medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Pancreatic cancer symptom trajectories from Danish registry data and free text in electronic health records

Jessica Xin Hjaltelin*[1], Sif Ingibergsdóttir Novitski*[1], Isabella Friis Jørgensen[1], Julia Sidenius Johansen[2], Inna M Chen[2], Troels Siggaard[1], Siri Vulpius[1], Lars Juhl Jensen[1], Søren Brunak[1,3]

[1] Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark, Blegdamsvej 3B, 2200 Copenhagen N, Denmark

[2] Department of Oncology, Copenhagen University Hospital - Herlev and Gentofte, Borgmester Ib Juuls Vej 1, 2730 Herlev, Denmark

[3] Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen Ø, Denmark

* joint first authors: jessica.hu@cpr.ku.dk, sif.novitski@cpr.ku.dk

Corresponding author: soren.brunak@cpr.ku.dk

Abstract

Pancreatic cancer is one of the deadliest cancer types with poor treatment options. Better detection of early symptoms and relevant disease correlations could improve pancreatic cancer prognosis. In this retrospective study, we used symptom and disease codes (ICD-10) from the Danish National Patient Registry (NPR) encompassing 8.1 million patients from 1977 to 2018, of whom 22,727 were diagnosed with pancreatic cancer. To complement and compare these diagnosis codes with deeper clinical data, we used a text mining approach to extract symptoms from free text clinical notes in electronic health records (4,418 pancreatic cancer patients and 44,180 controls). We used both data sources to generate and compare symptom disease trajectories to uncover temporal patterns of symptoms prior to pancreatic cancer diagnosis for the same patients. We show that the text mining of the clinical notes was able to capture richer statistically significant symptom patterns, in particular general pain, abdominal pain, and liver-related conditions. We also detected haemorrhages (p-value = $4.80 \cdot 10^{-08}$) and headache (p-value = $2.12 \cdot 10^{-06}$) to be linked as early symptoms of pancreatic cancer. Chaining symptoms together in trajectories identified patients with jaundice conditions having higher median survival (>90 days) compared to patients following trajectories that included haemorrhage, oedema or anaemia (≤90 days). Additionally, we discovered a group of cardiovascular patients that developed pancreatic cancer with a lower median survival (≤90 days). These results provide an overview of two types of pancreatic cancer symptom trajectories. The two approaches and data types complement each other to provide a fuller picture of the early risk factors for pancreatic cancer.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Pancreatic cancer has been predicted to become the second leading cause of cancer deaths, surpassing breast, colorectal, and prostate cancer(Rahib et al. 2021). It has few and generic symptoms resulting in late diagnosis(Kim and Ahuja 2015; Chari et al. 2015) and poor prognosis with a 5-year survival rate of 11%("American Cancer Society" 2020). Hence, improved knowledge of symptoms and diseases occurring early is of high importance to treat this cancer type at a curable stage and provide better prognosis and guide screening programs for pancreatic cancer(Risch et al. 2015). If the cancer is detected at an early stage, where surgical removal of the tumor is possible, the survival rate increases to 42%("American Cancer Society" 2020).

Symptoms of pancreatic cancer are often mistaken for signs of less severe illnesses and overlooked in clinical practice. Some of the most frequent symptoms linked to pancreatic cancer are weight loss, abdominal pain, and anorexia(Hidalgo 2010). Others include upper abdominal pain, cholestasis, nausea(Hidalgo 2010), and dark urine and thirst(Liao et al. 2021). Cholestatic symptoms are more commonly found when the tumor is located in the head of the pancreas(De La Cruz, Young, and Ruffin 2014)(Porta et al. 2005). New-onset diabetes has additionally been found to co-occur with pancreatic cancer when accompanied by weight loss(Yuan et al. 2020),(Hart et al. 2011),(Bruenderman and Martin 2015).

National or regional disease registries hold longitudinal data on disease development. The registries in the Nordic countries are of high quality and among the oldest covering treatment in one-payer health care systems(Laugesen et al. 2021). The National Danish Patient Registry (NPR) contains hospital diagnoses since 1977 and allows for large data-driven studies to detect temporal disease progression patterns relevant in the context of stratified medicine(Jensen et al. 2014; Siggaard et al. 2020). A recent example was the characterization of multimorbidity correlations across cancer types in 0.7 million patients(Hu et al. 2019). However, much of the deeper phenotypic patient information resides within the free text of the electronic health records (EHRs)(Soguero-Ruiz et al. 2016; Delespierre et al. 2017). A small-scale study using 4,080 mixed types of cancers attempted to build more general "event trajectories" using a text mining and a pooled analysis-approach (Jensen et al. 2017). A prospective study investigating initial symptoms and diagnostic interval (time from onset to diagnosis) for known pancreatic cancer symptoms found no difference between pancreatic cancer and patients suspected of having pancreatic cancer(Walter et al. 2016). It is also suggested that symptoms appear sporadically, adding to the complex nature of the disease manifestation(Evans et al. 2014). Other studies used primary care EHRs to detect pancreatic cancer symptoms and found jaundice(Stapley et al. 2012), back pain, lethargy, and new-onset diabetes to be linked to pancreatic cancer(Keane et al. 2014). These studies detected pancreatic cancer symptoms using a single-disease approach, not considering the temporal ordering of symptoms or diseases.

In this paper, we present a large-scale study to investigate pancreatic cancer symptoms longitudinally. We cover all symptoms included in the International Classification of Disease (ICD-10)

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

terminology symptom chapter 18. Additionally, we also include in the text mining vocabulary other known or suggested pancreatic cancer symptoms. We generate and compare disease and symptom trajectories using registry data and clinical notes in EHRs to characterize the temporal ordering of symptoms across data sources.

Results

Extracting patient-level data from the Danish National Patient Registry and free text electronic health records

The Danish National Patient Registry (NPR) data spans the period 1977 to 2018, while the electronic health records (EHRs) used here are from the 2006 to 2016 time-interval. The NPR includes 8,110,702 patients where 22,727 patients are diagnosed with pancreatic cancer (**Table 1**). A subset of 4,418 of these pancreatic cancer patients is included in the EHRs. Almost as many females as males are identified with pancreatic cancer both in NPR and the EHRs (**Table 1**).

Table 1. Data set and patient characteristics

Data set	The National health Registry (NPR)	The free text clinical notes
Year span	1977-2018	2006-2016
N pancreatic pancer patients	22,727	4,418
N controls	8.1M	44,180
Female	11,326 (49.8%)	2,138 (48.4%)
Male	11,401 (50.2%)	2,280 (51.6%)
Mean age at diagnosis (female/male)	73/70	72/70
Counts age at diagnosis		
<40	232	33
40-50	914	161
50-60	2,998	520
60-70	6,165	1,355
70-80	7,564	1,492
>80	4,854	857

We limited the pancreatic cancer patient symptom history in the NPR and clinical notes to five years prior to the diagnosis. The text-based approach was able to identify 132 unique symptoms in the clinical notes, not registered in NPR (**Fig. 1A**). The most frequent symptoms exclusively found in the

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

clinical notes were abnormal blood pressure, anaemias, and conditions related to the intestine (**Fig. 1B**). Within abnormal blood pressure, 1,656 patients had hypertension and 75 patients had low blood pressure. In addition, emotional states like unhappiness and worries were also found (**Fig. 1B**). NPR contained 122 ICD-10 symptoms not found by text mining (**Fig. 1A**). Frequent symptoms exclusively identified in the NPR data comprise abdominal pain, dyspnea, pain localized to the upper abdomen, and abnormal findings on medical images (**Fig. 1C**). Despite the clinical notes data set having much fewer pancreatic cancer patients (4,418 versus 22,727 in NPR), we found a significantly higher number of symptoms in these. In NPR and free text clinical notes, 185 symptoms were identified that occurred in both sources (**Fig. 1A**). Of these 185 symptoms, the top 10 most frequent symptoms from the free text clinical notes and the NPR were compared (**Fig. 1D**). Both sources agreed on six symptoms as being among the top 10 most frequent symptoms for pancreatic cancer patients. These six symptoms were jaundice, abnormalities of breathing, dizziness and giddiness, abdominal and pelvic pain, symptoms and signs concerning food and fluid intake, and pain. From the hierarchical structure of the ICD-10 chapters, different levels of coding detail can be retrieved. In the symptom group "symptom and signs concerning food and fluid intake", the majority of patients represent the subgroups "abnormal weight loss" (R63.4) and anorexia (R63.0). Nausea, vomiting, anorexia, and haemorrhage were frequent symptoms in the free-text clinical notes but have low occurrences in NPR. On the contrary, acute abdomen was frequent in NPR but barely found in the clinical notes (**Fig. 1D**).

