HAL open science

Patient healthcare trajectory. An essential monitoring tool: a systematic review

Jessica Pinaire, Jérôme Azé, Sandra Bringay, Paul Landais

To cite this version:

Jessica Pinaire, Jérôme Azé, Sandra Bringay, Paul Landais. Patient healthcare trajectory. An essential monitoring tool: a systematic review. Health Information Science and Systems, 2017, 5 (1), 10.1007/s13755-017-0020-2 . hal-01566986

HAL Id: hal-01566986

https://hal.science/hal-01566986

Submitted on 21 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Pinaire *et al. Health Inf Sci Syst (2017) 5:1*
https://doi.org/10.1007/s13755-017-0020-2

Health Information Science and Systems

REVIEW

Patient healthcare trajectory. An essential monitoring tool: a systematic review

Jessica Pinaire[1,2,3*], Jérôme Azé[3], Sandra Bringay[3,4] and Paul Landais[1,2]

Abstract

Background: Patient healthcare trajectory is a recent emergent topic in the literature, encompassing broad concepts. However, the rationale for studying patients' trajectories, and how this trajectory concept is defined remains a public health challenge. Our research was focused on patients' trajectories based on disease management and care, while also considering medico-economic aspects of the associated management. We illustrated this concept with an example: a myocardial infarction (MI) occurring in a patient's hospital trajectory of care. The patient follow-up was traced via the prospective payment system. We applied a semi-automatic text mining process to conduct a comprehensive review of patient healthcare trajectory studies. This review investigated how the concept of trajectory is defined, studied and what it achieves.

Methods: We performed a PubMed search to identify reports that had been published in peer-reviewed journals between January 1, 2000 and October 31, 2015. Fourteen search questions were formulated to guide our review. A semi-automatic text mining process based on a semantic approach was performed to conduct a comprehensive review of patient healthcare trajectory studies. Text mining techniques were used to explore the corpus in a semantic perspective in order to answer non-a priori questions. Complementary review methods on a selected subset were used to answer a priori questions.

Results: Among the 33,514 publications initially selected for analysis, only 70 relevant articles were semi-automatically extracted and thoroughly analysed. Oncology is particularly prevalent due to its already well-established processes of care. For the trajectory thema, 80% of articles were distributed in 11 clusters. These clusters contain distinct semantic information, for example health outcomes (29%), care process (26%) and administrative and financial aspects (16%).

Conclusion: This literature review highlights the recent interest in the trajectory concept. The approach is also gradually being used to monitor trajectories of care for chronic diseases such as diabetes, organ failure or coronary artery and MI trajectory of care, to improve care and reduce costs. Patient trajectory is undoubtedly an essential approach to be further explored in order to improve healthcare monitoring.

Keywords: Systematic reviews, Text mining, Healthcare trajectory, PPS, Semi-automated, Word cloud

Background

Patient healthcare trajectory is a recent emergent topic in the literature, encompassing broad concepts. Our research was focused on the patient trajectory based on disease management and care, while also considering medico-economic aspects of the associated management.

We approached patient care trajectories based on an example; the occurrence of a myocardial infarction (MI). As MI treatment is performed in a health facility, we were able to trace the patient trajectories through the national hospital financing system, using comprehensive hospital databases or registers, regularly collected for billing purposes.

The first prospective payment system (PPS), based on diagnosis-related groups (DRG), was established in the United States in 1983. The objective of this system was

*Correspondence: jessica.pinaire@chu-nimes.fr
[1] Biostatistics, Epidemiology and Public Health Department, Nîmes University Hospital, Place R Debré, 30 029 Nîmes, France
Full list of author information is available at the end of the article

© Springer International Publishing AG 2017.

Table 1 Search questions

Non-a priori questions	
Q1	Do studies on the patients' trajectories exist?
Q2	What are the topics in these studies? (support, treatment, costs, etc.)
Q3	Which diseases are studied by trajectories?
Q4	Is PPS explored in the search?
Q5	Is PPS used in studying trajectories?
Q6	Are there any studies on the trajectories of patients with MI?
Q7	What is studied in MI?
A priori questions	
Q8	What are the various concepts of the trajectory? (How is this concept defined?)
Q9	What is the interest in the subject: have many studies focused on patients' trajectories?
Q10	Which countries conduct studies on trajectories?
Q11	What are the objectives of patient trajectory studies?
Q12	What methods are used in patient trajectory studies?
Q13	What are the characteristics of the studies: number of patients involved, duration of follow-up?
Q14	What data is used in these studies: hospital or other?

to control the expenditures of health care institutions and streamline the costs [1]. Thereafter, similar medical information systems were adopted in many other industrialised countries. Others, like France, also adopted an anonymised database with unique patient identifiers (for instance, through cryptographic hash functions) to facilitate chaining hospital stays [2–4]. In addition, the gradual increase in fees-for-services enhanced the coding quality [5]. The introduction of these systems enabled new epidemiological and/or economic studies [6–9] using these databases, with temporal follow-up of patients allowing tracing of their trajectory of care. This review investigated how the trajectory concept is defined, studied and what it achieves.

We carried out a literature search on PubMed using keywords related to trajectory, PPS and MI concepts. We then proceeded in two steps: (1) a non-a priori search with text mining techniques; and (2) a more standard analysis of a sub-selection of documents.

Similar systematic reviews [10] have been performed before, but without using automatic procedures. However, conducting an automatic search is of considerable interest for processing a large number of documents. Text mining allows better targeting for information retrieval and reduces the search time [11], while also enabling users to prioritise searches.

Our reviewing strategy is presented in the "Methods" section; the search questions that guided our review, together with the various methods used to address them. The results are reported in the "Results" section. We end with the "Discussion" section, where we present answers to the search questions and comment on the results. To conclude this section, we discuss the different existing text mining techniques used in systematic reviews.

Methods

Search questions

Healthcare researchers currently explore the literature manually, and use statistical methods or models that require a priori extreme simplification of the processes. Data exploration methods such as text mining methods end by the interpretation and exploration of the processes, not a priori by knowledge discovery. We formulated practical questions to guide the review process (see Table 1). We identified seven types of non-a priori questions expressed in general terms that integrate thematic and medical oriented issues and satisfy scientific and medical aspects for health care professional expertise. We also identified seven additional specific a priori questions requiring in-depth analysis.

Step 1: document retrieval

In PubMed, we searched for documents examining patient healthcare trajectories, as well as PPS and MI. The review selection process is summarised in Fig. 1. The trajectory concept can be expressed with different words such as "trajectory", "pathway" or "path". For the PPS theme, keywords used are "Prospective Payment System" and "PMSI" (Programme de médicalisation du système d'information, the French PPS equivalent), in addition to "DRG" (Diagnosis-Related Group) databases. This theme also arises in the International Classification of Diseases (ICD), pricing for the activity via the "fee-for-service" or "activity-based payment" expressions, but also in the

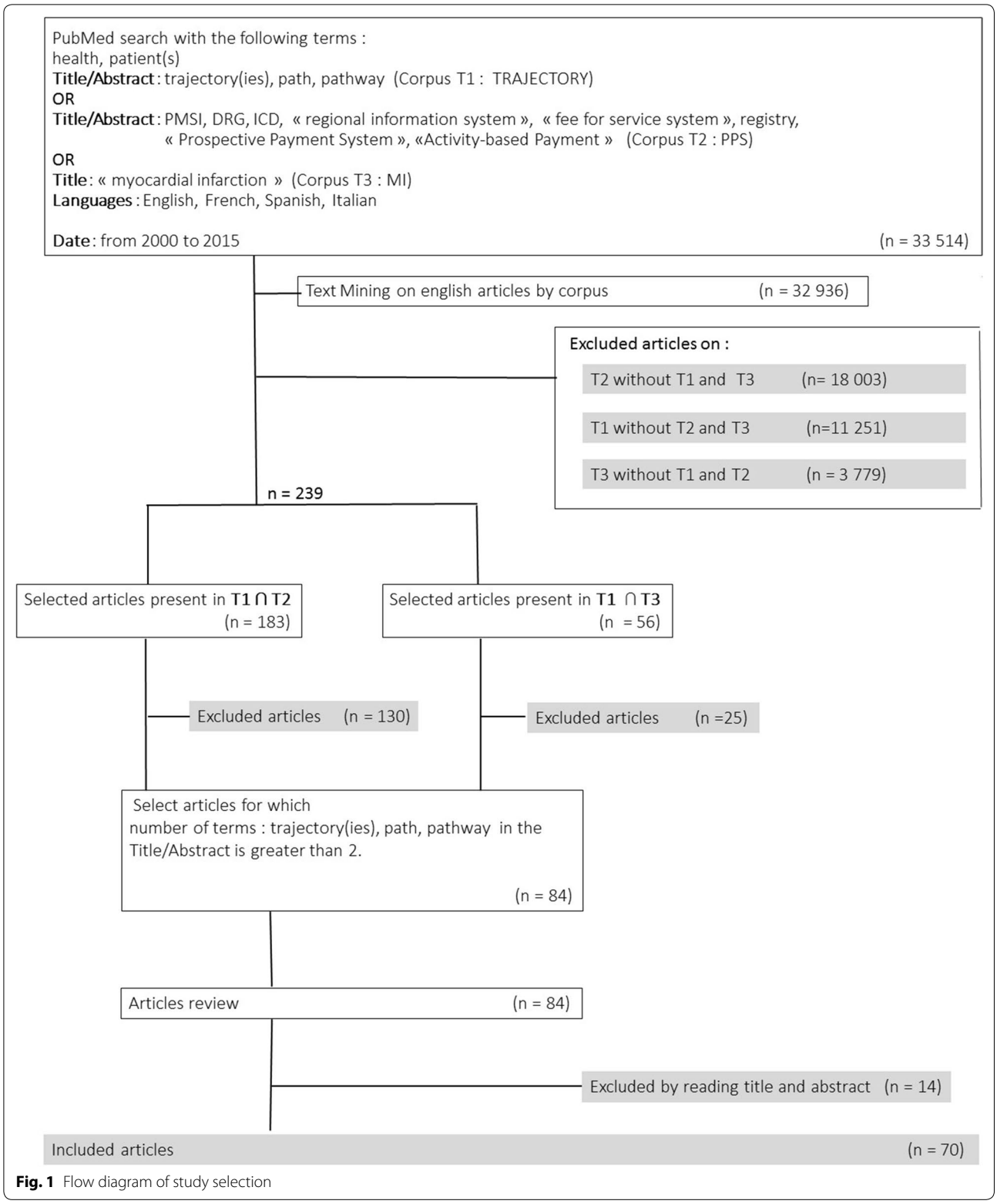

Fig. 1 Flow diagram of study selection

Table 2 Keywords used in document retrieval

Topics and constraints	Keywords
C1: Medical context	"health", "patient(s)"
T1: Trajectory	"trajectories", "trajectory", "path", "pathway(s)"
T2: PPS	"prospective payment system", "PMSI", "DRG", "ICD", "regional information system", "fee for service system", "registry", "Activity-based Payment"
T3: MI	"myocardial infarction" in the title
C2: Dates	January 1, 2000 to October 31, 2015
C3: Languages	English, French, Spanish and Italian

national health registry or hospital registry concepts. We conducted searches according to the themes and constraints summarised in Table 2.

