

# An approach for mining care trajectories for chronic diseases

Elias Egho, Nicolas Jay, Chedy Raïssi, Gilles Nuemi, Catherine Quantin, Amedeo Napoli

▶ **To cite this version:**

Elias Egho, Nicolas Jay, Chedy Raïssi, Gilles Nuemi, Catherine Quantin, et al.. An approach for mining care trajectories for chronic diseases. AIME 2013 - 14th Conference on Artificial Intelligence in Medicine, May 2013, Murcia, Spain. hal-00883117

**HAL Id: hal-00883117**
**https://inria.hal.science/hal-00883117**

Submitted on 9 Nov 2013

**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.

# An approach for mining care trajectories for chronic diseases

Elias Egho[1],Nicolas Jay[1], Chedy Raïssi[2], Gilles Nuemi[3], Catherine Quantin[3], Amedeo Napoli[1]

[1] Orpailleur Team, LORIA, Vandoeuvre-les-Nancy, France
{firstname.lastname}@loria.fr
[2] INRIA, Nancy Grand Est, France
{firstname.lastname}@inria.fr
[3] Service de Biostatistique et d'Information Médicale , CHU de Dijon, Dijon, France
{firstname.lastname}@chu-dijon.fr

**Abstract.** With the increasing burden of chronic illnesses, administrative health care databases hold valuable information that could be used to monitor and assess the processes shaping the trajectory of care of chronic patients. In this context, temporal data mining methods are promising tools, though lacking flexibility in addressing the complex nature of medical events. Here, we present a new algorithm able to extract patient trajectory patterns with different levels of granularity by relying on external taxonomies. We show the interest of our approach with the analysis of trajectories of care for colorectal cancer using data from the French casemix information system.

**Keywords:** datamining, chronic illness, claim data, sequential pattern mining, trajectory of care

## 1 Introduction

Chronic illnesses are a major burden in both developed and developing countries[5]. Patients with chronic conditions use more services and a greater array of services than other consumers. Multiple encounters of chronic patients with the healthcare system define a so-called "trajectory" of care . Lack of coordination along the trajectory of care, bad implementation of guidelines or inappropriate organization of the healthcare system may have a negative impact on quality and costs of care.

Due to the fragmentation of clinical information systems, little knowledge is readily available to describe and assess the actual processes involved in long-term care, especially in the scope of a cross-institutional analysis. However, in many countries, health information systems routinely collect medical and administrative data at regional or national scale. Among them, case-mix information systems were originally built for hospital activity report and billing purpose[2]. They hold valuable information that could help health care managers and professionals to develop inter-organizational knowledge and bring deeper insights

into inpatient care trajectories. In order to produce the expected knowledge and support decision making, case-mix systems have to be turned into longitudinal and patient-centered information systems. This requires the linkage of different stays of a same patient into a sequence that will be further processed. Because of the complex nature and extreme diversity of medical problems, patient care trajectories must be summarized and categorized for allowing meaningful inference about outcomes of particular interest.

Data-mining methods are especially adapted to the analysis of sequences and successfully used in biomedical domain [3, 4, 7, 1]. case-mix systems capture medical problems, procedures, demographic and administrative data using controlled vocabularies and standardized records. In that context, sequences of hospitalizations can be analyzed with sequential pattern mining algorithms[9]. Meanwhile, case-mix records have a multidimensional structure that traditional sequential patterns can not fully reflect. Moreover, the granularity of the initial data may be too fine to generate interesting patterns. The availability of classifications and ontologies used to code information in case-mix systems is an opportunity to integrate additional knowledge into the mining process and achieve better results. Although a few approaches have been developed to tackle the problems of granularity and multidimensionality in sequential pattern mining[8], they are still not adapted to the problem of mining care trajectories.

In this paper, we present an new algorithm, MMISP (Mining Multidimensional Itemsets Sequential Patterns). MMISP is able to extract patterns from care trajectories in a multidimensional temporal database, using external taxonomic knowledge at appropriate levels of granularity. We illustrate this approach in analysing care trajectories for colorectal cancer using data from the french case-mix information system.

