

Using Primary Care Data to Report Real-World Pancreatic Cancer Survival and Symptomatology

Nathan JEFFREYS^{a,1}, Hajira DAMBHA-MILLER^b, Xuejuan FAN^c, Filipa FERREIRA^c, Harshana LIYANAGE^c, Julian SHERLOCK^c, John WILLIAMS^c, Rebecca RICE^d, Ali STUNT^d, Sara FAITHFULL^e, Piers GATENBY^a, Agnieszka LEMANSKA^e and Simon de LUSIGNAN^c

^aRoyal Surrey County Hospital, Royal Surrey NHS Foundation Trust, Guildford, UK

^bDivision of Primary Care and Population Health, University of Southampton, UK

^cNuffield Department of Primary Care Health Sciences, University of Oxford, UK

^dPancreatic Cancer Action, UK

^eSchool of Health Sciences, Faculty of Health and Medical Sciences, University of Surrey, UK

Abstract: Pancreatic cancer is the 10th most common cancer diagnosed; despite recent advances in many areas of oncology, survival remains poor, in part owing to late diagnosis. Whilst primary care data are used widely for epidemiology and pharmacovigilance, they are less used for observing survival. In this study we extracted a pancreatic cancer cohort from a nationally representative English primary care database of electronic health records (EHRs) and reported on their symptom and mortality data. A total of 11,649 cases were identified within the Oxford Royal College of General Practitioners (RCGP) Clinical Informatics Digital Hub network. All-cause mortality data was recorded for 4623 (39.69%). Mean age at recording of cancer diagnosis was 71.4 years (SD 12.0 years). 1-year and 5-year survival was 22.06% and 3.27% respectively. Within a multivariate model, age had a significant impact on survival; those diagnosed under the age of 60 had the longest survival, as compared to those age 60 – 79 (HR: 1.36, 95% CI: 1.20 – 1.54, $p < 0.001$) and 80+ (HR: 2.13, 95% CI: 1.86 – 2.44, $p < 0.01$). Symptomatology was examined; at any time point abdominal pain was the most commonly reported symptom present in 5271 cases (45.2%), but within the 12 months preceding diagnosis jaundice was the most common feature, present in 2587 patients (22.2%). Future studies clarifying other contributing factors on survival outcomes and patterns of symptomatology are needed; primary care EHRs provide an opportunity to evaluate real-world cancer patient cohort data.

Keywords: Electronic health records (EHRs), pancreatic cancer, survival, health inequalities, real-world evidence

1. Introduction

Pancreatic cancer is the 10th most common cancer in the UK, with around 10,300 new cases annually [1]. Around 8 in 10 cases are diagnosed at a late stage where treatment options are limited; mortality is subsequently high, with 1-year survival of 25.4% and 5-

¹ Corresponding author: Nathan Jeffreys. Email: n.jeffreys@nhs.net

year survival of 7.3%. There is a clear need to detect early signals of pancreatic cancer in order to aid earlier case identification, diagnosis, and successful treatment.

Electronic health records (EHRs) have been previously used in primary care research to define disease cohorts, evaluate symptomatology, and develop decision support tools to aid clinicians in work-up [2,3]. However, less focus has been placed on using EHR data to provide up-to-date and population-specific survival estimates. We undertook this study to define the pancreatic cancer cohort within our EHR network, with the aim of evaluating current UK real-world survival and symptomatology within this nationally representative cohort.

2. Methods

2.1. Study design, data extraction

The Oxford Royal College of General Practitioners (RCGP) Clinical Informatics Digital Hub (ORCHID) network contains pseudonymised EHRs from over 500 general practices in England, representing over 10 million patients (>10% GP practices within UK). UK general practice is a registration-based system, where patients register with a single practice and have a unique NHS number which links primary and secondary care data. Patients are free to opt out of data sharing and are not included within subsequent extracts or analyses.

Events and person level data were extracted from the ORCHID July 2020 data extract that comprises 590 practices and over 10 million patient records. Patients were included within the pancreatic cancer cohort if they had any SNOMED-CT coding consistent with a diagnosis of pancreatic cancer of any type. For clinical features, we searched for candidate events selected from the NICE NG12 red flag symptoms [4] for the whole cohort at any time point and in the 12 months preceding diagnosis (coding list available on request). Such events were only included if the patient was over the age of 40 at the time of initial symptom recording; this approach aligns with NICE NG12 guidance for pancreatic cancer screening [4].

2.2. Statistical Analysis

Database extraction was performed in Structured Query Language (SQL) Server Management Studio version 18.3.1 and statistical analysis was performed in R version 3.5.1. For continuous variables, mean with standard deviation or median with Interquartile Range (IQR) is reported as appropriate. For categorical variables, proportions are reported as a percentage of the whole cancer cohort identified. Survival was the time from first recording of pancreatic cancer diagnosis to date of death. Univariate and multivariate analyses using Cox proportional hazard regression were performed; results were reported as hazard ratios (HR) with 95% confidence intervals. $P < 0.05$ was used as the level of significance.

