HAL open science

Recherche et visualisation de trajectoires dans les parcours de soins des patients ayant eu un infarctus du myocarde

Jessica Pinaire, Julien Rabatel, Jérôme Azé, Sandra Bringay, Paul Landais

To cite this version:

Jessica Pinaire, Julien Rabatel, Jérôme Azé, Sandra Bringay, Paul Landais. Recherche et visualisation de trajectoires dans les parcours de soins des patients ayant eu un infarctus du myocarde. SIIM: Symposium Ingénierie de l'Information Médicale, Jun 2015, Rennes, France. hal-01180416

HAL Id: hal-01180416

https://hal.science/hal-01180416v1

Submitted on 28 Jul 2015

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.

Recherche et visualisation de trajectoires dans les parcours de soins des patients ayant eu un infarctus du myocarde

Jessica Pinaire[1,3], Julien Rabatel[1], Jerôme Azé[1], Sandra Bringay[1,4], Paul Landais[2,3]

[1] LIRMM, UMR 5506, Université Montpellier, France
prenom.nom@lirmm.fr

[2] ÉQUIPE D'ACCUEIL 2415, Institut Universitaire de Recherche Clinique, Université Montpellier, Montpellier, France
paul.landais@umontpellier.fr

[3] CHU, Département d'information médicale, BESPIM, Nîmes, France
paul.landais@chu-nimes.fr

[4] AMIS, Université Paul Valéry, Montpellier, France
Sandra.Bringay@univ-montp3.fr

Résumé : La collecte des données hospitalières dans le cadre du PMSI génère sur le plan national des bases de données de l'ordre de 25 millions d'enregistrements par an. Ces données, qui sont recueillies à des fins médico-économiques, contiennent des informations qui peuvent avoir d'autres finalités : amélioration de la prise en charge du patient, prédictions de l'évolution des soins et de planification de leurs coûts, etc. Les travaux que nous présentons dans cet article exploitent ces données en se focalisant sur l'extraction de trajectoires de patients ayant eu un infarctus du myocarde. Les résultats obtenus sont restitués aux médecins via une interface dédiée qui permet une forte interaction avec les données.

Mots-clés : PMSI, Trajectoires de patients, Fouille de données.

1 Introduction

La collecte des données hospitalières dans le cadre du PMSI (Programme Médicalisé des Systèmes d'Information) génère sur le plan national des bases de données de l'ordre de 25 millions d'enregistrements (séjours) par an [1]. Ces données recueillies à des fins médico-économiques, peuvent *a posteriori* servir à des fins d'analyse et de recherche, pour examiner des questions médicales et épidémiologiques. Dans ces bases, grâce à un numéro anonyme de patient, il est possible de chaîner les séjours et suivre le parcours hospitalier d'un patient. Les enjeux associés à l'extraction de connaissances dans ces types de données sont importants : 1) au niveau de l'individu pour prédire sa trajectoire de soins Egho *et al.* (2013) ; 2) au niveau de la population pour prédire l'évolution et les coûts associés à la santé des populations (notamment dans le cas des maladies chroniques) Jensen *et al.* (2014).

Dans cet article, nous nous sommes intéressés aux patients atteints d'un infarctus du myocarde (IM), au cours de l'année 2009, ayant nécessité un acte de cardiologie interventionnelle, suivi durant trois années. Notre objectif est d'extraire des profils de parcours de soins fréquents pour l'IM qui prennent en compte des informations contextuelles fréquemment associées aux

1. Guide méthodologique de production des résumés de séjour du PMSI en médecine, chirurgie et obstétrique (fascicule spécial 2004/2 bis du Bulletin officiel)
http://www.atih.sante.fr/textes-officiels-du-pmsi-en-mco

