

Analyse de la trajectoire de soins en cancérologie : fouille de données et extraction des connaissances à partir du PMSI

Nicolas Jay

▶ **To cite this version:**

Nicolas Jay. Analyse de la trajectoire de soins en cancérologie : fouille de données et extraction des connaissances à partir du PMSI. Sciences du Vivant [q-bio]. 2004. hal-01731825

HAL Id: hal-01731825
https://hal.univ-lorraine.fr/hal-01731825

Submitted on 14 Mar 2018

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.

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

17640 1
Double

UNIVERSITÉ HENRI POINCARÉ, NANCY 1 FACULTÉ DE MÉDECINE DE NANCY

Année 2004 N° 93

SCD-UHP - NANCY 1
BIBLIOTHÈQUE
MÉDECINE

THÈSE

pour obtenir le grade de

DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement

dans le cadre du troisième cycle de Médecine Spécialisée

par

Nicolas JAY

le 8 octobre 2004

Analyse de la trajectoire de soins en cancérologie

Fouille de données et Extraction des Connaissances à partir du PMSI

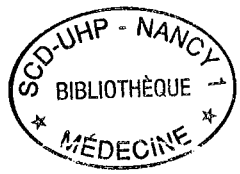
Examinateurs de la thèse :

Monsieur le Professeur	F. KOHLER	Président
Monsieur le Professeur	Y. MARTINET	Juge
Madame le Professeur	E. ALBUISSON	Juge
Monsieur le Professeur	A. NAPOLI	Juge

0 5 NOV. 2004

UNIVERSITÉ HENRI POINCARÉ, NANCY 1 FACULTÉ DE MÉDECINE DE NANCY

Année 2004 N°



THÈSE

pour obtenir le grade de

DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement

dans le cadre du troisième cycle de Médecine Spécialisée

par

Nicolas JAY

le 8 octobre 2004

Analyse de la trajectoire de soins en cancérologie

Fouille de données et Extraction des Connaissances à partir du PMSI

Examinateurs de la thèse :

Monsieur le Professeur	F. KOHLER	Président
Monsieur le Professeur	Y. MARTINET	Juge
Madame le Professeur	E. ALBUISSON	Juge
Monsieur le Professeur	A. NAPOLI	Juge

UNIVERSITÉ HENRI POINCARÉ, NANCY 1

FACULTÉ DE MÉDECINE DE NANCY

Président de l'Université : Professeur Jean-Pierre FINANCE

Doyen de la Faculté de Médecine : Professeur Patrick NETTER

Vice-Doyen de la Faculté de Médecine : Professeur Henry COUDANE

Assesseurs	
du 1er Cycle :	**Mme le Docteur Chantal KOHLER**
du 2ème Cycle :	**Mr le Professeur Jean-Pierre BRONOWICKI**
du 3ème Cycle :	**Mr le Professeur Hervé VESPIGNANI**
de la Vie Facultaire :	**Mr le Professeur Bruno LEHEUP**

DOYENS HONORAIRES

Professeur Adrien DUPREZ – Professeur Jean-Bernard DUREUX
Professeur Georges GRIGNON – Professeur Jacques ROLAND

PROFESSEURS HONORAIRES

Louis PIERQUIN – Etienne LEGAIT – Jean LOCHARD – René HERBEUVAL – Gabriel FAIVRE – Jean-Marie FOLIGUET
Guy RAUBER – Paul SADOUL – Raoul SENAULT – Marcel RIBON
Jacques LACOSTE – Jean BEUREY – Jean SOMMELET – Pierre HARTEMANN – Emile de LAVERGNE
Augusta TREHEUX – Michel MANCIAUX – Paul GUILLEMIN – Pierre PAYSANT
Jean-Claude BURDIN – Claude CHARDOT – Jean-Bernard DUREUX – Jean DUHEILLE – Jean-Pierre GRILLIAT
Pierre LAMY – Jean-Marie GILGENKRANTZ – Simone GILGENKRANTZ
Pierre ALEXANDRE – Robert FRISCH – Michel PIERSON – Jacques ROBERT
Gérard DEBRY – Georges GRIGNON – Pierre TRIDON – Michel WAYOFF – François CHERRIER – Oliéro GUERCI
Gilbert PERCEBOIS – Claude PERRIN – Jean PREVOT – Jean FLOQUET
Alain GAUCHER – Michel LAXENAIRE – Michel BOULANGE – Michel DUC – Claude HURIET – Pierre LANDES
Alain LARCAN – Gérard VAILLANT – Daniel ANTHOINE – Pierre GAUCHER – René-Jean ROYER
Hubert UFFHOLTZ – Jacques LECLERE – Francine NABET – Jacques BORRELLY
Michel RENARD – Jean-Pierre DESCHAMPS – Pierre NABET – Marie-Claire LAXENAIRE – Adrien DUPREZ – Paul VERT
Philippe CANTON – Bernard LEGRAS – Pierre MATHIEU – Jean-Marie POLU - Antoine RASPILLER – Gilbert THIBAUT
Michel WEBER – Gérard FIEVE – Daniel SCHMITT – Colette VIDAILHET

==========

PROFESSEURS DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

(Disciplines du Conseil National des Universités)

42ème Section : MORPHOLOGIE ET MORPHOGENÈSE

1ère sous-section : ***(Anatomie)***

Professeur Jacques ROLAND – Professeur Gilles GROSDIDIER
Professeur Pierre LASCOMBES – Professeur Marc BRAUN

2ème sous-section : ***(Cytologie et histologie)***

Professeur Bernard FOLIGUET

3ème sous-section : ***(Anatomie et cytologie pathologiques)***

Professeur François PLENAT - Professeur Jean-Michel VIGNAUD – Professeur Eric LABOUYRIE

43ème Section : BIOPHYSIQUE ET IMAGERIE MÉDICALE

1ère sous-section : ***(Biophysique et médecine nucléaire)***

Professeur Alain BERTRAND – Professeur Gilles KARCHER – Professeur Pierre-Yves MARIE

2ème sous-section : ***(Radiologie et imagerie médicale)***

Professeur Luc PICARD – Professeur Denis REGENT - Professeur Michel CLAUDON
Professeur Serge BRACARD – Professeur Alain BLUM - Professeur Jacques FELBLINGER
Professeur René ANXIONNAT

44ème Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLECULAIRE, PHYSIOLOGIE ET NUTRITION

1ère sous-section : *(Biochimie et biologie moléculaire)*

Professeur Jean-Pierre NICOLAS

Professeur Jean-Louis GUEANT – Professeur Jean-Luc OLIVIER

2ème sous-section : *(Physiologie)*

Professeur Jean-Pierre CRANCE – Professeur Jean-Pierre MALLIE

Professeur François MARCHAL – Professeur Philippe HAOUZI

3ème sous-section : *(Biologie cellulaire)*

Professeur Claude BURLET

4ème sous-section : *(Nutrition)*

Professeur Olivier ZIEGLER

45ème Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE

1ère sous-section : *(Bactériologie – virologie ; hygiène hospitalière)*

Professeur Alain LE FAOU – Professeur Alain LOZNIEWSKI

2ème sous-section : *(Parasitologie et mycologie)*

Professeur Bernard FORTIER

3ème sous-section : *(Maladies infectieuses ; maladies tropicales)*

Professeur Thierry MAY – Professeur Christian RABAUD

46ème Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ

1ère sous-section : *(Épidémiologie, économie de la santé et prévention)*

Professeur Philippe HARTEMANN – Professeur Serge BRIANÇON

Professeur Francis GUILLEMIN – Professeur Denis ZMIROU

2ème sous-section : *(Médecine et santé au travail)*

Professeur Guy PETIET – Professeur Christophe PARIS

3ème sous-section : *(Médecine légale et droit de la santé)*

Professeur Henry COUDANE

4ème sous-section : *(Biostatistiques, informatique médicale et technologies de communication)*

Professeur François KOHLER – Professeur Éliane ALBUISSON

47ème Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE

1ère sous-section : *(Hématologie ; transfusion)*

Professeur Christian JANOT – Professeur Thomas LECOMPTE – Professeur Pierre BORDIGONI

Professeur Pierre LEDERLIN – Professeur Jean-François STOLTZ

2ème sous-section : *(Cancérologie ; radiothérapie)*

Professeur François GUILLEMIN – Professeur Thierry CONROY

Professeur Pierre BEY – Professeur Didier PEIFFERT

3ème sous-section : *(Immunologie)*

Professeur Gilbert FAURE – Professeur Marie-Christine BENE

4ème sous-section : *(Génétique)*

Professeur Philippe JONVEAUX – Professeur Bruno LEHEUP

48ème Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE

1ère sous-section : *(Anesthésiologie et réanimation chirurgicale)*

Professeur Claude MEISTELMAN – Professeur Dan LONGROIS - Professeur Hervé BOUAZIZ

Professeur Paul-Michel MERTES

2ème sous-section : *(Réanimation médicale)*

Professeur Henri LAMBERT – Professeur Alain GERARD

Professeur Pierre-Edouard BOLLAERT – Professeur Bruno LÉVY

3ème sous-section : *(Pharmacologie fondamentale ; pharmacologie clinique)*

Professeur Patrick NETTER – Professeur Pierre GILLET

4ème sous-section : *(Thérapeutique)*

Professeur François PAILLE – Professeur Gérard GAY – Professeur Faiez ZANNAD

49ème Section : PATHOLOGIE NERVEUSE ET MUSCULAIRE, PATHOLOGIE MENTALE, HANDICAP et RÉÉDUCATION

1ère sous-section : *(Neurologie)*

Professeur Gérard BARROCHE – Professeur Hervé VESPIGNANI

Professeur Xavier DUCROCQ

2ème sous-section : *(Neurochirurgie)*

Professeur Jean-Claude MARCHAL – Professeur Jean AUQUE

Professeur Thierry CIVIT

3ème sous-section : *(Psychiatrie d'adultes)*

Professeur Jean-Pierre KAHN

4ème sous-section : *(Pédopsychiatrie)*

Professeur Daniel SIBERTIN-BLANC

5ème sous-section : *(Médecine physique et de réadaptation)*

Professeur Jean-Marie ANDRE

50ème Section : PATHOLOGIE OSTÉO-ARTICULAIRE, DERMATOLOGIE et CHIRURGIE PLASTIQUE

1ère sous-section : *(Rhumatologie)*

Professeur Jacques POUREL – Professeur Isabelle VALCKENAERE

2ème sous-section : *(Chirurgie orthopédique et traumatologique)*

Professeur Jean-Pierre DELAGOUTTE – Professeur Daniel MOLE

Professeur Didier MAINARD

3ème sous-section : *(Dermato-vénéréologie)*

Professeur Jean-Luc SCHMUTZ – Professeur Annick BARBAUD

4ème sous-section : *(Chirurgie plastique, reconstructrice et esthétique)*

Professeur François DAP

51ème Section : PATHOLOGIE CARDIORESPIRATOIRE et VASCULAIRE

1ère sous-section : *(Pneumologie)*

Professeur Yves MARTINET - Professeur Jean-François CHABOT

2ème sous-section : *(Cardiologie)*

Professeur Etienne ALIOT – Professeur Yves JUILLIERE – Professeur Nicolas SADOUL – Professeur Christian de CHILLOU

3ème sous-section : *(Chirurgie thoracique et cardiovasculaire)*

Professeur Jean-Pierre VILLEMOT

Professeur Jean-Pierre CARTEAUX – Professeur Loïc MACE

4ème sous-section : *(Chirurgie vasculaire ; médecine vasculaire)*

52ème Section : MALADIES DES APPAREILS DIGESTIF et URINAIRE

1ère sous-section : *(Gastroentérologie ; hépatologie)*

Professeur Marc-André BIGARD

Professeur Jean-Pierre BRONOWICKI

2ème sous-section : *(Chirurgie digestive)*

-

3ème sous-section : *(Néphrologie)*

Professeur Michèle KESSLER – Professeur Dominique HESTIN (Mme) – Professeur Luc FRIMAT

4ème sous-section : *(Urologie)*

Professeur Philippe MANGIN – Professeur Jacques HUBERT – Professeur Luc CORMIER

53ème Section : MÉDECINE INTERNE, GÉRIATRIE et CHIRURGIE GÉNÉRALE

1ère sous-section : *(Médecine interne)*

Professeur Francis PENIN – Professeur Denise MONERET-VAUTRIN – Professeur Denis WAHL

Professeur Jean-Dominique DE KORWIN – Professeur Pierre KAMINSKY

Professeur Athanase BENETOS - Professeur Gisèle KANNY – Professeur Abdelouahab BELLOU

2ème sous-section : *(Chirurgie générale)*

Professeur Patrick BOISSEL – Professeur Laurent BRESLER

Professeur Laurent BRUNAUD

54ème Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE, ENDOCRINOLOGIE ET REPRODUCTION

1ère sous-section : *(Pédiatrie)*

Professeur Daniele SOMMELET – Professeur Michel VIDAILHET – Professeur Pierre MONIN
Professeur Jean-Michel HASCOET – Professeur Pascal CHASTAGNER – Professeur François FEILLET

2ème sous-section : *(Chirurgie infantile)*

Professeur Michel SCHMITT – Professeur Gilles DAUTEL – Professeur Pierre JOURNEAU

3ème sous-section : *(Gynécologie-obstétrique ; gynécologie médicale)*

Professeur Michel SCHWEITZER – Professeur Jean-Louis BOUTROY
Professeur Philippe JUDLIN – Professeur Patricia BARBARINO

4ème sous-section : *(Endocrinologie et maladies métaboliques)*

Professeur Georges WERYHA – Professeur Marc KLEIN – Professeur Bruno GUERCI

5ème sous-section : *(Biologie et médecine du développement et de la reproduction)*

