PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A CASE-CONTROL STUDY COMPARING THE INCIDENCE OF EARLY SYMPTOMS IN PANCREATIC AND BILIARY TRACT CANCER
AUTHORS	Keane, Margaret; Horsfall, Laura; Rait, Greta; Pereira, Stephen

VERSION 1 - REVIEW

REVIEWER	Kazuo Inui, MD Professor Department of Gastroenterology, Second Teaching Hospital, Fujita Health University, Nagoya, Japan
REVIEW RETURNED	10-Jun-2014

GENERAL COMMENTS	The data had no major selection bias because it was extracted from primary care database. The results were reasonable compared with previous papers. Because the recognition of early alarm symptoms in PDAC and BTC leads to early detection of these cancers, the article is valuable for the readers of your journal. The manuscript was written clearly and comprehensive. The only problem is that the Odds Ratios of back pain and lethargy were low. However, the low ORs might be caused by large cohort and multiple variables. Please explain the reasons of the low ORs.

REVIEWER	Mitsugi Shimoda Second Department of Surgery Dokkyo Medical University
REVIEW RETURNED	17-Jun-2014

- The reviewer completed the checklist but made no further comments.

REVIEWER	Mike Hernandez M.D. Anderson Cancer Center Houston, TX
REVIEW RETURNED	02-Sep-2014

GENERAL COMMENTS	Thank you for giving me the opportunity to review your manuscript. Using early symptom alarms to identify individuals with a greater risk for developing cancer is important for both clinicians and patients. Early symptoms may be indicative of the disease process in its

	infancy when patients have the highest potentially for being treated successfully. Although the paper is well written, there are several major issues that need to be addressed. 1. The primary aim of the study was to determine early symptom profiles of PDAC and BTC. The authors make an attempt at this aim by using a case-control study, with up to six controls for each case. In the study design (line 44; page 7) section, four matching items are listed: age, sex, practice, and year of diagnosis. It should be made clearer that the stratified sampling used to acquire controls described in the study population section (line 42; page8) also ensured matching on the same four characteristics. 2. The tables would be better organized if Table 1 showed patient demographics and Table 2 showed the description of symptoms. 3. Table 2 shows the break-down of patient characteristics by PDAC and BTC. Although the controls were matched based on four characteristics, it would be helpful to include a break-down of control patients alongside the PDAC and BTC patients. 4. BMI is specified in Table 2, but no summary statistics are provided. 5. Table 2 as well as the covariates section (line 32; page 9) would benefit from having a brief description of the Townsend score. 6. The frequency and onset of symptoms in Table 1 seems to be for PDAC patients only. Why aren't BTC patients included as well? Since the manuscript refers to both outcomes, then summaries for both outcomes should be presented. 7. Rather than providing 95% CIs for the median days prior to symptom onset, the range may be more beneficial because the range along with the median allows for the interpolation of the interquartile range. 8. Table 1 shows the "Proportion of patients experiencing symptom" as a heading; however, both frequencies and percents are provided. Remove "Proportion of" and use "N = 296 (%)". Also, please indicate as a footnote that each symptom was assessed independently. Additionally, the table should include information on the average number (or median and range) of times the symptom was recorded to give the reader some idea of each symptom's intensity. Finally, there is no written description in the methods section to prompt the reader about the broader categorizations for the symptoms that were created. With the table in mind, a better way of categorizing the symptoms would be early versus late. As the data show, some patients had their symptoms recorded more than a year from diagnosis, while other patients have their earliest recorded symptoms within 3 months prior to diagnosis. 9. In the statistical analysis section, the authors use a multivariable logistic regression model. Given the data was matched, why wasn't a conditional logistic regression used? Please explain why the matching was disregarded, and why the analysis was run as a multivariable logistic regression model adjusted with the use of the matching covariates. 10. Table 3 showing the results of the multivariable logistic regression model needs to be organized. For example, the ORs showing association between PDAC and Controls should be rank ordered so the read can easily see that the OR for weight loss was highest (6.60), OR for abdominal pain (6.38), and etc. Please place jaundice at the very front or very end of the list with an asterisk indicating that only 0.2% of controls were recorded with jaundice. However, among PDAC and BCT patients, jaundice was highly prevalent occurring in 43.2% of BTC patients and 30.8% of PDAC patients which lead to extremely large OR estimates. So, a simple