données séquentielles. Dans un premier temps, nous avons utilisé des motifs séquentiels Srikant & Agrawal (1996) pour identifier des évènements chronologiques fréquents associés à des pathologies. Par exemple, dans le cas des séquences d'actes réalisées pour des patients dans un hôpital, nous allons pouvoir extraire les informations suivantes : fréquemment un patient ayant eu un IM subit une coronarographie, puis la pose d'un stent le cas échéant. L'extraction classique de motifs séquentiels se focalise sur les séries d'actes sans considérer des informations relatives aux patients comme leur sexe, leur âge, leur poids, etc. Or, en considérant le fait qu'un motif séquentiel est spécifique à un contexte donné (*e.g.* les jeunes hommes), un professionnel de santé pourra adapter sa stratégie de soin au contexte du patient et prendre les décisions adéquates. Nous avons pris en compte ces informations en utilisant la méthode d'extraction de motifs contextuels définie dans Rabatel *et al.* (2010). Ce type d'approches permet d'envisager trois types d'applications : le résumé et l'exploration des données par l'expert avec ici l'identification des trajectoires de patients, puis la classification, pour associer des parcours hospitaliers similaires et enfin la prédiction Fabregue *et al.* (2011), pour déterminer des probabilités de trajectoires pour un patient, et améliorer la prise en charge. Dans cet article, nous nous focalisons uniquement sur la première application.

Dans cet article, nous allons décrire le processus d'extraction de connaissances mis en place, en particulier la base de données et la hiérarchie de contextes utilisées en entrée de l'algorithme de fouille de données. Nous définirons une trajectoire fréquente sous la forme d'un motif séquentiel contextuel et nous évoquerons les techniques d'extraction de ces motifs fréquents ou discriminants. Enfin, nous analyserons les résultats obtenus et montrerons comment nous présentons ces résultats aux professionnels de santé via une visualisation interactive des motifs.

2 Matériel et méthodes

2.1 Données brutes

À partir des bases nationales de PMSI MCO [2] de 2009 à 2011, nous avons récupéré les informations sur les séjours hospitaliers du patient. Au préalable, nous avons effectué une étude de qualité sur les données. À l'aide d'une requête SQL, nous interrogeons la base de données, en retenant les patients ayant un séjour, en 2009, avec un code acte [3] de cardiologie interventionnelle. Nous collectons les informations concernant le patient : son âge, son sexe, sa région d'habitation, le nombre de séjours hospitaliers en trois ans et la distance entre sa commune de résidence et celle de l'établissement fréquenté, ainsi que des informations sur le séjour lui-même, le type d'établissement fréquenté (Privé/ Public), la durée de séjour, le décès, le GHM [4].

Le nombre de patients répondant au critère de sélection est de 86 318. La pathologie étudiée touche en majorité les hommes avec 75%, contre 25% pour les femmes (, alors que dans la base nationale il y a 44,4% d'hommes *vs* 55,6% de femmes). La moyenne d'âge de la population étudiée est de 46 ans (ET [5] : 12,4). Ici, on retrouve plutôt une population de seniors avec une médiane sur l'âge à 67 ans. Sur la période étudiée, nous observons peu de décès hospitaliers, 7

2. Médecine Chirurgie Obstétrique
http://www.atih.sante.fr/textes-officiels-du-pmsi-en-mco

3. Article L. 6113-8 du code de la santé publique

4. Groupe Homogène de Malades

5. Ecart Type

895, soit moins de 10%. Ce qui peut signifier que cette pathologie est bien prise en charge, ou que le patient décède avant son arrivée à l'hôpital.

Pour décrire la domiciliation d'origine des patients, nous avons utilisé un découpage de la France en territoires médicaux suivant la carte des inter-régions du CeNGEPS [6]. Les français, ayant cette pathologie, vivent en majorité dans le Sud, ce qui peut s'expliquer par le phénomène d'héliotropisme des seniors, puis dans l'Ouest et l'Ile de France.

Nous constatons également, que les patients, sont en moyenne à 25 minutes (ET : 19,8) de leur établissement de santé. Avec une médiane à 20 minutes, nous pouvons donc affirmer, qu'une majorité de la population vit, à proximité d'un établissement de santé.

La durée des séjours est plutôt courte avec une médiane à 3 jours. Notons également que le nombre de séjours sur trois ans est peu élevé une médiane à 3 jours. Pour chacune de ces grandeurs, l'intervalle inter-quartiles est de 3, soit une tendance, qui se situe autour de la médiane. En effet, étant donné la gravité de la maladie, soit les patients décèdent hors séjours hospitaliers, et nous n'avons pas d'information ; soit la maladie se soigne bien et il n'y a pas de pathologie lourde associée, qui nécessiterait des hospitalisations régulières. Il est important de noter que 20% des patients ne viennent qu'une fois en trois ans, 42,4% viennent au plus deux fois et 60% au plus trois fois.

