

Comments to Authors:

Reviewer 1

Comment: The manuscript is nicely written and deals with an important healthcare problem in Thailand.

Responses: We thank the reviewer for the positive comment.

Reviewer 2

Comment 1: The data regarding cholangiocarcinoma (CCA), liver fluke or *O. viverrini* (OV) have not been presented under the results of this study. Therefore, the descriptions on liver fluke under Conclusion should be deleted from Abstract.

Responses: The descriptions on liver fluke under the Conclusion section of the abstract has been deleted. The conclusion of the Abstract has been edited as follows:

“**CONCLUSION** Diabetes and chronic liver diseases may be associated with cholangiocarcinoma in the Thai population.”

Comment 2: Also, in this study, the definite relationship between the cause of CCA and OV infections have not directly explored. Hence, it is advisable to briefly discuss these ideas with comparisons to other studies. It could be misleading to provide lengthy descriptions about the relationship between CCA and OV infections when the prevalence of OV has not been researched in this study.

Responses: We thank the reviewer for this suggestion. To avoid misleading readers, we have removed the following sentences about relationship between CCA and OV infections in the Introduction section on page 7:

“People in Northeast Thailand become infected with liver fluke by eating uncooked freshwater fish as a staple food. Chronic OV infection induces chronic inflammation in the biliary tract, which leads to the development of CCA through several mechanisms[2]. Liver fluke infection has been shown to increase the risk of developing CCA by 10 folds ^[10].”

The following sentences in the Discussion section were also deleted:

“Similarly, the OV infection rate in freshwater fish, an intermediate host of OV, has dramatically declined from 70% to less than 1%.”

Comment 3: What are the related diseases which cause Cirrhosis? It is advisable to include data which explains whether the causes include hepatitis, alcohol or other factors.

Responses: We have abstracted data on underlying liver diseases related to cirrhosis including chronic viral hepatitis B and C infection and alcoholic liver disease. The numbers of CCA patients with the underlying liver diseases causing cirrhosis among cirrhotics and non-cirrhotics are shown in the below Table. In summary, chronic viral hepatitis B and C infection and alcoholic liver disease are related to cirrhosis.

	Cirrhotics (n=1,896)	Non-cirrhotics (n=37,525)	P
Hepatitis B infection	124 (6.5%)	132 (0.3%)	<0.0001
Hepatitis C infection	100 (5.3%)	84 (0.2%)	<0.0001
Alcoholic liver disease	197 (10.4%)	0 (0%)	<0.0001

We have added a sentence to describe underlying diseases of cirrhosis in the Result section on page 11 as follows:

“CCA patients with cirrhosis (n=1896) were significantly more likely to have hepatitis B infection, hepatitis C infection and alcoholic liver disease than CCA patients without cirrhosis (n=37,525), i.e., 124 (6.5%) vs. 132 (0.3%), 100 (5.3%) vs. 84 (0.2%), and 197 (10.4%) vs. 0 (0%) for hepatitis B infection, hepatitis C infection and alcoholic liver disease, respectively; $p < 0.0001$ for all.”

Comment 4: Was there a difference in morbidity between the liver cirrhosis (LC) with hepatitis B or C group and the liver cirrhosis (LC) without hepatitis group? If so, it is necessary to analyze whether hepatitis affects the difference in CCA morbidity as this seems to be more significant than analyzing LC.

Responses: We appreciate the reviewer’s comment for this thoughtful question. To determine whether etiology of cirrhosis or cirrhosis *per se* had a different impact on outcome of CCA patients, we have performed Cox proportional hazard analysis. Please note that, for this analysis, mortality was selected as an

indicator of morbidity because death was the most robust outcome. By multivariable analysis, we found that cirrhosis was independently associated with survival with adjusted hazard ratio (AHR) (95% confidence interval [95 CI] of 1.30 (1.22-1.38), $p < 0.001$; while hepatitis B and C and alcohol were not associated with survival, with AHRs (95% CI) of 0.99 (0.86-1.15), 0.91 (0.77-1.08) and 0.93 (0.78-1.12), $p = 0.93$, 0.27 and 0.43, respectively. In summary, cirrhosis but not etiology of cirrhosis was significantly associated with mortality.

Comment 5: The descriptions under Discussion seem be quite irrelevant and textual. Since there is no analysis regarding the treatment and survival rates of CCA, it is advisable that the following description be deleted on page 25: “CCA is an insidious disease, with median survival of only 8 months and 5-year survival of only 10%.in mortality among regions requires further investigation”.