Step 2: first text mining approach

The strategy was as follows:

(i) From the selection of articles gathered in step 1, we created a corpus of texts, divided into three parts, T1 to T3 (corresponding to Table 2 topics), consisting of the title and abstract, in which we removed the keywords (see Table 2) in order to only keep the other terms;
(ii) The three parts of the corpus were analysed separately with IRaMuteQ[1] software. This is an R interface for multidimensional analysis of texts and questionnaires [12], allowing statistical analysis of the text corpus [13];
(iii) We applied the following pre-processing techniques: (a) Lemmatization of texts, (b) Dictionary enrichment: we lemmatized unrecognised terms by TreeTagger[2] and added specific medical terms and well-known acronyms such as acute myocardial infarction (AMI). Subsequently, the analyses were conducted with the full forms (nouns, adjectives, adverbs and verbs);
(iv) We carried out conventional textual analysis, then similarity analysis and finally clustering. The various tools used were as follows:

Word cloud This is a synthetic representation of the terms distribution: the most recurrent words are in the centre with text size proportional to the number of occurrences. Thus, this kind of representation symbolises, by order of importance, the concepts covered in all of the articles. This method will provide an answer to Q1.

Similarity analysis This graph theory-based technique is conventionally used to describe social representations based on survey questionnaires [14]. Similarity analysis is applied to study the proximity and relationships between elements in a set, in the form of maximum trees. The objective is to reduce the numbers of links between two items, to obtain an acyclic connected graph. The maximum tree is therefore the tree created by the strongest edges of the graph, where the strength is measured by the occurrence of the linked terms. For each corpus, we selected the tree representation described in [15] and in the algorithm in [16], to describe communities via the shortest path, thus highlighting the most frequently associated words in the same sentence or text. The graph generates a more precise idea of the content of articles concerning the concepts and themes raised by linking important terms. This method will provide answers to Q3, Q5 and Q7.

Text clustering Reinert clustering [17] is a form of divisive hierarchical clustering (DHC) that is carried out in several stages, offering a global approach to the corpus. It identifies statistically independent word classes after partitioning the corpus. These classes may be interpreted by their profiles, which are characterised by specific correlated words. DHC summarises this through a dendrogram. This analysis generates a complementary vision with regard to similarity analysis by clustering articles according to concepts, partly identified by similarity analysis, characterised by word groups. This method will supplement the answer to Q7, and address Q2, Q4 and Q6.

Step 3: thorough analysis of the selected articles

We used the sub-selection technique derived from Moher's method described in [18], and crossed the sets of themes: T1 and T2, denoted $T1 \cap T2$, then T1 and T3, denoted $T1 \cap T3$. This selection was performed in the same manner as described in Table 2. We added an additional constraint to better target our study through counting the K occurrence number of the trajectory concept in each document and selecting those for which: $K \geq 2$. We counted each time the words "trajectories", "trajectory" or "pathway" appeared in the titles and summaries of the articles.

Our reading grid was based on that described in the PRISMA[3] guidelines. We selected items that could be used to address the a priori search questions (see Table 1), Q8 to Q14: publication year, country of study, number of patients, observation period, methods and objectives. Other items that were irrelevant to our study were not kept. We added the following items: pathologies

[1] http://www.iramuteq.org/.
[2] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.
[3] www.prisma-statement.og.

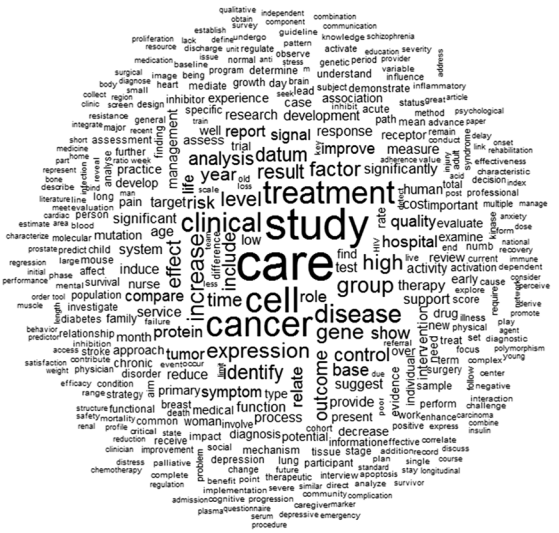

Fig. 2 Word cloud: trajectory, PPS, MI

studied, databases used and definition of the trajectory concept.

Results

Some results, not listed in this paper, can be viewed at the following address: http://www.lirmm.fr/~pinaire/.

Step 1: document retrieval

The document retrieval resulted in a total of 33,514 articles.

Step 2: first text mining approach

We present the results obtained by our method which combined different approaches of lexicographic analysis (see below) following the flow diagram (Fig. 1).

Word cloud For T1 (see Fig. 2), the salient terms were "care", "study", "cancer", "cell", "treatment" and "increase", while for T2 they were "study", "registry", "datum" and "cancer", and for T3 they were "AMI", "acute" and "hospital".

Similarity analysis In a maximum tree only the strongest edges of the graph are kept. An edge symbolises the

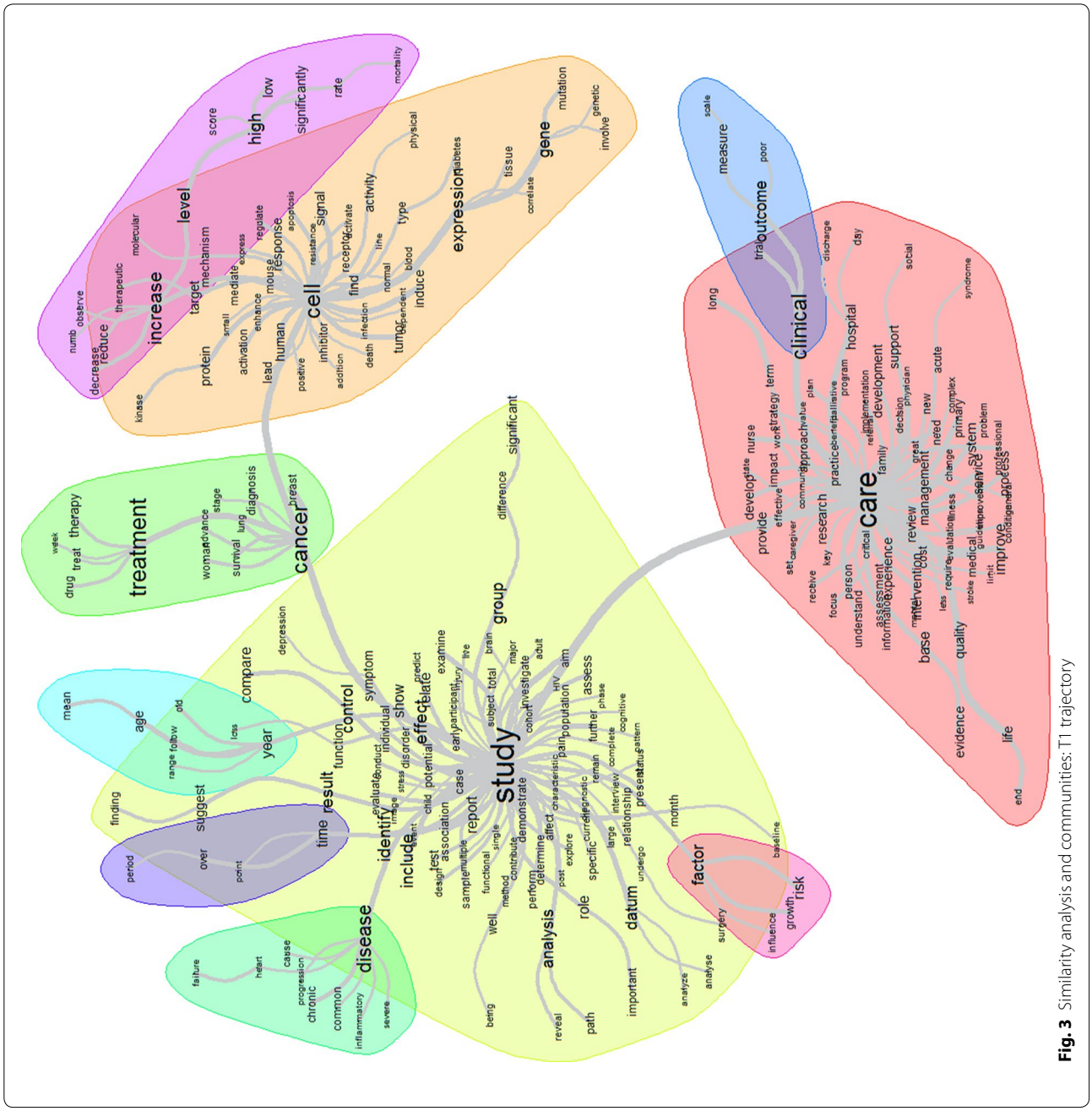

Fig. 3 Similarity analysis and communities: T1 trajectory

co-occurrence between vertices (i.e. terms), and its thickness represents the strength of the link. For instance, in Fig. 3, the link between "care" and "study" is thicker than the link between "significant" and "difference". For T1, Fig. 3 comprises three hubs: the bottom part with a large network characterised by "care", then a smaller contiguous network gathering the terms "clinical" and "outcome". On the right upper part, there is a network encompasses genetics terms, with "cell", "expression" and "gene", then a smaller connecting network gathering the terms "increase", "high" and "significantly". The central upper part contains the terms "cancer", "diagnosis" and "treatment". Finally, the left upper part has a large network containing "study", to which are attached several smaller clusters characterised by the terms "risk", "disease", "time", "year" and finally "disease". The most closely linked word communities are "genetics" with "cancer", "cancer" with "study", and "study" with "care".