Notre base de données comprend 266 établissements dont 97 en privé et 169 en public. 40,3% des actes de cardiologie interventionnelle sont réalisés dans le privé contre 59,7%, dans le public. Les établissements pratiquant ce type d'intervention, se situent en majorité dans le Sud-Ouest, puis le Sud et Nord-Ouest. Le nombre d'interventions suit une répartition analogue à celle des patients sur le territoire.

2.2 Données séquentielles et contextes

Pour notre étude, nous nous concentrons sur le GHM codé sur 5 caractères. Ce dernier renseigne les caractéristiques de l'hospitalisation. Il représente un mode de prise en charge spécifique à un motif d'hospitalisation, une classe d'âge et une durée de séjour. À ce mode de prise en charge est associé un tarif qui permet de rémunérer l'établissement de santé. Par exemple les GHM 05C08 (Autres interventions cardiothoraciques) et 05K13 (Actes thérapeutiques par voie vasculaire), correspondent à un mode de prise en charge identique de l'ordre de 23 300 euros. À chaque patient, est associé une série de GHM, de longueur égale au nombre de ses séjours effectués en hospitalisation complète [7] sur trois ans. Cette chaîne de caractères est la trajectoire du patient. 05K05-05C17 signifie que le patient a eu une endoprothèse avec IM, puis un autre séjour pour une ligature de veine et éveinage. Nous obtenons ainsi 43 250 séquences de GHM différentes qui correspondent aux trajectoires des patients.

Nous avons créé des contextes en utilisant les variables suivantes : l'âge, le nombre d'hospitalisations sur trois ans et le sexe. Dans un premier temps, nous avons discrétisé les variables quantitatives que sont l'âge et le nombre d'hospitalisations. Après concertation avec l'expert clinicien pour éviter de créer trop de sous-contextes, nous avons discrétisé l'âge en trois classes : les moins de 45 ans, entre 45 et 65 ans et les plus de 65 ans. De même, nous avons décomposé le nombre d'hospitalisation en trois classes : ceux qui ne viennent qu'une fois, entre deux et trois

6. `http://www.cengeps.fr/fr`
7. Le patient reste au moins une nuit à l'hôpital

fois, et ceux qui viennent plus de trois fois. Toutes les combinaisons faites, nous obtenons dix-huit contextes-minimaux. Un contexte minimal est une combinaison de modalités de chacune des variables entrant en jeu, *e.g* les femmes de moins de 45 ans hospitalisées une seule fois.

2.3 Recherche de motifs séquentiels contextuels

Considérons la base de données présentée dans le tableau 1 qui décrit différents GHM :
— 05M13 : Douleurs thoraciques ;
— 05M06 : Angine de poitrine ;
— 05M16 : Athérosclérose coronarienne ;
— 05M04 : IM aigu ;
— 05M09 : Insuffisance cardiaque.
associés au cours du temps par des professionnels de santé à des patients. Ces données sont séquentielles car elles présentent des événements (les GHM) disposés suivant un ordre (le temps). Par exemple, nous constatons que pour le patient P_1, les GHM $05M13$ et $05M04$ ont été réalisés ensemble en janvier puis le GHM $05M06$ en février. En examinant le tableau 1, nous constatons également que le motif $05M13$ suivi plus tard par $05M06$ est vérifié par plus de 50% des patients (8 sur 14). En supposant que le professionnel de santé précise qu'il est intéressé par des GHM qui apparaissent dans au moins 50% des cas (support minimum) présents dans la base alors le motif $< (05M13)(05M06) >$ est fréquent. Cet exemple considère la base comme un ensemble indivisible pour la recherche des motifs. Pourtant, les circonstances liées aux données, ce que l'on appelle les contextes, impliquent l'existence de sous-ensembles de données rassemblant des propriétés similaires. Pour notre cas d'étude, par exemple, nous pouvons intégrer des informations supplémentaires comme dans le tableau 2 qui associe à chaque patient son âge (*jeune* ou *âgé*) et son sexe (*homme* ou *femme*). Ces informations contextuelles peuvent avoir une influence non négligeable sur ce qui se produit dans les données et l'extraction de motifs doit rendre cette influence perceptible pour l'utilisateur afin de lui offrir une vue contextualisée des données.