Responses: We thank the reviewer for this comment. We have deleted these descriptions in the Discussion section as suggested by the reviewer.

Comment 6. The descriptions on page 17 are redundant as they have been previously presented in the manuscript. Therefore, it is advisable to summarize the following descriptions within two sentences: “Of note, we recognize that the proportions of patients with hepatitis B and C.....and determine the relationship between OV infection and CCA incidence by regions”.

Responses: We thank again the reviewer for this comment and suggestion. We have deleted the above descriptions and replace them with the following two sentences on page 16-17:

“Data on hepatitis B/C infection, diabetes, smoking, alcohol drinking, obesity, and metabolic syndrome were not available in the discharge summary notes of all participants, thus precluding us from estimating the effect of these conditions on CCA risk. Similarly, the prevalence of OV infection and the relationship between OV infection and CCA incidence cannot be determined because the diagnosis of OV infection was not available in the medical records.”

Reviewer 3

This is an informative article about the risk factors of cholangiocarcinoma in Thailand using nationwide database. The fact that, even in Thailand, diabetes and hepatitis are the risk factors of cholangiocarcinoma in the region where fluke infection is not endemic, is interesting.

Comment 1. In this study, C24.1 of the ICD-10 is included in cholangiocarcinoma (in Study population section). Because C24.1 demonstrates cancer of ampulla of Vater, it is necessary to explain the reason that the author included this in cholangiocarcinoma.

Responses: We thank the reviewer for this comment. Ampulla of Vater is considered to be a part of distal bile duct. Cancer of ampulla of Vater was therefore included in this study.

We have added a sentence to clarify this point in the Method section on page 8 as follows:

“Cancer of the ampulla of Vater was included in this study because it occurs within 2 cm of the distal end of the common bile duct.”

Comment 2. As the author mentioned it in the last of the manuscript, unfortunately they were not able to determine the relationship between fluke infection and cholangiocarcinoma. It is a little deviation from the main point of this paper, but many readers` concern may be association between fluke infection and cholangiocarcinoma genesis. It would be better to add brief topics about that in discussion.

Responses: We have added sentences to briefly describe the association between fluke infection and cholangiocarcinoma in the Discussion section on page 13 as follows:

“OV infection was shown to be associated with a 10-fold increased risk of CCA”

Reviewer 4

This is a valuable study using big data regarding to cholangiocarcinoma. The authors need to clarify or explain some points.

Comment 1. Please show the division of region (northeast, north, central, etc.) on the Thailand map.

Responses: We thank the reviewer for this comment. We have added Figure 1 illustrating the division of region and the number of CCA patients in each region on Thailand map.

Comment 2. Non-CCA patients include all patients without CCA in the data registry? Please more describe about non-CCA patients in Study population section and Table 1.

Responses: We apologize for the unclear description of the non-CCA patients. We have added sentences to describe non-CCA patients in the Method section on page 8-9 as follows:

“Controls: There were 2 control cohorts in this study. First, to compare demographic differences between CCA patients and non-CCA individuals, all Thai citizens were used as controls (n=64,076,033). Demographic data of the Thai citizens were obtained in the year 2011 (the midpoint of the study period) from the Department of Provincial Administration, Ministry of the Interior, Thailand (Table 1). Second, to explore the potential associations between cirrhosis, diabetes and chronic viral hepatitis B and C infections and CCA development in the Thai population, patients without CCA who were admitted to hospitals during the study period were used as controls (n=18,508,448). The data of non-CCA controls were obtained from discharge summary notes in the same manner as CCA cases.”

Table 1 has been edited as follows:

Table 1. Baseline characteristics of CCA patients diagnosed between 2009 and 2013

Variable	CCA cases (N=39,421)	Thai population (N=64,076,033)	<i>p</i> -value
Gender			
Male, n (%)	24,120 (61.2%)	31,529,148 (49.2%)	<0.001
Age in years, mean \pm SD	64.1 \pm 11.7	34.9*	
Geographic region, n (%)			<.0001
Northeast	24,239 (61.5%)	21,585,883 (33.7%)	<0.001
North	6,699 (17.0%)	11,783,311 (18.4%)	<0.001
Central	5,295 (13.4%)	16,060,141 (25.1%)	<0.001
Bangkok	1,114 (2.8%)	5,674,843 (8.8%)	<0.001
South	1,056 (2.7%)	8,971,855 (14.0%)	<0.001

*Weighted average age calculated using the midpoint; Standard deviation was undefined; thus, a t-test was not performed.