Text clustering Following this clustering, 80% of the articles of T1 were distributed in 11 disjointed clusters, 86% for T2 in five clusters, and 98% for T3 in five clusters. We then performed a second clustering on the subcorpus of each theme, consisting of articles that were not clustered during the first analysis.

For T1, Fig. 4 shows, from right to left, two clusters pooling the concepts of genetic organisation (cluster 5), signal organisation and cellular mediation (cluster 10). Cluster 8 pools concepts related to the immune system response in an inflammatory process. Cluster 1 pools dysfunctions related to diabetes and the consequences. Cluster 2 and 7 respectively symbolise time in the organization of hospital stays, and time in the trajectory. Cluster 6 concerns questionnaires and psychometric scales with depression. Cluster 11 contains concepts related to medical imagery. In the last branch, cluster 9, medicine is described with regard to its financial and regulatory aspects. Cluster 4 concerns the patient management including practices. Cluster 3 groups together terms pertaining to the way information is conveyed.

We performed a more detailed examination of clusters 3 and 4, while studying the similarity tree. The two clusters pool 1645 items between them. For cluster 3, there are three nodes (see Fig. 5), i.e. for "care", which is closely related to that for "study", which in turn is closely related to that for "cancer". For cluster 4 (see Fig. 5), there is only one node represented by "care", from which there are several branches for "research" and "process", and then higher there is a sub-node for "clinic", connecting "trial", "datum", "bases" and "identify".

In the second clustering of the 3160 non-clustered articles, we identified five clusters consisting of 99% of the articles. From right to left, cluster 1 pools the concept of studies from a methodological standpoint. Cluster 5 concerns end-of-life issues. Cluster 4 pools the macroscopic aspect of care with public support. Cluster 3 groups studies involving animal experiments. Finally, cluster 2 concerns genetic mutation and anomalies. Three articles could not be clustered due to a lack of information.

Step 3: thorough analysis of the selected articles

Through set crossing, we generated a sub-selection of 84 articles, including 53 for T1∩T2 and 31 for T1∩T3. After reading the abstracts, we eliminated 8 articles for T1∩T2, and 6 for T1∩T3 as they evaluated a protein or organisation trajectory rather than patient trajectory. For the majority of items, we created categories as detailed in Table 3 and the sources of associated references are indexed in Table 4.

Generally, the authors used several sources and methods in their studies. The results for these items are summarised in Fig. 6.

We grouped the countries according to continent. For T1∩T2, we noted strong representation from Europe (55% of articles) and the Americas (29%). There were some studies from Oceania (9%) and Asian countries (7%). For T1∩T3, the article distribution was essentially between three continents: Europe (36%), the Americas (28%) and Asia (24%). Australasia was marginal, with 4% of articles. There were some atypical studies with data from multiple continents (8%).

We next considered publication year. For T1∩T2, the results highlighted activity that began developing in 2013. While for T1∩T3, we noted a peak of activity in 2004, and increasing activity in 2012.

The number of patients involved was then analysed, showing that the number of patients ranged from 14 to 6.2 million T1∩T2 (vs 20–30.20 million for T1∩T3), with a median of 859 and an interquartile interval (IQ) of 3250 (vs 604.5 and IQ = 933.25 for T1∩T3), with missing data for three articles (vs five for T1∩T3).

We also focused on the observation duration, measured in months, available in more than 85% articles. Observation duration ranged from 5 to 180 months in T1∩T2 (vs 3 to 240 months in T1∩T3) with a median of 36 months and an IQ of 54 months (vs 12 and 99 months in T1∩T3).

Discussion

Our method is based on a semi-automatic approach of text mining. We used terms and concepts which emerged from classification techniques rather than the simple presence of words. This approach was structured into two main steps prior to a thorough text analysis of the selected articles. These two steps were based on document retrieval and text mining techniques.

