associated symptoms in PDAC and BTC were low. However, nearly all patients reported alarm symptoms and often attended their GP on several occasions in the year prior to diagnosis. To aid the earlier diagnosis of these cancers, this study would therefore support the development of CDSTs, which incorporate multiple early onset, alarm symptoms.

A recent validation study of an existing CDST for PDAC suggests it may over predict cancer risk in certain groups including older patients.¹⁰ This has the potential to cause unnecessary anxiety for patients and substantially increase workloads in hospital departments due to extra referrals for the investigation of patients with suspected cancer. Further refinement of existing tools to improve their diagnostic accuracy is therefore likely to be necessary.

Patients with PDAC or BTC in this study visited their GP more frequently in the 2 years prior to diagnosis. This reflected trends found in other primary care studies⁷ and large patient surveys (National Cancer Patient Experience Survey).¹³ A change in attendance behaviour should therefore be considered as an alarm feature for cancer, particularly if patients reattend with the same alarm symptom or a constellation of alarm symptoms.

Apart from one other retrospective secondary care study, the length of time symptoms are present in patients with PDAC has received little attention.⁸ By comparison, pruritus appeared to be reported earlier in this primary care cohort but other symptoms such as change in bowel habit and anorexia were present for a similar length of time. Previous studies measured symptom onset in accordance with the development of

Table 2 Patient characteristics

	Biliary tract cancer n=848 (%)	Pancreatic cancer n=2773 (%)	Comparator group n=15 395 (%)
Males (%)	390 (46)	1328 (48)	7233 (47)
Mean age (\pm SD)	71 (12)	72 (12)	71 (11)
Mean BMI (\pm SD)	27 (5)	27 (5)	27 (5)
Smoking status (%)			
Never	446 (53)	1345 (49)	8610 (56)
Ex	181 (21)	586 (21)	3635 (24)
Current	203 (24)	760 (27)	3090 (20)
Missing	18 (2)	82 (3)	60 (1)
Townsend score (%)			
(Most affluent) 1	194 (23)	739 (27)	4176 (27)
2	197 (23)	685 (25)	3611 (24)
3	168 (20)	542 (20)	3176 (21)
4	140 (17)	457 (17)	2536 (17)
(Most deprived) 5	128 (15)	290 (11)	1542 (10)
Missing	21 (2)	60 (2)	354 (2)

Body mass index (BMI)=weight (kg)/(height(m))².

BMI<18.5=underweight, BMI 18.5–25=normal, BMI 25–30=overweight, BMI >30=obese.

jaundice or abdominal pain rather than final diagnosis as in this study, which may account for some of the discrepancies observed.⁸

Identified early symptoms of PDAC (table 1) were similar to those identified in other primary and secondary care studies.^{6–8} However, dyspepsia and pruritus have not been identified as alarm symptoms for PDAC in primary care patients previously.^{6–7} Dysphagia was identified in another primary care study as an independent predictor of PDAC in men,⁶ however it was not found to be a common early symptom in this study. Significant overlap occurred in the early symptoms of PDAC and BTC and may account for why these tumours are often difficult to differentiate preoperatively. However, even in these two malignant conditions that are recognised to present similarly, certain symptoms such as back pain, lethargy and new onset diabetes were identified as unique features of PDAC. Hence, when designing future CDSTs, symptom overlap and the inclusion of unique symptoms should be a design consideration.

The frequency of alarm symptoms in this study was similar to other primary care studies^{6–7} but lower than those reported in retrospective secondary care studies.^{6–8} This trend has been reported before⁷ and may reflect that there are some symptoms for which patients do not seek medical advice. For example, anorexia and change in taste were seen with much greater frequency in secondary care than primary care studies. If CDSTs are to be used by patients directly in the future, further modification of these tools in line with their pre-presentation symptom profile will be necessary.

Commonly performed blood tests, in particular liver function tests (bilirubin, ALT, ALP), often became abnormal prior to diagnosis. These tests are frequently performed by GPs as part of drug monitoring and

routine health checks and therefore could be included in future algorithms.