Professeur Hubert GERARD

55ème Section : PATHOLOGIE DE LA TÊTE ET DU COU

1ère sous-section : *(Oto-rhino-laryngologie)*

Professeur Claude SIMON – Professeur Roger JANKOWSKI

2ème sous-section : *(Ophtalmologie)*

Professeur Jean-Luc GEORGE – Professeur Jean-Paul BERROD – Professeur Karine ANGIOI-DUPREZ

3ème sous-section : *(Chirurgie maxillo-faciale et stomatologie)*

Professeur Michel STRICKER – Professeur Jean-François CHASSAGNE

==========

PROFESSEURS DES UNIVERSITÉS

64ème Section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE

Professeur Daniel BURNEL

==========

MAÎTRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

42ème Section : MORPHOLOGIE ET MORPHOGENÈSE

1ère sous-section : *(Anatomie)*

Docteur Bruno GRIGNON – Docteur Jean-Pascal FYAD

2ème sous-section : *(Cytologie et histologie)*

Docteur Edouard BARRAT – Docteur Jean-Claude GUEDENET
Docteur Françoise TOUATI – Docteur Chantal KOHLER

3ème sous-section : *(Anatomie et cytologie pathologiques)*

Docteur Yves GRIGNON – Docteur Béatrice MARIE
Docteur Laurent ANTUNES

43ème Section : BIOPHYSIQUE ET IMAGERIE MÉDICALE

1ère sous-section : *(Biophysique et médecine nucléaire)*

Docteur Marie-Hélène LAURENS – Docteur Jean-Claude MAYER
Docteur Pierre THOUVENOT – Docteur Jean-Marie ESCANYE – Docteur Amar NAOUN

44ème Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLÉCULAIRE, PHYSIOLOGIE ET NUTRITION

1re sous-section : *(Biochimie et biologie moléculaire)*

Docteur Xavier HERBEUVAL – Docteur Jean STRACZEK – Docteur Sophie FREMONT

Docteur Isabelle GASTIN - Docteur Bernard NAMOUR – Docteur Marc MERTEN

2ème sous-section : *(Physiologie)*

Docteur Gérard ETHEVENOT – Docteur Nicole LEMAU de TALANCE – Docteur Christian BEYAERT

4ème sous-section : *(Nutrition)*

Docteur Didier QUILLIOT

45ème Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE

1re sous-section : *(Bactériologie – Virologie ; hygiène hospitalière)*

Docteur Francine MORY – Docteur Michèle WEBER – Docteur Christine LION

Docteur Michèle DAILLOUX – Docteur Véronique VENARD

2ème sous-section : *(Parasitologie et mycologie)*

Docteur Marie-France BIAVA – Docteur Nelly CONTET-AUDONNEAU

46ème Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ

1re sous-section : *(Epidémiologie, économie de la santé et prévention)*

Docteur François ALLA

4ère sous-section : *(Biostatistiques, informatique médicale et technologies de communication (type biologique)*

Docteur Pierre GILLOIS

47ème Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE

1ère sous-section : *(Hématologie ; transfusion)*

Docteur François SCHOONEMAN

3ème sous-section : *(Immunologie)*

Docteur Marie-Nathalie SARDA

4ème sous-section : *(Génétique)*

Docteur Christophe PHILIPPE

48ème Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE

1ère sous-section : *(Anesthésiologie et réanimation chirurgicale)*

Docteur Jacqueline HELMER – Docteur Gérard AUDIBERT

3ème sous-section : *(Pharmacologie fondamentale ; pharmacologie clinique)*

Docteur Françoise LAPICQUE – Docteur Marie-José ROYER-MORROT

Docteur Damien LOEUILLE

54ème Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE, ENDOCRINOLOGIE ET REPRODUCTION

5ème sous-section : *(Biologie et médecine du développement et de la reproduction)*

Docteur Jean-Louis CORDONNIER

==========

MAÎTRES DE CONFÉRENCES

19ème section : SOCIOLOGIE, DÉMOGRAPHIE

Madame Michèle BAUMANN

32ème section : CHIMIE ORGANIQUE, MINÉRALE, INDUSTRIELLE

Monsieur Jean-Claude RAFT

40ème section : SCIENCES DU MEDICAMENT
Monsieur [illegible] ZEAU

60ème section : MÉCANIQUE, GÉNIE MÉCANIQUE ET GÉNIE CIVILE
Monsieur Alain DURAND

64ème section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE
Madame Marie-Odile PERRIN – Mademoiselle Marie-Claire LANHERS

65ème section : BIOLOGIE CELLULAIRE
Mademoiselle Françoise DREYFUSS – Monsieur Jean-Louis GELLY – Madame Anne GERARD
Madame Ketsia HESS – Monsieur Pierre TANKOSIC – Monsieur Hervé MEMBRE

67ème section : BIOLOGIE DES POPULATIONS ET ÉCOLOGIE
Madame Nadine MUSSE

68ème section : BIOLOGIE DES ORGANISMES
Madame Tao XU-JIANG

======

MAÎTRES DE CONFÉRENCES ASSOCIÉS
Médecine Générale
Docteur Alain AUBREGE
Docteur Francis RAPHAEL

========

PROFESSEURS ÉMÉRITES

Professeur Georges GRIGNON – Professeur Michel PIERSON - Professeur Michel BOULANGE
Professeur Alain LARCAN - Professeur Michel WAYOFF – Professeur Daniel ANTHOINE
Professeur Hubert UFFHOLTZ – Professeur Pierre GAUCHER – Professeur Claude CHARDOT
Professeur Adrien DUPREZ - Professeur Paul VERT – Professeur Jean PREVOT – Professeur Jean-Pierre GRILLIAT
Professeur Philippe CANTON – Professeur Pierre MATHIEU – Professeur Gilbert THIBAUT
Professeur Daniel SCHMITT

==========

DOCTEURS HONORIS CAUSA

Professeur Norman SHUMWAY (1972)
Université de Stanford, Californie (U.S.A)
Professeur Paul MICHIELSEN (1979)
Université Catholique, Louvain (Belgique)
Professeur Charles A. BERRY (1982)
Centre de Médecine Préventive, Houston (U.S.A)
Professeur Pierre-Marie GALETTI (1982)
Brown University, Providence (U.S.A)
Professeur Mamish Nisbet MUNRO (1982)
Massachusetts Institute of Technology (U.S.A)
Professeur Mildred T. STAHLMAN (1982)
Wanderbilt University, Nashville (U.S.A)
Harry J. BUNCKE (1989)
Université de Californie, San Francisco (U.S.A)
Professeur Théodore H. SCHIEBLER (1989)
Institut d'Anatomie de Würtzburg (R.F.A)
Professeur Maria DELIVORIA-PAPADOPOULOS (1996)
Université de Pennsylvanie (U.S.A)
Professeur Mashaki KASHIWARA (1996)
Research Institute for Mathematical Sciences de Kyoto (JAPON)
Professeur Ralph GRÄSBECK (1996)
Université d'Helsinki (FINLANDE)
Professeur James STEICHEN (1997)
Université d'Indianapolis (U.S.A)
Professeur Duong Quang TRUNG (1997)
Centre Universitaire de Formation et de Perfectionnement des Professionnels de Santé d'Hô Chi Minh-Ville (VIÊTNAM)

À Lucie et Jeanne,

À mes parents,

À ma famille.

À notre maître et président de thèse,

Monsieur le Professeur François KOHLER,

Professeur de Biostatistiques et Informatique Médicale (option biologie)

Pour m'avoir fait l'honneur de présider le jury de cette thèse,

Pour la richesse du thème de recherche que vous m'avez proposé.

Vous m'avez guidé sur les chemins de l'informatique médicale et accordé votre soutien tout au long de mon internat,

Veuillez trouver ici l'expression de ma profonde gratitude et de mon plus grand respect.

À nos juges,

À Monsieur le Professeur Yves MARTINET,

Professeur de Pneumologie

Vous me faites l'honneur de juger ce travail,

Veuillez trouver ici l'expression de mon profond respect.

À Madame le Professeur Eliane ALBUISSON,

Professeur de Biostatistiques, Informatique Médicale et Technologies de la Communication

Pour l'intérêt que vous portez à mon travail présent et à venir,

Soyez assurée de ma reconnaissance.

À Monsieur Amedeo NAPOLI,

Directeur de recherche CNRS, LORIA

Pour ton ouverture d'esprit et ta disponibilité,

Pour l'accueil que tu m'as réservé dans ton équipe,

Tu me fais l'honneur de siéger dans le jury de cette thèse,

Trouve ici l'expression de mes sincères remerciements.

Je remercie Jeanne Fresson, Michèle de Gaspéri, Brigitte Lacour pour leur collaboration.

Merci également à Sylvie Monchovet et Christine Roussel pour la relecture attentive de ce document.

Merci à Boris Marchand pour son aide sur mécano.

Merci à Gilles, pour ces quatre années, avec toute mon amitié.

Merci à mes beaux-parents pour leur gentillesse et leur soutien.

À mes amis stéphanois, blainvillois, nancéiens, lyonnais, parisiens...

SCD-UHP - NANCY
BIBLIOTHÈQUE
MÉDECINE

SERMENT

"Au moment d'être admis à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité. Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux. Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité. J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences. Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis dans l'intimité des personnes, je tairai les secrets qui me sont confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs. Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré et méprisé si j'y manque".

Table des matières

Introduction **16**

1 Contexte **18**

1.1 Le concept de trajectoire de soins 18

1.1.1 Définition de la notion de trajectoire 19

1.1.2 Modélisation 20

1.1.3 État des lieux 21

1.2 L'Extraction de Connaissances dans les Bases de Données 23

1.3 Application au domaine de la cancérologie 25

1.4 Présentation du P.M.S.I. 25

1.4.1 Le RSS en détail 26

1.4.2 La classification en Groupes Homogènes de Malades (GHM) 28

2 Matériel et méthode **30**

2.1 Matériel 30

2.1.1 Sélection de l'échantillon 30

2.1.2 Chaînage de l'information 31

2.1.3 Constitution de la base de données 32

2.1.4 Nature des données collectées 33

2.2 Méthode 34

2.2.1 Constitution de files actives par localisation 34
2.2.2 Treillis de Galois 35
2.2.3 Motifs séquentiels fréquents 40
2.2.4 Implémentation 44

3 Résultats **45**

3.1 Prise en charge globale 45
3.1.1 Âge, sexe, hospitalisations 45
3.1.2 Localisation des tumeurs 49
3.2 Prise en charge par établissement 52
3.2.1 Prise en charge du cancer au CHU de Nancy 52
3.2.2 Prise en charge du cancer au Centre Alexis Vautrin 55
3.2.3 Prise en charge du cancer au SINCAL 57
3.2.4 Prise en charge du cancer à la Maternité Régionale de Nancy 58
3.3 Prise en charge partagée 59
3.3.1 Flux inter-établissements 59
3.3.2 Partage entre le CHU et le Centre Alexis Vautrin 63
3.3.3 Population 63
3.3.4 Hospitalisations 63
3.3.5 Localisation des tumeurs 63
3.3.6 Nature de la coopération CHU-CAV 66
3.4 Mise en œuvre des motifs séquentiels fréquents 68
3.4.1 Aspects pratiques 68
3.4.2 Résultats 69

4 Discussion **71**

4.1 Utilisation du PMSI dans l'étude de la trajectoire de soins 71

4.2 Apports de l'extraction de connaissances dans les bases de données 72
4.3 Limites 73
4.3.1 Limites liées aux données 73
4.3.2 Limites liées à la méthode 74
4.4 Perspectives 76

Conclusion **78**

Bibliographie **80**

SCD-UHP - NANCY
BIBLIOTHÈQUE
MÉDECINE

Introduction

Le Programme de Médicalisation des Systèmes d'Information (PMSI) a comme objectif de décrire l'activité de soins des hôpitaux. Cette analyse est fondée sur le recueil systématique et le traitement automatisé d'une information médico-administrative minimale et standardisée.

Les données collectées font l'objet d'un classement en un nombre volontairement limité de groupes de séjours présentant une similitude médicale et un coût voisin. Ces groupes forment la classification en Groupes Homogènes de Malades (GHM). Avec la mise en place de la Tarification À l'Activité (T2A), l'utilisation du PMSI prend une orientation résolument budgétaire : les fonds alloués aux hôpitaux par l'assurance maladie seront en grande partie calculés sur la mesure de leur activité par le biais des GHM.

Le recueil et l'exploitation de routine des données du PMSI sont centrés sur les séjours des patients et non sur les patients eux-mêmes. Ce modèle reflète mal l'activité de soins liée aux pathologies au long court comme les maladies cardio-vasculaires ou le cancer. Jusqu'en 2001, il était impossible de relier entre elles les hospitalisations d'un même patient dans des établissements différents. La récente mise en place d'une procédure de chaînage par patient de ces hospitalisations laisse entrevoir la perspective de faire du PMSI un outil de suivi longitudinal des patients. Les caractéristiques du parcours ou *trajectoire* qu'empruntent les patients dans un espace de soins sont aujourd'hui mal connues. La maîtrise d'un tel outil serait un progrès notable vers une meilleure compréhension de ces trajectoires. Elle permettrait de décrire, d'évaluer et d'améliorer la prise en charge de pathologies multidisciplinaires.

L'objectif de notre travail est d'évaluer le potentiel du PMSI pour expliciter la trajectoire de soins des patients dans le domaine de la cancérologie. Nous décrivons comment il peut être transformé en un outil de mesure de l'activité d'oncologie centré sur le patient. Nous introduisons enfin un modèle d'analyse des trajectoires de soins faisant appel à des méthodes informatiques d'extraction des connaissances dans les bases de données.