## 2 Problem Statement

The PMSI[4] is the french adaptation of the Diagnoses Related Groups[2]. In the PMSI database, each stay is a standardized record of administrative and clinical data, especially about the institution, the patient's principal diagnosis and the realized medical procedures. In order to formalize the problem, we accordingly model each hospitalization along three dimensions: (*i*) healthcare institution, (*ii*) diagnosis and (*iii*) medical procedure. Two dimensions, i.e. healthcare institutions and diagnosis, are considered as ordered sets with an associated subsumption relation (i.e. a partial ordering). The set of healthcare institutions H, the set of diagnosis DG and the set of medical procedures MP, are given:

- H=$\{t_h, uh, gh, uh_p, uh_n, gh_p, gh_l\}$.
- DG=$\{t_d, c, r, c_1, c_2, r_1, r_2\}$.
- MP=$\{mp_1, mp_2, mp_3, mp_4\}$.

The subsumption relation for H and DG is defined as below (Figure 1).

[4] Programme de Médicalisation des Sytèmes d'Information

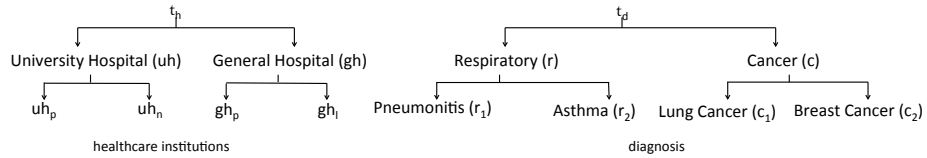



**Fig. 1.** The subsumption relation for the healthcare institutions set H and the diagnosis set DG

**Definition 1.** *A* **partially ordered set (poset)** *is a pair $(D, \leqslant)$, where $D$ is a set and $\leqslant$ is a partial order relation on $D$. For $x \in D$ the* **down set** *of $x$, denoted by $\downarrow x$, is a set of all specializations of $x$; $\downarrow x = \{y \in D | y \leqslant x\}$. The* **up-set** *of $x$ is $\uparrow x = \{y \in D | x \leqslant y\}$ .*

Among the three basic dimension, H and DG are posets. The hospitalization of a patient is then considered as a vector with 3 components, $(H, DG, MP)$.

*Example 1.* $(uh_p, c_1, \{mp_1, mp_2\})$ is an hospitalisation for a patient. It is a vector with three components $uh_p \in H, c_1 \in DG$ and $\{mp_1, mp_2\} \subseteq MP$.

**Definition 2.** *(Elementary vector) An elementary vector $v = (v_1, v_2, v_3)$ is a vector with 3 elements. Given two vectors $v = (v_1, v_2, v_3)$ and $v' = (v_1^{'}, v_2^{'}, v_3^{'})$, $v$ is more general than $v'$, denoted by $v' \leq_v v$, for every $i = 1...3$*

$v_i' \leqslant v_i$ *if $v_i, v_i'$ are elements in a poset*
$v_i \subseteq v_i^{'}$ *if $v_i, v_i'$ are sets*

*Example 2.* $v = (uh_p, c_1, \{mp_1, mp_2\})$ is a vector with 3 elements $uh_p$, $c_1$ and $\{mp_1, mp_2\}$. The vector $v^{'} = (uh, t_d, \{mp_1\})$ is more general than $v$, $v \leq_v v'$, because of:

- $uh_p \leqslant uh$; $uh_p, uh \in H$
- $c_1 \leqslant t_d$; $c_1, t_d \in DG$.
- $\{mp_1\} \subseteq \{mp_1, mp_2\}$; $\{mp_1\}, \{mp_1, mp_2\} \subseteq MP$.

**Definition 3.** *(Patient Trajectory) A patient trajectory is a pair ($V$,$<_t$), where $V$ is a set of elementary vectors and $<_t$ is a temporal order relation on $V$. The patient trajectory represents like $P = \langle P_1 P_2 ... P_l \rangle$, where $P_1, P_2, ..., P_l \in V$ and $P_1 <_t P_2 <_t P_3 ... <_t P_l$. Given two trajectories $P = \langle P_1 P_2 ... P_l \rangle$ and $T = \langle T_1 T_2 ... T_{l'} \rangle$, $P$ is more general than $T$, denoted by $T \leq_p P$, if there exist indices $1 \leq i_1 < i_2 < ... < i_l \leq l'$ such that $T_j \leq_v P_{i_j}$ for all $j = 1 \ldots l$ and $l \leqslant l'$. We say that $T$ is more specific than $P$.*

