



Mortality and extraintestinal cancers in patients with primary sclerosing cholangitis and inflammatory bowel disease

Ashwin N. Ananthakrishnan^{a,b,*}, Andrew Cagan^d, Vivian S. Gainer^d, Su-Chun Cheng^c, Tianxi Cai^c, Peter Szolovits^f, Stanley Y. Shaw^{b,g}, Susanne Churchill^h, Elizabeth W. Karlson^{b,i}, Shawn N. Murphy^{b,d,j}, Isaac Kohane^{b,e,h}, Katherine P. Liao^{b,i}

^a Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, United States

^b Harvard Medical School, Boston, MA, United States

^c Department of Biostatistics, Harvard School of Public Health, Boston, MA, United States

^d Research IS and Computing, Partners HealthCare, Charlestown, MA, United States

^e Children's Hospital Boston, Boston, MA, United States

^f Massachusetts Institute of Technology, Cambridge, MA, United States

^g Center for Systems Biology, Massachusetts General Hospital, Boston, MA, United States

^h i2b2 National Center for Biomedical Computing, Brigham and Women's Hospital, Boston, MA, United States

ⁱ Division of Rheumatology, Allergy and Immunology, Brigham and Women's Hospital, Boston, MA, United States

^j Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

Received 22 December 2013; received in revised form 20 January 2014; accepted 24 January 2014

KEYWORDS

Crohn's disease;
Ulcerative colitis;
Sclerosing cholangitis;
Cancer;
Colorectal cancer

Abstract

Introduction: Primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD) frequently co-occur. PSC is associated with increased risk for colorectal cancer (CRC). However, whether PSC is associated with increased risk of extraintestinal cancers or affects mortality in an IBD cohort has not been examined previously.

Methods: In a multi-institutional IBD cohort of IBD, we established a diagnosis of PSC using a novel algorithm incorporating narrative and codified data with high positive and negative predictive value. Our primary outcome was occurrence of extraintestinal and digestive tract cancers. Mortality was determined through monthly linkage to the social security master death index.

Results: In our cohort of 5506 patients with CD and 5522 patients with UC, a diagnosis of PSC was established in 224 patients (2%). Patients with IBD–PSC were younger and more likely to be male

* Corresponding author at: Crohn's and Colitis Center, Massachusetts General Hospital, 165 Cambridge Street, 9th Floor, Boston, MA 02114, United States. Tel.: +1 617 724 9953; fax: +1 617 726 3080.

E-mail address: aananthakrishnan@partners.org (A.N. Ananthakrishnan).

compared to IBD patients without PSC; three-quarters had UC. IBD–PSC patients had significantly increased overall risk of cancers compared to patients without PSC (OR 4.36, 95% CI 2.99–6.37). Analysis of specific cancer types revealed that a statistically significant excess risk for digestive tract cancer (OR 10.40, 95% CI 6.86–15.76), pancreatic cancer (OR 11.22, 95% CI 4.11–30.62), colorectal cancer (OR 5.00, 95% CI 2.80–8.95), and cholangiocarcinoma (OR 55.31, 95% CI 22.20–137.80) but not for other solid organ or hematologic malignancies.

Conclusions: PSC is associated with increased risk of colorectal and pancreatobiliary cancer but not with excess risk of other solid organ cancers.

© 2014 European Crohn's and Colitis Organisation. Published by Elsevier B.V. All rights reserved.

1. Introduction

As inflammatory bowel diseases (IBD) often have their onset during young adulthood and are associated with preserved life-expectancy, there is considerable interest in the long-term outcomes including risk of malignancy.^{1–4} In a large Danish cohort, both Crohn's disease (CD) and ulcerative colitis (UC) were associated with a modest increase in risk of cancer involving the gastrointestinal tract as well as extraintestinal cancers.⁵ However, while there was a temporal decrease in the incidence of gastrointestinal cancers, the incidence of extraintestinal cancers remained stable or proportionally increased over time. Other cohorts have variably demonstrated this increased extra-intestinal cancer risk, and emphasize the need for continued study.^{6,7}

Between 2 and 5% of patients with IBD have associated primary sclerosing cholangitis (PSC), an autoimmune inflammatory disease of the biliary system that is associated with progressive fibrosis, cirrhosis, and end-stage liver disease.^{8,9} In contrast, nearly 80% of patients with PSC have underlying IBD. Several studies have examined the natural history of patients with IBD–PSC and have described a distinct disease phenotype with more frequent occurrence of ulcerative pancolitis and a milder course of underlying IBD.^{8–12} It is well recognized that patients with PSC–IBD have a greater risk of colorectal cancer and cholangiocarcinoma.^{9,13} However, there has been only limited examination of whether co-existing PSC is associated with an increase in risk of extra-intestinal cancer.