TABLE 1 – Mise en valeur du motif $< (05M13)(05M06) >$ (en gras) soit le GHM *05M13* suivi du GHM *05M06*. Ce motif est fréquent dans la base pour un support minimum de 50%.

Patients	Janvier	Février	Mars	Avril
P_1		**05M13**, $05M04$		**05M06**
P_2	**05M13**,05M06		**05M06**	
P_3	**05M13**	05M13		**05M06**
P_4	05M16	**05M13**		**05M06**,05M16
P_5	05M04	**05M13**,05M06	**05M06**,05M16,05M04	
P_6		05M06		05M13
P_7		**05M13**	**05M06**	05M13
P_8	05M04	**05M13**		**05M06**,05M16
P_9		**05M13**,05M06	05M13	**05M06**,05M04
P_{10}			05M06,05M16,05M04	
P_{11}			05M06,05M04	05M13
P_{12}	05M09	05M06,05M16,05M04		05M13
P_{13}		05M06,05M04,05M09		
P_{14}	05M06		05M13	05M09

TABLE 2 – Mise en valeur du motif $< (05M13)(05M06) >$ (en gras) avec les informations contextuelles sur l'âge et le sexe. Ce motif est spécifique aux personnes âgées. Une seule personne jeune est concernée.

Patients	Age	Sexe	Janvier	Février	Mars	Avril
P_1	âgé	homme		**05M13**,05M04		**05M06**
P_2	âgé	homme	**05M13**,05M06		**05M06**	
P_3	âgé	homme	**05M13**	05M13		**05M06**
P_4	âgé	homme	05M16	**05M13**		**05M06**,05M16
P_5	âgé	homme	05M04	**05M13**,05M06	**05M06**,05M16,05M04	
P_6	âgé	femme		05M06		05M13
P_7	âgé	femme		**05M13**	**05M06**	05M13
P_8	âgé	femme	05M04	**05M13**		**05M06**,05M16
P_9	jeune	homme		**05M13**,05M06	05M13	**05M06**,05M04
P_{10}	jeune	homme			05M06,05M16,05M04	
P_{11}	jeune	homme			05M06,05M04	05M13
P_{12}	jeune	femme	05M09	05M06,05M16,05M04		05M13
P_{13}	jeune	femme		05M06,05M04,05M09		
P_{14}	jeune	femme	05M06		05M13	05M09

Considérons maintenant le motif $< (05M13)(05M06) >$ dans le tableau 2, nous constatons que :

— ces GHM sont fréquents dans la population âgée (7 personnes âgées sur 8) mais pas dans la population jeune (seulement 1 personne sur 6) ;

— ces GHM demeurent fréquents chez les âgés quel que soit leur sexe (5 hommes âgés sur 5 et 2 femmes âgées sur 3).

Savoir qu'un comportement est spécifique ou général à un contexte est alors utile pour une interprétation médicale. Dans cette approche, la propriété « d'être fréquent » dépend d'un contexte donné. La propriété « d'être spécifique » ou discriminant, se caractèrise par le fait d'être plus fréquent pour un contexte, par rapport à tous les autres contextes.