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

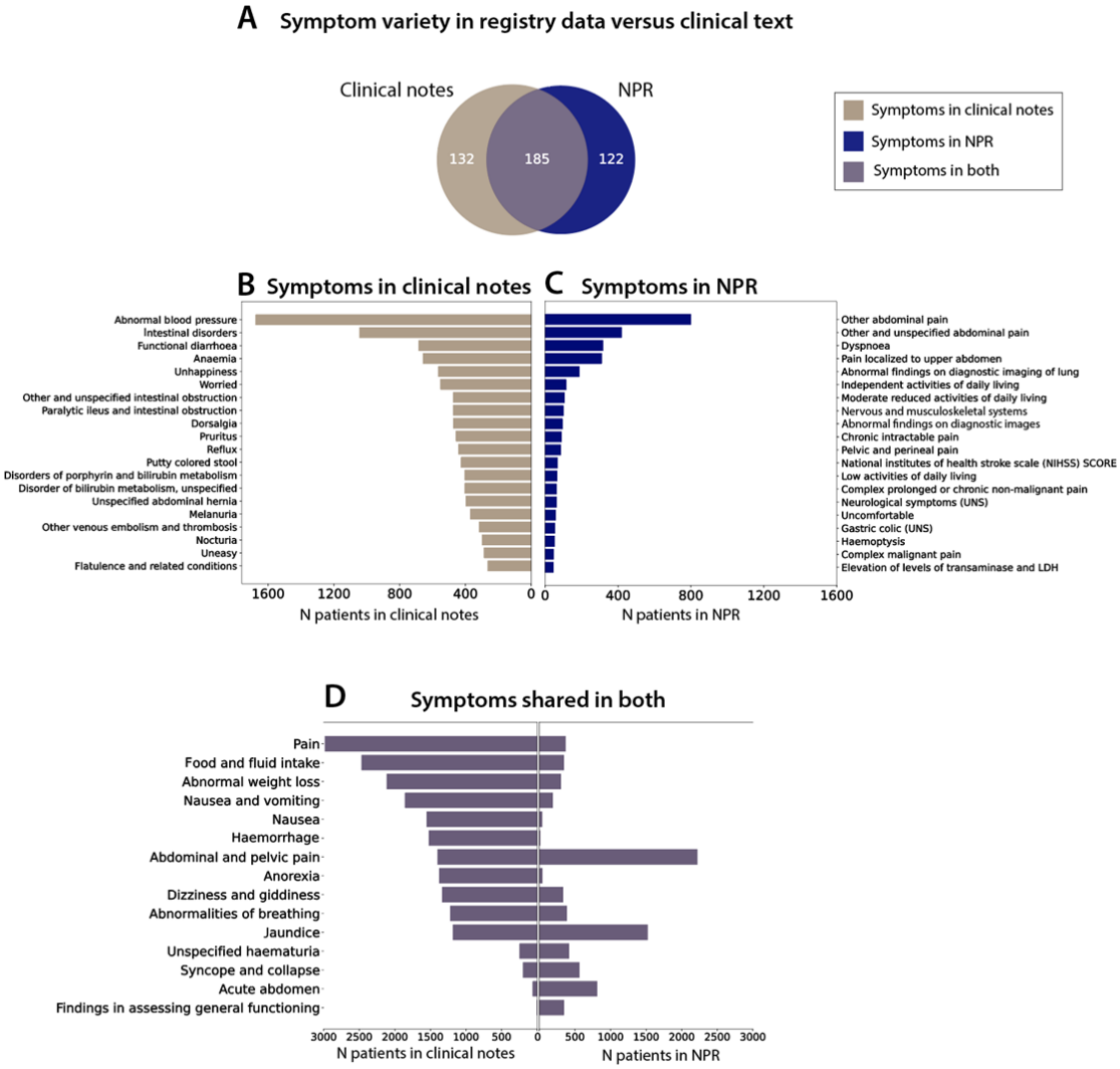

Fig. 1. Comparison of pancreatic cancer symptoms in the Danish National Patient Registry (NPR) and electronic health records (EHRs). (**A**) Symptoms previous to the pancreatic cancer diagnosis identified in NPR (N_{NPR}=122), by text mining of clinical notes from EHRs (N_{notes}=132) or in both data sources (N_{both}=185). (**B**) The top 20 most frequent symptoms that are only found in the clinical notes. (**C**) The top 20 most frequent symptoms only found in NPR. (**D**) The top 15 most frequent symptoms from the clinical notes and NPR, from the list of 185 overlapping symptoms. Some symptom names have been shortened for overview (see Supplementary Table S1).

Text mining validation

In total, 4,141 out of 4,418 (94%) of the pancreatic cancer patients in clinical notes have at least one symptom identified by text mining. A control group of 44,180 patients was generated by matching age and sex. From these, 38,503 patients (87%) had a match for at least one symptom. The performance of the text mining was validated using a test corpus comprising a random extraction of

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

200 clinical notes from 200 different patients. In these notes, a total of 807 symptoms were manually annotated and the text mining method was able to identify 675 correctly and 132 symptoms were not found. This yielded a sensitivity score of 83.4%. A specificity score of 99.8% was obtained since the majority of clinical notes comprise non-symptom words describing the patient's contact with the hospital. Symptoms incorrectly matched to the dictionary constitute a total of 53 words.

Pancreatic cancer trajectories from NPR

Longitudinal disease trajectories were generated for the pancreatic cancer patients, where significant directional diagnosis pairs were joined to represent patients that traverse a complete disease path. The width of the trajectories illustrates the size of a patient group that moves through a particular path and a patient can follow multiple paths. The ICD-10 disease codes from NPR were used to generate disease trajectories (**Fig. 2A, Supplementary Table S2**). ICD-10 chapters, which are associated with pancreatic cancer, are abdominal diseases, cardiovascular diseases, diseases relating to the ears and mastoid process, endocrine, nutritional and metabolic diseases, and the symptom chapter. Most of the codes from the symptoms chapter appear after the diagnosis of pancreatic cancer, except for abdominal and pelvic pain (**Fig. 2A**). This symptom is part of six different trajectories comprising 1,363 patients. One of the trajectory groups is formed by cardiovascular diseases (1,720 patients) including angina pectoris and acute myocardial infarction that traverse to angina pectoris, chronic ischemic heart disease, type 2 diabetes or heart failure and afterwards into malignant neoplasm of the pancreas. The patients following these trajectories have the shortest survival (median survival ≤90 days) compared to the patients following the other trajectories. Other disease trajectories that show short survival are the cataract trajectory and the gonarthrosis-hypertension trajectory, comprising 321 and 332 patients, respectively. The rest of the trajectories has median survival above >90 days and includes for example type 2 diabetes (780 patients) or abdominal pain (1,363 patients) before pancreatic cancer. Survival in months for the patients following a disease trajectory is shown in **Supplementary Fig. S1B**.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

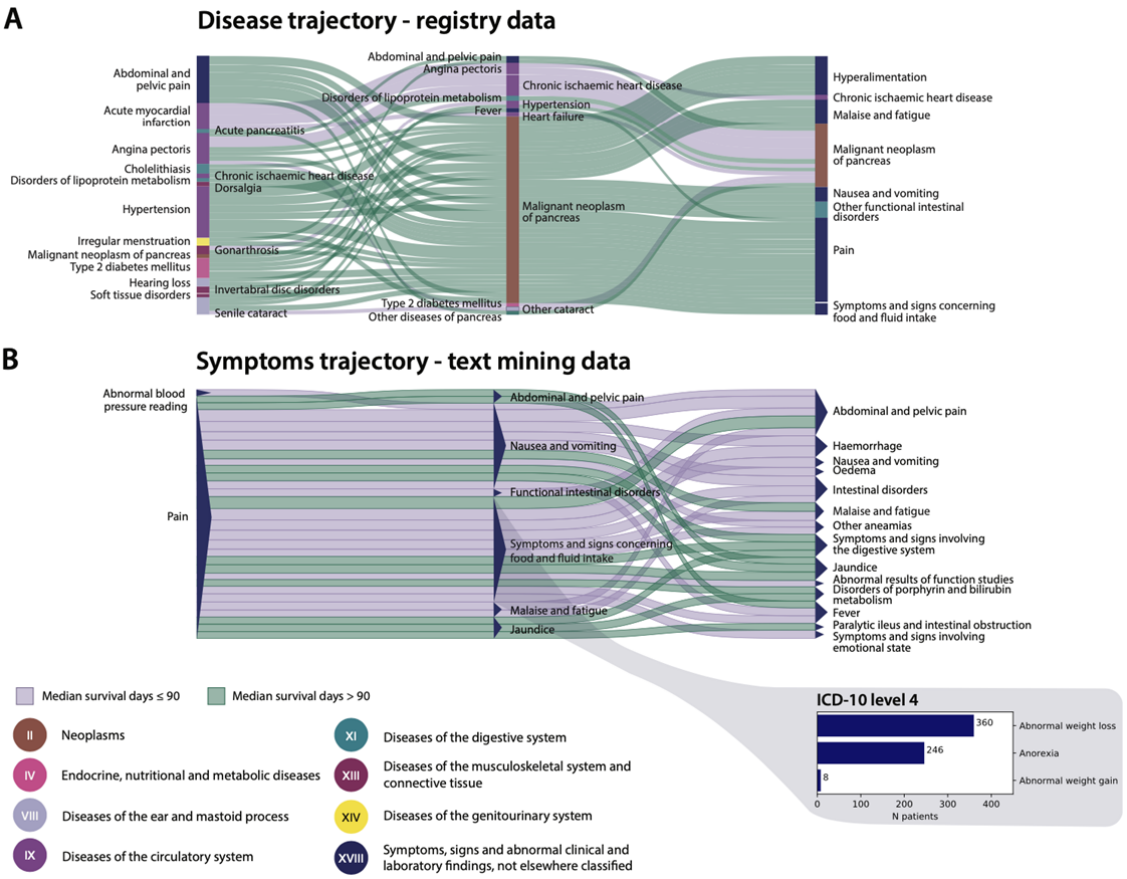

Fig. 2. Disease and symptoms trajectories before and after pancreatic cancer diagnosis. (A) The registry disease trajectories consist of significant disease pairs with a Relative Risk (RR) > 1 (Supplementary Table S1). Each trajectory has a minimum of 300 patients. **(B)** Symptom trajectories from clinical notes consisting of significant disease pairs with RR > 1 (Supplementary Table S3). Each trajectory has a minimum of 100 patients. The width of the trajectories indicates visually the actual number of patients. The purple-colored trajectories represent patient groups with median survival ≤90 days. The green-colored trajectories represent patient groups with median survival >90. Disease and symptoms in the trajectories are colored by the ICD-10 chapters. Some symptoms names have been shortened for overview.