1 Contexte

1.1 Le concept de trajectoire de soins

Dans les réflexions et les débats sur la régulation du système de santé, la dimension de l'organisation des soins prend depuis quelques années une place croissante. La mise en adéquation des ressources disponibles avec les besoins de la population et leur optimisation tiennent une place de premier ordre dans l'amélioration de la qualité des soins. Elles contribuent, au coté de la mise en place des pratiques médicales de référence, à maintenir un système de soins performant compatible avec une régulation des dépenses de santé.

Ainsi, il existe plusieurs niveaux sur lesquels on peut agir parallèlement pour gagner en qualité et en efficience médico-économique :

- à l'échelle des professionnels de santé avec les guides de bonnes pratiques,
- à l'échelle de la population avec l'éducation, l'information et la promotion de la santé,
- à l'échelle du système de soins.

Cette dernière préoccupation apparaît dans l'évolution de la planification hospitalière, et elle est au cœur des expérimentations de réseaux de soins. Ceux-ci tentent de décloisonner les institutions, d'améliorer la coordination et la continuité des soins rendus aux malades, tout en mutualisant les ressources. Elle prend toute sa mesure dans le domaine des pathologies chroniques qui représentent notamment les maladies cardio-vasculaires et le cancer les principaux enjeux de santé publique en France. Leur prise en charge doit se faire dans la durée et exige

des stratégies au long cours avec une intervention multidisciplinaire autour du patient.

Cependant, à l'échelle de la population, la véritable nature de cette prise en charge reste mal connue. Le cloisonnement des secteurs institutionnels a produit des systèmes d'information non communicants, à partir desquels on peut connaître la répartition macro-économique des ressources entre les différents segments de l'offre, mais qui ne sont pas orientés pour suivre le patient et comprendre son comportement dans le système de soins. Récemment, de nombreux travaux tentent de déchiffrer ce qu'il est convenu d'appeler la « trajectoire de soins » des patients au sein du système de santé. Ils tirent partie de la mise en place de méthodes de chaînage de l'identifiant patient. Ces dispositifs permettent de rattacher à un même malade des épisodes de soins dispensés par des prestataires différents et de reconstituer un historique médical et économique. En outre, le chaînage de l'information et la maîtrise de la trajectoire de soins sont au centre de la logique des réseaux et de leur évaluation [1].

1.1.1 Définition de la notion de trajectoire

Perrier [2] a relevé dans la littérature quatre définitions de la notion de trajectoire patient :

1. « *Une trajectoire de maladie peut être vue comme une imbrication et une succession de tâches dont l'ensemble constitue un arc de travail* »
2. « *On appelle trajectoire une succession dans le temps de volumes de type « prises en charge », à partir d'un instant initial, et pendant une durée donnée* »
3. « *Une trajectoire patient est une succession dans le temps d'évènements relatifs à un patient ou à un groupe de patients, relevant d'un problème ayant provoqué cette succession d'évènements* ».
4. « *La trajectoire d'un patient est une séquence ordonnée d'évènements pathologiques et d'interactions avec le système de soins* ».

Premier constat, la trajectoire est une *séquence* d'événements, donc une liste d'états *ordonnés* dans le temps. De ce fait peut-on considérer qu'un patient hospitalisé une première fois

pour infarctus, puis pour angor, suit la même trajectoire qu'un patient présentant ces deux épisodes dans l'ordre inverse ? De plus, cette liste d'évènements est bornée à la fois dans le temps et dans l'espace.

Deuxième constat, les définitions 2 et 4 intègrent la notion de trajectoire de « maladie », alors que les deux autres prennent en compte tous les événements de santé quels qu'ils soient. Enfin, la troisième définition fait référence à des groupes de patients, sous-entendant l'existence de *profils* types de trajectoire. Il ressort de ces définitions que la notion de trajectoire est « point-de-vue » dépendante et que plusieurs facteurs déterminent ce point-de-vue :

- la fenêtre d'observation détermine les bornes à la fois spatiales et temporelles à l'intérieur desquelles une trajectoire peut être reconstituée ;
- la qualité (au sens métier) de l'observateur détermine quels types d'événements de santé comptent dans la construction de la trajectoire ;
- l'information contenue dans une trajectoire est de caractère polymorphe, à la fois temporelle, géographique, démographique, médicale et économique.
- l'échelle d'observation peut être individuelle, ou collective impliquant l'existence d'une typologie des trajectoires.

1.1.2 Modélisation

Naiditch [3] oppose deux modèles de classification des trajectoires de soins, qui organisent de manière différente leur découpage en « épisodes » de soins. Le modèle des ACG (Adjusted Clinical Groups, cf 1.1.3) se fonde sur une approche populationnelle. Il est orienté patient, l'abordant dans sa globalité en intégrant l'ensemble de ses problèmes de santé sur une période donnée. A ce modèle s'opposent des classifications plus médicalisées, orientées vers une pathologie particulière. Elles sont plus complexes et confrontées à des choix théoriques souvent liés à la définition de l'épisode même. Pour Naiditch, les principaux problèmes que doivent résoudre ces systèmes sont les suivants :

- comment définir les entités médicales sur lesquelles s'appuient les épisodes de soins ?
- comment définir la longueur de ces épisodes, leur début, leur fin ?
- comment définir des règles de création de plusieurs épisodes ?
- comment gérer la nature des liaisons existant entre eux ?

Pour Riou [4], la représentation des trajectoires de soins nécessite trois éléments qualitatifs indispensables. La population concernée est-elle composée de patients tout venant ou sélectionnés sur une pathologie particulière ? Les bornes d'observation sont physiques et temporelles : la trajectoire part d'un service adresseur vers un service receveur et dure le temps de l'épisode pathologique. Enfin le contenu est lié à l'objectif poursuivi mais reste aussi fortement dépendant des capacités du système d'information.

1.1.3 État des lieux

Adjusted Clinical Groups

Les Adjusted Clinical Groups (ACG), ex Ambulatory Care Groups, sont apparus aux États-Unis dans les années 80, après les travaux de Starfield[5]. Ils sont nés de la volonté des compagnies d'assurances de se doter d'outils susceptibles d'expliquer et mieux encore de prédire le niveau de consommation médicale annuelle d'un assuré. Élaboré à partir du même algorithme que les Diagnoses Related Groups (DRG), la version américaine des GHM, ils classent des patients en 93 groupes différents selon leur profil pathologique. Chaque pathologie d'un patient est tout d'abord classée dans un des 32 All Diagnoses Group (ADG). Un patient peut donc appartenir à plusieurs ADG. Une deuxième étape oriente les patients vers un ACG final, en tenant compte de la polypathologie à partir des associations diagnostiques fréquentes.

Au total, les ACG sont bien un système de classification orienté patient, à la différence des DRG. Pour Rodrigues et al. [6], ils constituent un modèle tarifaire robuste pour un système de paiement à la capitation. Mais le modèle des ACG élaboré pour le *managed care*, est conçu pour classer des patients pris en charge dans un espace de soins fermé : les Health Maintenance

Organizations (HMO). Sa finalité est essentiellement tarifaire. Naiditch [3] le décrit comme faiblement médicalisé, ce qui pourrait limiter son utilisation pour l'évaluation des pratiques médicales.

Les expériences françaises

Les études menées en France sur la trajectoire de soins poursuivent plusieurs objectifs.

- comprendre les mécanismes qui génèrent des flux de patients entre plusieurs prestataires de soins [7], ou à l'intérieur d'un même établissement [8] ;
- évaluer les modalités médicales de prise en charge de certaines pathologies [9] ;
- évaluer le coût de la prise en charge globale de certaines pathologies [2, 10, 11] ;
- tester des outils de mesure de l'activité médicale ambulatoire [6].

Le cloisonnement des institutions et l'hétérogénéité des systèmes d'information limitent la construction d'une trajectoire globale de soins. Le projet Agora Data vise à analyser la trajectoire de soins des patients souffrant de lombalgies. Dans ce cadre, Le Duff [12] définissent un modèle d'intégration de données médicales existantes fournies par l'assurance maladie, les hôpitaux et les professionnels de santé. Ce modèle est conçu pour analyser la trajectoire de soins à l'aide de méthodes de fouille de données.

Quels outils pour analyser les trajectoires ?

La trajectoire de soins est un objet complexe :

- les contacts de soins surviennent de façon séquentielle, avec ou sans périodicité ;
- ces contacts mettent les patients en relation avec des prestataires de spécialités diverses, disséminés sur le plan géographique ;
- l'étendue des problèmes de santé et des procédures médicales est très vaste, même dans le cadre d'une pathologie particulière ;

Cette complexité et le manque de connaissances a priori limitent les possibilités qu'offrent les méthodes d'analyse statistique classiques. De nombreux travaux s'orientent vers la mise en œuvre de techniques particulières, issues des domaines de la fouille de données (Data mining) et de la représentation des connaissances [13, 12, 14, 8].

1.2 L'Extraction de Connaissances dans les Bases de Données

« We're drowning in information and starving for knowledge. » John Naisbitt

« Nous sommes submergés d'information et avides de connaissance. »

Durant ces dernières années, les techniques de production et de stockage de données ont connu un très grand développement, notamment dans le domaine médical. Parallèlement à la génération de ces grandes masses de données, les outils informatiques de stockage, de manipulation et de retrait de données ont connu des développements très rapides. Les techniques statistiques classiques montrent leurs limites face à des données d'un tel volume et une telle complexité.

L'Extraction de Connaissances dans les bases de Données (ECBD) (en anglais « Knowledge Discorvery in Data-bases (KDD) ») est un processus fortement interactif, itératif, qui s'inscrit dans une démarche d'aide à la décision : les données disponibles sont d'abord transformées en une information dont on va extraire une connaissance intelligible, utile, intéressante dans le contexte, tournée vers l'aide à décision [15].

L'ECBD est un processus *cyclique* qui regroupe plusieurs étapes :

- une étape de sélection de données cibles dans un entrepôt de données,
- une étape de pré-traitement,
- une étape de transformation,
- une étape de fouille de données ou data mining,
- une étape de visualisation/interprétation des résultats.

Ce processus conduit à l'extraction de connaissances nouvelles qui peuvent servir de point de départ à un nouveau cycle (figure 1.1).

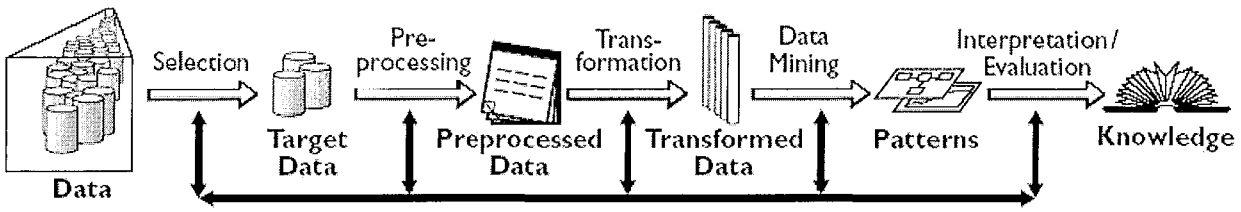


FIG. 1.1: Le processus d'Extraction de Connaissances dans les Bases de Données (extrait de [15])

Le data mining correspond à l'ensemble des techniques et des méthodes qui, à partir de gros volumes de données, permettent d'obtenir des connaissances exploitables. Il existe une distinction précise entre le concept d'ECBD et celui de data mining. En effet, ce dernier n'est que l'une des étapes du processus de découverte de connaissances correspondant à l'extraction des connaissances à partir des données.

Le data mining est de plus en plus employé dans la recherche bio-médicale, et ses applications touchent de nombreux domaines : analyse du génome, aide au diagnostic médical, construction d'ontologies médicales, imagerie, pharmacovigilance ... Dans le contexte actuel, l'analyse de la trajectoire de soins se prête particulièrement bien au data mining. En effet, il n'existe que rarement de recueil d'information dévolu aux trajectoires et les connaissances dont disposent les experts du domaine médical sont parcellaires. Les systèmes d'information en santé, comme ceux de l'Assurance Maladie ou le Programme de Médicalisation des Systèmes d'Information, produisent une masse énorme de données médico-économiques. Bien qu'initialement collectées à d'autres fins, ces données complexes sont susceptibles de receler des connaissances « cachées » sur la trajectoire de soins.

1.3 Application au domaine de la cancérologie

Avec 260 000 nouveaux cas par an et 150 000 décès [16], le cancer est un problème majeur de santé publique. Il relève d'une prise en charge multidisciplinaire et de nombreuses structures hospitalières, publiques ou privées y participent, accompagnées par la médecine de ville. Cependant, si la coordination des soins tend à se développer entre différentes structures, le cancer ne bénéficie pas encore d'un suivi particulier dans le paysage sanitaire français et reste noyé dans les spécialités médicales et chirurgicales [17]. Plusieurs études, essentiellement médico-économiques, ont tenté d'évaluer de manière globale le coût de la prise en charge du cancer [17, 11, 18, 19]. Bien qu'elles se soient déroulées avant la mise en place du chaînage des séjours, elles marquent l'intérêt que présentent les données du PMSI dans une approche globale du patient. En effet, le PMSI est le seul système d'information fournissant à l'échelon national, de manière standardisée, continue et exhaustive des informations médico-économiques sur la prise en charge hospitalière du cancer. Bien que, d'un point de vue épidémiologique, ces données soient moins précises que celles des registres, ou de l'enquête permanente cancer des centres de la FNCLCC[1], elles le demeurent suffisamment pour repérer et distinguer différentes localisations de cancers [20]. Le PMSI a d'ailleurs servi de support à une estimation de l'incidence du cancer colorectal dans la région de Saint-Etienne [21].

