Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth
A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-391

PaFloChar: An Innovating Approach to Characterise Patient Flows in Myocardial Infarction

Jessica PINAIRE[a,b,c,1], Jérôme AZé[b], Sandra Bringay[b,d], Pascal PONCELET[b], Christophe GENOLINI[e] and Paul LANDAIS[a]
[a] *UPRES EA2415, Montpellier University, France*
[b] *LIRMM, UMR 5506, Montpellier University, Montpellier, France*
[c] *University Hospital Nîmes, Nîmes, France*
[d] *AMIS, Paul-Valéry University, Montpellier, France*
[e] *Start'up Zebrys, Toulouse, France*

Abstract. A better knowledge of patient flows would improve decision making in health planning. In this article, we propose a method to characterise patients flows and also to highlight profiles of care pathways considering times and costs. From medico-administrative data, we extracted spatio-temporal patterns. Then, we clustered time between hospitalisations and cost trajectories in order to identify profiles of change over time. This approach may support renewed management strategies.

Keywords. Healthcare Trajectories, Prospective Payment System, Myocardial Infarction, Spatio-temporal Pattern Mining, Patient Flows, Longitudinal Data.

1. Introduction

Over the past 20 years, studies related to Myocardial Infarction (*MI*) have pointed out a shift in gender patterns that showed an increase in both cardiovascular risk and mortality among women [1]. In parallel, an increase in the number of hospitalisations has been observed, as has, the cost of health care following a *MI*, in France, this cost has multiplied by three over the last decade [2].

In this context, the analysis of patient flows appears timely: 1) to predict patient admissions; 2) to adapt the patient care pathway; and 3) to manage the availability of resources. Usually researchers use both data mining methods [3] combined with predictive models based on decision trees to model patient flows and build decision support tools [4]. In this article, we characterise the flows of patients with *MI* through an innovative approach. Inspired by spatio-temporal mining method, we propose to extract patterns from the French Prospective Payment System (PPS). Rather than considering spatiality for extracting patterns, we considered directly the codes from the International Classification of Diseases 10th Revision (ICD-10). By clustering these trajectories we

[1] Corresponding Author: Jessica Pinaire, UPRES EA2415, 641 avenue du Doyen Gaston Giraud, 34093 Montpellier, France; E-mail:jessica.pinaire@lirmm.fr.

could automatically identify some trends in time between hospitalisations and costs. Thus these information extracted enabled to considering health planning strategies.

Section 2 introduces the mining of spatio-temporal patterns and proposes an illustration using PPS data. The patient flow characterisation process is described in Section 3. The Section 4 focuses on the characterisation of patient flows for women aged 65+. A discussion on the results and the limits of the proposal appears in Section 5. Finally, Section 6 concludes the paper.

2. Spatio-temporal patterns

Spatio-temporal mining aims at extracting set of objects sharing the same behaviour during a period of time. Even if many different patterns can be extracted, in this paper we mainly focus on closed swarm patterns. Let $O=\{o_1,...,o_n\}$ be a group of moving objects and a set of timestamps $T=\{t_1,...,t_p\}$. For each object $o \in O$ at time t, we have its spatial informations $x_{t_i}^{o_j}, y_{t_i}^{o_j}$. Let min_o be a user-defined threshold standing for a minimum number of objects and min_t the minimum number of timestamps. Informally, a swarm is a group of moving objects O containing at least min_o individuals which are closed each other for at least min_t timestamps. A swarm can be formally defined as follows:

Definition 1 (Swarm and Closed Swarm) *A pair (O,T) is a swarm if:*

1) $\forall\, t_i \in T, \exists\ c$ *s.t.* $O \subseteq c$, *c is a cluster;*
2) $|O| \geq min_o$ *and* $|T| \geq min_t$

A swarm (O,T) is a closed swarm if:

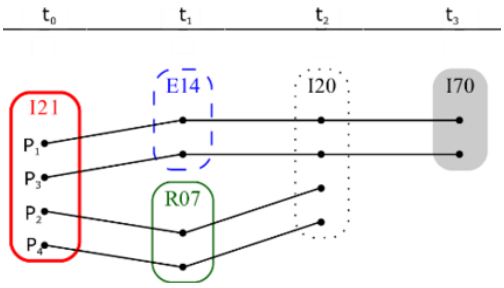
1) $\nexists\, O'$ *s.t. (O',T) is a swarm and* $O \subset O'$;
2) $\nexists\, T'$ *s.t. (O,T') is a swarm and* $T \subset T'$

According to this definition, the first condition is that $\forall\, t_i \in T, \exists\ c$ *s.t.* $O \subseteq c$, *c* is a cluster. In spatio-temporal data this cluster is obtained by grouping the objects that are sufficiently closed according their locations. In this paper rather than location we focus on the different kinds of hospitalisations for patients and we group them together if the share at one time the same hospitalisation reasons.

To illustrate, we consider the events of four patients. Time is divided into timestamps corresponding to one hospitalisation and the ICD-10 codes (I21: acute *MI*; R07: Chest pain; E14: Diabetes; I20: Angina pectoris; I70: Atherosclerosis) refer to the reasons of hospitalisation. Figure 1 illustrates an example of different trajectories that are followed by patients. For instance, we notice that patient P_2 has been first hospitalised for an acute *MI* (I21) at time t_0, then a Chest pain (R07) at time t_1 and finally for Angina pectoris (I20) at time t_2 (see Figure 1 (a)). Patients sharing the same code at one time can then be grouped together. Let us now assume that $min_o = 2$ and $min_t = 2$ and we are thus provided by the following swarms: $\{(P_1,P_3),(0,1)\}$, $\{(P_1,P_3),(1,2)\}$ and $\{(P_1,P_3),(0,1,2)\}$. We observe that these swarms are redundant because they can be grouped together in a closed swarm: $\{(P_1,P_3),(0,1,2)\}$.

	Timestamps			
Patients	t_0	t_1	t_2	t_3
P_1	I21	E14	I20	I70
P_2	I21	R07	I20	
P_3	I21	E14	I20	I70
P_4	I21	R07	I20	

(a)



(b)

Figure 1. Patient trajectories. (a) The sequential database with for each patient at each timestamp its hospitalisation code. (b) The different trajectories that when grouping together patients sharing the same code.

3. PaFloChar: Patient Flow Characterisation Process

In this section we present the patient flow characterisation process which has two main steps. The first one aims at extracting and then sorting spatio-temporal patterns from the PPS data. These patterns will correspond to the flow patterns of patients. The second step cluster the time between hospitalisations and cost trajectories in order to identify trends.

Step 1. Spatio-temporal pattern mining. We found 412,486 *MI* patients over the period of 2009 to 2014 from the PPS database. Each patient has a sequence of ICD-10 codes which length is equal to the number of stays over these six years. A first filtering is performed in order to remove irrelevant stays, i.e. stays that are not related to cardiopathology. Then, sequences are ordered according to the relative time corresponding to the occurrence of a stay. These final sequences are called the **patient trajectories**. Finally, we mined closed swarms using the Get_Move algorithm [5].

Step 2. Time and Cost Evolution Profiles. We clustered time [2] and cost trajectories with kmlShape [6]. This method, derived from k-means for longitudinal data (kml), is able to detect curve shapes in order to cluster curves having the same shape. To our knowledge, there no other cluster method to do this. Then, we established the assignment of the flow groups, created in step 1, in the time and tariff clusters.

4. Experiments

In this section, as an illustration of the process, we present the results related to women aged 65+.

Step 1. We highlighted three groups of patients flows in which first events were *Angina pectoris*, *Ischemic heart disease* and *MI*. Then, the flow was separated into several branches in which vertices were these events and *Death*. At the third stay, we observed new events: *Heart rhythm disorders* and *Heart failure*.