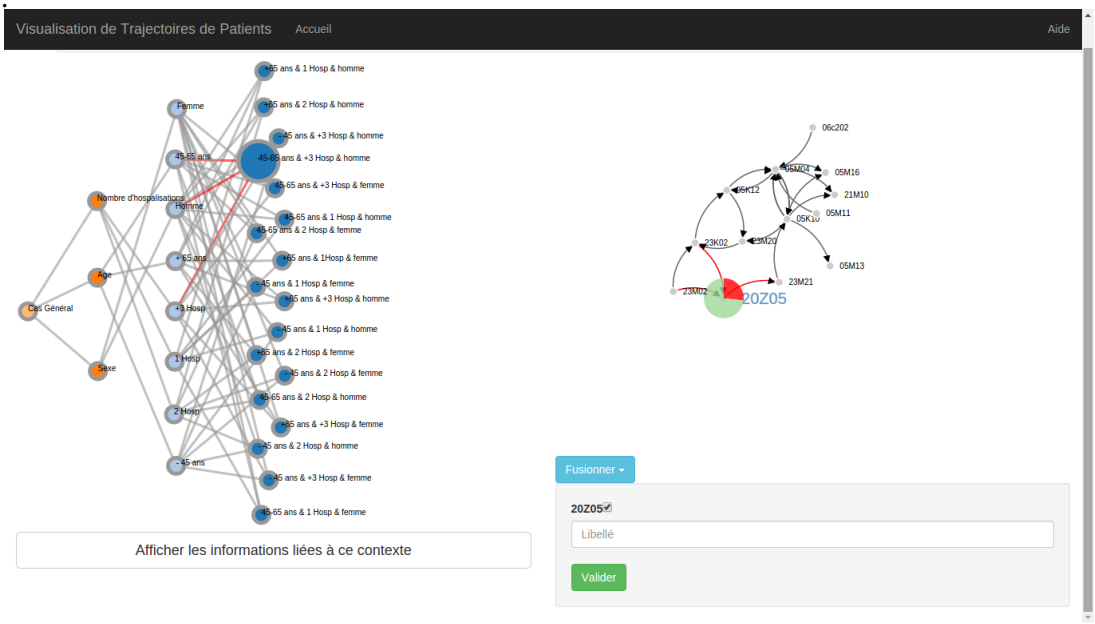
De manière générale, l'extraction de motifs est un problème difficile qui nécessite de naviguer dans un très grand espace de recherche. La prise en compte des contextes étend encore cet espace de recherche. Via des propriétés théoriques intéressantes basées sur la fréquence et sur les propriétés formelles associées au treillis formé par les contextes, nous avons développé un algorithme très efficace pour l'extraction de ces motifs contextuels Rabatel *et al.* (2010). Cette approche a été généralisée à d'autres mesures d'intérêt que la fréquence, utiles pour la sélection des motifs tels que le gain d'information, le taux d'émergence, la confiance dont l'objectif est de s'appuyer sur les caractéristiques statistiques des motifs pour isoler les plus intéressants au sens de critères experts.

2.4 La visualisation

Grâce aux méthodes décrites précédemment, nous extrayons des motifs spécifiques à une population (*e.g.* fréquent chez les âgés et non fréquents chez les jeunes) ou généraux (fréquents dans l'ensemble de la population). Ces résultats deviennent vite denses et difficiles à exploiter, c'est la raison pour laquelle il s'est avéré indispensable de créer un outil de visualisation des trajectoires de patients. L'interface graphique se présente de la manière suivante : sur la partie

gauche de l'écran, nous visualisons la hiérarchie des contextes. Chaque rond correspond à un contexte (par exemple, les hommes de moins de 45 ans). Le contexte le plus à gauche est le plus général et les contextes les plus à droite sont les plus spécifiques (voir figure 1). L'utilisateur peut ainsi naviguer dans la hiérarchie de contextes. Lorsque ce dernier choisit un contexte, il fait apparaître sur la partie droite de l'écran, un graphe correspondant à l'agrégation des motifs séquentiels contextuels associés au contexte sélectionné. La conception graphique a été réalisée à l'aide de la librairie D3. Les sommets sont des GHM, les arcs ordonnent les GHM. La taille des arcs est proportionnelle au nombre de patients concernés. Par survol des sommets à l'aide de la souris, l'utilisateur fait apparaitre leur libellé, s'il clique sur l'un d'eux, un graphique circulaire s'affiche, correspondant au nombre de personnes décédées. Ce choix d'affichage par multiples action nous permet de gagner en lisibilité. Nous envisageons également afficher d'autres informations paramétrées par les professionnels de santé. Ces derniers peuvent interagir avec l'interface, pour fusionner des GHM qu'ils considèrent comme similaires d'un point de vue médical et ainsi recalculer le graphe.

FIGURE 1 – Capture d'écran de l'interface graphique créée pour analyser et explorer les données.