1.4 Présentation du P.M.S.I.

Le Programme de Médicalisation Des Systèmes d'Information (PMSI) est le système qui permet une mesure médicalisée de la production hospitalière française, et l'adaptation de la rémunération des hôpitaux et des cliniques à ce produit. Il repose sur la nomenclature des Groupes Homogènes de Malades (GHM), qui est une adaptation de la classification américaine des Diagnosis Related Groups (DRG) élaborée par l'équipe du Professeur Robert Fetter (Uni-

[1]Fédération Nationale des Centres de Lutte Contre le Cancer

versité de Yale, États-Unis). Elle permet de classer les séjours hospitaliers dans des groupes présentant une double homogénéité en termes médical et économique.

Dans ce système, chaque séjour donne lieu à la réalisation d'un Résumé Standardisé de Sortie (RSS), constitué d'un petit nombre d'informations administratives (âge,durée de séjour...) et médicales (diagnostic principal, co-morbidités, actes médicaux et chirurgicaux) relatives au patient pendant ce séjour.

Les informations médicales sont des informations codées :

- au moyen de la Classification statistique Internationale des Maladies et problèmes de santé connexes (CIM10) pour les diagnostics,
- au moyen de la Classification Commune des Actes Médicaux pour les actes médicaux et chirurgicaux. Celle-ci a remplacé le Catalogue des Actes Médicaux utilisé jusqu'en 2003.

Selon l'organisation des établissements, ce codage peut être soit

- décentralisé : il est alors réalisé directement dans les services de soins, au plus près du malade
- centralisé : dans les Départements d'information médicale (DIM).

L'obtention de ces informations conditionne le fonctionnement du dispositif PMSI dans toutes ses implications et utilisations, qu'elles soient d'ordre budgétaire ou autre.

1.4.1 Le RSS en détail

Le RSS est divisé en Résumés d'Unité Médicale (RUM). On désigne par unité médicale un ensemble individualisé de moyens assurant des soins à des malades hospitalisés, repéré par un code spécifique dans une nomenclature déterminée par l'établissement [22].

La partie administrative du RUM comprend les données suivantes :

- le numéro de RSS : il permet de relier entre eux les différents RUM de l'hospitalisation d'un patient au sein de la même entité juridique,
- le sexe

- la date de naissance
- le code postal du lieu de résidence
- le numéro FINESS de l'entité juridique
- le numéro identifiant l'Unité Médicale(UM) responsable des soins du patient
- les dates d'entrée et de sortie de l'unité médicale
- les modes d'entrée et de sortie, ainsi que la provenance et la destination du patient. Ces éléments permettent de flécher le parcours du patient dans l'établissement
- le nombre de séances

La partie médicale du RUM se compose :

- de diagnostics, codés à partir de la $10^{ème}$ révision de la Classification Internationale des maladies.
- d'actes médicaux codés à partir de la Classification Commune des Actes Médicaux.
- poids à l'entrée dans l'unité médicale pour les nouveau-nés
- indice de gravité simplifié, pour les patients de plus de 15 anshospitalisés en secteur réanimation, soins intensifs ou surveillance continue.

L'information médicale contenue dans le RUM obéit à des règles de hiérarchisation bien définies. Ainsi, après le passage d'un patient dans une unité médicale, le médecin responsable détermine quel motif de soins a mobilisé l'essentiel de l'effort médical et soignant. Celle-ci devient le diagnostic principal (DP) [22]. Il doit être obligatoirement renseigné à chaque séjour en unité médicale. Il existe d'autres types de diagnostics, dont le renseignement, facultatif, permet de mieux caractériser le séjour d'un patient :

- le Diagnostic Relié (DR) a pour rôle, en association avec le DP et lorsque celui-ci n'y suffit pas, de rendre compte de la pathologie du patient.
- les Diagnostics Associés Significatifs (DAS) sont d'autres morbidités ayant donné lieu à une prise en charge diagnostique ou thérapeutique supplémentaire au cours du séjour. Ils peuvent influer sur la classification du séjour en GHM.

- les Données Associées Documentaires (DAD) peuvent recenser n'importe quel type d'information, codes d'actes ou de diagnostics ou langage naturel. Elles n'interviennent pas dans le groupage du séjour et ne sont pas transmises à la tutelle.

1.4.2 La classification en Groupes Homogènes de Malades (GHM)

La collecte et le codage de l'ensemble des informations de l'hospitalisation d'un patient permet le classement de celle-ci dans un Groupe Homogène de Malade au moyen d'un algorithme *ad-hoc*. La classification GHM est issue de la classification des DRG, classification médico-économique des séjours hospitaliers, élaborée par l'équipe du Professeur Fetter de l'université de Yale, au cours des années 70 [23]. Les DRG ont été construits à partir d'une base de données de plus d'un million de séjours hospitaliers. L'objectif était de décrire ces séjours selon deux axes, médical et économique. Ainsi, les séjours classés dans un même groupe ont engendré une consommation de ressources comparable, tout en préservant une cohérence médico-chirurgicale.Les DRG ont été importés en France au début des années 80, avec la naissance du PMSI, puis adaptés aux pratiques médicales françaises, tout en respectant la philosophie de la classification : analyse de la variance des coûts par séjour et intervention d'experts médicaux. Cependant, la version française de la classification GHM comporte deux particularités majeures :

- son application aux séjours très courts (moins de 72 heures),
- la création d'une catégorie de « séances » accueillant les séances de dialyse, radiothérapie et chimiothérapie.

Elle a subi, depuis, de nombreuses modifications pour tenir compte des changements de nomenclature, de l'évolution des pratiques professionnelles et des caractéristiques du système de santé français. Actuellement, la classification GHM, dans sa version 9, compte environ 600 Groupes Homogènes de Malades. Ceux-ci sont rassemblés en 26 Catégories Majeures de Diag-

nostics (CMD) qui correspondent le plus souvent à un découpage fonctionnel par appareil (ex : CMD 05 Affections de la l'appareil circulatoire). Par ailleurs, les séjours de moins de quarante-huit heures sont regroupés dans la Catégorie Majeure, dite CM24, puisque c'est la durée de séjour qui est alors déterminante. La CM 90 comprend quant à elle des « Erreurs et autres séjours inclassables » [24].

L'algorithme de classification des GHM repose sur un arbre de décision faisant appel à des listes prédéfinies de diagnostics ou d'actes et des variables discriminantes telles que la durée de séjour, l'âge, le sexe, le mode de sortie, la présence de comorbidités, d'actes classants...

2 Matériel et méthode

Nous avons analysé la trajectoire des patients atteints de cancer dans les établissements du site hautement spécialisé public de Nancy. Leur parcours a été reconstitué grâce aux données du PMSI, puis exploré par une combinaison de méthodes statistiques classiques et d'extraction de connaissances dans les bases de données.

2.1 Matériel

2.1.1 Sélection de l'échantillon

Population : Les patients ont été sélectionnés dans les bases PMSI 2003 des établissements suivants :

- le CHU de Nancy
- la Maternité Régionale A. Pinard, Nancy
- le Syndicat Nancéien de la Chirurgie de l'Appareil Locomoteur (SINCAL), Nancy
- le Centre Alexis Vautrin, Vandœuvre-les-Nancy

Critère d'inclusion : avoir eu, au cours de l'année 2003, au moins une hospitalisation avec un Diagnostic Principal (DP) ou Relié (DR) de cancer. La liste des codes marqueurs de cancer retenus pour cette étude a été élaborée par le Comité Technique Régional de l'Information Médicale (COTRIM) de Lorraine, à partir de la CIM10.

Critère d'exclusion : aucun.

2.1.2 Chaînage de l'information

Depuis 2001, le PMSI intègre un dispositif de chaînage. Il permet de relier entre elles, grâce à un numéro de chaînage anonyme, les différentes hospitalisations d'un même patient. Afin de permettre les opérations de chaînage et de garantir une totale confidentialité des données, différents fichiers informatiques sont produits et différentes actions logicielles sont exercées. Le principe du chaînage anonyme consiste en la création d'un numéro anonyme propre à même patient mais commun à toutes ses hospitalisations (quel que soit le secteur d'hospitalisation : public ou privé d'une part, mais aussi MCO ou soins de suite ou de réadaptation (SSR) ou même psychiatrie à l'avenir). Sur une période donnée, les différents épisodes d'hospitalisation d'un même patient peuvent ainsi être identifiés et liés entre eux.

Le numéro de chaînage anonyme du patient est créé à partir de variables identifiantes, par un procédé sécurisé dénommé Fonction d'Occultation des Informations Nominatives (FOIN). Les variables identifiantes utilisées pour construire le numéro anonyme sont :

- le numéro d'assuré social (ouvrant droit),
- la date de naissance du patient,
- le sexe du patient.

Il est anonyme car :

- il est impossible de retrouver l'identité d'un patient à partir de ce numéro (non réversibilité du cryptage) ;
- à partir des données nominatives de deux personnes, qui seraient approchantes, on n'obtient pas de numéros anonymes similaires.

Il est caractéristique d'un individu car :

- à partir des mêmes données d'identité on obtient le même numéro anonyme (reproductibilité) ;
- la probabilité d'obtenir le même numéro à partir de données nominatives différentes est très faible.

À noter, toutefois, que les jumeaux de même sexe ayant le même ouvrant droit ne peuvent être distingués.

Il est donc possible d'identifier par ce numéro anonyme les différentes hospitalisations d'un même malade, tant dans le secteur public que dans le secteur privé, sans connaître l'identité de ce patient. C'est cette opération qui est dénommée chaînage anonyme.

2.1.3 Constitution de la base de données

En pratique, le rassemblement des données s'est effectué en trois étapes :

1. Au sein de chaque établissement, l'investigateur local a sélectionné dans son entité juridique les patients respectant les critères d'inclusion. Ces patients ont été identifiés par leur numéro de chaînage anonymisé.
2. Chaque investigateur a alors remis en mains propres au coordonnateur la liste des numéros anonymes retenus, sur support adéquat.
3. Dans un troisième temps, le coordonnateur a remis en mains propres la liste des numéros anonymes de tous les patients sélectionnés (provenant des quatre établissements) à chaque investigateur local. Ce dernier a rassemblé, concernant chaque patient de la liste, les données PMSI 2003 d'éventuelles hospitalisations dans son établissement, quel qu'en soit le motif.

La décomposition du processus de recueil en trois étapes nous a permis de récupérer, dans un établissement donné, des informations sur un patient sélectionné par un autre établissement.

Pour automatiser le recueil, et garantir le maximum de confidentialité, nous avons développé deux applications en java, que chaque site a pu appliquer :

- Mecano : extrait, à partir d'une liste de codes CIM10 marqueurs, les numéros anonymes des patients chez qui on retrouve au moins un marqueur en Diagnostic Principal ou Diagnostic Relié.

- Anomec : extrait, à partir d'une liste de numéros anonymes de patients, une partie des informations contenues dans les RUMs de ces patients (cf 2.1.4.

Ces deux algorithmes prennent en entrée un fichier de RUM au format 108 (groupé) et les fichiers ANO-HOSP et HOSP-PMSI utilisés pour la constitution des Résumés de Sorties Anonymes.

Cette procédure présente deux avantages essentiels :

- elle est standardisée, fonctionne avec des fichiers normalisés et reste reproductible dans tout établissement possédant une activité MCO ;
- seule les données nécessaires à l'étude circulent. Les investigateurs locaux n'ont accès qu'à la liste de numéros anonymes chaînables. Le coordonnateur n'a pas accès aux bases de données de chaque établissement.

2.1.4 Nature des données collectées

Les quatre établissements de l'étude ont produit pour l'année 2003 leurs données PMSI en utilisant les mêmes standards :

- Fichier de RUM au format 108,
- Classification Commune des actes médicaux dans sa version V0,
- Version 7 de la classification des Groupes Homogènes de Malades

Les informations transmises au coordonnateur pour l'analyse sont extraites du RUM et comportent :

- l'identification de l'établissement par son numéro FINESS,
- le numéro de chaînage anonyme des patient,
- la date d'entrée dans l'entité juridique,
- le mode d'entrée dans l'entité juridique,
- la provenance,
- la date de sortie de l'entité juridique,
- le mode sortie de l'entité juridique,

- la destination,
- le sexe,
- l'âge en années (entières),
- le Groupe Homogène de Malade (GHM) version 7,
- les Diagnostics Principaux (DP) des Résumés d'Unité Médicale (RUM),
- les Diagnostics Reliés (DR) des RUM.

Pour plus de confidentialité, à l'issue de la collecte, le numéro de chaînage anonyme a été remplacé par un numéro généré aléatoirement et toute correspondance entre les deux a été supprimée. Les données collectées étant à caractère indirectement nominatif, le protocole a fait l'objet d'une procédure de déclaration à la Commission Nationale Informatique et Libertés.

2.2 Méthode

2.2.1 Constitution de files actives par localisation

Pour chaque patient, nous avons cherché dans chacune de ses hospitalisations, les DP ou DR indiquant une localisation maligne primitive. Quatorze grandes catégories de localisations ont été retenues sur la base de la classification utilisée par le logiciel DAC [25]. Cet outil, développé à l'Assistance Publique-Hôpitaux de Paris, permet une description des séjours de cancérologie d'un établissement à partir des données PMSI. Il regroupe les localisations tumorales dans les catégories suivantes :

- tumeurs malignes pleuro-pulmonaires,
- tumeurs malignes du sein,
- tumeurs malignes des organes génitaux féminins (hors sein),
- tumeurs malignes des organes génitaux masculins (y compris prostate),
- tumeurs malignes du système urinaire (hors prostate),
- tumeurs malignes de l'appareil digestif (hors système hépato-biliaire),

- tumeurs malignes du système hépato-biliaire,
- tumeurs malignes cutanées,
- tumeurs malignes des Voies aéro-digestives supérieures,
- tumeurs malignes du sang et des organes hématopoïétiques,
- tumeurs malignes des Os, tissu conjonctif et autres tissus mous,
- tumeurs malignes des glandes endocrines,
- tumeurs malignes du Système nerveux central et oeil,
- autres tumeurs malignes.