One challenge in addressing this question has been the need for a large cohort of patients with IBD–PSC and prolonged follow-up. It has been difficult to define PSC accurately in large datasets without prospectively recruited patients as many of the features supporting a diagnosis of PSC are non-specific (for example, elevated liver function tests). In addition, the utility of administrative billing codes that serve as a useful first-pass to define eligible cases are limited as the diagnosis code for PSC is used more commonly for cholangitis from other etiologies such as gallstone disease, and consequently has low specificity and positive predictive value.¹⁴ With increasing adoption of electronic medical records (EMR), there is an important unmet need for accurate definition of PSC in such data sources that may allow for efficient accrual of cohorts and definition of eligible cases. We have previously demonstrated that natural language processing (NLP) allows for the identification of free text phrases from within such EMR cohorts, improving the predictive value of case definition algorithms without compromising sensitivity.^{15,16}

We performed this study with the aims of (i) developing a case-definition algorithm for identifying patients with PSC with high accuracy in a validated EMR IBD cohort; (ii) examining the impact of co-existing PSC on the risk of gastrointestinal and extraintestinal cancers in patients with IBD; and (iii) defining the impact of PSC on mortality in patients with IBD and identifying risk factors for such outcomes.

2. Methods

2.1. Data source

The data source for our study was a validated EMR IBD cohort. The development of our cohort has been described in detail in our previous publications.^{15,17–19} In brief, from a cohort comprising the entire population receiving care at one of two major tertiary referral hospitals (Massachusetts General Hospital and Brigham and Women's Hospital) or affiliated hospitals and practices in the Greater Boston area, we identified all potential IBD patients with at least one International classification of diseases, 9th edition, clinical modification code for Crohn's disease (555.x) or ulcerative colitis (556.x). Using codified and narrative data including free text mentions of terms identified using natural language processing with the clinical Text Analysis and Knowledge Extraction System (cTAKES),²⁰ we developed a classification algorithm that identified patients with true diagnosis of CD or UC with a high specificity and positive predictive value. This resulted in a final cohort of 5522 UC and 5506 CD patients. The validity of our algorithm was confirmed in an independent random sample from our cohort.¹⁵

2.2. Determination of primary sclerosing cholangitis

Due to the lack of a specific ICD-9-CM code for PSC and poor specificity of the ICD-9-CM code for cholangitis (576.1) in identifying patients with PSC,¹⁴ we adopted a two-step approach to define patients with confirmed PSC (Fig. 1). First, we identified all patients with possible PSC through a preliminary screen which consisted of: 1) at least one ICD-9-CM code for any of the following – cholangitis (576.1), cholangiocarcinoma (155.1), liver transplantation (V42.7, 50.5), ICD-9-CM or current procedural terminology (CPT) codes for endoscopic retrograde cholangiopancreatography (ERCP) or liver biopsy, receiving at least one prescription for ursodeoxycholic acid (ursodiol), or 2) at least one narrative mentions of 'primary sclerosing cholangitis' 'PSC' 'sclerosing

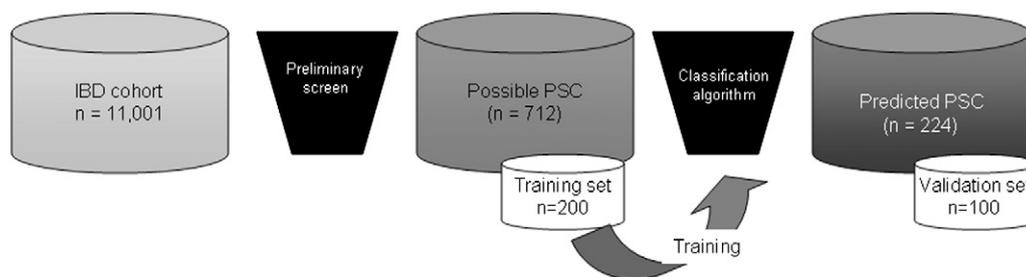


Figure 1 Defining primary sclerosing cholangitis in an electronic medical record cohort. IBD – inflammatory bowel disease, PSC – primary sclerosing cholangitis. Preliminary screen consistent of at least one ICD-9-CM code for cholangitis, cholangiocarcinoma, liver transplantation, ERCP, liver biopsy, prescription for ursodiol, or narrative concepts of “primary sclerosing cholangitis” or “sclerosing cholangitis”.

cholangitis’ within the EMR (Supplemental Table 1). NLP searches for narrative mentions were carried out in all outpatient office visit notes, endoscopic procedures, pathology reports, radiology reports, discharge summaries, and operative reports. Patients who had none of the variables included in our screen were determined to be unlikely to have PSC or have sufficient information to confirm/refute the diagnosis given the need for radiology, histologic, or endoscopy (ERCP) to establish the diagnosis.

A sample of 200 patients was selected at random from the cohort who passed the initial screen and chart review was performed by a board certified gastroenterologist (A.N.A.) to confirm or exclude a diagnosis of PSC based on accepted criteria for diagnosis.^{21,22} A penalized logistic regression model with the adaptive Lasso procedure was used to develop our final multivariate classification model incorporating both codified and narrative data. The classification algorithm assigned each patient in our cohort a probability of having a confirmed diagnosis of PSC ranging from 0 to 1. We then selected a probability threshold corresponding to a specificity of 95% and classified patients with probability equal to or greater than the threshold value as truly having PSC. The accuracy of our classification model was validated by reviewing the charts of an independent random sample of 100 patients classified by the model as having PSC.