3 Résultats

Dans un premier temps, après plusieurs analyses, on constate que les GHM les plus fréquents dans la base sont 05K05 et 05K06, qui correspondent aux endoprothèses posées par voie vasculaire, respectivement avec ou sans IM, (*i.e.*, pose de stent par exemple). Cela s'explique car il s'agit de l'un des critères de sélection des patients. Nous avons donc choisi d'écarter ces GHM puis de refaire les analyses en diminuant le support minimum. L'effectif de la base est alors réduit : 66 742 patients. En effet, les patients ayant effectué des mono-séjours ne contenant qu'un des GHM cité précédemment ne sont pas pris en considération. Compte tenu des différents tests

effectués pour déterminer le support minimum, nous avons retenu le seuil de 1% pour lequel nous obtenons un bon compromis entre le nombre de motifs extraits et leurs longueurs. Nous allons analyser quelques uns des résultats obtenus dans la suite de cet article.

3.1 Motifs fréquents

Comme l'on pouvait s'y attendre, les GHM fréquents sont essentiellement des GHM de cardiologie, *i.e* des GHM commençant par "05". Le critère de sélection SQL de la population (infarctus du myocarde GHM=05M04) est retrouvée avec ce motif fréquent dans la base, tous contextes confondus. La fréquence du GHM 05M04 est de 9,7%.

Lorsque nous nous focalisons sur l'analyse contextuelle des motifs, nous constatons que dans la sous-population des patients agés de plus de 65 ans (37 826 patients), les GHM les plus fréquents sont : les actes thérapeutiques par voie vasculaire (05K13) avec un support de 7,6%, puis la pose d'un stimulateur cardiaque (05C15) avec un support de 3,89%. Si nous considérons maintenant les patientes situées dans la même tranche d'âge, et hospitalisées plus de 3 fois (6 896 patients), nous observons le motif suivant 05M09-05M09 pour 5,8% de ces patientes. Ce motif traduit l'insuffisance cardiaque et un état de choc, ce même GHM est répété trois fois pour 4% de la population. En considérant uniquement les femmes de plus de 65 ans qui ne sont hospitalisées qu'une seule fois (308 patientes), les actes thérapeutiques par voie vasculaire (05K13) ont un support de 20,5%, l'IM aigu (05M04) a un support de 12,6% et l'IM avec décès (24Z09) un support de 9,7%. Ce dernier GHM est également fréquent dans le même contexte pour les hommes (552 patients) à 5,7%. Un autre motif fréquent chez les femmes de plus de 65 ans (12 362 patients) est celui de l'oedème pulmonaire avec détresse respiratoire (04M13) avec un support de 3,3%.

Un motif remarquable est celui de la répétition des séances de rééducation (23M02), pour les hommes, venant plus de 3 fois, jusqu'à 20 fois pour les moins de 45 ans (1 064 patients) avec un support de 1,32% et 17 fois chez les 45-65 ans (9 869 patients) avec un support de 1%.

Du côté des patients agés de 45 à 65 ans, venant entre 2 et 3 fois, nous observons quelques similarités dans les motifs fréquents entre les hommes (11 119 patients) et les femmes (1960 patients). Par exemple, les actes diagnostiques par voie vasculaire (05K10-05K10) ont un support de 1,18%, resp. 1,21%. Ce motif se retrouve en triplet pour ceux qui viennent plus de 3 fois, avec un support de 1,7% pour les hommes (9 862 patients) et 2,13% pour les femmes (2 061 patients). D'autres évènements fréquents aussi bien pour les hommes que pour les femmes de cette tranche d'âge, venant entre 2-3 fois sont : IM aigu (05M04) à 7,2% (resp. 7,5%), Angine de poitrine (05M06) à 6,6% (resp. 7,2%), et Endoscopie digestive (06K04) à 7,5% (resp. 7,1%).

Pour les patientes, de moins de 45 ans, venant entre 2 et 3 fois (2 59 patientes), nous trouvons les motifs suivants : IM aigu suivi des actes diagnostiques (05M04-05K10) à 2,3%, ou suivi d'athérosclérose coronarienne (05M04-05M16) à 1,5% ou encore douleur thoracique (05M04-05M13) avec le même support. Concernant celles qui viennent plus de 3 fois (206 patientes), nous observons que la pose d'un défibrillateur après actes diagnostiques (05K10-05C19) est fréquente avec un support de 1,69% et également les troubles vasculaires périphériques (05M12-05M12) à 1%. Chez les hommes (1 531 patients), c'est légèrement différent avec les motifs 05K10-05K10 et 05M04-05M16 à 1%. Nous observons également le diabète (10M02) associé à 05M04, 05M06 ou encore 05M16 (autres affections de l'appareil respiratoire) au seuil de 1%.