Comment 3. The data registry probably did not contain full information regarding cirrhosis, B/C hepatitis, and diabetes in all patients. What percentages were available in the data registry?

Responses: We thank the reviewer for this question. Because the diagnosis of disease was exclusively obtained from the discharge summary note, we were therefore not able to determine the number of patients with no diagnosis of cirrhosis, B/C hepatitis and diabetes who had performed investigations to confirm the absence of these conditions. We recognized this limitation and have added a sentence to describe this limitation in the Discussion section on page 16 as follows:

“Data on hepatitis B/C infection, diabetes, smoking, alcohol drinking, obesity, and metabolic syndrome were not available in the discharge summary notes of all participants”

Comment 4. Patients with B or C hepatitis do not have cirrhosis? If so, what are the causes of cirrhosis?

Responses: We thank the reviewer for this question. We have abstracted data on underlying liver diseases related to cirrhosis including chronic viral hepatitis B and C infection and alcoholic liver disease. The numbers of patients with the underlying liver diseases causing cirrhosis among cirrhotics and non-cirrhotics are shown in the below Table. In summary, chronic viral hepatitis B and C infection and alcoholic liver disease are causes of cirrhosis.

	Cirrhotics (n=1,896)	Non-cirrhotics (n=37,525)	P
Hepatitis B infection	124 (6.5%)	132 (0.3%)	<0.0001
Hepatitis C infection	100 (5.3%)	84 (0.2%)	<0.0001
Alcoholic liver disease	197 (10.4%)	0 (0%)	<0.0001

We have added a sentence to describe underlying diseases of cirrhosis in the Result section on page 11 as follows:

“CCA patients with cirrhosis (n=1896) were significantly more likely to have hepatitis B infection, hepatitis C infection and alcoholic liver disease than CCA patients without cirrhosis (n=37,525), i.e., 124 (6.5%) vs. 132 (0.3%), 100 (5.3%) vs. 84 (0.2%), and 197 (10.4%) vs. 0 (0%) for hepatitis B infection, hepatitis C infection and alcoholic liver disease, respectively; $p<0.0001$ for all.”

Comment 5. How do you calculate one-year mortality? Describe in Statistical analysis section.

Responses: We have added sentences to describe the calculation of one-year mortality in the Statistical analysis section on page 10 as follows:

“One-year mortality was calculated using the per-patient information in which the survival time was calculated from the subtraction between the date of death and the date of the first admission. One-year mortality was then defined as 1 if the survival time was less than a year and 0 otherwise.”

Comment 6. Please describe Ptrend method in Statistical analysis section.

Responses: We have added the following sentences to describe Ptrend method in the Statistical analysis section on page 10 as follows:

“An analysis of the trend in CCA incidence and mortality was performed using “nptrend”, the nonparametric test for trend across ordered groups developed by Cuzick (1985), which is an extension of the Wilcoxon rank-sum test.”

Reviewer 5

The study aimed to identify potential risk factors of cholangiocarcinoma patients among 5 different regions of Thailand, the author found that diabetes and chronic liver diseases may be associated with cholangiocarcinoma in the Thai population. The manuscript may be of some value.

There are some revision points.

Comment 1. There are also other factors that may attributed to this disease, such as age, the choledocholithiasis or hepatolithiasis, choledochal cysts, bile duct adenoma, Caroli's syndrome and so on, especially the choledocholithiasis or hepatolithiasis, which is more common in Asian countries than western countries, author should explain why did not involve these factors and how these factors affect the results.

Responses: We appreciate the reviewer's thorough thinking. We completely agree with the reviewer that these factors have been shown to be associated with CCA development. In this study, we did not include these factors in the analysis because we aimed to explore the factors that have recently been proposed to be related to CCA and may in part contribute to the rising incidence of CCA worldwide.

We have abstracted additional information on choledocholithiasis, hepatolithiasis, choledochal cyst, Calori's syndrome, and bile duct adenoma using the following ICD codes: K80.5 for choledocholithiasis or hepatolithiasis; Q44.4 for choledochal cyst or Calori's syndrome; and D13.4 and D13.5 for intrahepatic and extrahepatic bile duct adenoma. We found that, of the 39,421 CCA patients, there were 360 (0.91%), 17 (0.04%), and 8 (0.02%) patients who had choledocholithiasis or hepatolithiasis, choledochal cyst or Calori's syndrome and bile duct adenoma, respectively. Given the relatively small number of patients with these conditions, we believe these factors had a minimal impact on the results reported in this study.