	explanation like the one provided above would suffice in explaining the high ORs for jaundice, but the authors should consider using exact logistic regression or even a penalized likelihood approach. Prior to doing this, the authors should first establish why the data was not analyzed using the matching structure described in the methods section. 11. Table 3 needs a footnote indicating that each symptom was assessed independently. Also, Table 3 shows adjusted ORs, but it is unclear from the manuscript the association age, gender, time period, and social deprivation derived from the Townsend score have on the likelihood of being diagnosed with PDAC and BCT. Please provide the analysis was that used to derive estimates in the table's heading. 12. The authors report that biochemical tests near date of diagnosis were "substantially" higher for PDAC and BTC patients. However, the actual values, differences, and corresponding p-values from an appropriate statistical test should be provided. 13. Table 4 shows comparisons of biochemical tests between PDAC and BTC patients with patient controls. The Table heading should make it clear if multilevel modeling was used to obtain parameter estimates. Please explain why the data's matched structure was not incorporated into the multilevel modeling framework. 14. Figures 1a and 1b are visually difficult to assess. They show that symptoms intensify nearing diagnosis; however, the authors should consider using a simpler graph that would collapse symptoms data on the basis of 50 or 100 day increments prior to diagnosis into single points. This would allow for data from PDAC, BTC, and controls to be included in one graph. 15. Figure 2 is a scatterplot of biochemical measurements across days prior to diagnosis. The authors don't indicate the number of patients used to create each graph. These figures do not show relative comparisons between PDAC, BTC, and controls. Furthermore, these figures would be more informative if patient trajectories were plotted to assess the within and between patient variability of serum bilirubin, glucose, and haemoglobin.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name Kazuo Inui, MD
Institution and Country Professor
Department of Gastroenterology, Second Teaching Hospital, Fujita Health University,
Nagoya, Japan

The data had no major selection bias because it was extracted from primary care database. The results were reasonable compared with previous papers.

Because the recognition of early alarm symptoms in PDAC and BTC leads to early detection of these cancers, the article is valuable for the readers of your journal.

The manuscript was written clearly and comprehensive. The only problem is that the Odds Ratios of back pain and lethargy were low. However, the low ORs might be caused by large cohort and multiple variables. Please explain the reasons of the low ORs.

We thank the reviewer for their comments. The frequencies of alarm symptoms in this study were similar to other primary care studies but lower than those reported in retrospective secondary care studies. This trend has been observed by other authors as well and probably indicates that there are some symptoms for which patients don't seek medical advice. Lethargy and back pain are common symptoms in primary care and are only rarely associated with PDAC thus accounting for the low OR, however in combination with other symptoms they may still be a useful symptom for the early detection of PDAC as both appeared to be present for many months before the onset of the disease.

Reviewer: 2
Reviewer Name: Mitsugi Shimoda
Institution and Country: Second Department of Surgery, Dokkyo Medical University

There are no comments.

We thank the reviewer for reviewing the manuscript.

Reviewer: 3
Reviewer Name: Mike Hernandez
Institution and Country: M.D. Anderson Cancer Center, Houston, TX

Thank you for giving me the opportunity to review your manuscript. Using early symptom alarms to identify individuals with a greater risk for developing cancer is important for both clinicians and patients. Early symptoms may be indicative of the disease process in its infancy when patients have the highest potentially for being treated successfully.

Although the paper is well written, there are several major issues that need to be addressed.
1. The primary aim of the study was to determine early symptom profiles of PDAC and BTC. The authors make an attempt at this aim by using a case-control study, with up to six controls for each case. In the study design (line 44; page 7) section, four matching items are listed: age, sex, practice, and year of diagnosis. It should be made clearer that the stratified sampling used to acquire controls described in the study population section (line 42; page8) also ensured matching on the same four characteristics.

We thank the reviewer for their comments; this has been clarified in the updated manuscript as below:

"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 that 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." [Page 8]

2. The tables would be better organized if Table 1 showed patient demographics and Table 2 showed the description of symptoms.

This has been revised in the updated manuscript

3. Table 2 shows the break-down of patient characteristics by PDAC and BTC. Although the controls were matched based on four characteristics, it would be helpful to include a break-down of control patients alongside the PDAC and BTC patients.

This has now been added to (new) Table 1 (see below)

Table 1. Patient Characteristics

	Biliary Tract Cancer	Pancreatic cancer	Comparator group
	n=848 (%)	n=2,773 (%)	n=15,395 (%)
Males (%)	390 (46)	1328 (48)	7233 (47)
Mean age (±SD)	71 (12)	72 (12)	71 (11)
Mean BMI (±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.
SD=standard deviation

4. BMI is specified in Table 2, but no summary statistics are provided.

The mean BMI (+/- SD) is provided in line 3 of Table 2 (now Table 1). A definition of BMI has also been included. [Page 20 and above].