*Example 3.* $\langle (uh_p, c_1, \{mp_1, mp_2\})(gh_l, r_1, \{mp_2\}) \rangle$ represents a patient trajectory with two hospitalizations. It expresses the fact that a patient was admitted to the hospital $uh_p$ for a lung cancer $c_1$, and underwent procedures $mp_1$ and $mp_2$. Then he went to the hospital $gh_l$ for pneumonitis $r_1$ where he underwent procedure $mp_2$.

| Patients | Trajectories |
|---|---|
| $patient_1$ | $\langle(uh_p, c_1, \{mp_1, mp_2\})(uh_p, c_1, \{mp_1\})(gh_l, r_1, \{mp_3\})\rangle$ |
| $patient_2$ | $\langle(uh_n, c_1, \{mp_4\})(uh_p, c_2, \{mp_1, mp_2\})(gh_l, r_1, \{mp_2\})\rangle$ |
| $patient_3$ | $\langle(uh_p, c_1, \{mp_4\})(gh_l, r_2, \{mp_3\})\rangle$ |
| $patient_4$ | $\langle(uh_p, c_2, \{mp_1, mp_2\})(gh_p, r_2, \{mp_3\})(gh_l, r_2, \{mp_2\})\rangle$ |

**Table 1.** An example of a database of patient trajectories.

Let $P_{DB}$ be the patient trajectories for four patients $patient_1$, $patient_2$, $patient_3$ and $patient_4$, Table 1.

Let `supp(P)` be the number of trajectories that are more specific than $P$ in $P_{DB}$ and $\sigma$ be a minimum support threshold specified by the end-user. Let P be a trajectory, P is a frequent trajectory pattern in $P_{DB}$ if and only if $supp(P) \geq \sigma$.

Using the poset for some dimensions, we can extract a large number of frequent trajectory patterns. To avoid the patterns overfloading, our approach only extracts the set of all most specific frequent trajectory patterns in $P_{DB}$. Actually, frequency is anti-monotonic (i.e if $P = \langle(uh_p, c1, \{mp_1, mp_2\})\rangle$ is a frequent then $T = \langle(uh, c, \{mp_1\})\rangle$ which is more general than $P$ is also frequent). So, all the most specific frequent trajectory patterns can lead to some general frequent trajectory patterns.

**Definition 4.** *(Most Specific Frequent Trajectory) Let $P$ be a trajectory. $P$ is a most specific frequent trajectory, if and only if: $supp(P) \geq \sigma$ and for all $T$ such that $T \leq_p S$; $supp(T) \leqslant \sigma$*

*Example 4.* Let $\sigma = 0.75$ (i.e. a trajectory is frequent if it appears at least three times in $P_{DB}$). The trajectory $P = \langle(uh_p, c, \{mp_1, mp_2\})\rangle$ is frequent. $T = \langle(uh, c, \{mp_1\})\rangle$ is also frequent. Nevertheless, $T$ is not a most specific frequent trajectory pattern while $P$ is one.

## 3 Mining patient trajectory patterns

In this section, we present an approach for extracting all the most specific frequent trajectory patterns from patients trajectories. Our approach is called **MMISP** (*Mining Multidimensional Itemsets Sequential Patterns*). The basic idea of MMISP is finding a way to transfer the multidimensional itemsets sequential database into a classical sequential database (i.e. sequence of itemsets). So, MMISP is based on three steps:

1. Extract all the frequent elementary vector $v$ without taking into account the temporal relation between them in each trajectory.
2. Map the frequent elementary vectors which extracted in the first step to an alternate representation. Then, the patient trajectories are encoded by using the new representation of frequent elementary vectors.
3. Apply a standard sequential mining algorithm to enumerate frequent patient trajectories.