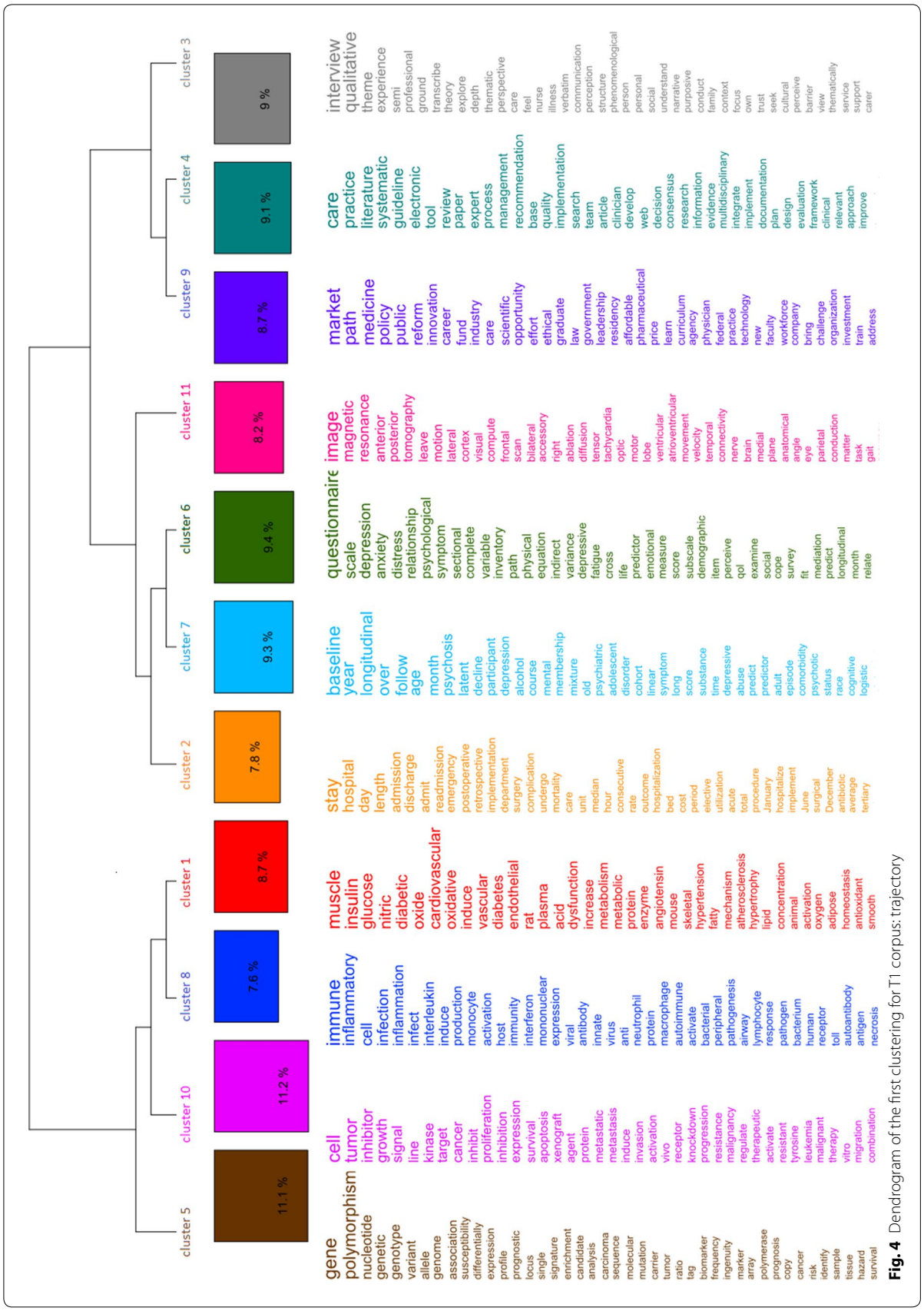

Fig. 4 Dendrogram of the first clustering for T1 corpus: trajectory

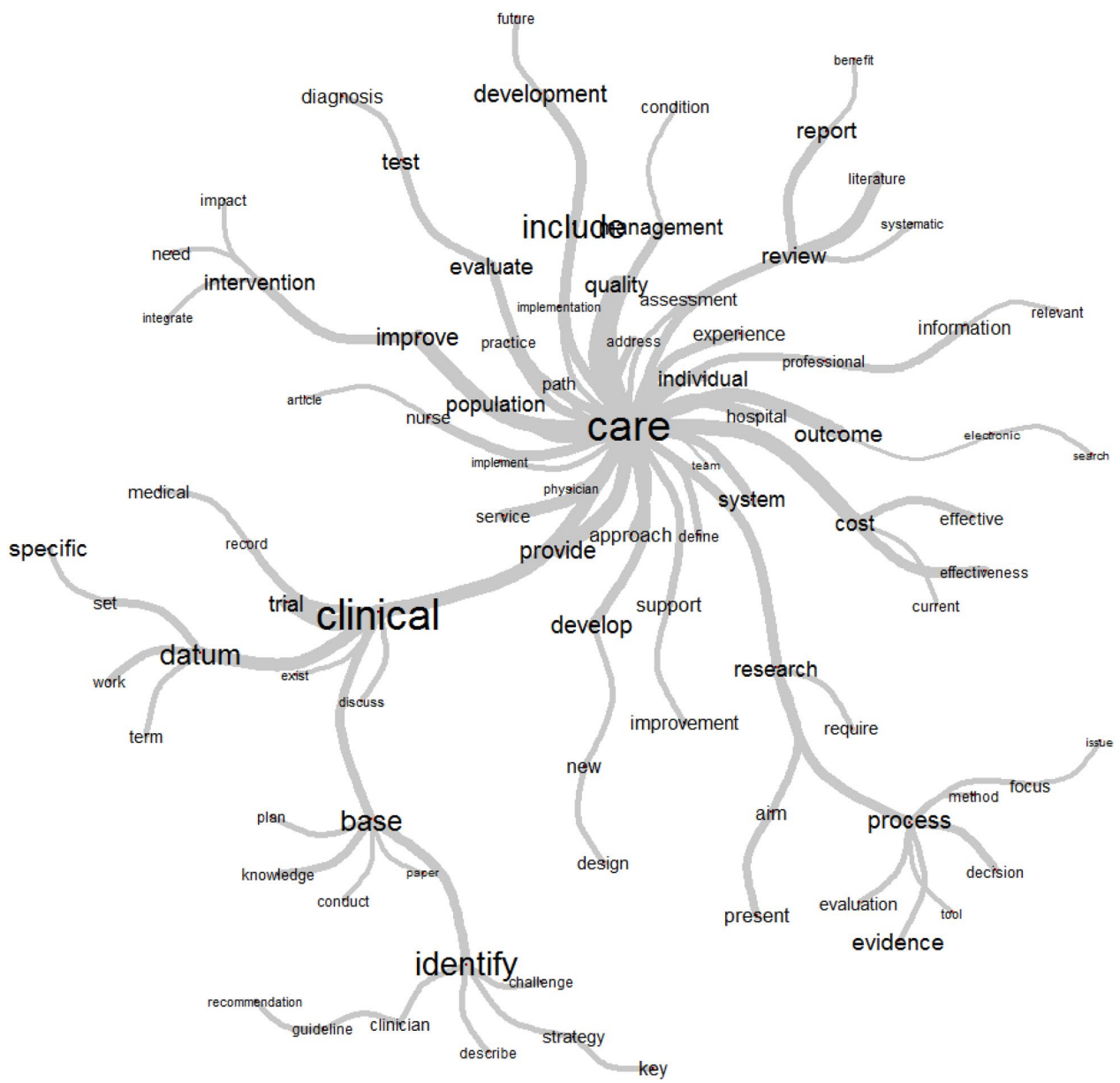

Fig. 5 Similarity analysis on cluster 3 from the first T1 clustering corpus

Step 1: document retrieval

For document retrieval, we chose to focus our study on PubMed. The search results are entirely dependent on the choice of keywords, making this a particularly delicate task when definitions may vary between authors and countries. Indeed, we encountered this difficulty for T2. As presented in Table 2, the keywords used were "Prospective Payment System", "PMSI", "DRG", "ICD", "regional information system", "fee for service system", "registry", "Activity-based Payment". However some documents used words not in our final selection, such as in [19], which contains the term "national case-mix system". Our objective was not to be exhaustive with regard to covering all the publications, but rather to define a general method of analysis. A way to improve our approach would be to implement an adaptive algorithm for keywords enrichment.

Step 2: first text mining approach

The lexicographic analysis was based on three combined tools.

Table 3 Description of the observed items and categories

Item	Category	Details
Objective	Medical advances	Better comprehension of a disease and subsequent adaptation of care
	Health recommendations	Medical instructions to improve the health status and/or avoid its impairment
	Cost/quality assessment	Comparison of treatments, care process, new drugs
	Health planning	Implementation of care process to improve support
	Data processing tool	Creation of a synthetic tool for data visualisation or algorithms formulations to gather and classify data derived from several sources
	Others	Other objectives
Databases	Registry	Registry databases
	Hospital & PPS	Hospital or prospective payment system databases
	Interviews	Any type of interview
	Questionnaires	Questionnaires and multiple choice questionnaire
	Others	Pharmacy databases, social security databases, data from GPs, blood bank databases, and patient diaries.
Diseases	Cardiovascular diseases	Stroke, MI, heart failure
	Diabetes	Diabetes case
	Cancer	Colorectal, prostate, breast, lung, bladder, cervical, endometrial
	Lung diseases	Chronic obstructive pulmonary disease, pulmonary embolism
	Kidney diseases	Renal failure
	Neurological diseases	Multiple sclerosis, schizophrenia, depression
	Others	Gout, osteoarthritis, scoliosis, craniotomy, colon penetration injury, pelvic fractures, pain
Methods	Survival models	Cox, Kaplan–Meier models
	Param and non-param tests	Chi Squared, Fisher, Student, Kruskal–Wallis, Mann–Whitney tests, etc.
	Clustering	
	ANOVA–ANCOVA	
	Linear or logistic models	Linear or logistic regression, GLM, logistic model
	Others	Latent variable models, Kappa coefficient, meta-analysis, etc.
Trajectory	Cost	Tracking costs in the case of a treatment or care process
	Care	Care trajectory, with the history of consultations, reasons for hospitalisation, care provided
	Care process	Series of steps through which the patient passes into an integrated care process, or a series of operational steps of a care team
	Health outcomes	Symptoms, clinical, cognitive developments
	Biological measures	Different types of measures
	Risk	Disease progression risk
	Survival	Survival pattern
	Others	Measurement of time physical activity, patient decision making

For the *word cloud* approach, the occurrence of the terms "study" and "care", for all of the studied fields, means that these articles cover care concept and studies on topics such as diseases or drugs. For T1, the terms "treatment" and "increase" reflect a focus on patient healthcare trajectories. Thus, there are many studies on patient trajectories. Here we have answered **Q1**.

For the *similarity analysis*, the results showed that for T1, studies were closely related to care, disease and more specifically to cancer. In response to **Q3**, the studied diseases were those causing severe and chronic organ dysfunction: heart, kidneys, or lungs. We noted that the cancer concept was also closely related to that of genetics. We found here that the use of the keyword "pathway" highlighted all articles pertaining to cell signalling or gene pathways [20–29].

For T2, cancer was closely related to the registry data. This highlights the descriptive aspect of the data information, i.e. registry data describing the patient's cancer history from its diagnosis. We noted that the study concept was related to the disease concept, i.e. cardiac or renal, but also to the various treatments and therapies. In response to **Q5**, T2 was thus related to research: in disease studies [30–37], to compare care and coding [38–40], but also in monitoring of patients over time and the survival rate [31, 41–44]. Survival rate forms part of a trajectory concept. This trajectory concept may also

Table 4 Review references sources by item reviewed

Item	Category	*T*1∩*T*2	*T*1∩*T*3
Trajectory concept	Cost	[58, 59, 86, 87]	
	Care	[30, 31, 38, 60, 81, 88–93]	[39, 40, 46, 47, 53, 54, 94, 95]
	Care process	[41, 96–107]	[48–51, 108]
	Health outcomes	[32–36, 42, 109–115]	[43, 45, 52, 55, 56, 116]
	Biological measures	[37]	
	Risk		[57]
	Survival	[117]	[44, 118]
	Others	[119, 120]	[121–123]
Objective	Medical advances	[31, 33, 35–37, 109, 113, 120]	[40, 43, 54]
	Health recommendations	[30, 32, 41, 106, 115]	[46, 47, 50, 52, 53, 56, 116, 121, 122]
	Cost/quality assessment	[42, 58, 59, 86, 87, 90, 96, 98–103, 107, 111, 114]	[39, 44]
	Health planning	[90, 92, 93, 104, 110, 112]	[48, 49, 51, 95, 108]
	Data processing tool	[60, 81, 88, 91]	
	Others	[34, 89, 97, 105, 117, 119]	[45, 55, 57, 94, 118, 123]
Continent	America	[31, 35–37, 41, 42, 91, 96, 98, 99, 103, 109, 120]	[48–50, 56, 57, 95, 123]
	Asia	[104, 105, 107]	[43, 45, 94, 118, 121, 122]
	Europe	[30, 32, 33, 38, 58–60, 81, 86–90, 93, 100, 102, 106, 110–115, 117, 119]	[39, 40, 47, 51–54, 108, 116]
	Intercontinental		[44, 55]
	Australasia	[34, 92, 97, 101]	[46]
Databases	Questionnaires	[30, 36, 89, 90, 102, 104, 109, 110, 112–115]	[50, 52, 55, 121]
	Interviews	[33, 96–98, 117]	[94, 95, 122, 123]
	Hospital & PPS	[33, 34, 38, 42, 81, 86–88, 91, 92, 99, 104, 105, 111, 117]	[40, 43, 45–52, 54, 57, 116, 118, 121, 122]
	Registry	[30–32, 34–37, 41, 58–60, 81, 87–89, 91, 93, 96–98, 100, 101, 103, 106, 107, 109, 113–115, 119, 120]	[39, 47, 57]
	Others	[31, 34, 35, 41, 58–60, 90, 99, 101, 105, 112]	[53, 56]
Methods	ANOVA–ANCOVA	[41, 105, 112]	[44–46, 49, 56, 121, 122]
	Clustering	[32, 35, 58, 91]	
	Linear or logistic models	[30, 35, 36, 38, 86, 93, 100, 102, 103, 109, 113–115, 117, 119]	[50, 55]
	Param and non param tests	[30, 32, 35, 41, 58, 96, 98, 101–104, 106, 107, 110–114, 119, 120]	[44, 46–49, 51, 53–55, 118, 121, 122]
	Survival models	[30, 31, 34, 37, 107, 120]	[44, 47, 52, 54, 118, 122]
	Others	[32, 33, 37, 42, 59, 60, 81, 86, 87, 89, 91, 92, 110, 111, 113, 117]	[40, 43, 48, 49, 51–53, 56, 57, 94, 116, 118, 121–123]
Diseases	Cardiovascular diseases	[38, 96, 98, 101, 105, 107, 113, 114]	[39, 40, 43–57, 94, 95, 108, 116, 118, 121–123]
	Cancer	[30, 31, 34, 42, 81, 87–91, 97, 99, 106, 112, 115, 117, 120]	
	Diabetes	[37, 60, 92, 100, 105]	
	Neurological diseases	[33, 35, 93, 104, 109, 110]	[57]
	Lung diseases	[32, 90, 111]	
	Kidney diseases	[59, 60, 119]	
	Others	[32, 41, 58, 86, 102, 103]	[47]

encompass the registry, i.e. a longitudinal concept containing many concepts related to longitudinality.[4]

The T3 graph highlights two standpoints regarding MI studies: firstly that of clinicians who study MI, its risks and aggravating factors to gain insight into preventing and, if necessary, managing these patients. Secondly, that of patients with coronary symptoms, which could progress to incidents, which could then progress to AMI requiring hospitalisation and with high risk of mortality depending on the patient's age. This partly answers **Q7**.