Lorsqu'une localisation primitive est retrouvée chez un patient, celui-ci est ajouté à la file active de la catégorie contenant cette localisation. Ainsi, une patiente ayant eu un Diagnostic Relié codé :

```
C540 :TUMEUR MALIGNE DE L'ISTHME DE L'UTERUS
```

est ajoutée à la file active *Tumeurs malignes des organes génitaux féminins (hors sein).*

Par ce mécanisme, les patients ayant eu des diagnostics appartenant à des catégories différentes peuvent participer à des files actives différentes. En revanche, ceux qui ont présenté des tumeurs malignes localisées dans le même appareil ne sont comptés qu'une seule fois dans sa file active (par exemple : une tumeur de l'œsophage et une autre du colon).

2.2.2 Treillis de Galois

Les treillis de Galois sont à la base d'une famille de méthodes de classification conceptuelle introduite par Barbut et Monjardet [26]. Wille a popularisé cette approche en l'utilisant comme support de l'analyse formelle de concepts [27]. Godin [28] propose une revue des méthodes de classification conceptuelle basées sur les treillis de Galois et leurs applications.

Un treillis de Galois peut être vu comme une hiérarchie de *concepts.* Chaque concept est une paire composée d'une *extension* et d'un *intension.* L'extension est un ensemble d'objets ou

instances du domaine étudié. L'intension est un ensemble de proprités communes aux instances de l'extension.

Aspects formels

Nous reprenons ici les aspects formels des treillis de Galois tels qu'ils sont décris par Godin [28].

Soient deux ensembles finis E et E', et une relation binaire $\mathcal{R} \subseteq E \times E'$. On peut représenter par un treillis le regroupement naturel des éléments de E et E' par rapport à $\mathcal{R}$. Ce treillis est appelé treillis de Galois. Chaque élément du treillis est un couple noté (X, X') appelé concept formel. X et X' sont des parties de E, respectivement E' vérifiant :

- $X' = f(X)$ où $f(X) = \{x' \in E' | \forall x \in X, x\mathcal{R}x'\}$
- $X = f'(X')$ où $f'(X') = \{x \in E | \forall x' \in X', x\mathcal{R}x'\}$

Bien que cette définition soit symétrique, E fait généralement référence aux objets d'une application et E' à leur propriétés ou attributs. Par définition, $f(\emptyset) = E'$ et $f'(\emptyset) = E$. Les couples (X, X') vérifiant les deux propriétés précédentes sont dits fermés.

Le treillis de Galois G de la relation $\mathcal{R}$ est l'ensemble de tous les couples fermés dérivés de $\mathcal{R}$. Si on munit $P(E)$ de l'inclusion $\subseteq$, on peut définir la fermeture $h = f \circ f' : P(E) \rightarrow P(E)$ vérifiant :

- $\forall (X, Y), X \subseteq Y \Rightarrow h(X) \subseteq h(Y)$
- $\forall X, X \subseteq h(X)$
- $\forall X, h(h(X)) = h(X)$

Si $h(X) = X$, X est dit fermé pour h. L'ensemble des éléments fermés pour h forme un treillis H isomorphe à G. Il représente les éléments X des couples complets de G. On peut symétriquement définir la fermeture $h' = f \circ f'$ de $P(E')$ dans $P(E')$ et le treillis H' des fermés pour h'. Ceux-ci représentent les éléments X' des couples complets de G et H' est isomorphe à

G. Une relation d'ordre est définie sur les concepts de G. Soient $C_1 = (X_1, X_1')$ et $C_2 = (X_2, X_2')$

$$C_1 < C_2 \Leftrightarrow X_1' \subset X_2' \Leftrightarrow X_2 \subset X_1$$

C_1 est appelé père de C_2, et C_2 fils de C_1. Cette relation d'ordre est utilisée pour générer le graphe (diagramme de Hasse) du treillis G. (C_1, C_2) est un arc du graphe si :

- $C_1 < C_2$
- il n'existe pas de concept C_3 tel que : $C_1 < C_3 < C_2$

Lorsqu'on dessine le graphe, les arcs sont par convention tous orientés dans le même sens, soit vers le haut soit vers le bas. Le diagramme de Hasse peut s'interpréter comme une généralisation/spécialisation de concepts où C_1 est plus général que C_2 si $C_1 < C_2$.

Exemple

Soient un ensemble $P = \{1, 2, 3, 4, 5, 6\}$ de 6 patients et un ensemble $E = \{A, B, C, D\}$ de 4 établissements. On désire représenter la manière dont les patients fréquentent ces établissements. Ces informations sont données par la relation binaire $\mathcal{R}$ « a fréquenté » du tableau 2.1. Ainsi, le patient 1 a fréquenté les établissements A et D.

Le treillis de Galois de cette relation comporte huit concepts consignés dans le tableau 2.2. Ces huit concepts peuvent être interprétés comme des « classes » de prise en charge des patients par les établissements.

	A	B	C	D
1	1	0	0	1
2	1	1	0	0
3	0	1	0	0
4	1	0	1	0
5	1	1	1	0
6	1	0	0	0

TAB. 2.1: Relation $\mathcal{R}$ « a fréquenté »

extension	intension
{1}	{A,D}
{2,5}	{A,B}
{2,3,5}	{B}
{4,5}	{A,C}
{5}	{A,B,C}
{1,2,4,5,6}	{A}
{1,2,3,4,5,6}	{}
{}	{A,B,C,D}

TAB. 2.2: Concepts du treillis de la relation $\mathcal{R}$ « a fréquenté »

Soit le sous-ensemble de patients $Q = \{1, 2, 5\}$. Pour déterminer le fermé de l'ensemble Q, on procède de la manière suivante :

- on détermine quel est l'ensemble des propriétés communes aux patients 1,2 et 5, c'est-à-dire une liste d'établissements que *tous* ont fréquenté ;
- ici, il s'agit de $\{A\}$;
- on recherche ensuite la liste de tous les patients qui ont fréquenté A ;
- ce qui donne l'ensemble $\{1, 2, 4, 5, 6\}$

Ainsi, pour chaque concept :

- l'intension est le plus « grand » ensemble d'établissements que les patients de ce concept ont fréquentés,
- l'extension est le plus « grand » ensemble de patients que les établissements de ce concept se sont partagé.

Le diagramme de Hasse du treillis de la relation « a fréquenté » est représenté dans figure 2.1.

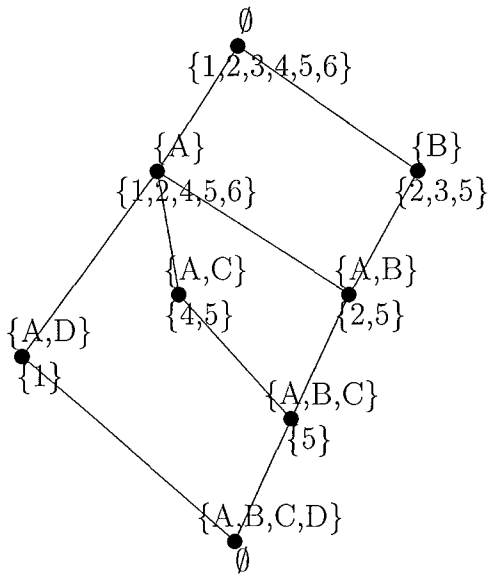


FIG. 2.1: Treillis de la relation $\mathcal{R}$

Représentation et variantes

La représentation précédente du treillis comporte des informations redondantes.

1. Un attribut d'un concept père est retrouvé dans tous ses descendants. Dans notre exemple, l'établissement A se retrouve dans les 5 concepts qui descendent de $(\{A\}; \{1, 2, 4, 5, 6\})$. On peut supprimer ces occurrences sans perte d'information.
2. De façon duale, un objet d'un concept fils est retrouvé dans tous ses ancêtres. Ainsi le patient 5 de notre exemple se retrouve dans les cinq ancêtres du concept $(\{A, B, C\}; \{5\})$

Il est possible de construire ce que l'on appelle un treillis d'*héritage* qui ne fait apparaître que les éléments non redondants de l'ensemble des concepts du treillis initial. Ne figurent alors dans l'intension d'un concept que les attributs qui n'apparaissent pas chez ses ancêtres. Les autres sont retrouvés par héritage. Par symétrie, ne figurent dans l'extension d'un concept que les objets n'apparaissant pas chez ses descendants.

Le treillis d'héritage de la relation R du tableau 2.1 est représenté dans la figure 2.2. Dans un treillis d'héritage, un objet n'est représenté que dans un et un seul concept. On peut donc définir dans notre exemple le concept dans lequel apparaît un patient comme sa « classe » de

prise en charge.

Une telle méthode de classification présente plusieurs intérêts :

- elle est non supervisée, c'est-à-dire qu'elle ne nécessite pas de connaissances à priopri sur les classes des objets ;
- elle permet d'établir une hiérarchie ;
- elle n'est pas réduite à un arbre ;
- sa représentation supporte une interface de navigation.

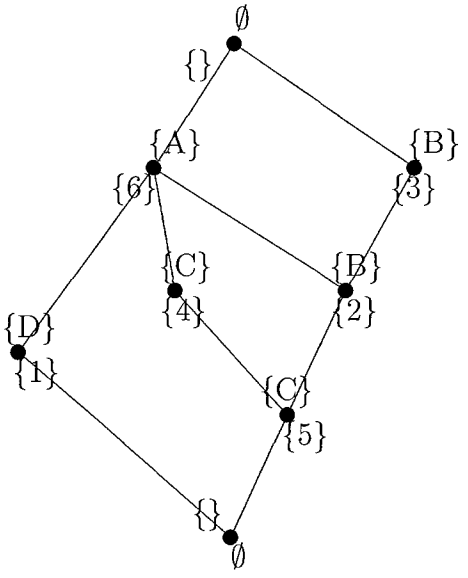


FIG. 2.2: Treillis d'héritage de la relation $\mathcal{R}$

2.2.3 Motifs séquentiels fréquents

Introduction

Introduite par Agrawal et Srikant [29], l'extraction de motifs séquentiels fréquents est une méthode de data mining destinée à découvrir de façon automatique des répétitions caractéristiques dans une base de données ordonnées. Ce peut être par exemple une succession de bases ATGC que l'on retrouve fréquemment dans le même ordre sur différents brins d'ADN, ou encore un comportement identique de lecteurs dans une bibliothèque qui emprunteraient

successivement et dans le même ordre *Les Trois Mousquetaires* puis *Vingt ans après* et enfin *Le Vicomte de Bragelonne* de la trilogie de Dumas.

Cadre formel

Définition 2.1 (Séquence) *Soit $\mathcal{I}$ un ensemble fini d'items. Une séquence engendrée par $\mathcal{I}$ est un liste ordonnée de parties de $\mathcal{I}$. On appelle taille d'une séquence S et on note $|S|$ le nombre d'items dans S. Une séquence de taille k est une k-séquence.*

Notation $(i_1 i_2 \dots i_n)$ est une *partie* de $\mathcal{I}$ dans lequel les i_j sont des *éléments* de $\mathcal{I}$. $S = \langle s_1 s_2 \dots s_m \rangle$ est une *séquence* dans laquelle les s_i sont des *parties* de $\mathcal{I}$. On dit également que les s_i sont des parties de S.

Exemple 2.1

Soit $\mathcal{I} = \{a, b, c, d, e\}$ un ensemble d'items.

- $S = \langle (a)(ae)(cde) \rangle$ est une séquence
- (ae) est une partie de S
- $|S| = 1 + 2 + 3 = 6$.

Définition 2.2 (Sous-séquence) *Une séquence $A = \langle a_1 a_2 \dots a_n \rangle$ est une sous-séquence de $B = \langle b_1 b_2 \dots b_m \rangle$ s'il existe n entiers $i_1, i_2, \dots i_n$ tels que :*

- $i_1 < i_2 < \dots < i_n$ *et*
- $a_1 \subset b_{i_1}, a_2 \subset b_{i_2} \dots a_n \subset b_{i_n}$

On note $A \subseteq B$. On dit également que B supporte A.

Exemple 2.2

Soit la séquence $S = \langle (c)(abc)(bd) \rangle$.

- $\langle (ab)(bd) \rangle$ est une sous-séquence de S,
- $\langle (c)(ad) \rangle$ n'est pas une sous-séquence de S.

Définition 2.3 (Super-séquence) *Une séquence S est une super-séquence de S' si et seulement si S' est une sous-séquence de S.*

Définition 2.4 (Base de séquences) *Soient $\mathcal{O}$ un ensemble fini d'objets, $\mathcal{I}$ un ensemble fini d'items et $\mathcal{S}$ l'ensemble des séquences engendrées par $\mathcal{I}$. Soit $\mathcal{R}$ une relation binaire entre $\mathcal{O}$ et $\mathcal{S}$ telle que $\forall o \in \mathcal{O}, \exists! S \in \mathcal{S} | (o, S) \in \mathcal{R}$. On appelle base de séquences le quadruplet $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{S}, \mathcal{R})$.*

Exemple 2.3

Soit $\mathcal{O} = \{p_1, p_2, p_3, p_4\}$ un ensemble de quatre patients. On reconstitue pour chacun la séquence de traitements qu'il a reçus parmi l'ensemble des traitements $\mathcal{I} = \{x, r, c\}$ où x est la chirurgie, r une cure de radiothérapie, c une cure de chimiothérapie. Les séquences de traitements sont les suivantes :

- p_1 a reçu x puis c puis c et à nouveau x,
- p_2 a reçu x et r de façon concomitante puis c,
- p_3 a reçu c puis c puis x,
- p_4 a reçu c puis c puis r puis r.