### 3.1 Generating frequent elementary vectors

MMISP starts by searching for the frequent elementary vectors in the trajectories. MMISP firstly studies the patient's trajectory like a set of elementary vectors without taking into account the temporal relation order between them. The support of elementary vector $v$ is defined as follows,

**Definition 5.** *(Support of elementary vector v, supp(v)) Let $P_{DB}$ be a database of patient trajectories with m patients and let $P =< P_1P_2...P_l >$ be a patient trajectory in $P_{DB}$. The support of elementary vector v is defined as follows*

$$supp(v) = \frac{|\{P \in P_{DB}; \exists j \in [1,..,l];\ P_j \leq_v v\}|}{m}$$

*Example 5.* In our example, the support of $(gh, r, \{mp_3\})$ is $\frac{3}{4}$, because of:

- $(gh_l, r_1, \{mp_3\}) \in patient_1$ where $(gh_l, r_1, \{mp_3\}) \leq_v (gh, r, \{mp_3\})$.
- $(gh_l, r_2, \{mp_3\}) \in patient_3$ where $(gh_l, r_2, \{mp_3\}) \leq_v (gh, r, \{mp_3\})$.
- $(gh_p, r_2, \{mp_3\}) \in patient_4$ where $(gh_p, r_2, \{mp_3\}) \leq_v (gh, r, \{mp_3\})$.

MMISP generates all the frequent elementary vectors by building a poset $(L, \leq_v)$. Building $(L, \leq_v)$ is done as follows:

- Firstly, we generate the most general elementary vector. In our running example, we have two dimensions with posets $H$ and $DG$ and one dimension with a set $MP$, so the most general elementary vector is $(t_h, t_d, \{\})$.
- Then, the recursive generation of the new elementary vectors continues by using each previously generated frequent elementary vector $(v)$. For each element $v_1, v_2, v_3 \in v$, we replace $v_k$, where $k \in [1, 3]$ with each of its specialization from the set $special(v_k)$. At each step, we take only the frequent elementary vector which has support greater than $\sigma$.

We define the set $special(v_i)$ as follows:

**Definition 6.** *Let $v_i$ be the $i^{th}$-element in the vector $v = (v_1, v_2, v_3)$ and let D be the ground set of the component $v_i$*

$$special(v_i) = \begin{cases} \{a \in D; a \leq v_i \text{ and } \nexists b \in D;\ a \leq b \text{ and } b \leq v_i\} & \text{if } D \text{ is a poset} \\ \{v_i \cup \{a\}; a \in D \setminus v_i\} & \text{if } D \text{ is a set} \end{cases}$$

*Example 6.* In our example, $special(t_h)$=$\{uh, gh\}$, $special(t_d)$=$\{r, c\}$ and $special(\{\})$ =$\{\{mp_1\},\ \{mp_2\},\ \{mp_3\}, \{mp_4\}\}$. With $\sigma = \frac{3}{4}$ we can generate new seven frequent elementary vectors from $(t_h, t_d, \{\})$. They are $(uh, t_d, \{\})$, $(gh, t_d, \{\})$, $(t_h, r, \{\})$ , $(t_h, c, \{\})$, $(t_h, t_d, \{mp_1\})$,$(t_h, t_d, \{mp_2\})$ and $(t_h, t_d, \{mp_3\})$. The first and the second are generated by replacing $t_h$ by $child(t_h)$, the third and the forth are generated by replacing $t_d$ by $special(t_d)$, and the rest are generated by replacing $\{\}$ by $special(\{\})$.

The objective of MMISP is to generate all the most specific frequent patient trajectories, thus it retains only the most specific frequent elementary vectors from $(L, \leq_v)$ .

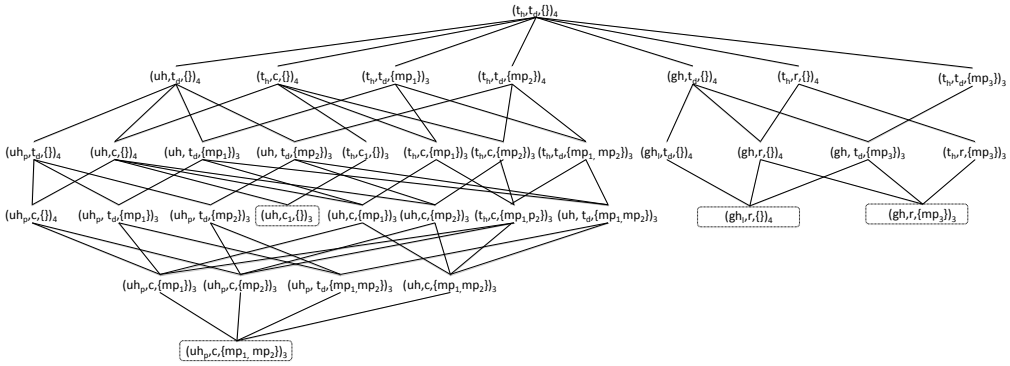


**Fig. 2.** The poset $(L, \leq_v)$ is generated by taking into account the two posets H and DG in Figure 1 and the set $MP = \{mp_1, mp_2, mp_3, mp_4\}$ with minsup= $\frac{3}{4}$.