The *text clustering* enabled summarisation of the results in order to list the topics studied, asked in **Q2**, in these

[4] November 6, 1995: Decree relative to the Registers National Committee.

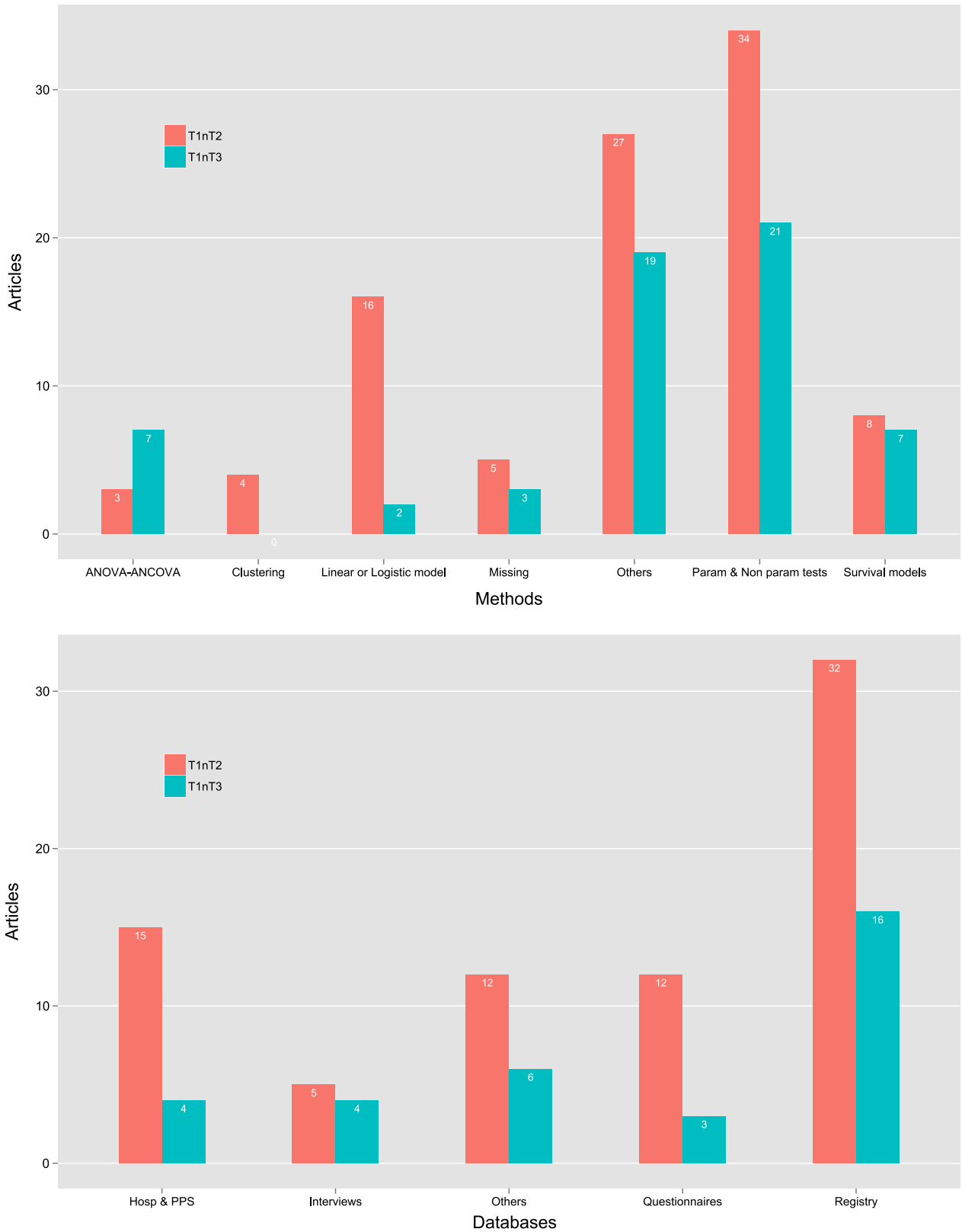

Fig. 6 Databases and methods used in trajectories studies

articles concerning patient trajectories. The first topic that we covered is disease with, for example, metabolic disorders such as diabetes and cardiovascular complications. Certain articles addressed patients' feelings, anxieties and disease experience. In the patient trajectory, there was support from the patient's immediate relations and family, but also health services, such as home nurses. Other articles focused on end-of-life situations, palliative care and processes set up to manage this last stage of the disease. Another topic was clinical research, involving developing cohorts, data collection, and methods used in different studies. Other studies concentrated on hospital organisation, various services, patient care staff, and associated costs. Other articles were focused on the health regulations and recommendations from guides of good practices.

As a response to **Q4**, our conclusion regarding the two T2 clusterings was that PPS is used in research primarily in the study of diseases, sometimes on disease onset, especially on disease management, associated costs, treatment and possible complications, but also in its coding. The studied diseases included neurological disorders, cancer, irregular heartbeat and cardiovascular diseases, the implantable medical devices to regulate these anomalies, traumas and wounds, infectious diseases, organ transplants, genetic and autoimmune diseases, and finally renal failure. Pregnancy and birth are also studied.

The T3 results, in reply to **Q6** and to complete response to **Q7**, showed that MI is studied from several aspects, with the first regarding the risk factors (socioeconomic, age, hypertension, diabetes). Then there are the biochemical and cardiocirculatory functioning aspects, the various mechanisms which lead to MI and genetic predisposition [45]. In addition, there is the psychological aspect of ill patients and the consequences. There is also emergency management before hospital admission, including transport and first aid. Then there is care at admission, medication management [44] and associated costs—here the trajectory concept emerges. There is also an aspect regarding the effectiveness of the measures implemented [46–52] and the different treatments [53–55]. Another investigated aspect concerns lifestyle, with regard to dietary habits, healthy lifestyle [56], comorbidities [43, 57] (smoking and/or alcohol), but also environmental factors like atmospheric pollution.

Step 3: thorough analysis of the selected articles

A thorough analysis of the selected articles was performed. Trajectory studies require, first and foremost, a definition of this concept, which is the focus of question **Q8**. The results showed that in most cases trajectory is characterised by care processes established for a specific disease to improve patient care, facilitate health planning within institutions, ensure prevention, predict the course of the disease and prevent the onset of symptoms.

In response to **Q9,** we found that interest in patient trajectory studies have increased in the last 5 years. The resurgence of studies in 2013 could be explained by the improvement in the quality of databases as of 2009 (ref), particularly in France, and the possibility of chaining hospital stays and reconstructing patient care trajectories throughout the country.

This interest in trajectories mainly stems from Europe and the Americas with 47 and 29% of studies, respectively. The PPS concept necessarily led to only including countries with a similar health system database organisation. This is a weakness of our study, since countries with a different information health system to the American model were not selected through this filter. Thus we have answered **Q10**.

Then we sought to determine the rationale for why these studies were conducted and provide a response to **Q11**. The six-category article distribution we defined showed that the aim of most of the studies was to compare treatments, techniques or care procedures. In each case, the aim was to reduce costs while improving the quality of care. Patient healthcare trajectory studies appeared beneficial in two ways: (1) First, the trajectory provides insight into the course of the disease following medical and surgical care. (2) Secondly, the trajectory may be highly informative regarding the medico-economic aspects so as to be able to streamline the patient's care management to avoid treatment dispersion.

In addition, the methods used underpinned the rationale of comparative studies as part of care techniques, treatment or care processes. These methods (Anova, comparisons tests, survival models, linear or logistic regression, etc.) are listed in the second part of Fig. 6 which solves question **Q12**.

We pursued this investigation by assessing the study characteristics, and answered **Q13**. In the studies, the number of recruited patients was estimated a priori for statistical analyses in good conditions with sufficient power. However, we identified a few studies that were conducted on the entire population, without sampling.

Overall, the study time was short, not more than a few years, which could be explained by economic considerations or a lack of data. For retrospective studies, for example, it was sometimes hard to trace back several years because the information is deleted after a certain period of time.

Next, we investigated the origin of the data used. For T1∩T3, registry data were mostly used. For T1∩T2, hospital databases and hospitalisation billing databases were used, so the studies were mostly hospital-based. Moreover, apart from hospital databases, some studies took

patients' feelings into account via interviews and questionnaires. Some studies required additional information on, for instance, medication [31, 58] through pharmacy databases or non-hospital care [59, 60] with social security databases for complete patient care monitoring.

To supplement previous findings concerning the list of diseases studied, for T1∩T2, the patient trajectories were closely focused on different cancers. Note that this brings us back to the results that were highlighted in step 2. We thus resolved question **Q14.**

Methods analysis

There are many different text mining techniques which are being constantly developed for literature searches and systematic review [6]. In systematic reviews, text mining techniques are used for four purposes [61]: (1) automatic terms recognition to identify and extract terms automatically from texts [62]; (2) document classification by generating subsets of documents focused on a specific topic [63–66]; (3) document clustering to group documents into topics. These correspond to topics shared by all the documents in the group they contain and by no other document in the collection [17, 67, 68]; (4) drafting abstracts by selecting sentences from each document based on the significance of its terms, which are combined via classification techniques [69]. Some authors used text mining for other purposes. For example, in [70] the authors created correspondence databases linking authors with the name abbreviations and processed a co-authorship analysis. In [71] the authors annotated abstracts in two ways, first the gene or protein of interest, then the protein interactions and/or gene functions. Ultimately, they categorised documents according to these annotations. Thus, combining text mining methods for systematic review is a hot topic [72–75].

While there is no consensus for a method in conducting a review with a huge number of documents, there are several techniques in text mining already used in various fields to explore text data [76–78]. Here, we wished to gain an overview of the document content in a recent developing field of inquiry, in order to provide general information and to respond to research questions. Our aim is to maximise the recall to ensure comprehensive study. We also aimed to better select publications, then reviewed them in a classical manner, by creating filters. With our method, searches are conducted based on the meaning of the words and concepts emerging from classification techniques rather than simply the presence of this term and concept. Thus, we conducted an in-depth study to explore the texts, starting by highlighting keywords, which were often used in the abstracts. Word cloud representation was most suited for this step, as it enabled a quick visual reading of the results. However, beyond the visual data display, word clouds do not provide much information.