Pancreatic cancer trajectories from clinical notes

The symptom trajectories were generated for the pancreatic cancer patients with available clinical notes 5-years prior to the diagnosis (**Fig. 2B**). All trajectories shown are significant in terms of direction with at least 100 patients following the complete path of three symptoms. A patient can follow several symptom paths. The majority of the trajectories begin with pain including 1,085 patients. From the pain symptom, a large group traverse into symptoms and signs concerning food and fluid intake (N=614). Of these 614 patients, 360 have abnormal weight loss, 246 have anorexia, and eight have abnormal weight gain. Several trajectories were also identified with pain that traverse into symptoms of nausea and vomiting or jaundice. At the end of the three symptom trajectories, 376 patients end up with abdominal pain, 272 patients with haemorrhage, 261 with

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

196 functional intestinal disorders, and 222 with jaundice. All the patients following trajectories that end
197 up with haemorrhage, oedema, and anemia have short survival (median survival ≤90 days) opposed
198 to for example disorders of porphyrin and bilirubin metabolism and jaundice (median survival >90
199 days). See **Supplementary Table S3** for trajectory details. The survival of the patients following the
200 symptom trajectories are presented in **Supplementary Fig. S1A.**
201
202
203 **Temporality of symptoms from clinical notes**
204
205 The distribution of symptoms registered over time can be seen for the 20 most frequent symptoms
206 that occur significantly more often in the pancreatic cancer patients opposed to the control group
207 (**Fig. 3**). The distribution is over a five-year period previous to the pancreatic cancer diagnosis. If a
208 symptom is registered multiple times in one hospital encounter it is included once. Otherwise, all
209 occurrences of a symptom during the five-year period are included for a patient. Some of the most
210 frequent symptoms identified are symptoms related to pain, weight loss, and jaundice. Additionally,
211 haemorrhage (p-value = $4.80 \cdot 10^{-08}$) and headache (p-value = $2.12 \cdot 10^{-06}$) were frequent and significant
212 among pancreatic cancer patients. **Supplementary Table S4** shows the number of patients and all p-
213 values relating to these symptoms.
214
215 Symptoms such as jaundice, anorexia, weight loss, and abdominal conditions were observed only
216 closer to the pancreatic cancer diagnosis and the median of these appear within 6 months before
217 pancreatic cancer diagnosis. The more general symptoms such as pain, abnormalities of breathing,
218 abnormal blood pressure, abnormalities of heart beat, and headache have a median higher than 6
219 months before diagnosis. Most of the symptoms can be identified earlier in the clinical notes for
220 patients with short survival (≤90 days) opposed to patients with longer survival (> 90 days).
221

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

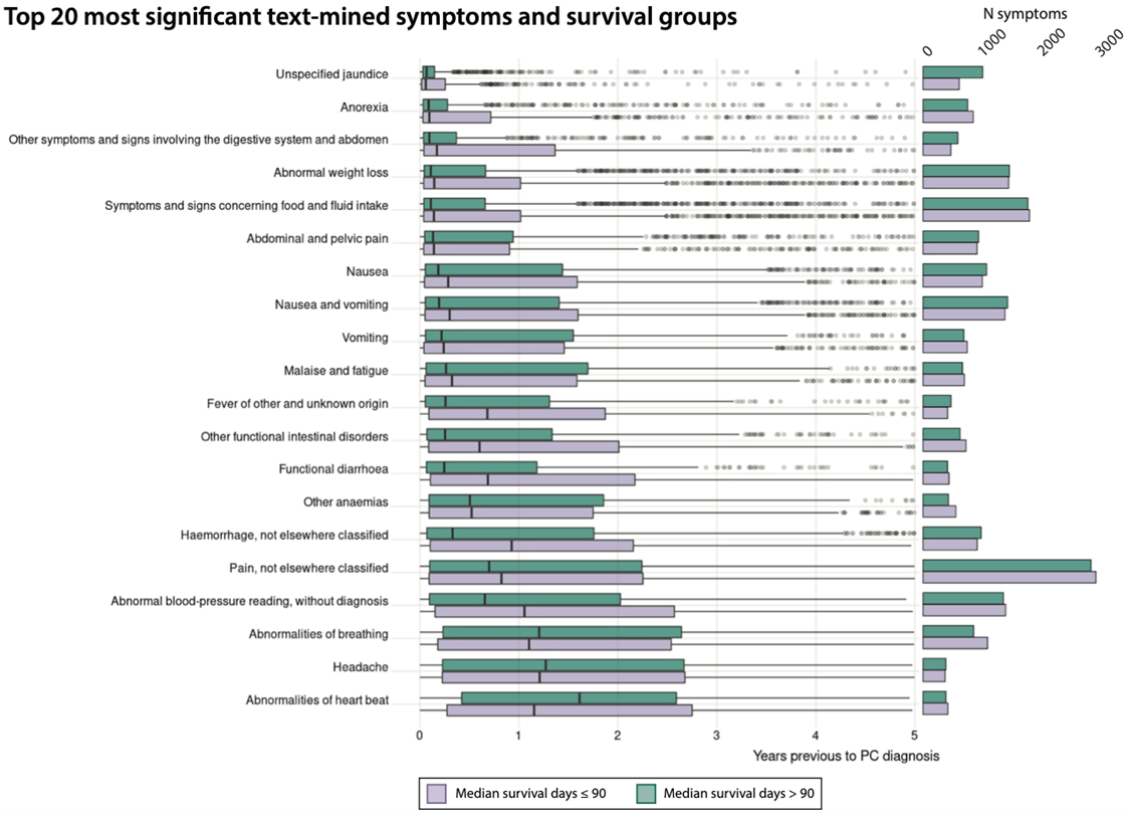

Fig. 3. Top 20 most frequent text mined symptoms from the clinical notes five years previous to pancreatic cancer diagnosis. The top 20 most common and significant (P<0.05) symptoms in the text mined clinical notes are shown with survival information and time to pancreatic cancer diagnosis (Supplementary Table S4 and S5). The symptoms are extracted over a five-year period up to the time of pancreatic cancer (PC) diagnosis. If a symptom is noted more than once in one hospital encounter, the symptom is counted once. The purple bars indicate patients with survival ≤90 days and the green bars indicate patients with a survival >90 days. Symptom names may have been shortened for overview (see Supplementary Table S1).

Discussion

This study uncovered statistically significant disease and symptom trajectories prior to pancreatic cancer that may be further assessed as early risk factors for pancreatic cancer for screening purposes. We complemented pancreatic cancer symptoms from hospital diagnosis codes with symptoms extracted by a text mining approach using free text in EHRs. From this, we discovered that symptoms were more abundant in the EHRs opposed to the NPR. The high variation in symptom frequency between the two sources was astonishing since the clinical notes were only available on a subset of the pancreatic cancer patients (4,418 out of 22,727). For example, more than 2,000 patients out of 4,418 had weight loss according to the free text clinical notes but in the registry, it was less than 500 patients out of 22,727. On the contrary, there were symptoms identified in the NPR that were not identified in the clinical notes which could be explained by the higher number of cases over a larger period of time in the NPR data set.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

The most common pancreatic cancer symptoms such as abdominal pain and weight loss (Porta et al. 2005) as well as jaundice(Walter et al. 2016) were detected in both the registry and the text mining approach. Another study found that 21% of pancreatic cancer patients had dyspnoea(Krech and Walsh 1991), which agrees with our findings in relation to abnormalities of breathing. However, we discovered that haemorrhage and headache were frequent and significant in this cohort which is usually not described as a typical symptom of pancreatic cancer. In the clinical notes, more than a third of the patients had haemorrhages and the incidence is reported to be only 3%-12% for advanced cancers(Harris and Noble 2009). Additionally, gastrointestinal bleeding has been found to occur rarely among pancreatic cancer patients(Lee et al. 1994; Richardson and Baldeo 2016). Headache has been found to occur in 11% of advanced cancer patients which fits roughly with our results(Walsh, Donnelly, and Rybicki 2000), but is usually not described in relation to pancreatic cancer. Treatments or other comorbidities are entities that could be considered in relation to these findings to check if it is caused by the cancer, external source or by other confounders.

We found that for patients with short survival (≤90 days), symptoms could be tracked further back in the clinical notes, where patients with longer survival showed symptoms appearing closer to diagnosis. This could indicate that the pancreatic cancer had metastasized further in the patients with short survival. Short survival is commonly defined in oncology studies as death within 90 days(Sgouros and Maraveyas 2008). Using this threshold could be a strict choice to distinguish patient groups and it is important to interpret survival of each patient group exactly.

Tables S2 and S3 provide an overview of all the trajectories identified in our study including how many patients follow them and the median survival of each trajectory group. Symptom trajectories with a shorter median survival (≤90 days) included anaemia, haemorrhages, and oedema which is related to advanced cancers(Caro et al. 2001), (Pereira and Phan 2004), (Tai et al. 2016). Jaundice was coupled to higher survival (>90 days), since in the case of pancreatic cancer, it contributes as a clearer cancer symptom compared to the other more generic pancreatic cancer symptoms(Strasberg et al. 2014). Another study found the diagnostic interval to be shorter for jaundice compared to for example weight loss for pancreatic cancer patients(Walter et al. 2016), (Gobbi et al. 2013). It could therefore indicate a faster diagnosis but not necessarily a longer survival. We additionally identified patients with heart diseases and a lower survival that developed pancreatic cancer and it has been suggested that thromboembolic events could play a role in pancreatic cancer(Bergqvist et al. 2006) (Bertero et al. 2018). Cardiovascular diseases have been indicated as a risk factor for certain cancers(Lau et al. 2021), and contrariwise it is discussed if cancer should be included in cardiovascular risk prediction tools(Blaes and Shenoy 2019). Though the aim in our study was to identify patient groups that follow specific symptom trajectories there seem to be no prominent groups that stand out. All patients begin with general symptoms such as pain or abnormal blood pressure and afterwards more cancer specific symptoms such as weight loss, nausea, and jaundice appear.