**Definition 7.** *(Most specific frequent elementary vector, MSFV ) Let $v$ be an elementary vector, $v$ is a most specific frequent elementary vector, if and only if $supp(v) \geq \sigma$ and $\nexists\ v'$ an elementary vector, where $supp(v) = supp(v')$ and $v' \leqslant_v v$.*

| id | *MSFV* |
|---|---|
| 1 | $(uh_p, c, \{mp_1, mp_2\})$ |
| 2 | $(uh, c_1, \{\})$ |
| 3 | $(gh_l, r, \{\})$ |
| 4 | $(gh, r, \{mp_3\})$ |

**Table 2.** The most specific frequent elementary vectors extracted from $(L, \leq_v)$ in Figure 2.

*Example 7.* Figure 2 illustrates the generation of all frequent elementary vectors on our example with $\sigma = \frac{3}{4}$. Table 2 shows the hash table of all *MSFV* extracted from $(L, \leq_v)$ .

### 3.2 Mining patient trajectory

The next step of MMISP is studying the temporal relation between the most specific frequent elementary vectors extracted in previously step. This is done by taking each patient trajectory $P = \langle P_1 P_2, ..., P_l \rangle$ from the database of patient trajectories $P_{DB}$, then replacing each elementary vector $P_i \in P;\ i \in [1..l]$ with all elementary vectors $v \in MSFV$ where $P_i \leq_v v$.

*Example 8.* In our example, the trajectory of $patient_3$,$\langle(uh_p, c_1, \{mp_4\})(gh_l, r_2, \{mp_3\})\rangle$, is transformed into $\langle\{(uh, c_1, \{\})\}\{(gh_l, r, \{\}), (gh, r, \{mp_3\})\}\rangle$ because the first elementary vector of $patient_3$, $(uh_p, c_1, \{p_4\})$, can only be replaced with $(uh, c_1, \{\})$ from the *MSFV* set where $(uh_P, c_1, \{p_4\}) \leqslant_v (uh, c_1, \{\})$ and the second elementary vector of $patient_3$, $(gh_l, r_2, \{mp_3\})$, can be replaced by $(gh_l, r, \{\})$ and $(gh, r, \{mp_3\})$ from the *MSFV* set.

Table 3 shows the transformation of patient trajectories in $P_{DB}$ by using the set of all most specific frequent elementary vector *MSFV* in Table 2.

| Patients | Trajectories |
|---|---|
| $patient_1$ | $\langle\{(uh_p, c, \{mp_1, mp_2\}), (uh, c_1, \{\})\}\{(uh, c_1, \{\})\}\{(gh_l, r, \{\}), (gh, r, \{mp_3\})\}\rangle$ |
| $patient_2$ | $\langle\{(uh, c_1, \{\})\}\{(uh_p, c, \{mp_1, mp_2\})\}\{(gh_l, r, \{\})\}\rangle$ |
| $patient_3$ | $\langle\{(uh, c_1, \{\})\}\{(gh_l, r, \{\}), (gh, r, \{mp_3\})\}\rangle$ |
| $patient_4$ | $\langle\{(uh_p, c, \{mp_1, mp_2\})\}\{(gh, r, \{mp_3\})\}\{(gh_l, r, \{\})\}\rangle$ |

**Table 3.** Transforming a patient trajectories in Table 1 by using the set of all most specific frequent elementary vector in Table 2.

We apply a classical sequential pattern mining algorithm (e.g. [6, 11, 10]) to extract the frequent sequential patterns. This extraction has be done as follows: firstly we transform each patient trajectory into a sequence simple (i.e sequence of itemset like $< \{a, b\}\{a, d\} >$) and then we apply a CloSpan [10] on the transformation patient trajectories. The transformation has be done as follows:

- Each elementary vector in the *MSFV* set is assigned a unique id which will be used during the mining operation. This is illustrated in Table 2 .
- For each elementary vector $v$ in a patient trajectory in Table 3, we replace $v$ with its id in Table 2.