Further guidance on the best methods of managing patients with suspected PDAC or BTC in primary care is also urgently needed. Current UK guidance on referring patients with suspected PDAC or BTC from primary care is incorporated within guidance for all upper gastrointestinal cancers.¹⁴ These recommendations focus on the exclusion of oesophagogastric cancer through urgent gastroscopy, which is often normal in patients with PDAC or BTC. The lack of specific information about recognising PDAC or BTC and optimal methods for further investigation has the potential to cause further delays in diagnosis. The inclusion of more information about alarm symptoms for PDAC and BTC, the use of CDST in routine practice and thresholds for investigation would be particularly valuable to GPs. Ultimately alignment of these tools to rapid assessment pathways could prevent outpatients and diagnostic services becoming overwhelmed.

CONCLUSIONS

Referrals for investigation of suspected PDAC or BTC from primary care is currently dependent on symptom recognition. Further definition of early alarm symptoms associated with these two cancers by this study will support GPs in identifying patients with suspected PDAC or BTC. The information will also inform the future modification of current symptom-based CDSTs. Widespread use of these tools in primary care is expected to lead to patients being diagnosed at an earlier stage when curative therapy is possible. Subsequent improvements in overall survival are expected.

Table 3 Results from multivariable logistic regression models outlining the frequency and adjusted ORs of symptoms and signs of pancreatic cancer and BTC presenting to primary care in the 2 years prior to diagnosis compared with a control population without a cancer diagnosis

Symptom (%)	Biliary tract cancer n=829	Pancreatic cancer n=2790	Controls n=17 192	Pancreatic cancer vs control			Biliary tract cancer vs control			Pancreatic cancer vs biliary tract cancer		
				OR*	95% CI	p Value	OR*	95% CI	p Value	OR*	95% CI	p Value
Weight loss	46 (5.5)	294 (10.5)	302 (1.8)	6.6	5.54 to 7.86	<0.001	3.17	2.32 to 4.34	<0.001	2.00	1.49 to 2.68	<0.001
Abdominal pain	309 (37.3)	1225 (43.9)	1946 (11.3)	6.38	5.81 to 7.02	<0.001	4.68	4.01 to 5.47	<0.001	1.35	1.15 to 1.59	<0.001
Nausea and vomiting	126 (15.2)	463 (16.6)	978 (5.7)	3.43	3.00 to 3.91	<0.001	2.99	2.44 to 3.66	<0.001	1.12	0.89 to 1.39	0.33
Bloating	27 (3.3)	113 (4.1)	229 (1.3)	3.1	2.48 to 3.89	<0.001	2.35	1.57 to 3.53	<0.001	1.25	0.82 to 1.89	0.298
Dyspepsia	118 (14.2)	559 (20)	1597 (9.3)	2.56	2.30 to 2.85	<0.001	1.7	1.40 to 2.08	<0.001	1.51	1.23 to 1.86	<0.001
New onset diabetes	48 (5.8)	380 (13.6)	1037 (6)	2.46	2.16 to 2.80	<0.001	0.92	0.68 to 1.23	0.559	2.59	1.91 to 3.52	<0.001
Change in bowel habit	194 (23.4)	764 (27.4)	2557 (14.9)	2.17	1.98 to 2.39	<0.001	1.77	1.51 to 2.09	<0.001	1.23	1.04 to 1.46	0.014
Pruritus	91 (11)	147 (5.3)	526 (3.1)	1.73	1.43 to 2.10	<0.001	3.75	2.96 to 4.74	<0.001	0.44	0.34 to 0.59	<0.001
Lethargy	71 (8.6)	293 (10.5)	1308 (7.6)	1.42	1.25 to 1.61	<0.001	1.08	0.83 to 1.40	0.57	1.28	0.97 to 1.68	0.086
Back pain	111 (13.4)	446 (16)	2111 (12.3)	1.33	1.18 to 1.49	<0.001	1.01	0.82 to 1.26	0.913	1.28	1.01 to 1.61	0.037
Dysphagia	10 (1.2)	51 (1.8)	254 (1.5)	1.21	0.90 to 1.64	0.206	0.81	0.43 to 1.51	0.498	1.51	0.76 to 3.03	0.241
Non-cardiac chest pain	114 (13.8)	335 (12)	2055 (12)	1.02	0.91 to 1.16	0.699	1.19	0.97 to 1.46	0.092	0.86	0.70 to 1.07	0.175
Shoulder pain	47 (5.7)	137 (4.9)	1052 (6.1)	0.78	0.65 to 0.93	0.006	0.86	0.64 to 1.14	0.294	0.87	0.63 to 1.18	0.367
Jaundice *	358 (43.2)	860 (30.8)	36 (0.2)	246.00	172 to 351	<0.001	445	302 to 658	<0.001	0.57	0.48 to 0.66	<0.001

Models were adjusted for age, gender, time period and social deprivation (each symptom was assessed independently).