La base de séquences $\mathcal{D}$ des traitements des patients de $\mathcal{O}$ est représentée dans le tableau 2.3.

Définition 2.5 (Image d'une séquence) *Soit f la fonction qui associe à une séquence S de $\mathcal{S}$ l'ensemble des objets o de $\mathcal{O}$ qui sont associés à une super-séquence de S*

$$\begin{aligned} f : \Sigma &\rightarrow \mathcal{O} \\ S &\mapsto \{o / \exists S' \in \mathcal{S}, (o, S') \in \mathcal{R}, S \subseteq S'\} \end{aligned}$$

patients	séquences
p_1	$\langle(x)(c)(c)(x)\rangle$
p_2	$\langle(xr)(c)\rangle$
p_3	$\langle(c)(c)(x)\rangle$
p_1	$\langle(c)(c)(r)(r)\rangle$

TAB. 2.3: Base $\mathcal{D}$ des séquences de traitements reçus par les patients de l'ensemble $\mathcal{O}$

On dit qu'un objet o supporte une séquence S si $o \in f(S)$. *La séquence vide* $\emptyset = \langle\rangle$ *est supportée par tous les objets de* $\mathcal{O}$.

Exemple 2.4

Soit la séquence $S = \langle(x)(c)\rangle$. Dans la base de séquences du tableau 2.3, $f(S) = \{p_1, p_2\}$ représente tous les patients qui ont bénéficié d'une chirurgie, puis d'une chimiothérapie.

Définition 2.6 (Support d'une séquence) *On appelle support d'une séquence S dans* $\mathcal{D}$ *et on note* $\sigma_{\mathcal{D}}(S)$ *la proportion d'objets de* $\mathcal{O}$ *qui supportent S.*

$$\sigma_{\mathcal{D}}(S) = \frac{|f(S)|}{|\mathcal{O}|}$$

Exemple 2.5

$S = \langle(x)(c)\rangle$ est supportée par deux des quatre patients de $\mathcal{O}$. $\sigma_{\mathcal{D}}(S) = \frac{2}{4} = 0.5$

Définition 2.7 (Motif séquentiel fréquent) *Étant donné un seuil* $\epsilon \in [0, 1]$ *et une base de séquences* $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{S}, \mathcal{R})$, *une séquence S est un motif séquentiel fréquent si :*

$$\sigma_{\mathcal{D}}(S) \geq \epsilon$$

Définition 2.8 (Extraction de motifs séquentiels fréquents) *Le processus d'extraction de motifs séquentiels fréquents dans une base de séquences* $\mathcal{D} = (\mathcal{O}, \mathcal{I}, \mathcal{S}, \mathcal{R})$ *consiste à trouver toutes les séquences de* $\mathcal{S}$ *vérifiant la propriété précédente.*

Exemple 2.6

Si l'on fouille la base $\mathcal{D}$ du tableau 2.3 avec un support minimum de 0.5 on obtient les motifs séquentiels fréquents suivants :

- $\langle(x)\rangle$
- $\langle(r)\rangle$
- $\langle(c)\rangle$
- $\langle(x)(c)\rangle$
- $\langle(c)(x)\rangle$
- $\langle(c)(c)\rangle$
- $\langle(c)(c)(x)\rangle$

2.2.4 Implémentation

Les données ont été importées dans une base MySQL. Les statistiques descriptives, ainsi que les tests de comparaison de moyennes sont réalisés sous SAS V8.0. Comme dans beaucoup de travaux, la phase de préparation des données en ECBD demande un temps considérable. Pour automatiser certaines tâches et notamment la gestion des formats de données d'une application à l'autre, nous avons écrit des scripts en perl, SQL et java.

Les treillis de Galois sont construits et visualisés grâce à Concept Explorer [30] mais aussi JaLaBA [31]. Nous avons utilisé l'algorithme SLPMiner [32] pour extraire les motifs séquentiels fréquents.

3 Résultats

3.1 Prise en charge globale

3.1.1 Âge, sexe, hospitalisations

8032 patients ont été sélectionnés, 4101 hommes et 3931 femmes. La prise en charge de leur cancer représente 55065 RUM, soit 52290 RSS. 66,6% de ces hospitalisations ont eu lieu au Centre Alexis Vautrin , 33,1% au CHU et moins de 1% au SINCAL et à la Maternité. La majorité de ces hospitalisations dure moins de 24 heures (75,7%), 15,9 % durent entre un et 7 jours, 3,1 % entre 8 et 14 jours, et 5,2% durent plus de deux semaines. La figure 3.1 montre la répartition des durées de séjours par établissement. On note que 85% des hospitalisations au Centre Alexis Vautrin durent moins de 24 heures.

Le case-mix, qui est la répartition des RSS à travers les différents GHM, est très concentré. Neuf GHM représentent 80% des hospitalisations. En particulier, les GHM de radiothérapie (688 et 689) regroupent 44% d'entre elles. 25% des hospitalisations sont classées dans les GHM de chimiothérapie (681 et 587). La figure 3.2 montre la distribution des 28 GHM les plus fréquents, soit 90% des hospitalisations.

715 patients sont décédés pendant l'année. L'âge moyen à la première hospitalisation est de 57,49 ans.(tableau 3.1) Les patients sont hospitalisés pour leur cancer en moyenne 16,66 jours dans l'année, au cours de 6,51 hospitalisations. Chaque hospitalisation dure 4,8 jours. Les hommes sont significativement plus âgés (+ 1,4 an). Ils sont également hospitalisés plus

longtemps (+ 2,5 jours dans l'année) sans doute parce que la durée moyenne de chaque hospitalisation est significativement plus longue (+ 0,67 jour). En revanche il n'y a pas de différence en ce qui concerne le nombre total d'hospitalisations.

La figure 3.3 montre la distribution de l'âge des patients par sexe. Plus de 16% des femmes ont entre 40 et 50 ans, contre moins de 10% des hommes. 27,5% de ceux-ci ont entre 60 et 70 ans, contre 22,3% des femmes.

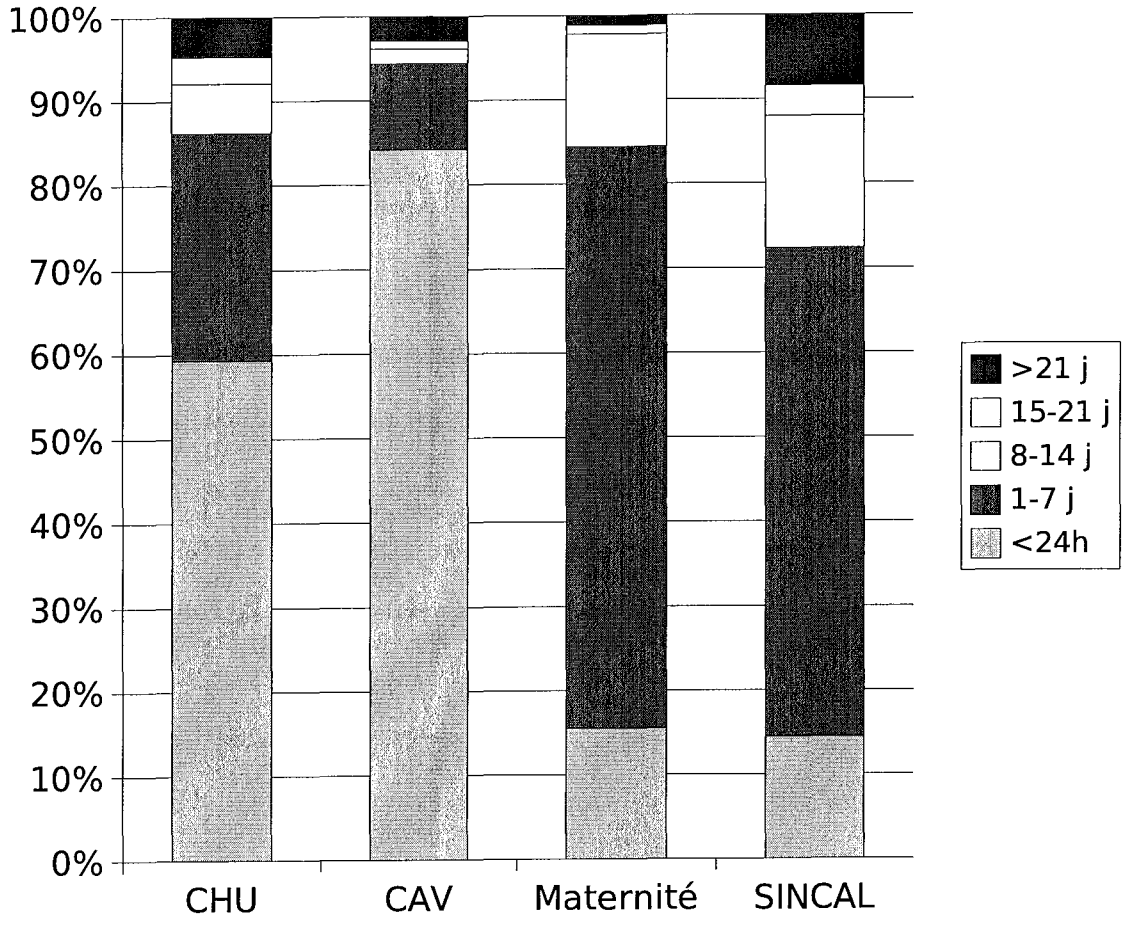


FIG. 3.1: Répartition des durées de séjour par établissement

Sexe	Hommes (n=4101)		Femmes (n=3931)		Total (n=8032)		p
	Moy	(E-T)	Moy	(E-T)	Moy	(E-T)	
Âge	58,18	(18,75)	56,77	(18,60)	57,49	(18,69)	0,0007
Durée de séjour totale	17,88	(24,79)	15,39	(21,74)	16,66	(23,38)	<,0001
Durée de séjour moyenne	5,13	(9,42)	4,46	(9,82)	4,80	(9,62)	0,0018
Nombre total d'hospitalisations	6,54	(9,61)	6,48	(9,18)	6,51	(9,40)	*ns*