5. Table 2 as well as the covariates section (line 32; page 9) would benefit from having a brief description of the Townsend score.

This has been included in the main text of the revised manuscript [Page 9 and below]:

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 two-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.”

6. The frequency and onset of symptoms in Table 1 seems to be for PDAC patients only. Why aren’t

BTC patients included as well? Since the manuscript refers to both outcomes, then summaries for both outcomes should be presented.

This was a very time consuming task due to the richness and complexity of THIN data and since cancer decision support tools have only thus far been developed for pancreatic cancer we elected to only perform this analysis on pancreatic cancer patients, where the data would be most applicable to the refinement of and development of existing tools.

7. Rather than providing 95% CIs for the median days prior to symptom onset, the range may be more beneficial because the range along with the median allows for the interpolation of the interquartile range.

This has been included in the revised manuscript [Page 20 and below].

8. Table 1 shows the “Proportion of patients experiencing symptom” as a heading; however, both frequencies and percents are provided. Remove “Proportion of” and use “N = 296 (%)”. Also, please indicate as a footnote that each symptom was assessed independently. Additionally, the table should include information on the average number (or median and range) of times the symptom was recorded to give the reader some idea of each symptom’s intensity. Finally, there is no written description in the methods section to prompt the reader about the broader categorizations for the symptoms that were created. With the table in mind, a better way of categorizing the symptoms would be early versus late. As the data show, some patients had their symptoms recorded more than a year from diagnosis, while other patients have their earliest recorded symptoms within 3 months prior to diagnosis.

The table has been revised in line with the reviewers very helpful comments.

Table 2: Frequency and onset of common and biologically plausible symptoms in a 10% cohort of patients with PDAC

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 GASTROINTESTINAL SYMPTOMS			
Dyspepsia / reflux	77 (26%)	136 [12-1095]	0.52 [0-15]
Nausea and vomiting	58 (20%)	73 [1-1073]	0.3 [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 GASTROINTESTINAL 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]
Steattorhoea 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

Outcomes
“Alarm symptoms and laboratory tests were selected based on clinical knowledge and the existing literature. [10 11 23-32] 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 2]. 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 2].”

9. In the statistical analysis section, the authors use a multivariable logistic regression model. Given the data was matched, why wasn’t a conditional logistic regression used? Please explain why the matching was disregarded, and why the analysis was run as a multivariable logistic regression model adjusted with the use of the matching covariates.

The comparator group was frequency matched and not pair matched and so conditional regression is not generally necessary. The frequency matched variables were included in all regression models as is appropriate to limit bias toward the null.

10. Table 3 showing the results of the multivariable logistic regression model needs to be organized. For example, the ORs showing association between PDAC and Controls should be rank ordered so the read can easily see that the OR for weight loss was highest (6.60), OR for abdominal pain (6.38), and etc. Please place jaundice at the very front or very end of the list with an asterisk indicating that only 0.2% of controls were recorded with jaundice. However, among PDAC and BTC patients, jaundice was highly prevalent occurring in 43.2% of BTC patients and 30.8% of PDAC patients which lead to extremely large OR estimates. So, a simple explanation like the one provided above would suffice in explaining the high ORs for jaundice, but the authors should consider using exact logistic regression or even a penalized likelihood approach. Prior to doing this, the authors should first establish why the data was not analyzed using the matching structure described in the methods section.

See answer to point 9 above. We decided to use the same analytical methods across symptoms to keep the manuscript and interpretation of results simple. However, we appreciate these comments and did attempt an exact regression on jaundice using Stata to check the odds ratios were similar but the memory required to store the conditional distribution exceeded that available despite using a fairly powerful desktop. Table 3 has been reordered and the caption clarified to make our methodology clear [Page 21].

11. Table 3 needs a footnote indicating that each symptom was assessed independently. Also, Table 3 shows adjusted ORs, but it is unclear from the manuscript the association age, gender, time period, and social deprivation derived from the Townsend score have on the likelihood of being diagnosed with PDAC and BCT. Please provide the analysis that was used to derive estimates in the table's heading.

Age, gender, time period and social deprivation derived from the Townsend score have the potential to affect the likelihood of being diagnosed with PDAC and BCT. A fuller evaluation of these potential associations is outside the scope of this manuscript. A definition of the Townsend score has been included in the manuscript and Table 3 has been amended to clarify our methodology [Page 21].