2.3. Determination of variables and outcomes

Our primary outcome was occurrence of extraintestinal and digestive tract cancers, defined as the occurrence of one or more of the ICD-9-CM codes for the cancer subtypes of interest. We examined only the occurrence of cancers after the first diagnosis code for CD, UC, or PSC. Mortality was determined through monthly linkage to the social security master death index. Chart review was performed by randomly selecting 25 patients with each cancer subtype and confirmed high positive predictive value (>85%).

Demographic information including age, gender, race, age at first diagnosis code for CD or UC was obtained from our datamart. Medication use was classified as ever or never exposure using the electronic prescription function of our medical record. Medications of interest included systemic corticosteroids, immunomodulators (azathioprine, 6-mercaptopurine, methotrexate), and anti-tumor necrosis factor biologic agents (infliximab, adalimumab,

certolizumab pegol). Information was also obtained on laboratory tests including alkaline phosphatase, total bilirubin, hemoglobin, serum creatinine and albumin. Each lab value was modeled in quartiles.

2.4. Statistical analysis

Data analysis was performed by using Stata 12.0 (StataCorp, College Station, TX). Continuous variables were summarized by using medians and interquartile ranges (IQR) while categorical variables were expressed as proportions and compared by using the chi-square test. Multivariate logistic regression was performed to identify the association of PSC with overall cancer, metastatic cancer, as well as each cancer subtype. The final multivariate model additionally adjusted for age, gender, race, duration of disease, and immunosuppressant use. We also performed a logistic regression analysis with mortality as the outcome in the entire IBD cohort to examine the association between PSC diagnosis and mortality. Within the subgroup of patients with IBD–PSC, we analyzed whether demographic or clinical variables predicted outcome. A two-sided p-value <0.05 indicated independent statistical significance in the final multivariate model. This study was approved by the institutional review board of Partners Healthcare.

3. Results

3.1. Derivation of the PSC cohort

A total of 5522 UC and 5506 CD patients were included in our study, from which 712 patients were categorized as *possible PSC* based upon passing our preliminary screen (Fig. 1). A random sample of 200 charts was reviewed from the screen positive cohort. Fifty-eight (29%) patients who were screen positive were confirmed to have PSC on chart review. Patients with chart-review confirmed PSC were similar in age, more likely to be male (72% vs. 42%), had a higher mean number of ICD-9-CM codes for cholangitis (5.9 vs. 0.2), and had more frequent narrative mentions of ‘sclerosing cholangitis’ or ‘primary sclerosing cholangitis’ than patients who were found not to have PSC on chart review, but were less likely to have competing diagnoses such as cholelithiasis (0.9 vs. 0.4) ($p < 0.05$).

We then developed a classification algorithm incorporating demographics, codified, and narrative data (Fig. 2). The strongest predictors of a confirmed PSC diagnosis was narrative mention of primary sclerosing cholangitis or the number of ICD-9-CM codes for cholangitis followed by use of ursodiol and undergoing an abdominal magnetic resonance imaging or magnetic resonance cholangiopancreatography (MR/MRCP) examination. In contrast, a diagnosis of Crohn's disease, female gender, and number of ICD-9-CM codes for cholelithiasis were negative predictors of confirmed PSC. The final model incorporating both codified and narrative data had accuracy (area under the receiver operating curve) of 0.975 at a specificity of 95% (false positive rate of 5%). At this cut-off, the positive (PPV) and negative predictive values (NPV) of our algorithm were 88% and 97% respectively. The best performing algorithm incorporating only codified data had lower PPV (86%) and NPV (92%) at a similar specificity. The algorithm classified 224 patients from our screen positive group as truly having PSC, representing 2% of our IBD cohort. A random validation set of 100 patients were again selected from this cohort and chart review confirmed the performance of our algorithm (PPV = 95%).

3.2. Characterization of patients with PSC

Table 1 compares the characteristics of IBD patients, stratified by diagnosis of PSC. Patients with IBD–PSC were younger and more likely to be male compared to IBD patients without PSC. Three-quarters of patients with IBD–PSC had ulcerative colitis compared to an even distribution in the non-PSC IBD cohort. Among the subgroup of PSC patients with CD, 14% had stricturing disease and 16% had evidence of penetrating disease characterized by abdominal abscesses or internal fistulae. Twenty percent had a diagnosis of perianal fistula or abscess. The proportion with stricture or perianal disease was similar to the non-PSC CD group while penetrating disease was slightly more common in PSC-CD (Odds ratio (OR)

Table 1 Comparison of characteristics of patients with inflammatory bowel disease, stratified by diagnosis of primary sclerosing cholangitis.