TAB. 3.1: Caractéristiques des hospitalisations par sexe

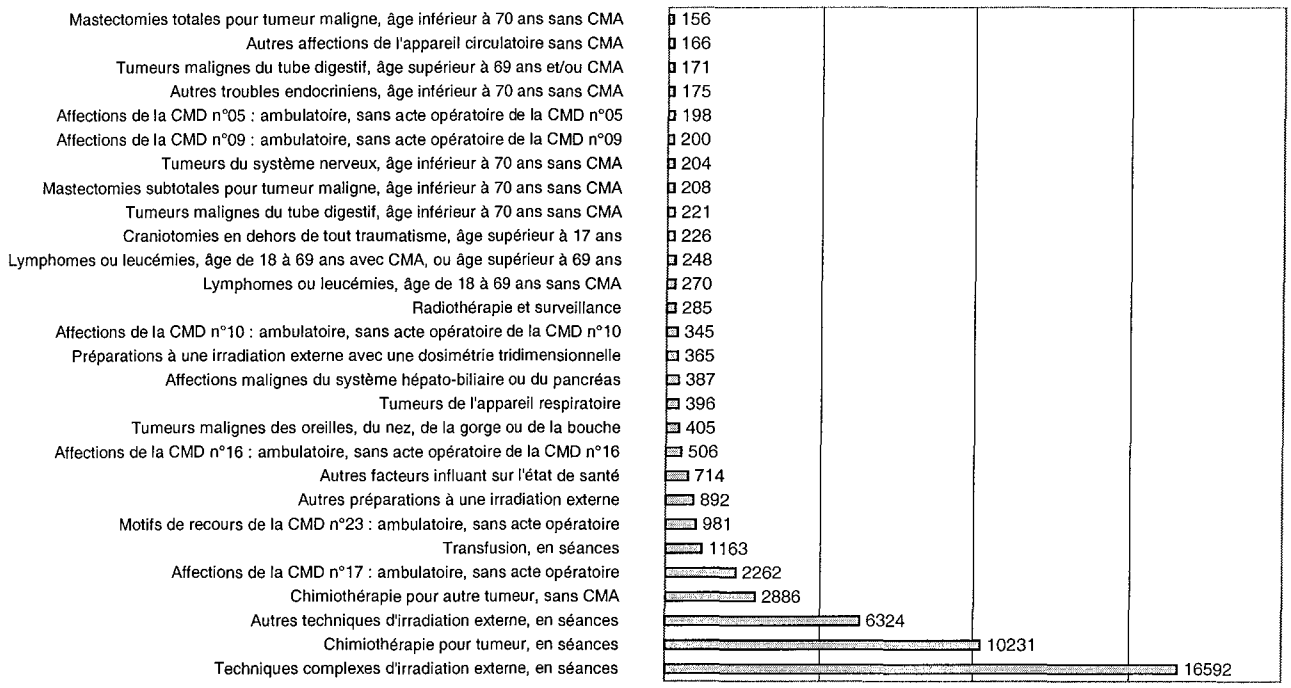


FIG. 3.2: Case-Mix

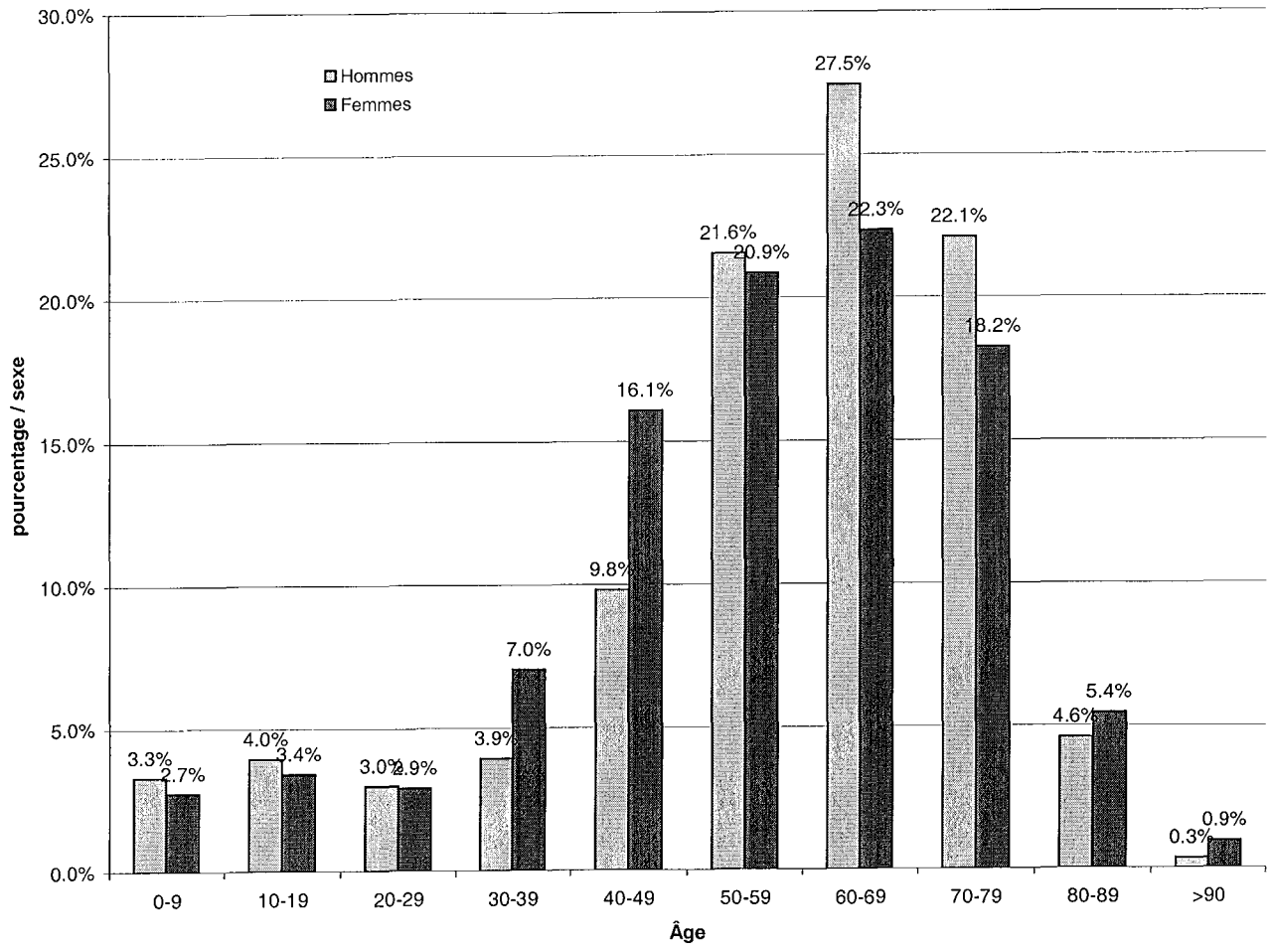


FIG. 3.3: Distribution de l'âge par sexe

3.1.2 Localisation des tumeurs

Localisations connues

7340 localisations tumorales (hors métastases) ont été retrouvées. Elles concernent 7003 patients dont 3455 hommes. Les deux sexes confondus, les localisations les plus fréquentes sont le sang et les organes hématopoïétiques (n=1014), le sein (n=1014), l'appareil digestif (hors système hépato-biliaire) (n=866), l'appareil pleuro-pulmonaire (n=624). (Tableau 3.2)

Chez l'homme, les localisations les plus fréquemment identifiées sont le sang et les organes hématopoïétiques (n=571), l'appareil digestif (n=499), l'appareil pleuro-pulmonaires (n=490) et les voies aéro-digestives supérieures (VADS)(n=469). Elles représentent plus de 55% des localisations.

Chez la femme, les localisations les plus fréquentes sont le sein (n=1001), le sang et des organes hématopoïétiques (n=471) et les organes génitaux (n=422). Elles représentent 51% des localisations. La présence d'une tumeur de l'appareil génital masculin correspond sans doute à une erreur de codage, soit du sexe, soit du diagnostic.

La figure 3.4 illustre la distribution des localisations par sexe.

Localisations inconnues

Chez 1029 patients, aucune localisation primitive en Diagnostic Principal ou Diagnostic Relié n'a été retrouvée. Cette situation se présente chez 9,8% des femmes et 15,8% des hommes ($p<0.0001$). Leur prise en charge a été analysée grâce aux diagnostics principaux de leurs différentes hospitalisations. Les motifs les plus fréquemment retrouvés sont les séances d'irradiations, les examens de dépistage, les tumeurs malignes secondaires et les séances de chimiothérapie. Le tableau 3.3 reprends la liste de ces motifs.

Localisations	Hommes n	Hommes %	Femmes n	Femmes %	Total) n	Total) %
Sang et des organes hématopoïétiques	571	15,6	471	12,8	1042	14,2
Sein	13	0,4	1001	27,2	1014	13,8
Appareil digestif (hors système hépato-biliaire)	499	13,6	367	10,0	866	11,8
Poumon	490	13,4	134	3,6	624	8,5
VADS	469	12,8	115	3,1	584	8,0
Glandes endocrines	139	3,8	349	9,5	488	6,6
Système nerveux central et oeil	249	6,8	213	5,8	462	6,3
Organes génitaux féminins (hors sein)	0	0,0	422	11,5	422	5,7
Autres	211	5,8	212	5,8	423	5,8
Cutanées	179	4,9	174	4,7	353	4,8
Organes génitaux masculins (y compris prostate)	340	9,3	1	0,0	341	4,6
Système urinaire (hors prostate)	230	6,3	86	2,3	316	4,3
Os, Tissu conjonctif et autres tissus mous	124	3,4	95	2,6	219	3,0
Système hépato-biliaire	145	4,0	41	1,1	186	2,5
Total	3659	100,0	3681	100,0	7340	100,0

TAB. 3.2: Nombre de patients distincts par localisation et par sexe

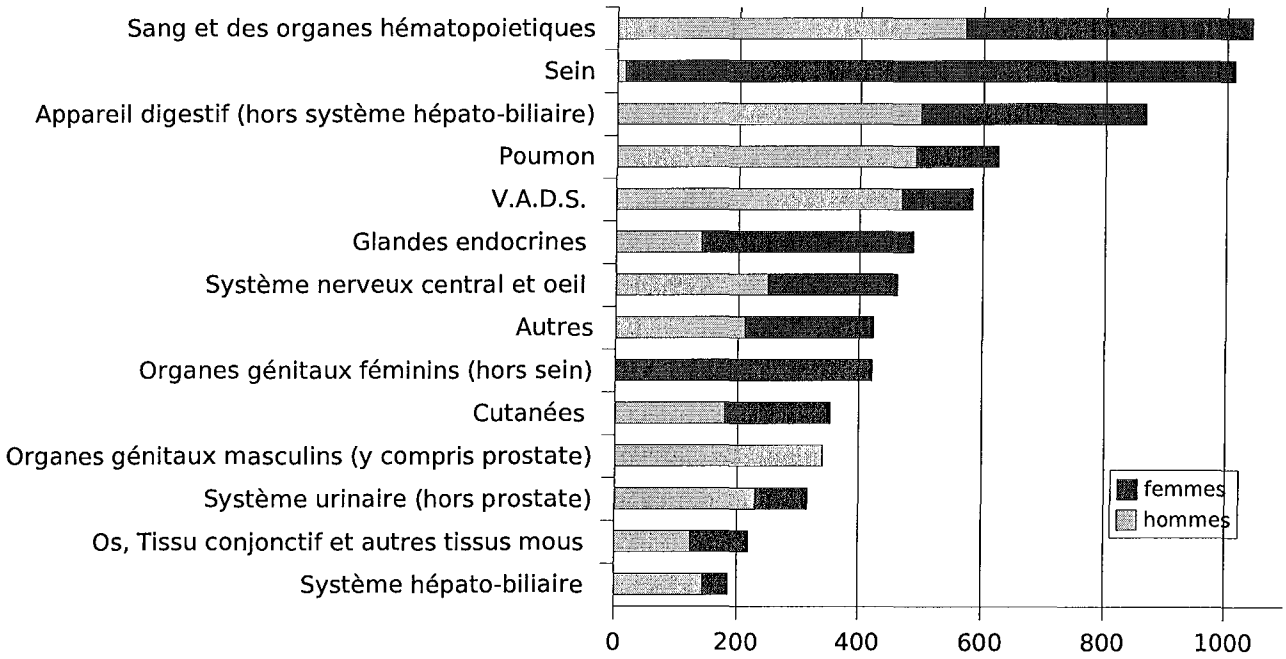


FIG. 3.4: Nombre de patients distincts par localisation et par sexe

Diagnostic Principal	n
Séance d'irradiation	213
Examen spécial de dépistage de tumeur de la prostate	177
Préparation a une radiothérapie	135
Examen spécial de dépistage de tumeur de l'intestin	117
Mise en observation pour suspicion de tumeur maligne	80
Séance de chimiothérapie pour tumeur	68
Tumeur maligne secondaire des os et de la moelle osseuse	43
Examen de contrôle après d'autres traitements pour tumeur maligne	42
Tumeur maligne secondaire du poumon	35
Tumeur maligne secondaire du cerveau et des méninges cérébrales	35
Tumeur maligne secondaire du foie	31
Soin de contrôle chirurgical, sans précision	20
Tumeur maligne secondaire et non précisée des ganglions lymphatiques (tête,face,cou)	18
Examen de contrôle après traitements combinés pour tumeur maligne	18
Tumeur maligne secondaire du rétro-péritoine et du péritoine	16
Examen de contrôle après traitement chirurgical d'une tumeur maligne	16
Antécédents personnels de tumeur maligne des organes digestifs	16
Tumeur maligne secondaire de la plèvre	14
Soins palliatifs	11
Hydrocéphalie au cours de maladies tumorales	11
Examen de contrôle après chimiothérapie pour tumeur maligne	10
Autres	194

TAB. 3.3: Nombre de patients sans localisation tumorale connue, pris en charge par Diagnostic Principal

3.2 Prise en charge par établissement

3.2.1 Prise en charge du cancer au CHU de Nancy

Âge, sexe, hospitalisations

3136 hommes et 2061 femmes ont été hospitalisés au CHU de Nancy pour leur cancer. L'âge moyen à la première hospitalisation est de 56,4 ans. On compte 12,9 jours passés au CHU dans l'année, au cours de 3,3 hospitalisations. La durée moyenne d'une hospitalisation est de 6,7 jours. Là encore, les hommes sont en moyenne plus âgés (+ 2,4 ans) que les femmes. Ils sont hospitalisées plus longtemps au cours de l'année (+ 1,3 jour). En revanche, on ne remarque pas de différence significative pour le sexe en ce qui concerne la durée moyenne de séjour ni le nombre d'hospitalisations (Tableau 3.4).

Localisation des tumeurs

4610 localisations tumorales (hors métastases) ont été retrouvées. Elles concernent 4416 patients, dont 2597 hommes. Les deux sexes confondus, les localisations les plus fréquentes sont le sang et les organes hématopoïétiques (n=1013), l'appareil digestif (hors système hépato-biliaire) (n=594) et les glandes endocrines (n=481). (Tableau 3.5 et figure 3.5)

Chez l'homme, les plus fréquemment identifiées sont les tumeurs du sang et des organes hématopoïétiques (n=571), de l'appareil pleuro-pulmonaire (n=381), de l'appareil digestif (n=358) et des voies aéro-digestives supérieures (n=254).

Chez la femme, on retrouve tout d'abord les tumeurs malignes du sang et des organes hématopoïétiques (n=458), des glandes endocrines (n=347) et de l'appareil digestif (hors système hépato-biliaire) (n=236).