We have added sentences to explain why these factors were not included in the study in the Discussion section on page 17 as follows:

“Choledocholiathiasis, hepatolithiasis, choledochal cysts, Caroli's syndrome and bile duct adenoma have been shown to be associated with CCA^[1,14]. These conditions were not included in the analysis because this study aimed to explore the conditions that have recently been proposed to be related to CCA and may in part contribute to the increasing incidence of CCA worldwide.”

Comment 2. Author in the introduction part wrote that Cholangiocarcinoma is the second most common primary liver tumor, in fact Cholangiocarcinoma is a malignant disease different from liver tumor, the surgical procedure and prognosis was totally different.

Responses: We have deleted the phrase “is the second most common primary liver tumor” in the Introduction section accordingly.

Comment 3. There are some spelling mistakes, language should be strengthened.

Responses: We thank the reviewer for the thorough review. We have carefully checked the manuscript and corrected the spelling mistakes. The manuscript was also re-reviewed and edited by the professional language editing company. The certificate from the company is attached.

Comment 4. We are wondering the causes that lead to the Cirrhosis in the current study, please explain the reasons.

Responses: We have abstracted data on underlying liver diseases related to cirrhosis including chronic viral hepatitis B and C infection and alcoholic liver disease. The numbers of patients with the underlying liver diseases causing cirrhosis among cirrhotics and non-cirrhotics are shown in the below Table. In summary, chronic viral hepatitis B and C infection and alcoholic liver disease are causes of cirrhosis.

	Cirrhotics (n=1,896)	Non-cirrhotics (n=37,525)	P
Hepatitis B infection	124 (6.5%)	132 (0.3%)	<0.0001
Hepatitis C infection	100 (5.3%)	84 (0.2%)	<0.0001
Alcoholic liver disease	197 (10.4%)	0 (0%)	<0.0001

We have added a sentence to describe underlying diseases of cirrhosis in the Result section on page 11 as follows:

“CCA patients with cirrhosis (n=1896) were significantly more likely to have hepatitis B infection, hepatitis C infection and alcoholic liver disease than CCA patients without cirrhosis (n=37,525), i.e., 124 (6.5%) vs. 132 (0.3%), 100

(5.3%) vs. 84 (0.2%), and 197 (10.4%) vs. 0 (0%) for hepatitis B infection, hepatitis C infection and alcoholic liver disease, respectively; $p < 0.0001$ for all.”

Comment 5. We also want to know about the differences between hepatitis-related Cirrhosis and non-hepatitis-related Cirrhosis in affecting the prevalence of cholangiocarcinoma.

Responses: There were 224 (1.44%) of the 15,544 patients with viral hepatitis B/C cirrhosis who developed CCA; while there were 1,672 (1.06%) of the 156,577 patients with non-viral hepatitis B/C-related cirrhosis who developed CCA ($p < 0.001$).

Comment 6. May be some figures are needed.

Responses: Thank you for this suggestion. We have added a figure of Thailand map displaying difference in CCA prevalence among 5 regions.

Comment 7. In fact, cholangiocarcinoma includes three types, intrahepatic, hilar and distal. Each with a different surgical procedure, and the prognosis was also different, it would be better if the authors analyze the characteristics and outcomes of these three types of classification respectively.

Responses: We again appreciate this insightful comment from the reviewer. We completely agree with the reviewer that each subtype has different characteristics and outcomes and should be analyzed separately. Unfortunately, in this study, it is not feasible to perform analysis by subtype because the diagnosis of cholangiocarcinoma was obtained using ICD codes. The ICD codes group hilar and distal cholangiocarcinoma together as extrahepatic cholangiocarcinoma. We have acknowledged this limitation in the Discussion section on page 17 as follows:

“The diagnosis of CCA was obtained from ICD codes, which group hilar and distal CCA together as extrahepatic CCA; thus, an analysis by CCA subtypes could not be performed.”

Comment 8. The author in this passage did not involve the treatment options of all the 39399 patients, since different treatment choices may contribute to different survival outcomes, because authors also presented the total survival

data, it would be more meaningful if authors can present the survival data according to different treatment options.

Responses: We have abstracted data on treatment options and found that only 756 (1.9%) received surgical treatment. Because majority of patients (98.1%) of the entire cohort did not receive surgical treatment, the survival of the entire cohort was chosen for presentation in this study to reflect the overall picture of survival outcome of CCA patients in Thailand.