One limitation of this study is that our data is strictly hospital contacts, which may result in a set of identified symptoms and diagnoses occurring not at the earliest disease stage. By including data from general practitioners, future studies could potentially identify even earlier symptom patterns prior to pancreatic cancer. Causal patterns reflected from trajectories can be difficult to interpret for

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

example if an event serves as an actual cause or is the result of confounders(Jensen et al. 2014). Confounders could for example be medication or other risk factors that are not accounted for in the analyses. In this type of trajectory study, we can only assess the association between diseases, but actual causal relationships will need to be validated in future studies. Furthermore, the available period for the clinical notes (2006-2016) was shorter compared to the NPR (1977-2018). A direct comparison between the two data sources can be challenging, since registry terms are not necessarily written as in the free text clinical notes. One example is acute abdomen, which describes a condition with severe abdominal pain that demands immediate medical attention. A clinical note might contain text that the patient was admitted with severe stomach pain but register it as the ICD-10 code "R10.0 - Acute abdomen" in NPR which essentially would mean the same. Also, information bias may exist in the coding procedure within NPR and it might not always be trivial to code the correct symptom or diagnosis(Schmidt et al. 2015), (Lynge, Sandegaard, and Rebolj 2011). From our study, we show that the inclusion of symptoms text mined from clinical notes largely complemented ICD-coded symptoms from the patient registry.

This study showed that deep phenotypical information stored in registry data and free text EHRs can be useful for detecting temporal patterns. The sequence of events that leads up till the pancreatic cancer diagnosis supports that symptoms may appear in a messy and complex order. Patients start experiencing general and unspecific symptoms, which then become more specific and severe as the cancer advances. The methodology presented here can be used for external data sets and may, when validated, serve as a clinical tool to stratify patient groups so the correct option of patient care can be offered at a personalized level.

Methods

Study design

The health data used was from the Danish NPR and from clinical notes. For the generation of the trajectories, we used disease codes from the hierarchical International Classification of Diseases (ICD) system. ICD version 10 codes at level 3 were used for this study. The case cohort of pancreatic cancer patients were defined based on the pancreatic cancer code C25. If patients were registered with another cancer type before pancreatic cancer, they were not included as cases.

Patient cohorts

From the Danish NPR, all hospital encounters in Denmark during the period 1977 to 2018 were used in our analyses, comprising 235,648,697 encounters for 8,172,661 patients. The ICD-10 Chapter 18 (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified) was used as vocabularies for extraction of symptoms from both NPR and the clinical notes. The EHR data comprising clinical notes included 1,906,769 patients from the Capital Region of Denmark and

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Region Zealand from the period of 2006 to 2016 with a total of 7,831,920 unique clinical notes. A clinical note contains information regarding a hospital encounter for a patient such as the reason for admission, findings, symptoms, operations, and treatments. Only patients with at least one clinical note within a five-year time interval before diagnosis were included. A control group was sampled for the patients having clinical notes using the same age and sex distribution with 10 control patients per pancreatic cancer patient and the same filtering criteria.
The registries could be linked through the Central Personal Register (CPR) number, which all Danish citizens possess.

Preprocessing of patient cohorts

Status codes 01 and 90 have been included in the analysis, excluding the rest of the status codes to make sure only active residents in Denmark (01) and inactive dead (90) are a part of the cohort. Furthermore, diagnosis types H and M (referral and temporary diagnosis) were excluded from the data for the extraction of disease and symptom frequencies and for the construction of trajectories. This was to ensure that the analysis is based on the main biologically relevant diagnoses in NPR.

Text mining clinical notes

A dictionary was constructed by including all symptoms from the ICD-10 symptom chapter 18. Known pancreatic cancer symptoms that were not already a part of chapter 18, were added to the vocabulary. The final symptom dictionary consisted of 691 symptoms. The symptoms were initially written in their singular forms and afterwards suffixes were added to ensure different variants of a symptom such as plural forms. Other word endings like “condition” or “symptom” were also added since these are sometimes put behind a symptom in Danish (**Fig. 4**). Afterwards, the extended dictionary and the clinical notes from the patients and the control group were tokenized. The unique tokens from each were then compared to extract spelling errors on words longer or equal to five characters. In order to extract words with spelling errors, the tokens from the dictionary were fuzzy matched against the tokens from the clinical notes using the Python package fuzzysearch(Einat 2020).

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

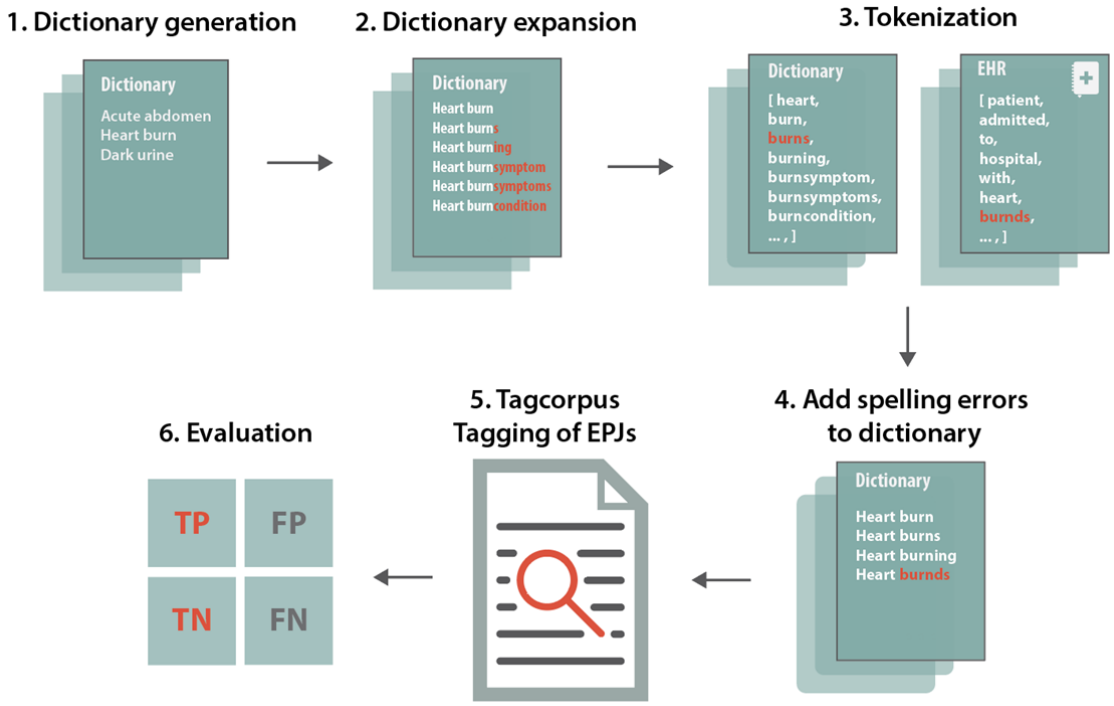

Fig. 4. Text mining pipeline. A dictionary was generated with symptoms and expanded with word endings to capture multiple forms of the same symptom. Afterwards, the dictionary and the corpus (clinical notes) were tokenized in order to extract spelling errors. The spelling errors were then added to the dictionary. Finally, the program tagcorpus (Pafilis et al. 2013) was used to tag the symptoms in the corpus. The text mining performance was then evaluated.

The metric to measure similarities was Levenshtein's distance. In this case, a maximum value of distance between two words was set to 1 to allow for either one substitution, one deletion or one addition of a character to the word. When the fuzzy matching was complete, all outcomes were further assessed and rechecked so that wrong matches could be removed.

After the extraction of variations for the synonyms, the program Tagcorpus(Pafilis et al. 2013) was used to tag the corpus (the clinical notes) for the pancreatic cancer patients and the control patients. It is a fast program written in C++, which for example is able to process thousands of PubMed abstracts per second((Pafilis and Jensen 2016), (L. J. Jensen 2016),(Pafilis et al. 2013)). A post-processing step was applied that removed sentences with negations and mentioning of other persons than the patient(Eriksson et al. 2014). This ensures that whenever there is a negation word in a sentence the word will not be tagged. For example, not, never, and no are a part of the negation dictionary. From the other person's filter, the symptom will not be tagged if a sentence contains for example mother, brother etc. since it might relate to another person and not to the patient in question. The evaluation of the text mining was done using a confusion matrix of 200 randomly selected clinical notes. These were checked manually for incorrectly and correctly matched symptoms. Afterwards, the metrics sensitivity and specificity were calculated. The sensitivity is defined as

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$= \frac{}{+} \quad (1)$$

where TP are the true positives and FN the false negatives. The sensitivity describes the fraction of correctly matched symptoms opposed to all symptoms. Specificity is the fraction of correctly matched words that are not symptoms opposed to all the non-symptom words in the clinical notes and is defined by

$$= \frac{}{+} \quad (2)$$

Generating pancreas cancer trajectories

We calculated significant disease trajectories for the population of pancreatic cancer patients using the methodology from Jensen *et al.(A. B. Jensen et al. 2014)*. Here, the Relative Risk (RR) was calculated (**Eq. 3**) for the strength of an association between two diseases in the exposed group compared to a comparison group. The exposed group was matched with the same age and sex group as the comparison group and seasonal changes were accounted for by taking samples from the comparison group disease discharge to have the same week as the disease 1 (D1) discharge in the exposed group. The count for the exposed group is denoted $C_{exposed}$ and the corresponding i'th comparison group as C_i where i ϵ {1,..,N} and N is the number of comparison groups.

$$= \frac{}{\frac{1}{}\sum_i \quad C_i} \quad (3)$$

P-values for the RR were calculated using binomial tests, where the average probability of sampling a control patient with D2 was compared with $C_{exposed}$. Afterwards, the diagnosis pairs (D1, D2) were tested for directionality with binomial tests to decide if the direction is significant for the significant disease pairs found. The patients having the direction (D1→ D2) or the other direction (D2→ D1) were counted. Patients having D1 and D2 at the same time were also counted and the total count constitutes N samples with 50 % probability of having either one of the directions. The p-values were afterwards Bonferroni corrected.