One way to gain further insight is to highlight a lexical universe attached to those keywords. Thus, the same word may be interpreted differently depending on the terms associated with it. Similarity analysis best addresses this issue. Its tree construction approach connects highly co-occurring networks of terms and allows a better understanding of the most frequently discussed themes through the various items making up each corpus.

The last step in the exploration process is to determine whether it is possible to classify these articles in the topics highlighted by similarity analysis. We compared these results by using Reinert clustering because it has the advantage of respecting the text construction. It is also offers more flexibility than latent Dirichlet allocation (LDA), for example, where the researcher has to pre-determine the number of clusters. Although some authors have proposed solutions for the "optimal" number of topics in topic modelling [79, 80], it is not possible to verify, making this method even harder to apply.

The text mining methods that we selected have proven to be effective in exploring the corpora without a priori and with open-ended questions, allowing us to quickly identify documents associated with subjects beyond genetics. This facilitated the filtering of articles to apply methods with a priori to answer specific questions. Although existing methods for exploring text data to conduct rapid reviews are good, we hoped to validate a non-traditional methodology to conduct more extensive systematic reviews for future research.

Conclusion

In this article, a semi-automatic text mining methodology was applied to investigate patient healthcare trajectory. Patterns were extracted and identified semi-automatically from the published articles in PubMed. With text mining techniques we could analyse large amounts of text data, which would have not been possible otherwise. The originality of our approach lies in assisting a research review on the basis of a semantic approach, from research questions to targeted documents which will be then thoroughly analysed. This method is well-adapted for complex review questions or hard to define topics such as those addressed in public health and more particularly in the context of patient healthcare trajectory literature. Finally, our strategy enabled us to explore the concept of trajectory in the care domain.

We illustrated our search using a frequent cause of hospital stay, the occurrence of a MI. We chose to trace the follow-up of these MI patients through the PPS. We addressed open-ended questions by determining the topics covered in each area, to explore areas transversely,

while highlighting studies dealing with patient trajectories with regard to MI, based on PPS data. This semantic approach was demonstrated to be well-tailored for addressing our issues.

Document retrieval on the patient trajectories was combined with two major themes, i.e. PPS databases and MI. The findings showed that this type of study is of interest in the biomedical community; for comparative trajectories of drug prescriptions and costs Sundberg et al. concluded [58] that: "Drug prescriptions and costs of analgesics increased following conventional care and decreased following integrative care, indicating potentially fewer adverse drug events and beneficial societal cost savings with integrative care". Similarly, with regards to access to the appropriate treatment in time for cancer patients, Defossez et al. affirmed [81] that "There is in particular a need to describe and analyse cancer care trajectories and to produce waiting time indicators… The evaluation shows the ability of an integrated regional information system to formalise care trajectories and automatically produce indicators for time-lapse to care instatement, of interest in the planning of care in cancer." Our study revealed that the trajectory concept, regardless of its form, is being explored, analysed and exploited, especially in oncology through the oncology communicative medical file and multidisciplinary meetings.

To complete this research, it would now be interesting to include studies on patient trajectories in electronic health records. Some recent studies have focused on the use of these new technologies in order to offer patients with mobility difficulties integrated care by pooling electronic records from patients, caregivers or healthcare teams as well as doctors' follow-ups [82]. However, the implementation of such processes requires considerable organisation and adequate resources [83] and can lead to technical interoperability problems [84].

We were also studying patient trajectories in a health environment with MI. We obtained DRG sequences by chaining hospital stays. These sequences represent the chronological pattern of hospital healthcare of patients. We have characterised patient trajectories by such DRG sequences. We have applied sequential pattern mining techniques [85] to our trajectories in order to highlight frequent hospital trajectory patterns. To our knowledge, this is the first time that this type of approach, by applying sequential patterns to hospital data or registry data, has been used. Our ultimate goal is to build a predictive model of MI trajectories to simulate disease progress in the coming years so as to help anticipate health needs.

Abbreviations

AMI: Acute myocardial infarction; DRG: Diagnosis related group; DHC: Divisive hierarchical clustering; ICD: International classification of diseases; IQ: Interquartile interval; LDA: Latent Dirichlet allocation; MI: Myocardial infarction; PMSI: Programme de médicalisation du système d'information; PPS: Prospective payment system.

Author details

[1] Biostatistics, Epidemiology and Public Health Department, Nîmes University Hospital, Place R Debré, 30 029 Nîmes, France. [2] UPRES EA 2415, Clinical Research University Institute, 641 av du Doyen Gaston Giraud, 34 093 Montpellier, France. [3] LIRMM, UMR 5506, Montpellier University, 860 rue de Saint Priest – Bât 5, 34 095 Montpellier Cedex 5, France. [4] AMIS, Paul Valéry University, Montpellier, France.

Authors' contributions

Jessica Pinaire (JP) performed data acquisition and conducted analysis. Interpretation was led by Paul Landais (PL). Jérôme Azé (JA) and Sandra Bringay (SB) contributed to the conception and design of the study. JP drafted the manuscript, PL, JA, SB critically revised the manuscript. We confirm that the final version of the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all authors.

Acknowledgements

We warmly thank Sarah Kabani for her expert editing of the manuscript.

Competing interests

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article.

Funding section

This research was partly supported by Nîmes University Hospital, and LIRMM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in publishedmaps and institutional affiliations.

Received: 2 March 2017 Accepted: 29 March 2017
Published online: 12 April 2017

References

1. Grant JB, Hayes RP, Pates RD, Elward KS, Ballard DJ. HCFA's health care quality improvement program: the medical informatics challenge. J Am Med Inf Assoc. 1996;3:15–26.
2. Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust N Z J Public Health. 1999;23:453–9.
3. Kendrick S, Clarke J. The Scottish record linkage system. Health Bull. 1993;51:72–9.
4. Saranummi N, Ensio A, Laine M, Nykänen P, Itkonen P. National health IT services in Finland. Methods Inf Med. 2007;46:463–9.
5. Le Bihan-Benjamin C, Landais P, Chatellier G. Linking hospital stays in the national PMSI MCO database improved between 2006 and 2009: analysis and consequences. J Écon Méd. 2012;30:17–30.
6. Le Manach Y, Collins G, Rodseth R, Le Bihan-Benjamin C, Biccard B, Riou B, et al. Preoperative score to predict postoperative mortality (POSPOM): derivation and validation. J Am Soc Anesthesiol. 2016;124:570–9.
7. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: what interest for medical research? Rev Médecine Interne. 2015;36:411–7.