Characteristic	No PSC (n = 10,777)	PSC (n = 224)	p-value
Age in years (mean (SD))	47 ¹⁹	45 ¹⁹	0.05
Age at first IBD diagnosis code (mean (SD))	41 ¹⁸	36 ¹⁸	0.01
Female (%)	54	26	<0.001
Non-white race (%)	15	18	0.13
Modified Charlson score (mean (SD))	2 ³	4 ⁴	<0.001
Duration of follow-up in years (mean (SD))	9 ⁷	10 ⁷	0.33
Ulcerative colitis	50	78	<0.001
≥ 1 IBD-related hospitalization (%)	31	42	<0.001
≥ 1 IBD-related surgery (%)	12	18	0.004
Ever 5-aminosalicylates use (%)	47	53	0.11
Ever steroids use (%)	36	52	<0.001
Ever immunomodulator use (%)	24	31	0.02
Ever anti-TNF biologic use (%)	12	7	0.03

IBD – Inflammatory bowel disease; PSC – primary sclerosing cholangitis; Anti-TNF – anti-tumor necrosis factor (infliximab, adalimumab, certolizumab pegol).

2.80, 95% CI 1.31–6.02). Patients with IBD–PSC were more likely to have required IBD-related hospitalization, surgery, steroids, or immunomodulators, but less likely to have received anti-TNF biologic therapy. Among those with PSC, the median value for highest bilirubin was 1.7 mg/dL (interquartile range (IQR) 0.8–5.5) and median alkaline phosphatase was

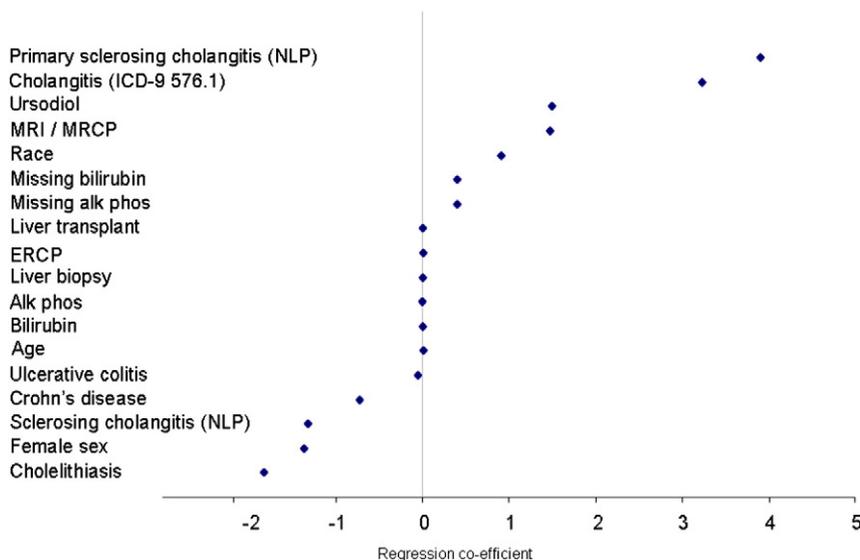


Figure 2 Variables predicting diagnosis of primary sclerosing cholangitis in an electronic medical record cohort.

342 IU/L (IQR 193–563). A total of 96 PSC patients underwent at least one ERCP (range 1–9), and 30 underwent liver transplantation (14%).

3.3. Cancer risk in patients with PSC

IBD–PSC patients had significantly increased overall risk of cancers compared to IBD patients without PSC (multivariate odds ratio (OR) 4.36, 95% confidence interval (CI) 2.99–6.37) (Table 2), and a similar increase in risk of metastatic cancer (OR 4.88, 95% CI 2.95–8.09). Analysis of specific cancer types revealed a statistically significant excess risk for digestive tract cancer (OR 10.40, 95% CI 6.86–15.76), pancreatic cancer (OR 11.22, 95% CI 4.11–30.62), colorectal cancer (OR 5.00, 95% CI 2.80–8.95), and cholangiocarcinoma (OR 55.31, 95% CI 22.20–137.80) but not for lymphoma or other solid organ malignancies. The excess cancer risk in PSC remained significant after excluding patients who had undergone a liver transplantation, as well as excluding cancers that were diagnosed within 6 months of the first ICD-9-CM code for cholangitis. The risk also remained similar when performing subgroup analysis among those with CD alone or UC alone. We additionally adjusted for intensity of healthcare utilization (defined as number of distinct facts which included office visits, radiology or laboratory tests, procedures, or hospitalizations) and observed no change in our estimates suggesting that the excess risk in PSC is not explained by more frequent healthcare contact in this group.

We then performed an analysis examining if severity of PSC assessed by quartiles of bilirubin, alkaline phosphatase, or presence of cirrhosis was associated with increased risk for solid tumor or hematologic malignancies. On preliminary analysis, both highest quartile of bilirubin (OR 3.44, 95% CI 1.24–9.83)

and alkaline phosphatase (OR 2.96, 95% CI 0.97–9.07) appeared to be associated with increased risk of cancer compared to the lowest quartile. However this association was predominantly driven by cholangiocarcinoma. Excluding hepatobiliary cancers neutralized the association between markers of PSC severity and overall risk of solid tumors. Markers of PSC severity were also not associated with risk of hematologic cancers.