To find significant text mined symptoms, we sampled 10 control patients for every pancreatic cancer case patient. The 44,418 control patients were stratified based on sex, birth year and age at diagnosis. We ensured that the control patients were diagnosed with another disease at the same age as the cases and extracted their clinical notes five years previous to that diagnosis. This was done to make sure the notes for the control and the case group originated at a similar time period during their lifetime and up to a diagnosis.

For the construction of the symptom trajectories the RR was calculated using the counts for the co-occurrence of all possible symptom pair (See **Table 2**) and the χ^2 was used to test the significance of the co-occurrence.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 2. Relative risk for symptoms pair

Relative Risk	$C_{Exposed}$	C_i
S_1S_2	N patients in $C_{Exposed}$ with S_1 and S_2	N patients in C_i with S_1 and S_2
Not S_1S_2	N patients in $C_{Exposed}$ without S_1 and S_2	N patients in C_i without S_1 and S_2

Subsequently, each symptom pair, with RR>1 and a significant p-value < 0.05, was tested for directionality and Bonferroni-corrected using the method from Jensen *et al.(A. B. Jensen et al. 2014)*. For the text mined symptoms, the day of admission was used as the time registered for the symptom, since capturing the symptom as early as possible is crucial for the purpose of this study.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Approvals

The work was approved as a registry study that does not require ethical permissions in Denmark as well as patient consent. Access to the data was approved by the Danish Data Protection Agency (ref: 514-0255/18-3000, 514-0254/18-3000, SUND-2016-50), the Danish Health Data Authority (ref: FSEID-00003724 and FSEID-00003092), and the Danish Patient Safety Authority (3-3013-1731/1/).

Data availability

Permission to access the person-sensitive data used for this study can be obtained through the Danish Data Protection Agency, the Danish Health Authority, and the Danish Health regions (Capital Region and Region Zealand). As the raw electronic patient records and registry information are individual-level data they are person sensitive and cannot be made publicly available but only analyzed in closed, secure environments. In the paper we have only provided diagnosis and co-occurrence information when grouped to at least five patients. All patient information published here is non-person-sensitive summary level data and have been shared in the Supplementary materials.

Code availability

The methodology and the data analysis have been carried out using Python software (version 3.8) and R version 3.6. The code for the text mining method is available at https://github.com/larsjuhljensen/tagger. The key algorithm for creation of disease trajectories have been described in details in the published studies Jensen *et al.*(Jensen et al. 2014) and *Siggaard et al.*(Siggaard et al. 2020).

Acknowledgements

We would like to acknowledge funding from the Novo Nordisk Foundation (grant agreements NNF14CC0001 and NNF17OC0027594), the Braindrugs (R279-2018-1145) as well as the Danish Innovation Fund (5184-00102B).

Author contributions

J.X.H. and SB conceptualized the study design and supervised the project. J.X.H. and S.I.N. ran the analyses together and wrote the manuscript draft together. T.S. helped run the text mining analysis. I.F.J. contributed in the decision making of the analysis and S.V. validated the text mining and translated the ICD-10 symptom chapter. J.S.J. and I.M.C. contributed to clinical expertise on pancreatic cancer oncology in Denmark. SB edited the draft version that was approved by all authors.

Competing interests

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

S.B. reports ownerships in Intomics A/S (now acquired by ZS Inc.), Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, ALK-Abello A/S and managing board memberships in Proscion A/S and Intomics A/S outside the submitted work. I.M.C. reported receiving research funding and hotel/airfare reimbursement to attend global health meetings from Roche, BMS, Celgene, Genis, and an advisory relationship with Amgen and AstraZeneca. L.J.J. reports ownerships in Amgen Inc, AstraZeneca PLC and Novo Nordisk A/S. All other authors declare no competing interests.

References

"American Cancer Society." 2020. *Cancer Facts & Figures 2020. Atlanta, Ga: American Cancer Society*.

Bergqvist, David, Karin Wåhlander, Henry Eriksson, Nils Sternby, and Mats Ögren. 2006. "Trousseau's Syndrome – What Is the Evidence?" *Thrombosis and Haemostasis*. https://doi.org/10.1160/th05-10-0694.

Bertero, Edoardo, Marco Canepa, Christoph Maack, and Pietro Ameri. 2018. "Linking Heart Failure to Cancer: Background Evidence and Research Perspectives." *Circulation* 138 (7): 735–42.

Blaes, Anne H., and Chetan Shenoy. 2019. "Is It Time to Include Cancer in Cardiovascular Risk Prediction Tools?" *The Lancet*.

Bruenderman, Elizabeth, and Robert C. G. Martin 2nd. 2015. "A Cost Analysis of a Pancreatic Cancer Screening Protocol in High-Risk Populations." *American Journal of Surgery* 210 (3): 409–16.

Caro, J. Jaime, J. Jaime Caro, Maribel Salas, Alexandra Ward, and Glenwood Goss. 2001. "Anemia as an Independent Prognostic Factor for Survival in Patients with Cancer." *Cancer*. https://doi.org/10.1002/1097-0142(20010615)91:12<2214::aid-cncr1251>3.0.co;2-p.

Chari, Suresh T., Kimberly Kelly, Michael A. Hollingsworth, Sarah P. Thayer, David A. Ahlquist, Dana K. Andersen, Surinder K. Batra, et al. 2015. "Early Detection of Sporadic Pancreatic Cancer: Summative Review." *Pancreas* 44 (5): 693–712.

De La Cruz, Maria Syl D., Alisa P. Young, and Mack T. Ruffin. 2014. "Diagnosis and Management of Pancreatic Cancer." *American Family Physician* 89 (8): 626–32.

Delespierre, T., P. Denormandie, A. Bar-Hen, and L. Josseran. 2017. "Empirical Advances with Text Mining of Electronic Health Records." *BMC Medical Informatics and Decision Making* 17 (1): 127.

Einat, Tal. 2020. "Fuzzysearch." *GitHub Repository*, June. https://github.com/taleinat/fuzzysearch.

Eriksson, Robert, Thomas Werge, Lars Juhl Jensen, and Søren Brunak. 2014. "Dose-Specific Adverse Drug Reaction Identification in Electronic Patient Records: Temporal Data Mining in an Inpatient Psychiatric Population." *Drug Safety*. https://doi.org/10.1007/s40264-014-0145-z.

Evans, Julie, Alison Chapple, Helen Salisbury, Pippa Corrie, and Sue Ziebland. 2014. "'It Can't Be Very Important Because It Comes and Goes'--Patients' Accounts of Intermittent Symptoms Preceding a Pancreatic Cancer Diagnosis: A Qualitative Study." *BMJ Open* 4 (2): e004215.

Gobbi, Paolo G., Manuela Bergonzi, Mario Comelli, Lara Villano, Donatella Pozzoli, Alessandro Vanoli, and Paolo Dionigi. 2013. "The Prognostic Role of Time to Diagnosis and Presenting Symptoms in Patients with Pancreatic Cancer." *Cancer Epidemiology* 37 (2): 186–90.

Harris, Dylan G., and Simon I. R. Noble. 2009. "Management of Terminal Hemorrhage in Patients With Advanced Cancer: A Systematic Literature Review." *Journal of Pain and Symptom Management*. https://doi.org/10.1016/j.jpainsymman.2009.04.027.

Hart, Phil A., Pratima Kamada, Kari G. Rabe, Sunil Srinivasan, Ananda Basu, Gaurav Aggarwal, and Suresh T. Chari. 2011. "Weight Loss Precedes Cancer-Specific Symptoms in Pancreatic Cancer-Associated Diabetes Mellitus." *Pancreas* 40 (5): 768–72.

Hidalgo, Manuel. 2010. "Pancreatic Cancer." *The New England Journal of Medicine* 362 (17): 1605–17.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Hu, Jessica X., Marie Helleberg, Anders B. Jensen, Søren Brunak, and Jens Lundgren. 2019. "A Large-Cohort, Longitudinal Study Determines Precancer Disease Routes across Different Cancer Types." *Cancer Research* 79 (4): 864–72.

Jensen, Anders Boeck, Pope L. Moseley, Tudor I. Oprea, Sabrina Gade Ellesøe, Robert Eriksson, Henriette Schmock, Peter Bjødstrup Jensen, Lars Juhl Jensen, and Søren Brunak. 2014. "Temporal Disease Trajectories Condensed from Population-Wide Registry Data Covering 6.2 Million Patients." *Nature Communications* 5 (June): 4022.

Jensen, Kasper, Cristina Soguero-Ruiz, Karl Oyvind Mikalsen, Rolv-Ole Lindsetmo, Irene Kouskoumvekaki, Mark Girolami, Stein Olav Skrovseth, and Knut Magne Augestad. 2017. "Analysis of Free Text in Electronic Health Records for Identification of Cancer Patient Trajectories." *Scientific Reports* 7 (April): 46226.

Jensen, Lars Juhl. 2016. "One Tagger, Many Uses: Illustrating the Power of Ontologies in Dictionary-Based Named Entity Recognition." https://doi.org/10.1101/067132.