8. Le Manach Y, Collins G, Bhandari M, Bessissow A, Boddaert J, Khiami F, et al. Outcomes after hip fracture surgery compared with elective total hip replacement. JAMA. 2015;314:1159–66.
9. Colas S, Collin C, Piriou P, Zureik M. Association between total hip replacement characteristics and 3-year prosthetic survivorship: a population-based study. JAMA Surg. 2015;150:979–88.
10. Van Hecke A, Heinen M, Fernandez-Ortega P, Graue M, Hendriks J, Høy B, et al. Access to effective healthcare: effective self-management support intervention for patients with a chronic condition and a low social economic status: a systematic review. BMC Nurs. 2015;14:1–2.
11. Cohen AM, Hersh WR, Peterson K, Yen P-Y. Reducing workload in systematic review preparation using automated citation classification. J Am Med Inf Assoc. 2006;13:206–19.
12. Ratinaud P, Déjean F. IRaMuTeQ : implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre. MASHS2009, Toulouse; 2009.
13. Ratinaud P, Marchand P. Application de la méthode ALCESTE à de « gros » corpus et stabilité des « mondes lexicaux » : analyse du « CableGate » avec IRaMuTeQ ». Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles; 2012. p. 835–844.
14. Flament C. Similarity analysis: a technique for researches in social representations. Cah Psychol Cogn. 1981;1:375–95.
15. Frutcherman TMJ, Reingold EM. Graphed drawing by force directed placement. Softw Pract Exp. 1991;21:1129–64.
16. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol. 2001;25:163–77.
17. Reinert A. Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. Cah. L'analyse Données. 1983;8(2):187–98.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
19. Jay N, Nuemi G, Gadreau M, Quantin C. A data mining approach for grouping and analyzing trajectories of care using claim data: the example of breast cancer. BMC Med Inf Decis Mak. 2013;13:130.
20. Burke T, Manglani Y, Altawil Z, Dickson A, Clark R, Okelo S, et al. A safe-anesthesia innovation for emergency and life-improving surgeries when no anesthetist is available: a descriptive review of 193 consecutive surgeries. World J Surg. 2015;39:2147–52.
21. Cresci S, Wu J, Province MA, Spertus JA, Steffes M, McGill JB, et al. Peroxisome proliferator-activated receptor pathway gene polymorphism associated with extent of coronary artery disease in patients with type 2 diabetes in the bypass angioplasty revascularization investigation 2 diabetes trial. Circulation. 2011;124:1426–34.
22. Davis LA, Polk B, Mann A, Wolff RK, Kerr GS, Reimold AM, et al. Folic acid pathway single nucleotide polymorphisms associated with methotrexate significant adverse events in United States veterans with rheumatoid arthritis. Clin Exp Rheumatol. 2014;32:324–32.
23. Park JY, Lee S-H, Shin M-J, Hwang G-S. Alteration in metabolic signature and lipid metabolism in patients with angina pectoris and myocardial infarction. PLoS ONE. 2015;10:e0135228.
24. Pedersen ER, Tuseth N, Eussen SJPM, Ueland PM, Strand E, Svingen GFT, et al. Associations of plasma kynurenines with risk of acute myocardial infarction in patients with stable angina pectoris. Arterioscler Thromb Vasc Biol. 2015;35:455–62.
25. Peters BJM, Pett H, Klungel OH, Stricker BHC, Psaty BM, Glazer NL, et al. Genetic variability within the cholesterol lowering pathway and the effectiveness of statins in reducing the risk of MI. Atherosclerosis. 2011;217:458–64.
26. Suresh R, Li X, Chiriac A, Goel K, Terzic A, Perez-Terzic C, et al. Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction. J Mol Cell Cardiol. 2014;74:13–21.
27. Nubukpo P. Place of the opioid system in biology and treatment of alcohol use disorder. L'Encéphale. 2014;40:457–67.
28. Tada Y, Hiroshima K, Shimada H, Morishita N, Shirakawa T, Matsumoto K, et al. A clinical protocol to inhibit the HGF/c-Met pathway for malignant mesothelioma with an intrapleural injection of adenoviruses expressing the NK4 gene. SpringerPlus. 2015;4:358.
29. Zhang Y, Nester CM, Martin B, Skjoedt M-O, Meyer NC, Shao D, et al. Defining the complement biomarker profile of C3 glomerulopathy. Clin J Am Soc Nephrol CJASN. 2014;9:1876–82.
30. Guldbrandt LM, Fenger-Grøn M, Rasmussen TR, Jensen H, Vedsted P. The role of general practice in routes to diagnosis of lung cancer in Denmark: a population-based study of general practice involvement, diagnostic activity and diagnostic intervals. BMC Health Serv Res. 2015;15:21.
31. Harlos C, Musto G, Lambert P, Ahmed R, Pitz MW. Androgen pathway manipulation and survival in patients with lung cancer. Horm Cancer. 2015;6:120–7.
32. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat Commun. 2014;5:4022.
33. Danielsson U, Bengs C, Lehti A, Hammarström A, Johansson EE. Struck by lightning or slowly suffocating-gendered trajectories into depression. BMC Fam Pract. 2009;10:56.
34. Jiwa M, Maujean E, Spilsbury K, Threlfal T. The trajectory of lung cancer patients in Western Australia—a data linkage study: still a grim tale. Lung Cancer. 2010;70:22–7.
35. Schwartz CE, Quaranto BR, Healy BC, Benedict RH, Vollmer TL. Cognitive reserve and symptom experience in multiple sclerosis: a buffer to disability progression over time? Arch Phys Med Rehabil. 2013;94:1971–81.
36. Sieberg CB, Simons LE, Edelstein MR, DeAngelis MR, Pielech M, Sethna N, et al. Pain prevalence and trajectories following pediatric spinal fusion surgery. J Pain. 2013;14:1694–702.
37. Gebregziabher M, Egede LE, Lynch CP, Echols C, Zhao Y. Effect of trajectories of glycemic control on mortality in type 2 diabetes: a semiparametric joint modeling approach. Am J Epidemiol. 2010;171:1090–8.
38. Palmer WL, Bottle A, Davie C, Vincent CA, Aylin P. Meeting the ambition of measuring the quality of hospitals' stroke care using routinely collected administrative data: a feasibility study. Int J Qual Health Care. 2013;25:429–36.
39. Aeyels D, Van Vugt S, Sinnaeve PR, Panella M, Van Zelm R, Sermeus W. Lack of evidence and standardization in care pathway documents for patients with ST-elevated myocardial infarction. Eur J Cardiovasc Nurs. 2015;15:45–51.
40. Kesavan S, Kelay T, Collins RE, Cox B, Bello F, Kneebone RL, et al. Clinical information transfer and data capture in the acute myocardial infarction pathway: an observational study. J Eval Clin Pract. 2013;19:805–11.
41. Biffl WL, Smith WR, Moore EE, Gonzalez RJ, Morgan SJ, Hennessey T, et al. Evolution of a multidisciplinary clinical pathway for the management of unstable patients with pelvic fractures. Ann Surg. 2001;233:843–50.
42. Diaz RJ, Laughlin S, Nicolin G, Buncic JR, Bouffet E, Bartels U. Assessment of chemotherapeutic response in children with proptosis due to optic nerve glioma. Child's Nerv Syst. 2008;24:707–12.
43. Myers V, Drory Y, Gerber Y. Israel study group on first acute myocardial infarction. Clinical relevance of frailty trajectory post myocardial infarction. Eur J Prev Cardiol. 2014;21:758–66.
44. Wang W, McKinnie SMK, Patel VB, Haddad G, Wang Z, Zhabyeyev P, et al. Loss of Apelin exacerbates myocardial infarction adverse remodeling and ischemia-reperfusion injury: therapeutic potential of synthetic Apelin analogues. J Am Heart Assoc. 2013;2:e000249.
45. Ginzburg K, Solomon Z, Koifman B, Keren G, Roth A, Kriwisky M, et al. Trajectories of posttraumatic stress disorder following myocardial infarction: a prospective study. J Clin Psychiatry. 2003;64:1217–23.
46. Kinsman LD, Rotter T, Willis J, Snow PC, Buykx P, Humphreys JS. Do clinical pathways enhance access to evidence-based acute myocardial infarction treatment in rural emergency departments? Aust J Rural Health. 2012;20:59–66.
47. Kristoffersen DT, Helgeland J, Waage HP, Thalamus J, Clemens D, Lindman AS, et al. Survival curves to support quality improvement in hospitals with excess 30-day mortality after acute myocardial infarction, cerebral stroke and hip fracture: a before-after study. BMJ Open. 2015;5:e006741.
48. Bestul MB, McCollum M, Stringer KA, Burchenal J. Impact of a critical pathway on acute myocardial infarction quality indicators. Pharmacotherapy. 2004;24:173–8.
49. Kucenic MJ, Meyers DG. Impact of a clinical pathway on the care and costs of myocardial infarction. Angiology. 2000;51:393–404.

50. Mazzini MJ, Stevens GR, Whalen D, Ozonoff A, Balady GJ. Effect of an American Heart Association get with the guidelines program-based clinical pathway on referral and enrollment into cardiac rehabilitation after acute myocardial infarction. Am J Cardiol. 2008;101:1084–7.
51. Pelliccia F, Cartoni D, Verde M, Salvini P, Mercuro G, Tanzi P. Critical pathways in the emergency department improve treatment modalities for patients with ST-elevation myocardial infarction in a European hospital. Clin Cardiol. 2004;27:698–700.
52. Smith ORF, Kupper N, Denollet J, de Jonge P. Vital exhaustion and cardiovascular prognosis in myocardial infarction and heart failure: predictive power of different trajectories. Psychol Med. 2011;41:731–8.
53. Hagiwara MA, Bremer A, Claesson A, Axelsson C, Norberg G, Herlitz J. The impact of direct admission to a catheterisation lab/CCU in patients with ST-elevation myocardial infarction on the delay to reperfusion and early risk of death: results of a systematic review including meta-analysis. Scand J Trauma Resusc Emerg Med. 2014;22:67.
54. O'Donnell S, Condell S, Begley C, Fitzgerald T. Prehospital care pathway delays: gender and myocardial infarction. J Adv Nurs. 2006;53:268–76.
55. Lewis EF, Li Y, Pfeffer MA, Solomon SD, Weinfurt KP, Velazquez EJ, et al. Impact of cardiovascular events on change in quality of life and utilities in patients after myocardial infarction: a VALIANT study (Valsartan in acute myocardial infarction). JACC Heart Fail. 2014;2:159–65.
56. Rankin SH, de Leon JF, Chen J-L, Butzlaff A, Carroll DL. Recovery trajectory of unpartnered elders after myocardial infarction: an analysis of daily diaries. Rehabil Nurs. 2002;27:95–102.
57. Dharmarajan K, Hsieh AF, Kulkarni VT, Lin Z, Ross JS, Horwitz LI, et al. Trajectories of risk after hospitalization for heart failure, acute myocardial infarction, or pneumonia: retrospective cohort study. BMJ. 2015;350:h411.
58. Sundberg T, Petzold M, Kohls N, Falkenberg T. Opposite drug prescription and cost trajectories following integrative and conventional care for pain—a case-control study. PLoS ONE. 2014;9:e96717.
59. Couchoud C, Couillerot A-L, Dantony E, Elsensohn M-H, Labeeuw M, Villar E, et al. Economic impact of a modification of the treatment trajectories of patients with end-stage renal disease. Nephrol Dial Transpl. 2015;30:2054–68.
60. Bossuyt N, Van Casteren V, Goderis G, Wens J, Moreels S, Vanthomme K, et al. Public Health Triangulation to inform decision-making in Belgium. Stud Health Technol Inf. 2015;210:855–9.
61. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Res Synth Methods. 2011;2:1–14.
62. Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms: the C-value/NC-value method. Int J Digit Libr. 2000;3:115–30.
63. Frunza O, Inkpen D, Matwin S, Klement W, O'Blenis P. Exploiting the systematic review protocol for classification of medical abstracts. Artif Intell Med. 2011;51:17–25.
64. Joachims T. Text categorization with support vector machines: learning with many relevant. In: ECML-98; 1998. p. 137–142.
65. Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv. 2002;34:1–47.
66. Mo Y, Kontonatsios G, Ananiadou S. Supporting systematic reviews using LDA-based document representations. Syst Rev. 2015;4:175.
67. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003;3:993–1022.
68. Bada M. Mapping of biomedical text to concepts of lexicons, terminologies, and ontologies. Biomed Lit Min. 2014;1159:33–45.
69. Bollegala D, Okazaki N, Ishizuka M. A bottom-up approach to sentence ordering for multi-document summarization. Inf Process Manag. 2010;46:89–109.
70. Lin JM, Bohland JW, Andrews P, Burns GA, Allen CB, Mitra PP. An analysis of the abstracts presented at the annual meetings of the Society for Neuroscience from 2001 to 2006. PLoS ONE. 2008;3:e2052.
71. Leitner F, Valencia A. A text-mining perspective on the requirements for electronically annotated abstracts. FEBS Lett. 2008;582:1178–81.
72. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4:5.
73. Paynter R, Bañez LL, Berliner E, Erinoff E, Lege-Matsuura J, Potter S, et al. EPC methods: an exploration of the use of text-mining software in systematic reviews. Rockville: Agency for Healthcare Research and Quality (US); 2016.
74. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev. 2015;4:78.
75. Lefebvre C, Glanville J, Wieland LS, Coles B, Weightman AL. Methodological developments in searching for studies for systematic reviews: past, present and future? Syst Rev. 2013;2:78.
76. Teich E, Fankhauser P. Exploring a corpus of scientific texts using data mining. Lang Comput. 2009;71:233–47.
77. Lebart L, Salem A, Berry L. Exploring textual data. Text Speech and Language Technology, vol. 4. Dordrecht: Kluwer; 1998.
78. Van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. ISSI NewLetter. 2011;4:51–4.
79. Greene D, O'Callaghan D, Cunningham P. How many topics? Stability analysis for topic models. In: Machine learning and knowledge discovery in databases. Springer, New York; 2014. p. 498–513.
80. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinform. 2015;16:S8.
81. Defossez G, Rollet A, Dameron O, Ingrand P. Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer. BMC Med Inf Decis Mak. 2014;14:24.
82. Skinner I, Smith C, Jaffray L. Realist review to inform development of the electronic advance care plan for the personally controlled electronic health record in Australia. Telemed J E-Health. 2014;20:1042–8.
83. Dent M, Tutt D. Electronic patient information systems and care pathways: the organisational challenges of implementation and integration. Health Inf J. 2014;20:176–88.
84. Waterson P, Eason K, Tutt D, Dent M. Using HIT to deliver integrated care for the frail elderly in the UK: current barriers and future challenges. Work. 2012;41(Suppl 1):4490–3.
85. Rabatel J, Bringay S, Poncelet P. Mining sequential patterns: a context-aware approach. Advanced knowledge discovery management. New York: Springer; 2013. p. 23–41.
86. Popp AJ, Scrime T, Cohen BR, Feustel PJ, Petronis K, Habiniak S, et al. Factors affecting profitability for craniotomy. Neurosurg Focus. 2002;12:e4.
87. Ricciardi A, Largeron N, Giorgi Rossi P, Raffaele M, Cohet C, Federici A, et al. Incidence of invasive cervical cancer and direct costs associated with its management in Italy. Tumori. 2009;95:146–52.
88. Bettencourt-Silva JH, Clark J, Cooper CS, Mills R, Rayward-Smith VJ, de la Iglesia B. Building data-driven pathways from routinely collected hospital data: a case study on prostate cancer. JMIR Med Inf. 2015;3:e26.
89. Jensen H, Sperling C, Sandager M, Vedsted P. Agreement between patients and general practitioners on quality deviations during the cancer diagnostic pathway and associations with time to diagnosis. Fam Pract. 2015;32:329–35.
90. Palmer J, Bozas G, Stephens A, Johnson M, Avery G, O'Toole L, et al. Developing a complex intervention for the outpatient management of incidentally diagnosed pulmonary embolism in cancer patients. BMC Health Serv Res. 2013;13:235.
91. Thompson CA, Kurian AW, Luft HS. Linking electronic health records to better understand breast cancer patient pathways within and between two health systems. eGEMS. 2015;3:1127.
92. Ellis E, Ballance K, Lunt H, Lewis D. Diabetes outpatient care before and after admission for diabetic foot complications. J Wound Care. 2010;19:150–2.
93. Myklebust LH, Sørgaard KW, Bjorbekkmo S, Eisemann MR, Olstad R. Time-trends in the utilization of decentralized mental health services in Norway—a natural experiment: the VELO-project. Int J Ment Health Syst. 2010;4:5.
94. Song L, Yan H, Hu D, Yang J, Sun Y. Pre-hospital care-seeking in patients with acute myocardial infarction and subsequent quality of care in Beijing. Chin Med J (Engl). 2010;123:664–9.
95. Young W, McShane J, O'Connor T, Rewa G, Goodman S, Jaglal SB, et al. Registered nurses' experiences with an evidence-based home care pathway for myocardial infarction clients. Can J Cardiovasc Nurs. 2004;14:24–31.
96. Arko FR, Bohannon WT, Mettauer M, Lee SD, Patterson DE, Manning LG, et al. Retroperitoneal approach for aortic surgery: is it worth it? Cardiovasc Surg Lond Engl. 2001;9:20–6.