Nous obtenons ici des motifs fréquents cohérents avec les connaissances actuelles en cardio-

logie.

3.2 Motifs discriminants

Nous avons également extrait des motifs discriminants, c'est-à-dire toujours fréquents mais également spécifiques à une classe. Les résultats sont similaires à ceux de la section précédente et en accord avec les connaissances en cardiologie.

Les motifs discriminants, nous permettent de reconstruire, l'évolution de cette pathologie avec l'IM suivi de l'angine de poitrine et/ou insuffisance cardiaque mais encore autre sténose coronarienne, le plus souvent pour les femmes quelle que soit la classe d'âge et un nombre de venues supérieur à 2. Un autre type de trajectoire se repère, en particulier chez les hommes de plus de 65 ans venant plus de 3 fois, avec une athérosclérose coronarienne, suivi d'une angine de poitrine, pour finir avec un IM.

Notons également, que la recherche de motifs discriminants permet d'identifier des sous-populations avec des GHM tels que l'intervention sur la rétine (02C05), pour les plus de 65 ans, ou encore le diabète (10M02) pour les plus de 3 hospitalisations comme facteur à risque pour les maladies coronariennes.

4 Conclusion

L'utilisation de techniques de fouille de données implantées dans une interface, offre la possibilité aux professionnels de santé de visualiser les connaissances extraites des données et de les manipuler. Les prochains travaux, pour la phase de description, consisteront à enrichir les informations existantes, à la fois sur le séjour, avec des détails d'ordre médical tels que le motif d'hospitalisation et les actes pratiqués, mais aussi d'augmenter la durée d'observation en ajoutant les années 2012 et 2013. D'autres types de trajectoires seront étudiées notamment pour prendre en compte la gradualité telle que décrite dans Hai *et al.* (2013). Dans une étape suivante, nous ferons l'analyse des résultats obtenus en collaboration avec l'expert clinicien. Nous nous appuierons sur cette partie du travail pour la phase de classification des trajectoires. L'efficacité des motifs en terme de prédiction sera également étudiée au niveau de l'individu et de la population. Nous pourrions ainsi prédire les futurs coûts associés aux maladies chroniques par exemple.

Références

EGHO E., JAY N., RAÏSSI C., NUEMI G., QUANTIN C. & NAPOLI A. (2013). An approach for mining care trajectories for chronic diseases. In *Artificial Intelligence in Medicine - 14th Conference on Artificial Intelligence in Medicine, AIME 2013, Murcia, Spain, May 29 - June 1, 2013. Proceedings*, p. 258–267.

FABREGUE M., BRINGAY S., PONCELET P., TEISSEIRE M. & ORSETTI B. (2011). Mining microarray data to predict the histological grade of a breast cancer. *Journal of Biomedical Informatics*, **44**(Suppl 1), S12–S16.

HAI P. N., PONCELET P. & TEISSEIRE M. (2013). Get_move : fouille de données d'objets mobiles. *Ingénierie des Systèmes d'Information*, **18**(4), 145–169.

JENSEN A. B., MOSELEY P., OPREA T., ELLESØE S. G., ERIKSSON R., SCHMOCK H., JENSEN P. B., JENSEN L. J. & BRUNAK S. (2014). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. volume 5 of *Nature Communications*, p. 4022.

RABATEL J., BRINGAY S. & PONCELET P. (2010). Aide à la décision pour la maintenance ferroviaire préventive. In *Extraction et Gestion des Connaissances*, EGC'10, Revue des Nouvelles Technologies de l'Information, p. 363–368 : Cépaduès-Éditions.

SRIKANT R. & AGRAWAL R. (1996). Mining sequential patterns : Generalizations and performance improvements. In P. M. G. APERS, M. BOUZEGHOUB & G. GARDARIN, Eds., *Advances in Database Technology - EDBT'96, 5th International Conference on Extending Database Technology, Avignon, France, March 25-29, 1996, Proceedings*, volume 1057 of *Lecture Notes in Computer Science*, p. 3–17 : Springer.