Step 2. We found three clusters of times (see Figure 2 (a)). *B* cluster (medium thick blue curve), *M* cluster (thin magenta curve) and *V* cluster (very thick green curve).

[2] Time between hospitalisations were calculated as the number of days between the last day of a stay and the first day of the subsequent stay.

Curves[3] characterise different trends: *B* cluster represents patients having short times (less than four months) that increase and then decrease; *M* cluster represents patients presenting with spaced stays at an early stage, then later more frequent; *V* cluster represents patients having long sequences. The assignment of group flows in these clusters shows that, in the *MI* group, most of them (were split into cluster *B* and *V*) have short times at the end of their pathway. We found three clusters of costs (see Figure 2 (b)): *B* cluster represents patients having increasing tariffs over time and decreasing slightly; *M* cluster represents patients having highly increasing tariffs; *V* cluster represents patients having first increasing tariff and then decreasing tariffs. The assignment of group flows in these clusters showed that, in the *MI* group, most patients (were in cluster *V*) had initially high tariff which decreased thereafter.

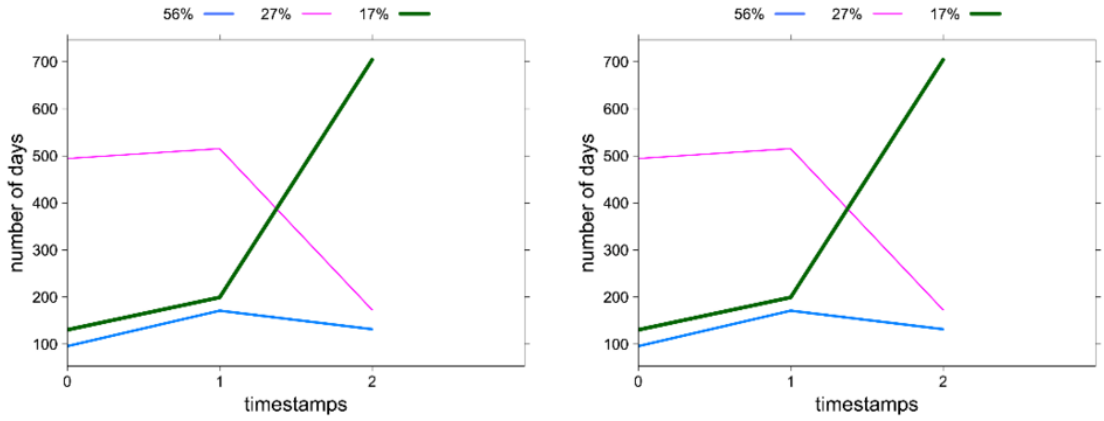


Figure 2. (a) Times between hospitalisations and (b) Costs kmlShape clusters.

5. Discussion

Patient flows. We highlighted three key steps in patient flow patterns: *MI*, *Angina pectoris* and *Ischemic heart disease*. The majority of patients showed signs of recurrence of coronary artery disease in the form of angina pectoris. Many of them experienced *MI* relapse and/or another manifestation of their ischemic heart disease. These results could be integrated in a decision-making tool. Indeed, this could be useful for a clinician to compare patient profile to similar profiles and warn them about risk of *MI* relapse.

Times and tariffs profiles. The times profiles provide information on the future hospitalisations related to cardiac pathology. In most cases, after an *MI*, hospitalisations are increasingly closed in time (on average three months). The majority of patient flow initiated by *Angina pectoris* or *Ischemic heart disease* have an upward trend in tariffs (cluster *B* in Figure 2 (b)). In contrast, patient flows initiated by *MI* largely show a downward trend in tariffs. Moreover, this work raises questions about the rhythms of hospitalisation frequency and the tariff over time.

[3] The graph reads as follows, at t_0 represents the time between the first stay and the second stay. For instance, the thin magenta curve, has an ordinate equal to 500: it means the time between the first and second stay is about 500 days in this cluster.