3.4. Mortality in PSC

IBD–PSC patients had higher mortality when compared to IBD patients without PSC (OR 3.51, 95% CI 2.30–5.36). Table 3 presents a multivariate analysis of predictors of mortality in IBD–PSC. Associated co-morbidity was the strongest predictor of mortality. Patients with 3 or more Charlson co-morbidity measures had significantly elevated mortality (OR 12.02, 95% CI 2.96–48.91) compared to those with fewer than three co-morbidities. Interestingly, patients who were on immunomodulator therapy had lower adjusted odds of death while patients in the highest quartile of bilirubin level had a tenfold increased in mortality (OR 10.50, 95% CI 1.62–68.03 compared to the lowest quartile).

4. Discussion

There is scarce literature on the risk of extraintestinal cancers in IBD–PSC. Since the majority of patients with PSC have underlying IBD, it is important to examine the excess risk associated with PSC not just in comparison to the general population, but also compared to IBD patients without underlying PSC as IBD itself has been variably associated with increased risk of gastrointestinal and extraintestinal cancers.^{1,5–7} We demonstrate that by using a combination

Table 2 Association between primary sclerosing cholangitis and various cancers in an inflammatory bowel disease cohort.

Cancer type	Number in PSC–IBD (n = 224)	Number in IBD (n = 10,777)	Multivariate OR [†]	95% CI
Solid tumor	42	688	4.36	2.99–6.37
Metastatic cancer	20	255	4.88	2.95–8.09
Breast	2	130	1.66	0.39–7.09
Melanoma	2	68	1.74	0.42–7.25
Prostate	6	110	2.19	0.92–5.22
Lung	3	98	1.84	0.57–5.96
Pancreatic	5	26	11.22	4.11–30.62
Oral	0	33	*	
Digestive tract	34	225	10.40	6.86–15.76
Esophageal	1	14	3.30	0.42–25.93
Gastric	1	21	2.80	0.37–21.44
Small intestine	0	22	*	
Respiratory	3	112	1.55	0.48–5.01
Bladder	2	41	2.74	0.64–11.65
Renal	4	70	2.67	0.95–7.53
Genitourinary	7	238	1.38	0.63–3.08
Neuroendocrine	0	8	*	
Colorectal	14	153	5.00	2.80–8.95
Cholangio Ca.	11	10	55.31	22.20–137.80
Lymphoproliferative malignancies	9	213	1.70	0.85–3.39
Death	34	765	3.51	2.30–5.36

[†] Each outcome analysis is adjusted for age, gender, race, duration of follow-up, use of immunomodulators, biologic use.

* Could not be estimated due to lack of events in the PSC–IBD group.

Table 3 Predictors of mortality in patients with PSC-IBD.

Characteristic	Odds ratio	Confidence interval
Charlson co-morbidity		
<3	1.0	
3 or more	12.02	2.96–48.91
Immunomodulator use		
No	1.0	
Yes	0.15	0.04–0.52
Highest bilirubin value		
Quartile 1	1.0	
Quartile 2	3.18	0.42–24.02
Quartile 3	1.97	0.32–12.19
Quartile 4	10.50	1.62–68.03

+ Adjusting additionally for quartile of highest alkaline phosphatase.

of administrative and narrative free text data, it is possible to define patients with PSC accurately in an EMR population, allowing for efficient accrual of cohorts. Second, by using this large multi-institutional cohort of patients with IBD-PSC with long duration of follow-up, we confirm previously demonstrated associations with colorectal cancer and hepatobiliary cancer, but reassuringly demonstrate that there is no increase in risk of extraintestinal cancers in patients with IBD-PSC compared to IBD patients without PSC.

Autoimmune diseases are frequently associated with increased risk of cancer including outside the target organ(s) affected. While the elevated risk of colorectal and small intestinal cancer in patients with IBD is well established, recent cohorts have additionally suggested that IBD patients are also at an elevated risk for extraintestinal cancers, particularly hematologic malignancies, melanoma, and smoking-related cancers.^{1,5,7,12} A meta-analysis by Pedersen et al. did not find a statistically significant elevation in overall extra-intestinal cancer risk, but identified increased risk of cancers of the lung, urinary bladder, upper gastrointestinal tract, and skin in patients with CD and leukemia in patients with UC.⁷ An earlier study by Bernstein et al. found an increased risk of lymphoma in patients with CD unrelated to immunosuppressive therapy.²³ Corollaries exist in other autoimmune diseases. Rheumatoid arthritis (RA) is associated with increased incidence of cancer, in particular lung cancer, Hodgkin's and non-Hodgkin's lymphoma²⁴ but it remains challenging to differentiate the effect of disease from that of immunosuppression that correlates with severity of inflammation. Systemic lupus erythematosus, RA, and Sjogren syndrome are all associated with increased risk of lymphoma²⁵ while systemic sclerosis is associated with increased risk of breast cancer.²⁶ With this background, definition of extraintestinal cancer risk in autoimmune diseases is important to define quantitatively in order to distinguish the risk associated with immunosuppressive therapies from the risk associated with the disease itself. Furthermore, it can also indicate the need for routine screening for early detection of such cancers.