*Example 9.* In our example, the patient trajectory $patient_3 = \langle\{(uh, c_1, \{\})\}\{(gh_l, r, \{\}), (gh, r, \{mp_3\})\}\rangle$ in Table 3 is transformed into $\langle\{2\}, \{3, 4\}\rangle$, because $(uh, c_1, \{\})$ has an id 2, $(gh_l, r, \{\})$ has an id 3 and $(gh, r, \{mp_3\})$ has an id 4.

| Patients | Trajectories |
|---|---|
| $patient_1$ | $\langle\{1, 2\}\{2\}\{3, 4\}\rangle$ |
| $patient_2$ | $\langle\{2\}\{1\}\{3\}\rangle$ |
| $patient_3$ | $\langle\{2\}\{3, 4\}\rangle$ |
| $patient_4$ | $\langle\{1\}\{4\}\{3\}\rangle$ |

**Table 4.** Transformed database in Table 3

Table 5 displays all frequent sequences in their transformed format and the frequent patient trajectories in which identifiers are replaced with their actual values with minsup=$\frac{3}{4}$.

| Frequent sequential patterns | Frequent patient trajectory patterns | Support |
|---|---|---|
| $\langle\{3\}\rangle$ | $\langle(gh_l, c)\rangle$ | 1 |
| $\langle\{2\}\{3\}\rangle$ | $\langle(uh, c_1)(gh_l, c)\rangle$ | 0.75 |
| $\langle\{4\}\rangle$ | $\langle(gh, r, \{mp_3\})\rangle$ | 0.75 |
| $\langle\{1\}\{3\}\rangle$ | $\langle(uh_p, c, \{mp_1, mp_2\})(gh_l, c)\rangle$ | 0.75 |

**Table 5.** Frequent patient trajectory patterns with minsup=$\frac{3}{4}$.

## 4 Results

This section describes the results obtained with MMISP on a set of 2618 trajectories of care of patients from the Burgundy region in France. Using data from the PMSI, the so-called french case mix system, we reconstituted the sequence of hospitalizations of patients having undergone surgery for colorectal cancer between 2006 and 2008, with a one year follow-up. Each event in a sequence was characterized by the following dimensions : hospital, principal diagnosis, procedures delivered during the stay. The hospital dimension was associated with a geographical taxonomy of 4 levels : root (France), administrative region, administrative department, hospital. Principal Diagnosis could be described at 5 levels of the $10^{th}$ International classification of Diseases (ICD10): root , chapter, block, 3-character, 4-character, terminal nodes. Procedures were represented by their first CCAM[5] code.

Figure 3 shows the number of discovered patterns at different thresholds according to their length. The total number of patterns grows exponentially for support below 34%. However, the increase is extremely variable considering the length of patterns and the number of short patterns (length<6) is still manageable. The high number of length 7 patterns can probably be explained by a combinatorial effect resulting from a high number of sequences of length 14-15 in the database. They correspond to the patients who underwent chemotherapy and usually had around 14 and 15 stays for 1 cycle.

Table 6 shows the items appearing in the Principal Diagnoses dimension of patterns for which support is over 32%. It can be noticed that the ICD10 tree has been mined at different levels. In the neoplasm branch, the most specific observed item is of depth 3, Malignant neoplasm of colon. In the branch of "Factors influencing ...", items of depth 4 (chemotherapy session for neoplasm) have been extracted. Children of "Malignant neoplasms of colon" are not frequent enough

[5] Classification Commune des Actes Médicaux : the french classification of medical and surgical procedures

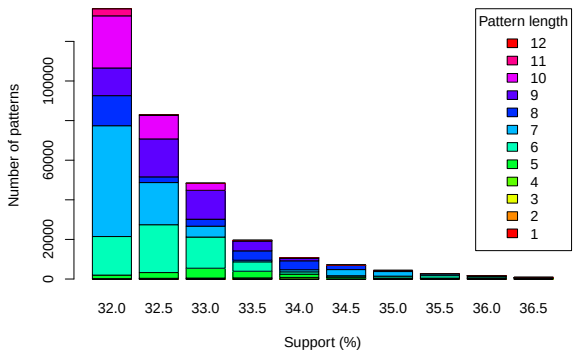



**Fig. 3.** Number of sequential patterns by support and length (stacked bars).

to be extracted, but "chemotherapy session" appears in a sufficient proportion of trajectories to be seen. Such results could not have been obtained by representing items at an arbitrary pre-determined level.