Keane, M. G., L. Horsfall, G. Rait, and S. P. Pereira. 2014. "A Case-Control Study Comparing the Incidence of Early Symptoms in Pancreatic and Biliary Tract Cancer." *BMJ Open* 4 (11): e005720.

Kim, Victoria M., and Nita Ahuja. 2015. "Early Detection of Pancreatic Cancer." *Chinese Journal of Cancer Research = Chung-Kuo Yen Cheng Yen Chiu* 27 (4): 321–31.

Krech, R. L., and D. Walsh. 1991. "Symptoms of Pancreatic Cancer." *Journal of Pain and Symptom Management* 6 (6): 360–67.

Lau, Emily S., Samantha M. Paniagua, Elizabeth Liu, Manol Jovani, Shawn X. Li, Katherine Takvorian, Navin Suthahar, et al. 2021. "Cardiovascular Risk Factors Are Associated with Future Cancer." *JACC. CardioOncology* 3 (1): 48–58.

Laugesen, Kristina, Jonas F. Ludvigsson, Morten Schmidt, Mika Gissler, Unnur Anna Valdimarsdottir, Astrid Lunde, and Henrik Toft Sørensen. 2021. "Nordic Health Registry-Based Research: A Review of Health Care Systems and Key Registries." *Clinical Epidemiology* 13 (July): 533–54.

Lynge, Elsebeth, Jakob Lynge Sandegaard, and Matejka Rebolj. 2011. "The Danish National Patient Register." *Scandinavian Journal of Public Health* 39 (7 Suppl): 30–33.

Pafilis, Evangelos, Sune P. Frankild, Lucia Fanini, Sarah Faulwetter, Christina Pavloudi, Aikaterini Vasileiadou, Christos Arvanitidis, and Lars Juhl Jensen. 2013. "The SPECIES and ORGANISMS Resources for Fast and Accurate Identification of Taxonomic Names in Text." *PloS One* 8 (6): e65390.

Pafilis, Evangelos, and Lars Juhl Jensen. 2016. "Real-Time Tagging of Biomedical Entities." https://doi.org/10.1101/078469.

Pereira, Jose, and Tien Phan. 2004. "Management of Bleeding in Patients with Advanced Cancer." *The Oncologist*. https://doi.org/10.1634/theoncologist.9-5-561.

Porta, Miquel, Xavier Fabregat, Núria Malats, Luisa Guarner, Alfredo Carrato, Ana de Miguel, Laura Ruiz, et al. 2005. "Exocrine Pancreatic Cancer: Symptoms at Presentation and Their Relation to Tumour Site and Stage." *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 7 (5): 189–97.

Rahib, Lola, Mackenzie R. Wehner, Lynn M. Matrisian, and Kevin T. Nead. 2021. "Estimated Projection of US Cancer Incidence and Death to 2040." *JAMA Network Open* 4 (4): e214708.

Risch, H. A., H. Yu, L. Lu, and M. S. Kidd. 2015. "Detectable Symptomatology Preceding the Diagnosis of Pancreatic Cancer and Absolute Risk of Pancreatic Cancer Diagnosis." *American Journal of Epidemiology*. https://doi.org/10.1093/aje/kwv026.

Schmidt, Morten, Sigrun Alba Johannesdottir Schmidt, Jakob Lynge Sandegaard, Vera Ehrenstein, Lars Pedersen, and Henrik Toft Sørensen. 2015. "The Danish National Patient Registry: A Review of Content, Data Quality, and Research Potential." *Clinical Epidemiology* 7 (November): 449–90.

Sgouros, Joseph, and Anthony Maraveyas. 2008. "Excess Premature (3-Month) Mortality in Advanced Pancreatic Cancer Could Be Related to Fatal Vascular Thromboembolic Events. A Hypothesis Based on a Systematic Review of Phase III Chemotherapy Studies in Advanced Pancreatic Cancer." *Acta Oncologica* 47 (3): 337–46.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Siggaard, Troels, Roc Reguant, Isabella F. Jørgensen, Amalie D. Haue, Mette Lademann, Alejandro Aguayo-Orozco, Jessica X. Hjaltelin, Anders Boeck Jensen, Karina Banasik, and Søren Brunak. 2020. "Disease Trajectory Browser for Exploring Temporal, Population-Wide Disease Progression Patterns in 7.2 Million Danish Patients." *Nature Communications* 11 (1): 4952.

Soguero-Ruiz, Cristina, Kristian Hindberg, Inmaculada Mora-Jiménez, José Luis Rojo-Álvarez, Stein Olav Skrøvseth, Fred Godtliebsen, Kim Mortensen, et al. 2016. "Predicting Colorectal Surgical Complications Using Heterogeneous Clinical Data and Kernel Methods." *Journal of Biomedical Informatics*. https://doi.org/10.1016/j.jbi.2016.03.008.

Stapley, S., T. J. Peters, R. D. Neal, P. W. Rose, F. M. Walter, and W. Hamilton. 2012. "The Risk of Pancreatic Cancer in Symptomatic Patients in Primary Care: A Large Case-Control Study Using Electronic Records." *British Journal of Cancer* 106 (12): 1940–44.

Tai, Shu-Yu, Chung-Yin Lee, Chien-Yi Wu, Hui-Ya Hsieh, Joh-Jong Huang, Chia-Tsuan Huang, and Chen-Yu Chien. 2016. "Symptom Severity of Patients with Advanced Cancer in Palliative Care Unit: Longitudinal Assessments of Symptoms Improvement." *BMC Palliative Care* 15 (March): 32.

Walsh, D., S. Donnelly, and L. Rybicki. 2000. "The Symptoms of Advanced Cancer: Relationship to Age, Gender, and Performance Status in 1,000 Patients." *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer* 8 (3): 175–79.

Walter, Fiona M., Katie Mills, Silvia C. Mendonça, Gary A. Abel, Bristi Basu, Nick Carroll, Sue Ballard, et al. 2016. "Symptoms and Patient Factors Associated with Diagnostic Intervals for Pancreatic Cancer (SYMPTOM Pancreatic Study): A Prospective Cohort Study." *The Lancet. Gastroenterology & Hepatology* 1 (4): 298–306.

Yuan, Chen, Ana Babic, Natalia Khalaf, Jonathan A. Nowak, Lauren K. Brais, Douglas A. Rubinson, Kimmie Ng, et al. 2020. "Diabetes, Weight Change, and Pancreatic Cancer Risk." *JAMA Oncology* 6 (10): e202948.Bruenderman, E., & Martin, R. C. G. (2015). A cost analysis of a pancreatic cancer screening protocol in high-risk populations. *The American Journal of Surgery*, *210*(3), 409–416. https://doi.org/10.1016/J.AMJSURG.2014.11.017

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary Tables

Supplementary Table S1

Complete names of ICD-10 codes that has been shortened for overview

Abbreviation	Complete name
Nervous and musculoskeletal systems	Other and unspecified symptoms and signs involving the nervous and musculoskeletal systems
Abnormal findings on diagnostic images	Abnormal findings on diagnostic images of other body structures
Disorders of lipoprotein metabolism	Disorders of lipoprotein metabolism and other lipidaemias
Hypertension	Essential primary hypertension
Irregular menstruation	Excessive, frequent and irregular menstruation
Hearing loss	Other hearing loss
Digestive system and abdomen	Systems and signs involving digestive system and abdomen
Fever	Fever of other and unknown origin
Hyperalimentation	Other hyperalimentation
Pain	Pain, not elsewhere classified
Intestinal disorders	Other functional intestinal disorders
Oedema	Oedema, not elsewhere classified
Anaemia	Other anaemias
Jaundice	Unspecified jaundice
Paralytic ileus and intestinal obstruction	Paralytic ileus and intestinal obstruction without hernia
Haemorrhage	Haemorrhage, not elsewhere classified
Intestinal disorders	Other functional intestinal disorders
Abnormal blood pressure	Abnormal blood-pressure reading, without diagnosis
Soft tissue disorders	Other soft tissue disorders, not elsewhere classified
Elevation of levels of transaminase and LDH	Elevation of levels of transaminase and lactic acid dehydrogenase (LDH)
Food and fluid intake	Symptoms and signs concerning food and fluid intake

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary Table S2

Disease trajectories listing the number of patients following each trajectory and the median survival in days for each group.

Trajectories	Number of patients	Median survival (days)
C25_R50_R52	330	395
E11_C25_R11	311	161
E11_C25_R52	607	156
E11_C25_R53	379	177
E11_C25_R67	442	159,5
E78_C25_R52	389	158,5
H25_C25_R52	448	116
H25_C25_R53	304	110
H25_C25_R67	378	110,5
H25_H26_C25	321	66
H91_C25_R52	394	136
H91_C25_R67	301	130
I10_C25_K59	558	156
I10_C25_R11	645	156
I10_C25_R52	1073	156
I10_C25_R53	737	161
I10_C25_R63	484	160
I10_C25_R67	831	146,5
I20_C25_R52	441	161
I20_C25_R67	312	140,5
I20_E11_C25	302	84
I20_E78_C25	413	91,5
I20_I25_C25	888	73
I20_R10_C25	301	92
I21_I20_C25	599	73
I21_I20_I25	373	78
I21_I25_C25	909	65
I21_I50_C25	363	48
I25_C25_R52	372	133
K80_C25_R52	458	187
K80_C25_R67	322	190
K85_K86_C25	340	96,5
M17_C25_R52	348	189
M17_I10_C25	332	68
M51_C25_R52	320	182

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

M51_I10_C25	301	91,5
M54_C25_R52	353	170
M79_C25_R52	306	187
N92_C25_R52	394	185
N92_R10_C25	304	102
R10_C25_K59	530	156
R10_C25_R11	600	176
R10_C25_R52	1029	174
R10_C25_R53	656	177
R10_C25_R63	438	196
R10_C25_R67	738	169

Supplementary Table S3

Symptom trajectories listing the number of patients following each trajectory and the median survival in days for each group. The table lists all symptom trajectories found with a minimum of 50 patients.