There have been few studies systematically examining cancer risk in patients with PSC, particularly those occurring outside the gastrointestinal tract.^{27,28} De Valle et al., by utilizing a cohort of 199 PSC patients in Sweden, found a sixfold increase in risk of cancer in patients with PSC compared

to the general population, similar in magnitude to the excess risk we identified compared to non-PSC IBD patients.²⁹ Their cohort, as did ours, confirmed the increase in risk of hepatobiliary, colorectal, and cholangiocarcinoma in PSC. However, neither that study nor the larger study by Bergquist et al. including 604 PSC patients specifically examined occurrence of extraintestinal cancers.³⁰ In a large series of patients with PSC, Claessen et al. described the occurrence of extrahepatic cancers but lacked a control population.³¹ Ngu et al. found an elevated risk of extrahepatic cancer in PSC but had a small number of cases limiting their ability to examine the spectrum of extraintestinal cancers in detail.²⁷

There are a few possible mechanisms of increased risk of cancers in patients with PSC. First, since the entire cohort (PSC and non-PSC) consisted of patients with underlying IBD, a significant portion were either currently or previously on immunosuppressive therapy. Immunosuppressive therapy, particularly thiopurines, has been associated with increased risk of cancers owing to impaired immune surveillance.³² Second, chronic inflammation in PSC and IBD and associated immune activation has been associated with increased risk of cancers independent of the effect of immunosuppressive therapy as has also been observed in other chronic inflammatory diseases like rheumatoid arthritis.³³ A third possibility is common genetic or environmental factors that predispose both to underlying PSC and cancer development. Finally, heightened medical surveillance in patients with PSC may lead to increased detection of cancers through the association that persisted after adjusting for intensity of health care utilization.

The increase in mortality risk in PSC-IBD patients compared to non-PSC IBD patients in our study is similar in magnitude to the standardized mortality ratio identified in other cohorts.^{27–29} Risk factors for mortality in the study by de Valle et al. were age, female gender, and bilirubin,²⁹ consistent with some of our findings. We did not observe an independent effect of age which could be due to the fact that we also adjusted for non-liver co-morbidity which was not adjusted for in the prior cohort study. Other studies have similarly consistently demonstrated an association between bilirubin levels and death or transplantation^{12,34–37} while the predictive value of other laboratory parameters have been less consistent.

There are a few implications for our findings. First, we demonstrate that the use of codified and free text data allows for accurate ascertainment of patients with PSC in an EMR cohort. A prior attempt at validation of administrative data for PSC found a positive predictive value of only 7%, which could only be improved to 41% by using a refined model. In contrast, we demonstrated that addition of free text phrases identified by using natural language processing and competing diagnoses allows us to reach a PPV of 90% and a specificity of 95%, both substantial improvements to the existing literature.¹⁴ We have also demonstrated that such methods are portable between different institutions despite the use of distinct EMR systems.^{38,39} This has great utility in rare diseases like PSC where assembly of cohorts of sufficient power for phenotypic or genetic analysis often requires multi-institutional or international collaboration. When linked to biospecimen repositories, such EMR based methods have proven very useful in replicating genetic associations or identifying novel hypotheses.^{40–44}

We readily acknowledge several limitations to our study. First, since the two hospitals contributing a majority of patients to our cohort were major academic referral centers, the acuity in our cohort is likely skewed towards severe disease. However, since both our IBD-PSC and non-PSC IBD controls were drawn from the same population pool, this is unlikely to have differentially affected our results. Furthermore, any excess risk identified in the PSC cohort was not entirely due to differential healthcare utilization as our findings were robust to adjusting for this variable. Second, we did not consistently have information on other important factors predictive of mortality such as liver histology. Third, we were unable to separately examine the risk associated with large-duct compared to small-duct PSC. However, given the different natural history of disease between the two disease subtypes, this is an important question that merits further exploration in future studies.

In conclusion, by using a large multi-institutional cohort of patients with PSC-IBD, we demonstrate that PSC is associated with an increased risk of mortality and cancer in patients with IBD. We confirmed previously reported associations with colorectal, hepatobiliary, and pancreatic cancer and reassuringly identified no additional associations with extraintestinal tumors in patients with PSC. We also present a novel method to accurately define patients with PSC in an EMR cohort, potentially allowing for inter-institutional portability and rapid and efficient accrual of multi-institutional cohorts for collaborative research.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.crohns.2014.01.019>.