2.3. Approvals

This study was approved by the RCGP board (data request number RSC_0420). Further ethical approval was not required. Patients consent to enrolment within the network on an opt-out basis.

3. Results

3.1. Cohort Demographics

Within our search we identified 11, 649 unique EHR records with coding mapped to a definite pancreatic cancer. This group had a mean age of 71.4 years (SD 12.0 years) and median age of 73.0 years (IQR 64.0 – 80.0 years) at first recording of pancreatic cancer. 5831 women and 5818 men were included within the cohort. Ethnicity data was available for 6700 (57.5%) patients; the majority of the cohort had ethnicity coding consistent with White British (6317, 94.3%), followed by Asian British (194, 2.9%), Black/African/Caribbean/Black British (130, 1.9%), Mixed/Multiple ethnic groups (29, 0.4%) and Other ethnicities (30, 0.4%). IMD data was available for 11,192 (96.1%); the cohort showed a bias towards the lower IMD quintiles (Q1: 7612, 68.0%; Q2: 2677, 23.9%; Q3: 706, 6.3%; Q4: 183, 1.6%; Q5: 14, 0.1%).

3.2. All-Cause Mortality Data

All-cause mortality data was recorded for 4623 (39.69%). The median survival after diagnosis was 0.33 years (IQR 0.12 to 0.97 years). 22.1% and 3.3% of people in the cohort survived 1 and 5 years respectively.

There was a significant effect of age on survival (Table 1). Those diagnosed under the age of 60 had the longest survival, with a significantly higher adjusted hazard ratio than both the 60 – 79 (adjusted HR: 1.36, 95% CI: 1.20 – 1.54, $p < 0.001$) and 80+ (adjusted HR: 2.13, 95% CI: 1.86 – 2.44, $p < 0.01$) age categories. There were no significant associations of gender, IMD status or ethnicity with the length of survival.

Table 1: Univariate and Multivariate Cox Regression for survival within pancreatic cancer cohort

	Number	Univariate Analysis			Multivariate Analysis		
		HR	95% CI	P value	HR	95% CI	P value
Age Category							
18-59	1840	1 (Reference)			1 (Reference)		
61 – 79	6600	1.43	1.31 – 1.57	< 0.001	1.36	1.20 – 1.54	< 0.001
80+	3206	2.17	1.97 – 2.40	< 0.001	2.13	1.86 – 2.44	< 0.001
Missing	3						
Sex							
Female	5831	1 (Reference)			1 (Reference)		
Male	5818	0.99	0.94 – 1.05	0.81	0.96	0.88 – 1.04	0.28
Missing	0						
IMD Quintile							
1 (Most deprived)	7612	1 (Reference)			1 (Reference)		
2	2677	1.03	0.97 – 1.11	0.38	1.02	0.93 – 1.13	0.64
3	706	1.06	0.93 – 1.20	0.40	1.14	0.97 – 1.34	0.12
4	183	1.26	0.96 – 1.66	0.10	1.35	0.98 – 1.87	0.07
5 (Least Deprived)	14	1.46	0.65 – 3.25	0.36	2.62	0.98 – 6.99	0.06
Missing	457						
Ethnicity							
White	6317	1 (Reference)			1 (Reference)		
Asian	194	0.85	0.65 – 1.11	0.22	0.84	0.63 – 1.10	0.20
Black	130	0.84	0.60 – 1.17	0.29	0.89	0.63 – 1.24	0.49
Mixed / Multiple	29	0.85	0.48 – 1.50	0.58	0.94	0.53 – 1.66	0.83
Other	30	0.79	0.41 – 1.52	0.48	0.90	0.47 – 1.75	0.76
Missing	4949						

3.3. Symptom Coding with Pancreatic Cancer Cohort

Symptom coding is reported in Table 2. Abdominal pain was the most common symptom recorded at any time point within the EHRs of our cohort, presenting in 5271 cases (45.2%). Jaundice was present in 3137 patients (26.9%) and a recording of newly diagnosed T2 diabetes was reported in 3070 patients (26.4%).

In the 12 months preceding diagnosis, jaundice was the most common symptom recorded for 2587 patients (22.2%), with abdominal pain being the next most common symptom, reported in 1891 patients (16.2%).