*Only 0.2% of controls were recorded with jaundice. However, among patients with PDAC and BCT, jaundice was highly prevalent occurring in 43.2% of patients with BTC and 30.8% of patients with PDAC which lead to extremely large OR estimates.

BTC, biliary tract cancer; PDAC, pancreatic ductal adenocarcinoma.

Table 4 Liver function tests, haemoglobin levels and BMI in patients with pancreatic cancer and biliary tract cancer compared with a control population without a cancer diagnosis

	Biliary tract cancer	Pancreatic cancer	Controls	Biliary tract cancer vs control			Pancreatic cancer vs control			Pancreatic cancer vs biliary tract cancer		
				Coefficient*	95% CI	p Value	Coefficient*	95% CI	p Value	Coefficient*	95% CI	p Value
Bilirubin (µmol/L)												
Number	628	2262	9227									
Mean (±SD)	26.1 (26.5)	20.7 (23.1)	10.9 (6.2)	15.3	12.99 to 17.60	<0.001	15.3	12.99 to 17.60	<0.001	-7.4	-13.7 to -1.1	0.021
ALP (U/L)												
Number	615	2220	9412									
Mean (±SD)	183.7 (107)	152.2 (95.9)	93.8 (48)	89.8	81.0 to 98.5	<0.001	89.8	81.0 to 98.5	<0.001	-63.6	-95.4 to -31.8	<0.001
ALT (U/L)												
Number	490	1719	7730									
Mean (±SD)	37.3 (24.5)	31.4 (20.7)	24.4 (12.8)	13.39	11.14 to 15.64	<0.001	13.39	11.14 to 15.64	<0.001	-3.1	-11.7 to 5.6	0.486
Haemoglobin (g/dL)												
Number	727	2602	9331									
Mean (±SD)	12.9 (1.8)	13 (1.7)	13.5 (1.6)	-0.62	-0.75 to -0.48	<0.001	-0.62	-0.75 to -0.48	<0.001	0.4	-0.4 to 1.3	0.33
BMI												
Number	608	2297	9316									
Mean (±SD)	26.6 (5.2)	25.6 (4.9)	27.7 (5.5)	-0.97	-1.42 to -0.53	<0.001	-0.97	-1.42 to -0.53	<0.001	-4.8	-7.9 to -1.7	0.003

The measurements used in the analysis are those taken closest to the date of diagnosis or a random consultation date for the control population.

*Adjusted for age, gender, time period and social deprivation.

ALP, alkaline phosphatase; ALT, alanine aminotransferase; BMI, body mass index.

Correction notice The license of this article has changed since publication to CC BY 4.0.

Contributors MGK performed the case note review and wrote the article. LH retrieved the data from the THIN database and performed the statistical analysis. GR and SPP had the idea for the study and reviewed and edited the manuscript.

Funding SPP was supported by National Institutes of Health (NIH) program grant PO1CA84203 and a Pancreatic Cancer UK project grant. The work was undertaken at UCLH/UCL, which receives a proportion of funding from the Department of Health's National Institute for Health Research (NIHR) Biomedical Research Centres funding scheme. MGK was supported by a Cancer Research UK research training bursary.

Competing interests None.

Ethics approval The THIN scheme was approved by the National Health Service South-East Multi-centre Research Ethics Committee and the present study was approved by the Scientific Review Committee at University College London.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement A list of Read Codes for each symptom is available from the corresponding author on request. All other data are included within the manuscript.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