Sources of funding

The study was supported by NIH U54-LM008748. A.N.A. is supported by funding from the American Gastroenterological Association and from the US National Institutes of Health (K23 DK097142). K.P.L. is supported by NIH K08 AR060257 and the Harold and Duval Bowen Fund. E.W.K. is supported by grants from the NIH (K24 AR052403, P60 AR047782, R01 AR049880).

Author contributions

An author may list more than one contribution, and more than one author may have contributed to the same element of the work

Study concept – Ananthakrishnan

Study design – Ananthakrishnan

Data collection – Ananthakrishnan, Gainer, Cagan, Cai, Cheng, Churchill, Kohane, Shaw, Liao, Szolovits, Murphy

Analysis – Ananthakrishnan, Cai, Cheng,

Preliminary draft of the manuscript – Ananthakrishnan

Approval of final version of the manuscript – Ananthakrishnan, Gainer, Cagan, Cai, Cheng, Churchill, Kohane, Shaw, Liao, Szolovits, Murphy

Conflict of interest

None.

References

- Jess T, Frisch M, Simonsen J. Trends in overall and cause-specific mortality among patients with inflammatory bowel disease from 1982 to 2010. *Clin Gastroenterol Hepatol* 2013;11:43–8.
- Romberg-Camps M, Kuiper E, Schouten L, Kester A, Hesselink-van de Kruijs M, Limonard C, et al. Mortality in inflammatory bowel disease in the Netherlands 1991–2002: results of a population-based study: the IBD South-Limburg cohort. *Inflamm Bowel Dis* 2010;16:1397–410.
- Selinger CP, Andrews J, Dent OF, Norton I, Jones B, McDonald C, et al. Cause-specific mortality and 30-year relative survival of Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis* 2013;19:1880–8.
- Selinger CP, Leong RW. Mortality from inflammatory bowel diseases. *Inflamm Bowel Dis* 2012;18:1566–72.
- Kappelman MD, Farkas DK, Long MD, Erichsen R, Sandler RS, Sorensen HT, et al. Risk of Cancer in Patients With Inflammatory Bowel Diseases: A Nationwide Population-based Cohort Study With 30 Years of Follow-up Evaluation. *Clin Gastroenterol Hepatol* 2014;12(2):265–73.
- Jess T, Horvath-Puho E, Fallingborg J, Rasmussen HH, Jacobsen BA. Cancer Risk in Inflammatory Bowel Disease According to Patient Phenotype and Treatment: A Danish Population-Based Cohort Study. *Am J Gastroenterol* 2013;108(12):1869–76.
- Pedersen N, Duricova D, Elkjaer M, Gamborg M, Munkholm P, Jess T. Risk of extra-intestinal cancer in inflammatory bowel disease: meta-analysis of population-based cohort studies. *Am J Gastroenterol* 2010;105:1480–7.
- Boonstra K, van Erpecum KJ, van Nieuwkerk KM, Drenth JP, Poen AC, Witteman BJ, et al. Primary sclerosing cholangitis is associated with a distinct phenotype of inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18:2270–6.
- Loftus Jr EV, Harewood GC, Loftus CG, Tremaine WJ, Harmsen WS, Zinsmeister AR, et al. PSC-IBD: a unique form of inflammatory bowel disease associated with primary sclerosing cholangitis. *Gut* 2005;54:91–6.
- Navaneethan U, P.G.K.V, Lashner BA, Lopez R, Kiran RP, Shen B. Severity of primary sclerosing cholangitis and its impact on the clinical outcome of Crohn's disease. *J Crohns Colitis* 2012;6:674–80.
- Navaneethan U, Venkatesh PG, Mukewar S, Lashner BA, Remzi FH, McCullough AJ, et al. Progressive primary sclerosing cholangitis requiring liver transplantation is associated with reduced need for colectomy in patients with ulcerative colitis. *Clin Gastroenterol Hepatol* 2012;10:540–6.
- Garioud A, Seksik P, Chretien Y, Corpechot C, Poupon R, Poupon RE, et al. Characteristics and clinical course of primary sclerosing cholangitis in France: a prospective cohort study. *Eur J Gastroenterol Hepatol* 2010;22:842–7.
- Torres J, Pineton de Chambrun G, Itzkowitz S, Sachar DB, Colombel JF. Review article: colorectal neoplasia in patients with primary sclerosing cholangitis and inflammatory bowel disease. *Aliment Pharmacol Ther* 2011;34:497–508.
- Molodecky NA, Myers RP, Barkema HW, Quan H, Kaplan GG. Validity of administrative data for the diagnosis of primary sclerosing cholangitis: a population-based study. *Liver Int* 2011;31:712–20.
- Ananthakrishnan AN, Cai T, Savova G, Cheng SC, Chen P, Perez RG, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis* 2013;19:1411–20.
- Ananthakrishnan AN, Cagan A, Gainer VS, Cai T, Cheng SC, Savova G, et al. Normalization of plasma 25-hydroxy vitamin D is associated with reduced risk of surgery in Crohn's disease. *Inflamm Bowel Dis* 2013;19:1921–7.