Symptom 1	Symptom 2	Symptom 3	Number of patients	Median survival (days)
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Other anaemias	57	93
Pain, not elsewhere classified	Nausea and vomiting	Other anaemias	102	71
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Other anaemias	56	57
Pain, not elsewhere classified	Other anaemias	Unspecified jaundice	60	93
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Other anaemias	57	87
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Other anaemias	57	41
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Other anaemias	51	66
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other anaemias	112	63
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other anaemias	53	58

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Pain, not elsewhere classified	Abdominal and pelvic pain	Disorders of porphyrin and bilirubin metabolism	70	135
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Disorders of porphyrin and bilirubin metabolism	54	124
Pain, not elsewhere classified	Nausea and vomiting	Disorders of porphyrin and bilirubin metabolism	94	118
Abnormal blood-pressure reading, without diagnosis	Unspecified jaundice	Disorders of porphyrin and bilirubin metabolism	63	152
Abnormalities of breathing	Unspecified jaundice	Disorders of porphyrin and bilirubin metabolism	52	76
Nausea and vomiting	Unspecified jaundice	Disorders of porphyrin and bilirubin metabolism	65	99
Pain, not elsewhere classified	Unspecified jaundice	Disorders of porphyrin and bilirubin metabolism	115	121
Symptoms and signs concerning food and fluid intake	Unspecified jaundice	Disorders of porphyrin and bilirubin metabolism	62	85
Pain, not elsewhere classified	Other symptoms and signs involving the digestive system and abdomen	Disorders of porphyrin and bilirubin metabolism	57	79
Pain, not elsewhere classified	Malaise and fatigue	Disorders of porphyrin and bilirubin metabolism	55	119
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Disorders of porphyrin and bilirubin metabolism	53	148
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Disorders of porphyrin and bilirubin metabolism	104	131
Pain, not elsewhere classified	Nausea and vomiting	Unspecified abdominal hernia	68	53
Pain, not elsewhere classified	Abdominal and pelvic pain	Paralytic ileus and intestinal obstruction without hernia	74	78
Pain, not elsewhere classified	Nausea and vomiting	Paralytic ileus and intestinal obstruction without hernia	81	131
Abnormal blood-pressure reading, without diagnosis	Unspecified jaundice	Paralytic ileus and intestinal obstruction without hernia	63	137
Abnormalities of breathing	Unspecified jaundice	Paralytic ileus and intestinal obstruction without hernia	53	62
Nausea and vomiting	Unspecified jaundice	Paralytic ileus and intestinal obstruction without hernia	55	107
Pain, not elsewhere classified	Unspecified jaundice	Paralytic ileus and intestinal obstruction without hernia	105	98
Symptoms and signs concerning food and fluid intake	Unspecified jaundice	Paralytic ileus and intestinal obstruction without hernia	58	85
Pain, not elsewhere classified	Other symptoms and signs involving the digestive system and abdomen	Paralytic ileus and intestinal obstruction without hernia	60	99
Pain, not elsewhere classified	Malaise and fatigue	Paralytic ileus and intestinal obstruction without hernia	57	62
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Paralytic ileus and intestinal obstruction without hernia	93	118

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abnormal blood-pressure reading, without diagnosis	Other functional intestinal disorders	Abdominal and pelvic pain	73	66
Nausea and vomiting	Other functional intestinal disorders	Abdominal and pelvic pain	76	88
Pain, not elsewhere classified	Other functional intestinal disorders	Abdominal and pelvic pain	122	89
Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	Abdominal and pelvic pain	70	53
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Other functional intestinal disorders	83	84
Pain, not elsewhere classified	Nausea and vomiting	Other functional intestinal disorders	155	84
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Other functional intestinal disorders	76	55
Nausea and vomiting	Other functional intestinal disorders	Unspecified jaundice	61	115
Pain, not elsewhere classified	Other functional intestinal disorders	Unspecified jaundice	87	89
Abnormal blood-pressure reading, without diagnosis	Other functional intestinal disorders	Other symptoms and signs involving the digestive system and abdomen	52	71
Nausea and vomiting	Other functional intestinal disorders	Other symptoms and signs involving the digestive system and abdomen	52	71
Pain, not elsewhere classified	Other functional intestinal disorders	Other symptoms and signs involving the digestive system and abdomen	82	71
Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	Other symptoms and signs involving the digestive system and abdomen	52	66
Abnormal blood-pressure reading, without diagnosis	Malaise and fatigue	Other functional intestinal disorders	76	93
Nausea and vomiting	Malaise and fatigue	Other functional intestinal disorders	58	88
Pain, not elsewhere classified	Malaise and fatigue	Other functional intestinal disorders	108	86
Symptoms and signs concerning food and fluid intake	Malaise and fatigue	Other functional intestinal disorders	58	48
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	86	68
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	73	42
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	82	59
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	151	65

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other functional intestinal disorders	66	52
Pain, not elsewhere classified	Other functional intestinal disorders	Abnormal results of function studies	59	103
Pain, not elsewhere classified	Abdominal and pelvic pain	Pruritus	52	112
Pain, not elsewhere classified	Nausea and vomiting	Pruritus	55	194
Pain, not elsewhere classified	Unspecified jaundice	Pruritus	72	187
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Pruritus	77	193
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Abdominal and pelvic pain	100	86
Pain, not elsewhere classified	Nausea and vomiting	Abdominal and pelvic pain	185	88
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Abdominal and pelvic pain	99	80
Pain, not elsewhere classified	Abdominal and pelvic pain	Flatulence and related conditions	54	87
Nausea and vomiting	Abdominal and pelvic pain	Unspecified jaundice	54	94
Pain, not elsewhere classified	Abdominal and pelvic pain	Unspecified jaundice	103	132
Pain, not elsewhere classified	Abdominal and pelvic pain	Ascites	70	32
Pain, not elsewhere classified	Other symptoms and signs involving the digestive system and abdomen	Abdominal and pelvic pain	80	104
Nausea and vomiting	Abdominal and pelvic pain	Fever of other and unknown origin	65	104
Pain, not elsewhere classified	Abdominal and pelvic pain	Fever of other and unknown origin	106	95
Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	Fever of other and unknown origin	61	88
Abnormal blood-pressure reading, without diagnosis	Malaise and fatigue	Abdominal and pelvic pain	80	112
Nausea and vomiting	Malaise and fatigue	Abdominal and pelvic pain	60	57
Pain, not elsewhere classified	Malaise and fatigue	Abdominal and pelvic pain	123	88
Haemorrhage, not elsewhere classified	Malaise and fatigue	Abdominal and pelvic pain	60	84
Symptoms and signs concerning food and fluid intake	Malaise and fatigue	Abdominal and pelvic pain	60	80
Nausea and vomiting	Haemorrhage, not elsewhere classified	Abdominal and pelvic pain	77	52
Symptoms and signs concerning food and fluid	Haemorrhage, not elsewhere classified	Abdominal and pelvic pain	64	87

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

intake				
Nausea and vomiting	Oedema, not elsewhere classified	Abdominal and pelvic pain	68	63
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	84	79
Cough	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	67	59
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	82	61
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	85	68
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	184	93
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Abdominal and pelvic pain	60	47
Pain, not elsewhere classified	Abdominal and pelvic pain	Abnormal results of function studies	68	143
Pain, not elsewhere classified	Nausea and vomiting	Flatulence and related conditions	53	78
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Unspecified jaundice	66	118
Pain, not elsewhere classified	Nausea and vomiting	Unspecified jaundice	124	116
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Unspecified jaundice	53	63
Pain, not elsewhere classified	Nausea and vomiting	Ascites	63	36
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Other symptoms and signs involving the digestive system and abdomen	71	92
Pain, not elsewhere classified	Nausea and vomiting	Other symptoms and signs involving the digestive system and abdomen	117	91
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Other symptoms and signs involving the digestive system and abdomen	63	93
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Fever of other and unknown origin	72	105
Pain, not elsewhere classified	Nausea and vomiting	Fever of other and unknown origin	123	83
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Fever of other and unknown origin	68	61
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Malaise and fatigue	72	128

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Pain, not elsewhere classified	Nausea and vomiting	Malaise and fatigue	133	118
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Malaise and fatigue	57	63
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Haemorrhage, not elsewhere classified	76	76
Pain, not elsewhere classified	Nausea and vomiting	Haemorrhage, not elsewhere classified	157	77
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Haemorrhage, not elsewhere classified	85	57
Abnormal blood-pressure reading, without diagnosis	Nausea and vomiting	Oedema, not elsewhere classified	77	115
Pain, not elsewhere classified	Nausea and vomiting	Oedema, not elsewhere classified	131	89
Symptoms and signs concerning food and fluid intake	Nausea and vomiting	Oedema, not elsewhere classified	68	83
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Nausea and vomiting	79	80
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Nausea and vomiting	71	65
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Nausea and vomiting	63	70
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Nausea and vomiting	157	85
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Nausea and vomiting	62	59
Pain, not elsewhere classified	Nausea and vomiting	Abnormal results of function studies	80	110
Other functional intestinal disorders	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	53	90
Abnormal blood-pressure reading, without diagnosis	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	59	136
Abnormalities of breathing	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	56	110
Abdominal and pelvic pain	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	59	109
Nausea and vomiting	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	88	117
Dizziness and giddiness	Unspecified jaundice	Other symptoms and signs involving the digestive system	67	90