REFERENCES

- Gudjonsson B. Cancer of the pancreas 50 years of surgery. *Cancer* 1987;60:2284–303.
- Khan SA, Davidson BR, Goldin RD, *et al.* Guidelines for the diagnosis and treatment of cholangiocarcinoma: an update. *Gut* 2012;61:1657–69.
- CRUK. Cancer Research UK Cancer Stats Incidence 2008. Secondary Cancer Research UK Cancer Stats Incidence 2008. April 2011. info.cancerresearchuk.org/cancerstats
- Bruno MJ. Early diagnosis of pancreatic cancer; looking for a needle in a haystack? *Gut* 2013;62:955–6.
- Holly EA, Chaliha I, Bracci PM, *et al.* Signs and symptoms of pancreatic cancer: a population-based case-control study in the San Francisco Bay area. *Clin Gastroenterol Hepatol* 2004;2:510–17.
- Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2012;62:e38–45.
- Stapley S, Peters TJ, Neal RD, *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;106:1940–4.
- Gullo L, Tomassetti P, Migliori M, *et al.* Do early symptoms of pancreatic cancer exist that can allow an earlier diagnosis? *Pancreas* 2001;22:210–13.
- Macmillan. Early diagnosis programme. Secondary Early diagnosis programme. 2014. <http://www.macmillan.org.uk/Aboutus/Healthandsocialcareprofessionals/Macmilliansprogrammesandservices/Earlydiagnosisprogramme.aspx>
- Collins GS, Altman DG. Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer(Pancreas). *Br J Gen Pract* 2013;63:e636–42.
- PCUK. Study for survival. Secondary Study for survival. 2011. http://www.pancreaticcancer.org.uk/media/100292/report_final_for_web.pdf
- Evans J, Chapple A, Salisbury H, *et al.* "It can't be very important because it comes and goes"—patients' accounts of intermittent symptoms preceding a pancreatic cancer diagnosis: a qualitative study. *BMJ Open* 2014;4:e004215.
- Lyratzopoulos G, Neal RD, Barbiere JM, *et al.* Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* 2012;13:353–65.
- NICE. Clinical Guideline 27: Referral guidelines for suspected cancer. Secondary Clinical Guideline 27: Referral guidelines for suspected cancer. 2005. <http://www.nice.org.uk/nicemedia/live/10968/29814/29814.pdf>
- Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Inform Prim Care* 2004;12:171–7.
- Blak BT, Thompson M, Dattani H, *et al.* Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care* 2011;19:251–5.
- Booth N. What are the Read Codes? *Health Libr Rev* 1994;11:177–82.
- Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010;60:e128–36.
- Dave S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. *Pharmacoepidemiol Drug Saf* 2009;18:704–7.
- Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. *Pharmacoepidemiol Drug Saf* 2009;18:76–83.
- Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf* 2013;22:64–9.
- Hassan MM, Bondy ML, Wolff RA, *et al.* Risk factors for pancreatic cancer: case control study. *Am J Gastroenterol* 2007;102:2696–707.
- Pitt HA, Dooley WC, Yeo CJ, *et al.* Malignancies of the biliary tree. *Curr Probl Surg* 1995;32:1–90.
- de Groen PC, Gores GJ, LaRusso NF, *et al.* Biliary tract cancers. *N Engl J Med* 1999;341:1368–78.
- Chapman RW. Risk factors for biliary tract carcinogenesis. *Ann Oncol* 1999;10:308–11.
- Silverman DT, Dunn JA, Hoover RN, *et al.* Cigarette smoking and pancreas cancer: a case-control study based on direct interviews. *J Natl Cancer Inst* 1994;86:1510–16.
- Fuchs CS, Colditz GA, Stampfer MJ, *et al.* A prospective study of cigarette smoking and the risk of pancreatic cancer. *Arch Intern Med* 1996;156:2255–60.
- Muscat JE, Stellman SD, Hoffmann D, *et al.* Smoking and pancreatic cancer in men and women. *Cancer Epidemiol Biomarkers Prev* 1997;6:15–19.
- Bonelli L, Aste H, Bovo P, *et al.* Exocrine pancreatic cancer, cigarette smoking, and diabetes mellitus: a case-control study in northern Italy. *Pancreas* 2003;27:143–9.
- Larsson SC, Permert J, Hakansson N, *et al.* Overall obesity, abdominal adiposity, diabetes and cigarette smoking in relation to the risk of pancreatic cancer in two Swedish population-based cohorts. *Br J Cancer* 2005;93:1310–15.
- Tyson GL, El-Serag HB. Risk factors for cholangiocarcinoma. *Hepatology* 2011;54:173–84