17. Ananthakrishnan AN, Cheng SC, Cai T, Cagan A, Gainer VS, Szolovits P, et al. Association Between Reduced Plasma 25-hydroxy Vitamin D and Increased Risk of Cancer in Patients with Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol* 2013 (iOct 23. pii: S1542-3565(13)01644-3).
18. Ananthakrishnan AN, Gainer VS, Perez RG, Cai T, Cheng SC, Savova G, et al. Psychiatric co-morbidity is associated with increased risk of surgery in Crohn's disease. *Aliment Pharmacol Ther* 2013;37:445–54.
19. Chapman R, Fevery J, Kalloo A, Nagorney DM, Boberg KM, Shneider B, et al. Diagnosis and management of primary sclerosing cholangitis. *Hepatology* 2010;51:660–78.
20. Singh S, Talwalkar JA. Primary sclerosing cholangitis: diagnosis, prognosis, and management. *Clin Gastroenterol Hepatol* 2013;11:898–907.
21. Bernstein CN, Blanchard JF, Kliewer E, Wajda A. Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer* 2001;91:854–62.
22. Mercer LK, Davies R, Galloway JB, Low A, Lunt M, Dixon WG, et al. Risk of cancer in patients receiving non-biologic disease-modifying therapy for rheumatoid arthritis compared with the UK general population. *Rheumatology (Oxford)* 2013;52:91–8.
23. Zintzaras E, Voulgarelis M, Moutsopoulos HM. The risk of lymphoma development in autoimmune diseases: a meta-analysis. *Arch Intern Med* 2005;165:2337–44.
24. Colaci M, Giuggioli D, Vacchi C, Lumetti F, Iachetta F, Marcheselli L, et al. Breast cancer in systemic sclerosis: Results of a cross-linkage of an Italian Rheumatologic Center and a population-based Cancer Registry and review of the literature. *Autoimmun Rev* 2014;13(2):132–7.
25. Ngu JH, Gearry RB, Frampton CM, Stedman CA. Mortality and the risk of malignancy in autoimmune liver diseases: a population-based study in Canterbury, New Zealand. *Hepatology* 2012;55:522–9.
26. Card TR, Solaymani-Dodaran M, West J. Incidence and mortality of primary sclerosing cholangitis in the UK: a population-based cohort study. *J Hepatol* 2008;48:939–44.
27. de Valle MB, Bjornsson E, Lindkvist B. Mortality and cancer risk related to primary sclerosing cholangitis in a Swedish population-based cohort. *Liver Int* 2012;32:441–8.
28. Bergquist A, Ekbohm A, Olsson R, Kornfeldt D, Loof L, Danielsson A, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. *J Hepatol* 2002;36:321–7.
29. Claessen MM, Vleggaar FP, Tytgat KM, Siersema PD, van Buuren HR. High lifetime risk of cancer in primary sclerosing cholangitis. *J Hepatol* 2009;50:158–64.
30. Wiesner RH, Grambsch PM, Dickson ER, Ludwig J, MacCarty RL, Hunter EB, et al. Primary sclerosing cholangitis: natural history, prognostic factors and survival analysis. *Hepatology* 1989;10:430–6.
31. Kim WR, Therneau TM, Wiesner RH, Poterucha JJ, Benson JT, Malinchoc M, et al. A revised natural history model for primary sclerosing cholangitis. *Mayo Clin Proc* 2000;75:688–94.
32. Talwalkar JA, Lindor KD. Natural history and prognostic models in primary sclerosing cholangitis. *Best Pract Res Clin Gastroenterol* 2001;15:563–75.
33. Dickson ER, Murtaugh PA, Wiesner RH, Grambsch PM, Fleming TR, Ludwig J, et al. Primary sclerosing cholangitis: refinement and validation of survival models. *Gastroenterology* 1992;103:1893–901.
34. Carroll RJ, Thompson WK, Eyster AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012;19:e162–9.
35. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V, et al. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am J Hum Genet* 2011;88:57–69.
36. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med* 2011;3:79re1.
37. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011;4:13.
38. Fullerton SM, Wolf WA, Brothers KB, Clayton EW, Crawford DC, Denny JC, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. *Genet Med* 2012;14:424–31.
39. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 2013;15:761–71.
40. Liao KP, Diogo D, Cui J, Cai T, Okada Y, Gainer VS, et al. Association between low density lipoprotein and rheumatoid arthritis genetic factors with low density lipoprotein levels in rheumatoid arthritis and non-rheumatoid arthritis controls. *Ann Rheum Dis* 2013 [May 28, Epub ahead of print].
41. Liao KP, Kurreeman F, Li G, Duclos G, Murphy S, Guzman R, et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. *Arthritis Rheum* 2013;65:571–81.
42. Fullerton SM, Wolf WA, Brothers KB, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. *Genet Med* 2012;14:424–31.
43. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 2013;15:761–71.
44. Liao KP, Kurreeman F, Li G, et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non-rheumatoid arthritis controls. *Arthritis Rheum* 2013;65:571–81.