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

		and abdomen		
Pain, not elsewhere classified	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	112	117
Oedema, not elsewhere classified	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	51	108
Symptoms and signs concerning food and fluid intake	Unspecified jaundice	Other symptoms and signs involving the digestive system and abdomen	96	125
Pain, not elsewhere classified	Unspecified jaundice	Fever of other and unknown origin	72	126
Abnormal blood-pressure reading, without diagnosis	Malaise and fatigue	Unspecified jaundice	57	125
Pain, not elsewhere classified	Malaise and fatigue	Unspecified jaundice	87	104
Nausea and vomiting	Haemorrhage, not elsewhere classified	Unspecified jaundice	60	93
Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	Unspecified jaundice	52	166
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	66	134
Cough	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	56	98
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	72	81
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	58	85
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	130	130
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Unspecified jaundice	61	79
Pain, not elsewhere classified	Unspecified jaundice	Abnormal results of function studies	63	110
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Ascites	71	36
Pain, not elsewhere classified	Other symptoms and signs involving the digestive system and abdomen	Fever of other and unknown origin	58	103
Abnormal blood-pressure reading, without diagnosis	Malaise and fatigue	Other symptoms and signs involving the digestive system and abdomen	63	111
Pain, not elsewhere classified	Malaise and fatigue	Other symptoms and signs involving the digestive system	87	107

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

		and abdomen		
Symptoms and signs concerning food and fluid intake	Malaise and fatigue	Other symptoms and signs involving the digestive system and abdomen	57	96
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Other symptoms and signs involving the digestive system and abdomen	73	106
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Other symptoms and signs involving the digestive system and abdomen	68	101
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Other symptoms and signs involving the digestive system and abdomen	63	93
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other symptoms and signs involving the digestive system and abdomen	127	113
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other symptoms and signs involving the digestive system and abdomen	53	61
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	76	80
Cough	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	54	44
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	61	58
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	62	69
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	127	77
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Symptoms and signs involving emotional state	53	48
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Fever of other and unknown origin	55	96
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Fever of other and unknown origin	56	75
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Fever of other and unknown origin	54	66
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Fever of other and unknown origin	118	80
Nausea and vomiting	Haemorrhage, not elsewhere classified	Malaise and fatigue	68	76
Symptoms and signs	Haemorrhage, not elsewhere	Malaise and fatigue	54	66

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

concerning food and fluid intake	classified			
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	80	52
Cough	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	58	64
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	70	55
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	63	58
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	143	68
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Malaise and fatigue	56	45
Pain, not elsewhere classified	Malaise and fatigue	Abnormal results of function studies	65	88
Abnormal blood-pressure reading, without diagnosis	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	90	94
Cough	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	67	43
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	89	89
Dizziness and giddiness	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	81	58
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	177	89
Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Haemorrhage, not elsewhere classified	70	40
Nausea and vomiting	Oedema, not elsewhere classified	Symptoms and signs concerning food and fluid intake	59	97
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Other abnormal findings in urine	57	180
Abnormalities of breathing	Symptoms and signs concerning food and fluid intake	Abnormal results of function studies	53	78
Pain, not elsewhere classified	Symptoms and signs concerning food and fluid intake	Abnormal results of function studies	100	85

632

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

633

634 **Supplementary Table S4**

635

Symptom	Number of patients	P-value
Jaundice	1184	$0.00 \cdot 10^{00}$
Anorexia	1374	$0.00 \cdot 10^{00}$
Digestive system and abdomen	1033	$1.44 \cdot 10^{-146}$
Abnormal weight loss	2111	$0.00 \cdot 10^{00}$
Food and fluid intake	2468	$0.00 \cdot 10^{00}$
Abdominal and pelvic pain	1400	$1.54 \cdot 10^{-250}$
Nausea	1551	$4.32 \cdot 10^{-91}$
Nausea and vomiting	1858	$9.21 \cdot 10^{-93}$
Vomiting	1085	$1.06 \cdot 10^{-44}$
Malaise and fatigue	1171	$2.73 \cdot 10^{-29}$
Fever	796	$1.45 \cdot 10^{-20}$
Intestinal disorders	1045	$1.63 \cdot 10^{-24}$
Functional diarrhoea	685	$2.47 \cdot 10^{-16}$
Other anaemias	658	$4.44 \cdot 10^{-09}$
Haemorrhage	1520	$4.80 \cdot 10^{-08}$
Pain	2982	$2.21 \cdot 10^{-15}$
Abnormal blood-pressure	1677	$7.28 \cdot 10^{-03}$
Abnormalities of breathing	1221	$1.90 \cdot 10^{-04}$
Headache	619	$2.12 \cdot 10^{-06}$

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Abnormalities of heart beat	641	$1.33 \cdot 10^{-07}$

Table caption: The 20 most frequent symptoms found by text mining clinical notes. Significance is tested by a χ^2 test statistics.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

678

679 **Supplementary Table S5**

680

Disease name	Survival group	Number of symptoms
Abdominal and pelvic pain	Less than 90d	918
Abdominal and pelvic pain	More than 90d	944
Abnormal blood-pressure reading, without diagnosis	Less than 90d	1405
Abnormal blood-pressure reading, without diagnosis	More than 90d	1365
Abnormal weight loss	Less than 90d	1454
Abnormal weight loss	More than 90d	1467
Abnormalities of breathing	Less than 90d	1096
Abnormalities of breathing	More than 90d	861
Abnormalities of heart beat	Less than 90d	422
Abnormalities of heart beat	More than 90d	389
Anorexia	Less than 90d	854
Anorexia	More than 90d	758
Fever of other and unknown origin	Less than 90d	419
Fever of other and unknown origin	More than 90d	476
Functional diarrhoea	Less than 90d	442
Functional diarrhoea	More than 90d	417
Haemorrhage, not elsewhere classified	Less than 90d	920
Haemorrhage, not elsewhere classified	More than 90d	988
Headache	Less than 90d	375
Headache	More than 90d	388
Malaise and fatigue	Less than 90d	702
Malaise and fatigue	More than 90d	670
Nausea	Less than 90d	1008
Nausea	More than 90d	1081
Nausea and vomiting	Less than 90d	1391
Nausea and vomiting	More than 90d	1437
Other anaemias	Less than 90d	557
Other anaemias	More than 90d	432
Other functional intestinal disorders	Less than 90d	730
Other functional intestinal disorders	More than 90d	629
Other symptoms and signs involving the digestive system and abdomen	Less than 90d	473
Other symptoms and signs involving the digestive system and abdomen	More than 90d	592

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Pain, not elsewhere classified	Less than 90d	2937
Pain, not elsewhere classified	More than 90d	2854
Symptoms and signs concerning food and fluid intake	Less than 90d	1809
Symptoms and signs concerning food and fluid intake	More than 90d	1781
Unspecified jaundice	Less than 90d	614
Unspecified jaundice	More than 90d	1013
Vomiting	Less than 90d	752
Vomiting	More than 90d	693

medRxiv preprint doi: https://doi.org/10.1101/2023.02.13.23285861; this version posted February 16, 2023. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Supplementary Figure

Supplementary Figure S1

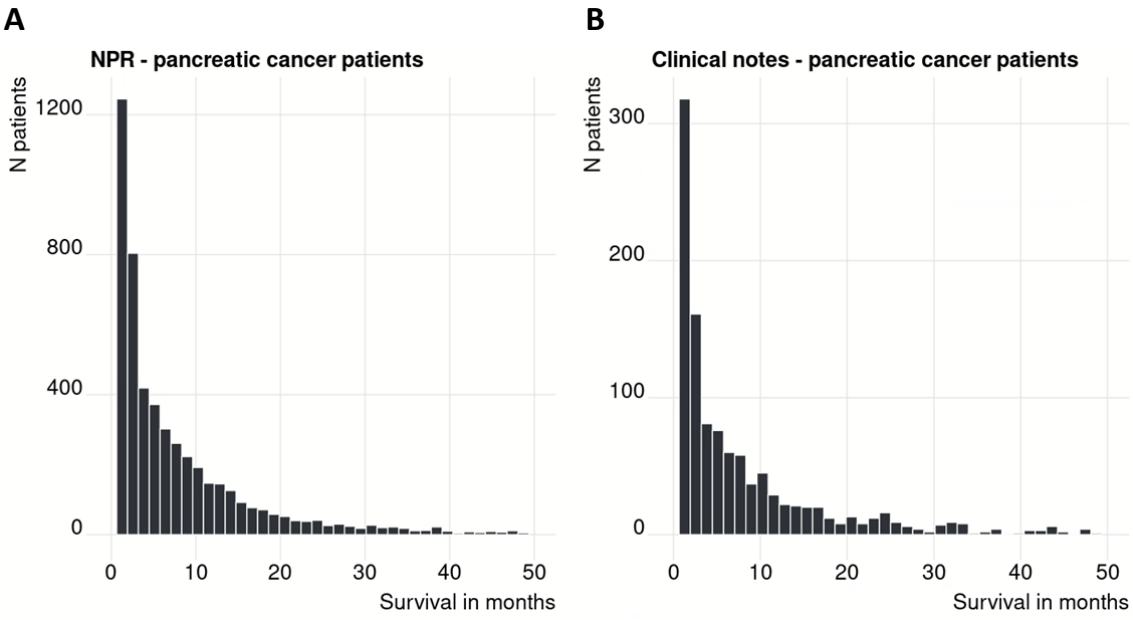

Figure S1. (A) Survival in months for patients in the registry disease trajectories. (B) Survival in months for patients in the text mining symptom trajectories. The pancreatic cancer patient cohort from is a subset of the pancreatic cancer patient cohort in A.