Table 2. Recording of Symptom Coding within Primary Care EHR of Pancreatic Cancer Cohort at any time point and in 12 months preceding diagnosis

<i>Clinical Features</i>	<i>Recorded at any time point</i>	<i>Percentage</i>	<i>Recorded in 12 months preceding cancer diagnosis</i>	<i>Percentage</i>
<i>Jaundice</i>	3137	26.9%	2587	22.2%
<i>Abdominal Pain</i>	5271	45.2%	1891	16.2%
<i>New Diabetes T2</i>	3070	26.4%	694	6.0%
<i>Constipation</i>	2768	23.8%	679	5.8%
<i>Weight Loss</i>	1197	10.3%	586	5.0%
<i>Finding of Vomiting</i>	1889	16.2%	558	4.8%
<i>Diarrhoea</i>	2761	23.7%	558	4.8%

4. Discussion

By using routine primary care EHR data, we have been able to utilize searches based on SNOMED-CT coding to define key clinical features of a real-world pancreatic cancer cohort, including cancer symptomatology and survival. Despite limiting our search to ‘red flag’ symptoms of pancreatic cancer, i.e. those suggested by NICE guidance, only a minority of patients had recorded EHR entries with these symptoms in primary care, with even fewer recorded in the twelve months prior to diagnosis. This suggests either low levels of symptom burden in primary care preceding diagnosis or poor coding of reported symptoms. Pancreatic cancer is often diagnosed in secondary healthcare settings which has implications for symptom coding within primary care databases.

Although mortality data was only available for a subset of this cohort, we show that survival data can be extracted from primary care EHRs and, despite utilising all-cause mortality instead of disease-specific mortality, this corresponds well to published survival estimates in the current literature [1]. Stratification by age, sex, IMD quintile and ethnicity in our analysis suggested age to be the only included baseline demographic that significantly impacted on survival. It is possible that there are disparities in other demographic factors or differences in symptomatology between demographic categories which should be investigated further.

A significant limitation of this study relates to missing or incomplete data within the EHR extract. Within our baseline characteristics data, ethnicity was missing for 42.5%, IMD quintile data in 3.9%, and mortality data in 60.3%. Analysis was limited to the demographic information and coding readily available within the network. There is also likely a delay between the date of initial cancer diagnosis and first registration of this within primary care records, which we have been unable to address in this study. SNOMED-CT coding consistent with pancreatic cancer was used for case identification;

an additional planned search aiming for further subcategorization based on tumour subtype or anatomical location was not feasible due to poor code utilisation (i.e. 0 coding entries relating to the main specific subtypes of pancreatic cancer, such as primary pancreatic adenocarcinoma and primary pancreatic malignant neuroendocrine tumour) .

There is a clear challenge with primary care EHR completeness and granularity evidenced within our study, as has been previously reported elsewhere [5]. Future work may be able to improve information yield by integrating ‘free-text’ data as well as clinical coding. Additional data relevant to oncological informatics research may be stored within large registries either part of national registry systems (e.g. National Cancer Registration and Analysis Service database in the UK) or trial-specific registries; linkage of primary care and secondary care records has elsewhere been shown to be feasible and provides greater accuracy and granularity of information for the disease of interest [6]. Such linkage would include information such as tumour subtype and anatomical location, which we found to be poorly recorded in primary care EHRs in this study.

5. Conclusion

Utilising the RCGP-RSC database, we were able to identify and define a pancreatic cancer patient cohort and subsequently evaluate important parameters relating to symptomatology and mortality. This study provides clinicians with updated UK-based survival estimates and symptom prevalence data. Future analysis should clarify high risk ‘clusters’ of concurrent symptoms to aid earlier diagnosis and utilise primary and secondary care data linkage to improve information yield and granularity.

Acknowledgements

This work was supported by a grant from Pancreatic Cancer Action, a registered Charity in England & Wales (1137689) and Scotland (SC049777).

References

- [1] Pancreatic cancer statistics. Cancer Research UK. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer>, Accessed March 10, 2021.
- [2] Schmidt-Hansen M, Berendse S, Hamilton W. Symptoms of Pancreatic Cancer in Primary Care: A Systematic Review. Vol. 45, *Pancreas*. Lippincott Williams and Wilkins; 2016; p. 814–8.
- [3] Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W. The risk of pancreatic cancer in symptomatic patients in primary care: A large case-control study using electronic records. *British Journal of Cancer*. 2012 Jun 5;106(12):1940–4.
- [4] Overview | Suspected cancer: recognition and referral | Guidance | NICE. Nice.org.uk. Available at: <http://www.nice.org.uk/guidance/ng12>, Accessed March 10, 2021.
- [5] Thiru K. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ*. 2003;326(7398):1070-0.
- [6] Hoopes M, Voss R, Angier H, Marino M, Schmidt T, DeVoe J et al. Assessing Cancer History Accuracy in Primary Care Electronic Health Records Through Cancer Registry Linkage. *JNCI: Journal of the National Cancer Institute*. 2020.