Limits. This study has several limitations : 1) *The choice of the database:* PPS is a budget allocation tool, so it presents limits in epidemiology studies [7]; 2) *The comparison with others studies:* Time is mostly investigated [8] as a precise event such as readmission for heart failure but less frequently for the more general event of heart disease. Moreover, most direct cost studies take into account emergency cost and drug consumption [9]. Access to the SNIIRAM (National health insurance system of inter-scheme information) database would allow a similar analysis to be carried out.

6. Conclusion

We looked for spatio-temporal patterns in the *MI* patients trajectories. The originality of the approach is that spatiality was assimilated to proximity of pathologies and the temporal aspect was related to the occurrence of a stay. Then we clustered their time and tariff trajectories to determine trends. Such an approach can be used to improve care by providing, for example, an integrated care pathway with scheduled visits proposed in the case of cancer [10]. In future work, we plan to investigate the co-evolution of the time and cost trajectories in order to establish whether they are related [11].

References

[1] Dégano, I. R., V. Salomaa, G. Veronesi, J. Ferriéres, I. Kirchberger, T. Laks, A. S. Havulinna, J.-B. Ruidavets, M. M. Ferrario, C. Meisinger, Twenty-five-year trends in myocardial infarction attack and mortality rates, and case-fatality, in six european populations, *Heart* 2015:**101**(17):1413–1421.

[2] Philippe, F., P. Blin, S. Bouée, C. Laurendeau, E. Torreton, J. Gourmelin, M. Velkovski-Rouyer, L. Levy-Bachelot, and G. Steg, Costs of healthcare resource consumption after a myocardial infarction in France: An estimate from a medicoadministrative database (GSB), *International Journal of Cardiology* 2016:**219**:387–393.

[3] Pagnoni, A., S. Parisi, and S. Lombardo, Analysis of patient flows via data mining, Studies in *Health Technology and Informatics* 2001:**2**:1379–1383.

[4] Dart, T., Y. Cui, G. Chatellier, and P. Degoulet, Analysis of hospitalised patient flows using data-mining, *Studies in Health Technology and Informatics* 2003:**95**:263–268.

[5] Phan, N., P. Poncelet, and M. Teisseire, All in one: mining multiple movement patterns, *International Journal of Information Technology & Decision Making* 2016:**15**(05):1115–1156.

[6] Genolini, C., R. Ecochard, M. Benghezal, T. Driss, S. Andrieu, and F. Subtil, kmlShape: An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes, *Plos one* 2016:**6**:e0150738

[7] Schneeweiss, S., and J. Avorn, A review of uses of health care utilization databases for epidemiologic research on therapeutics *Journal of Clinical Epidemiology* 2005:**58**(4):323–337.

[8] Miller, A. L., D. Simon, M. T. Roe, M. C. Kontos, D. Diercks, E. Amsterdam, and D. L. Bhatt, Comparison of Delay Times from Symptom Onset to Medical Contact in Blacks Versus Whites With Acute Myocardial Infarction, *The American Journal of Cardiology* 2017:**8**:1127–1134.

[9] Blin, P., C. Dureau-Pournin, R. Lassalle, J. Jové, F. Thomas-Delecourt, C. Droz-Perroteau, N. Danchin, and N. Moore, Outcomes, health care resources use, and costs in patients with post-myocardial infarction: The horus cohort study in the egb french claims and hospital database, *Value in Health* 2016:**19**(3):A16.

[10] Defossez, G., A. Rollet, O. Dameron, and P. Ingrand, Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer, *BioMed Central Medical Informatics and Decision Making* 2014:**14**(1):24.

[11] Genolini, C., J.-B. Pingault, T. Driss, S. Côté, R. E. Tremblay, F. Vitaro, C. Arnaud, and B. Falissard, Kml3d: a non-parametric algorithm for clustering joint trajectories, *Computer Methods and Programs in Biomedicine* 2013:**109**(1):104–111.