12. The authors report that biochemical tests near date of diagnosis were "substantially" higher for PDAC and BTC patients. However, the actual values, differences, and corresponding p-values from an appropriate statistical test should be provided.

Apologies, we neglected to refer to Table 4 at the end of this sentence. This has now been added and these paragraphs amended accordingly [page 13 and below]. Table 4 contains the actual values for liver function tests and haemoglobin levels in Pancreatic cancer and Biliary Tract cancer patients compared with a control population without a diagnosis. The measurements used in the analysis are those taken closest to the date of diagnosis or a random consultation date for the control population. All p-values were <0.001 and this has now been added to Table 4.

"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 to controls (P<0.001) [Table 4]. Mean serum bilirubin levels in BTC (26.1 µmol/L) and PDAC (20.7 µmol/L) were higher than in controls (10.2 µmol/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 to PDAC (P<0.001), there was no significant difference in routinely performed blood tests between the two cancer types [Table 4].

Body Mass Index (BMI) was significantly lower in PDAC patients compared to patients with BTC or control patients. However, adjustments for BMI and smoking status had no meaningful effect on any of the relationships reported." [Page 13]

13. Table 4 shows comparisons of biochemical tests between PDAC and BTC patients with patient controls. The Table heading should make it clear if multilevel modeling was used to obtain parameter estimates. Please explain why the data's matched structure was not incorporated into the multilevel modeling framework.

See answer to point 9 above and we have added multilevel modelling to the title of Table 4.

14. Figures 1a and 1b are visually difficult to assess. They show that symptoms intensify nearing diagnosis; however, the authors should consider using a simpler graph that would collapse symptoms data on the basis of 50 or 100 day increments prior to diagnosis into single points. This would allow for data from PDAC, BTC, and controls to be included in one graph.

Thank you for the suggestion. The data within figures 1a and b were generated as part of a preliminary study which was intended to determine the length of the analysis prior to diagnosis that should be used in the subsequent cohort study. The primary purpose of the graphs was simply to generate a visual of symptom patterns and to decide how far back we should extract data prior to the diagnosis date rather than comparisons between PDAC and BTC.

15. Figure 2 is a scatterplot of biochemical measurements across days prior to diagnosis. The authors don't indicate the number of patients used to create each graph. These figures do not show relative comparisons between PDAC, BTC, and controls. Furthermore, these figures would be more informative if patient trajectories were plotted to assess the within and between patient variability of serum bilirubin, glucose, and haemoglobin.

The revised graphs now include the number of patients and blood tests represented in each graph. We elected to present data in this manor as we expected the graphs would become noisy and difficult to interpret if all patient trajectories are plotted. As with the 10% cohort study [Table 2] we elected to concentrate on pancreatic cancer, as this is the tumour that cancer decision tools have been developed for thus far and further information is required to inform and develop current tools.

VERSION 2 – REVIEW

REVIEWER	Mike Hernandez MD Anderson Cancer Center Houston, TX USA
REVIEW RETURNED	25-Oct-2014

GENERAL COMMENTS	Thank you for addressing or commenting on my concerns. The manuscript looks more polished. Please see my additional comments below. Comment1: (Page 10 of 41 Line 15): The estimated linear regression coefficients represent the adjusted mean difference in clinical measurements between cancer and non-cancer patients. Please change the sentence to read more like: Linear regression was used to estimate adjusted mean differences in clinical measures between patients with and without cancer. Comment2: (Page 21 and 22 of 41): The titles for Tables 3 and 4 are very busy. Comment to the authors for future manuscripts, the table title should be short, but the extra details can be included at the bottom of the table. With that said, please consider the slight changes: Table 3's title should be, "Results from multivariable logistic regression models outlining the frequency and adjusted odds ratios of symptoms and signs of pancreatic cancer and biliary tract cancer presenting to primary care in the two years prior to diagnosis compared with a control population without a cancer diagnosis. Models were adjusted for age, gender, time period and social deprivation (each symptom was assessed independently). Comment3:

	Please add a couple of sentences mentioning the potential limitations of the analysis presented. For example, “Although conditional logistic regression was not utilized to take advantage of our matched data, the study’s sample size was large enough to make losses in statistical power negligible.” And, “the figures present rough visualizations, but are sufficient to show general trends.”

VERSION 2 – AUTHOR RESPONSE

We would like to thank Dr Hernandez and the other reviewers for taking the time to review our manuscript. We agree with all of the recent comments completely and have incorporated them in to our revised manuscript.