97. Baade PD, Youl PH, English DR, Mark Elwood J, Aitken JF. Clinical pathways to diagnose melanoma: a population-based study. Melanoma Res. 2007;17:243–9.
98. Buckley CJ, Lee SD, Arko FR, Bohannon WT, Mettauer M, Patterson DE, et al. Economic considerations for aortic surgery: retroperitoneal approach—is it worth it? Acta Chir Belg. 2000;100:247–50.
99. Ghosh K, Downs LS, Padilla LA, Murray KP, Twiggs LB, Letourneau CM, et al. The implementation of critical pathways in gynecologic oncology in a managed care setting: a cost analysis. Gynecol Oncol. 2001;83:378–82.
100. Goderis G, Van Casteren V, Declercq E, Bossuyt N, Van Den Broeke C, Vanthomme K, et al. Care trajectories are associated with quality improvement in the treatment of patients with uncontrolled type 2 diabetes: a registry based cohort study. Prim Care Diabetes. 2015;9:354–61.
101. Kinsman LD, Buykx P, Humphreys JS, Snow PC, Willis J. A cluster randomised trial to assess the impact of clinical pathways on AMI management in rural Australian emergency departments. BMC Health Serv Res. 2009;9:83.
102. Krummenauer F, Guenther K-P, Kirschner S. Cost effectiveness of total knee arthroplasty from a health care providers' perspective before and after introduction of an interdisciplinary clinical pathway—is investment always improvement? BMC Health Serv Res. 2011;11:338.
103. Miller PR, Fabian TC, Croce MA, Magnotti LJ, Elizabeth Pritchard F, Minard G, et al. Improving outcomes following penetrating colon wounds: application of a clinical pathway. Ann Surg. 2002;235:775–81.
104. Naqvi HA, Hussain S, Zaman M, Islam M. Pathways to care: duration of untreated psychosis from Karachi, Pakistan. PLoS ONE. 2009;4:e7409.
105. Tang W, Sun X, Zhang Y, Ye T, Zhang L. How to build and evaluate an integrated health care system for chronic patients: study design of a clustered randomised controlled trial in rural China. Int J Integr Care. 2015;15:e007.
106. van Hoeve J, de Munck L, Otter R, de Vries J, Siesling S. Quality improvement by implementing an integrated oncological care pathway for breast cancer patients. Breast Edinb Scotl. 2014;23:364–70.
107. Park YS, Chung SP, Chung HS, Lee HS, You JS. Implementation of a clinical pathway based on a computerized physician order entry system for ischemic stroke attenuates off-hour and weekend effects in the ED. Am J Emerg Med. 2014;32:884–9.
108. Laut KG, Foldspang A. The effects on length of stay of introducing a fast track patient pathway for myocardial infarction: a before and after evaluation. Health Serv Manag Res. 2012;25:31–4.
109. Ahmed S, Mayo N, Scott S, Kuspinar A, Schwartz C. Using latent trajectory analysis of residuals to detect response shift in general health among patients with multiple sclerosis article. Qual Life Res Int J. 2011;20:1555–60.
110. Cocchi A, Meneghelli A, Erlicher A, Pisano A, Cascio MT, Preti A. Patterns of referral in first-episode schizophrenia and ultra high-risk individuals: results from an early intervention program in Italy. Soc Psychiatry Psychiatr Epidemiol. 2013;48:1905–16.
111. Dely C, Sellier P, Dozol A, Segouin C, Moret L, Lombrail P. Preventable readmissions of "community-acquired pneumonia": usefulness and reliability of an indicator of the quality of care of patients' care pathways. Presse Médicale. 1983;2012(41):e1–9.
112. Klinkhammer-Schalke M, Lindberg P, Koller M, Wyatt JC, Hofstädter F, Lorenz W, et al. Direct improvement of quality of life in colorectal cancer patients using a tailored pathway with quality of life diagnosis and therapy (DIQOL): study protocol for a randomised controlled trial. Trials. 2015;16:460.
113. Mastenbroek MH, Denollet J, Versteeg H, van den Broek KC, Theuns DAMJ, Meine M, et al. Trajectories of patient-reported health status in patients with an implantable cardioverter defibrillator. Am J Cardiol. 2015;115:771–7.
114. Strömberg A, Fluur C, Miller J, Chung ML, Moser DK, Thylén I. ICD recipients' understanding of ethical issues, ICD function, and practical consequences of withdrawing the ICD in the end-of-life. Pacing Clin Electrophysiol PACE. 2014;37:834–42.
115. Veloso AG, Sperling C, Holm LV, Nicolaisen A, Rottmann N, Thayssen S, et al. Unmet needs in cancer rehabilitation during the early cancer trajectory—a nationwide patient survey. Acta Oncol Stockh Swed. 2013;52:372–81.
116. Martens EJ, Smith ORF, Winter J, Denollet J, Pedersen SS. Cardiac history, prior depression and personality predict course of depressive symptoms after myocardial infarction. Psychol Med. 2008;38:257–64.
117. Noble SI, Nelson A, Fitzmaurice D, Bekkers M-J, Baillie J, Sivell S, et al. A feasibility study to inform the design of a randomised controlled trial to identify the most clinically effective and cost-effective length of anticoagulation with low-molecular-weight heparin in the treatment of Cancer-Associated Thrombosis (ALICAT). Health Technol Assess Winch Engl. 2015;19:1–94.
118. Gerber Y, Benyamini Y, Goldbourt U, Drory Y. Israel Study Group on First Acute Myocardial Infarction. Prognostic importance and long-term determinants of self-rated health after initial acute myocardial infarction. Med Care. 2009;47:342–9.
119. Jayanti A, Wearden AJ, Morris J, Brenchley P, Abma I, Bayer S, et al. Barriers to successful implementation of care in home haemodialysis (BASIC-HHD):1. Study design, methods and rationale. BMC Nephrol. 2013;14:197.
120. Sverrisson EF, Zens MS, Fei DL, Andrews A, Schned A, Robbins D, et al. Clinicopathological correlates of Gli1 expression in a population-based cohort of patients with newly diagnosed bladder cancer. Urol Oncol. 2014;32:539–45.
121. Myers V, Drory Y, Gerber Y. Israel Study Group on First Acute Myocardial Infarction. Sense of coherence predicts post-myocardial infarction trajectory of leisure time physical activity: a prospective cohort study. BMC Public Health. 2011;11:708.
122. Gerber Y, Myers V, Goldbourt U, Benyamini Y, Scheinowitz M, Drory Y. Long-term trajectory of leisure time physical activity and survival after first myocardial infarction: a population-based cohort study. Eur J Epidemiol. 2011;26:109–16.
123. Rosenfeld AG. Treatment-seeking delay among women with acute myocardial infarction: decision trajectories and their predictors. Nurs Res. 2004;53:225–36.