| ICD10 level – Item |
|---|
| 0– Root |
| 1– Neoplasms |
| 2– Malignant neoplasms of digestive organs |
| 3– Malignant neoplasm of colon |
| 1– Factors influencing health status and contact with health services |
| 2– Persons encountering health services for specific procedures and health care |
| 3– Other medical care |
| 4– Chemotherapy session for neoplasm |

**Table 6.** Items extracted in the Principal Diagnosis dimension, (minsupp=32%)

Multidimensional sequential patterns can be analysed per se. For example, the pattern $\langle$(Root,C15-C26, {Colectomy}), (Burgundy,Z00-Z99,{})$\rangle$ shows that 69% of patients had a colectomy for a digestive cancer and a subsequent stay in the Burgundy region for complementary treatments and follow-up. This kind of information can help healthcare managers and deciders in planning and organizing healthcare resources at a regional level. Besides, sequential patterns can be seen as condensed representations of the care trajectories. As such, they can be reused as new variables to distinguish subgroups of patients in subsequent analysis. As an illustrative example, we selected a subset of frequent patterns to analyze the relationship between accessibility of care facilities and trajectories of care, as it has been shown that geographical disparities might be related to less favourable outcome in terms of survival. The cumulative driving distance trav-

elled by patients to access facilities along their care trajectory was used to fit a classification tree with patterns as predictors. As expected, longer distances were associated to trajectories involving chemotherapy sessions. However differences were observed according to the occurrence of hospitalizations in specific places. In particular, patients initially treated outside of the burgundy region travelled longer distances. Theses findings can bring experts to investigate specific hypothesis regarding the links between organization of care and health outcomes.

## 5 Conclusion

Care trajectories of chronic patients can be analysed using administrative databases and sequential pattern mining. The MMISP algorithm relies on external knowledge to enrich the mining process and produces results with appropriate levels of granularity. Experiments on data from the french case-mix information system show that MMISP is flexible enough to reflect both the relational and temporal structure of the care trajectories.

## References


1. Iyad Batal, Lucia Sacchi, Riccardo Bellazzi, and Milos Hauskrecht. A temporal abstraction framework for classifying clinical temporal data. *AMIA Annu Symp Proc*, 2009:29–33, 2009.
2. RB Fetter, Y Shin, JL Freeman, RF Averill, and JD Thompson. Case mix definition by diagnosis-related groups. *Med Care*, 18(2):1–53, Feb 1980.
3. Zhengxing Huang, Xudong Lu, and Huilong Duan. On mining clinical pathway patterns from medical behaviors. *Artif Intell Med*, 56(1):35–50, Sep 2012.
4. Mingoo Kim, Hyunjung Shin, Tae Su Chung, Je-Gun Joung, and Ju Han Kim. Extracting regulatory modules from gene expression data by sequential pattern mining. *BMC Genomics*, 12 Suppl 3:S5, Nov 2011.
5. Ellen Nolte and Martin McKee, editors. *Caring for people with chronic conditions : A health system perspective*. Open University Press, 2008.
6. Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Meichun Hsu. Prefixspan: Mining sequential patterns by prefix-projected growth. In *ICDE*, pages 215–224, 2001.
7. François Petitjean, Florent Masseglia, Pierre Gançarski, and Germain Forestier. Discovering significant evolution patterns from satellite image time series. *Int J Neural Syst*, 21(6):475–489, Dec 2011.
8. Marc Plantevit, Anne Laurent, Dominique Laurent, Maguelonne Teisseire, and Yeow WEI Choong. Mining multidimensional and multilevel sequential patterns. *ACM Trans. Knowl. Discov. Data*, 4:4:1–4:37, January 2010.
9. Ramakrishnan Srikant and Rakesh Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '96, pages 3–17, London, UK, UK, 1996. Springer-Verlag.
10. Xifeng Yan, Jiawei Han, and Ramin Afshar. Clospan: Mining closed sequential patterns in large datasets. In *In SDM*, pages 166–177, 2003.
11. Mohammed J. Zaki. Spade: An efficient algorithm for mining frequent sequences. *Mach. Learn.*, 42(1-2):31–60, January 2001.