Sexe	Hommes (n=3136) Moy	(E-T)	Femmes (n=2061) Moy	(E-T)	Total (n=5197) Moy	(E-T)	p
Âge	57,30	(19,86)	54,95	(21,92)	56,36	(20,73)	<,0001
Durée de séjour totale	13,42	(21,17)	12,10	(21,18)	12,90	(21,18)	0,0282
Durée de séjour moyenne	6,77	(11,07)	6,48	(14,09)	6,65	(12,36)	*ns*
Nombre total d'hospitalisations	3,30	(4,94)	3,37	(5,34)	3,33	(5,10)	*ns*

TAB. 3.4: Caractéristiques des hospitalisations au CHU de Nancy, par sexe

	Hommes		Femmes		Total	
Localisations	n	%	n	%	n	%
Sang et des organes hématopoïétiques	555	20,4	458	24,3	1013	22,0
Appareil digestif (hors système hépato-biliaire)	358	13,1	236	12,5	594	12,9
Glandes endocrines	134	4,9	347	18,4	481	10,4
Poumon	381	13,0	94	4,0	475	10,3
Système nerveux central et oeil	229	8,4	187	9,9	416	9,0
Autres	193	7,1	190	10,1	383	8,3
V.A.D.S.	254	9,3	63	3,4	317	6,9
Système urinaire (hors prostate)	177	6,5	62	3,3	239	5,2
Cutanées	107	3,9	116	6,2	223	4,8
Système hépato-biliaire	143	5,2	40	2,1	183	4,0
Organes génitaux masculins (y compris prostate)	137	5,0	1	0,1	138	3,0
Os, Tissu conjonctif et autres tissus mous	58	2,1	48	2,6	106	2,3
Organes génitaux féminins (hors sein)	0	0,0	30	1,6	30	0,7
Sein	1	0,0	11	0,6	12	0,3
Total	2727	100,0	1883	100,0	4610	100,0

TAB. 3.5: Nombre de patients distincts pris en charge par localisation et par sexe au CHU de Nancy

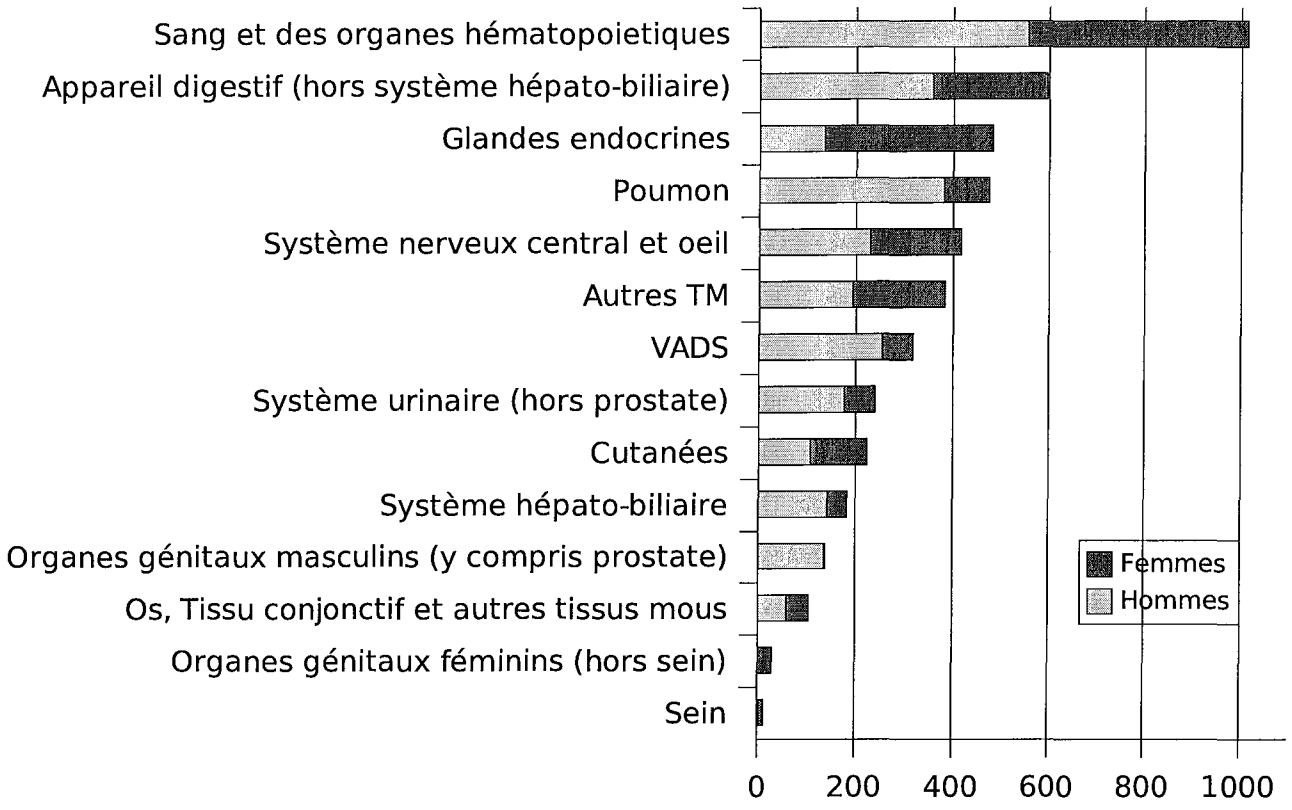


FIG. 3.5: Distribution des localisations au CHU

3.2.2 Prise en charge du cancer au Centre Alexis Vautrin

Âge, sexe, hospitalisations

1348 hommes et 2030 femmes ont été hospitalisés au Centre Alexis Vautrin pour leur cancer. L'âge moyen à la première hospitalisation est de 59,2 ans. On compte 19,5 jours passés au CAV dans l'année, au cours de 10,3 hospitalisations. La durée moyenne d'une hospitalisation est de 4,0 jours. Là encore, les hommes sont en moyenne plus âgés (+ 1,3 ans) que les femmes. Ils sont hospitalisés plus longtemps au cours de l'année (+ 5,6 jour), et plus souvent (+ 3,1 hospitalisations). En revanche, on ne remarque pas de différence significative les sexes concernant la durée moyenne de séjour ni le nombre d'hospitalisations (Tableau 3.6).

Localisation des tumeurs

3057 localisations tumorales (hors métastases) ont été retrouvées. Elles concernent 3022 patients, dont 1176 hommes. Les deux sexes confondus, les localisations les plus fréquentes sont le sein (n=999), les voies aéro-digestives supérieures (n=361), les organes génitaux féminins (n=356), l'appareil digestif (hors système hépato-biliaire) (n=334) (Tableau 3.7 et figure 3.6).

Chez l'homme, les plus fréquemment identifiées sont les voies aéro-digestives supérieures (n=295), les organes génitaux (y compris prostate) (n=221), l'appareil digestif (hors système hépato-biliaire) (n=184) et l'appareil pleuro-pulmonaire (n=162).

Chez la femme, on retrouve tout d'abord le sein (n=987), les organes génitaux (n=356), l'appareil digestif (hors système hépato-biliaire)(n=150) puis les voies aéro-digestives supérieures (n=66).

Sexe	Hommes (n=1348) Moy	(E-T)	Femmes (n=2030) Moy	(E-T)	Total (n=3378) Moy	(E-T)	p
Âge	59,97	(15,04)	58,65	(14,20)	59,18	(14,55)	0,0108
Durée de séjour totale	22,93	(23,90)	17,18	(19,30)	19,48	(21,44)	<,0001
Durée de séjour moyenne	4,20	(8,59)	3,86	(5,83)	3,99	(7,06)	*ns*
Nombre total d'hospitalisations	12,16	(12,00)	9,07	(10,32)	10,31	(11,12)	<,0001

TAB. 3.6: Caractéristiques des hospitalisations au Centre Alexis Vautrin, par sexe

	Hommes		Femmes		Total	
Localisations	n	%	n	%	n	%
Sein	12	1,0	987	53,0	999	32,7
Organes génitaux féminins (hors sein)	0	0,0	356	19,1	356	11,7
V.A.D.S.	295	24,7	66	3,5	361	11,8
Appareil digestif (hors système hépato-biliaire)	184	15,4	150	8,1	334	10,9
Poumon	162	13,6	59	3,2	221	7,2
Organes génitaux masculins (y compris prostate)	221	18,5	0	0,0	221	7,2
Cutanées	73	6,1	58	3,1	131	4,3
Os, Tissu conjonctif et autres tissus mous	66	5,5	46	2,5	112	3,7
Sang et des organes hématopoïétiques	56	4,7	54	2,9	110	3,6
Système urinaire (hors prostate)	65	5,4	28	1,5	93	3,0
Système nerveux central et oeil	40	3,4	37	2,0	77	2,5
Autres	10	0,8	14	0,8	24	0,8
Glandes endocrines	9	0,8	6	0,3	15	0,5
Système hépato-biliaire	2	0,2	1	0,1	3	0,1
Total	1195	100,0	1862	100,0	3057	100,0

TAB. 3.7: Nombre de patients distincts pris en charge par localisation et par sexe au Centre Alexis Vautrin

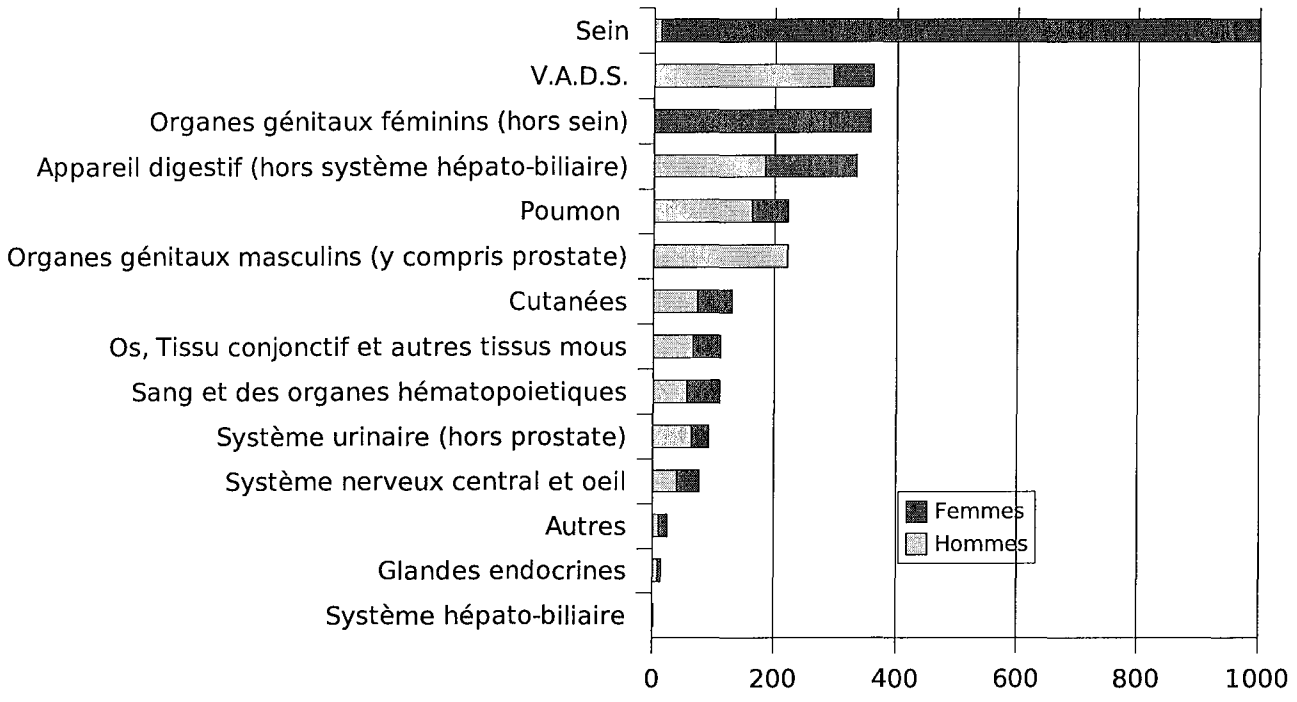


FIG. 3.6: Distribution des localisations au Centre Alexis Vautrin

3.2.3 Prise en charge du cancer au SINCAL

Âge, sexe, hospitalisations

39 hommes et 34 femmes ont été hospitalisés au SINCAL pour leur cancer. L'âge moyen à la première hospitalisation est de 54,3 ans. On compte 8,1 jours passés au SINCAL dans l'année, au cours de 1,1 hospitalisation. La durée moyenne d'une hospitalisation est de 6,9 jours. Aucune différence significative n'est attribuée au sexe pour ces variables (Tableau 3.8).

Localisation des tumeurs

46 localisations tumorales (hors métastases) ont été retrouvées. Elles concernent 46 patients, dont 24 hommes. Les deux sexes confondus, on retrouve par ordre de fréquence les os, le tissu conjonctifs et autres tissus mous (n=18), les localisations autres (n=17), la peau (n=7), le sang et des organes hématopoïétiques (n=3) et enfin le sein (n=1)(Tableau 3.9).

Sexe	Hommes (n=39) Moy	(E-T)	Femmes (n=34) Moy	(E-T)	Total (n=73) Moy	(E-T)	p
Âge	56,36	(16,08)	51,97	(17,87)	54,32	(16,96)	*ns*
Durée de séjour totale	7,77	(8,55)	8,47	(10,88)	8,10	(9,64)	*ns*
Durée de séjour moyenne	6,54	(7,08)	7,37	(9,01)	6,93	(7,99)	*ns*
Nombre total d'hospitalisations	1,15	(0,37)	1,12	(0,41)	1,14	(0,38)	*ns*

TAB. 3.8: Caractéristiques des hospitalisations au SINCAL, par sexe

	Hommes		Femmes		Total	
Localisations	n	%	n	%	n	%
Os, Tissu conjonctif et autres tissus mous	9	37,5	9	40,9	18	39,1
Autres	8	33,3	9	40,9	17	37,0
Cutanées	6	25,0	1	4,6	7	15,2
Sang et des organes hématopoïétiques	1	4,2	2	9,1	3	6,5
Sein	0	0,0	1	4,6	1	2,2
Total	24	100,0	22	100,0	46	100,0

TAB. 3.9: Nombre de patients distincts pris en charge par localisation et par sexe au SINCAL

3.2.4 Prise en charge du cancer à la Maternité Régionale de Nancy

Âge, sexe, hospitalisations

77 patientes ont été hospitalisées à la Maternité Régionale pour un cancer. L'âge moyen à la première hospitalisation est de 51,1 ans. On compte 5,1 jours passés à la Maternité Régionale dans l'année, au cours de 1,2 hospitalisation. La durée moyenne d'une hospitalisation est de 4,2 jours (Tableau 3.8). Notons la présence d'un patient de sexe masculin hospitalisé pour une tumeur des glandes endocrines.

Localisation des tumeurs

On dénombre 76 localisations primitives chez 76 femmes. 65 patientes ont été hospitalisées pour une tumeur des organes génitaux (hors sein), et 9 pour un cancer du sein. On retrouve également une localisation digestive et une localisation des os, du tissu conjonctif ou autres tissus mous (3.11).

(n=77)	Moy	(E-T)
Âge	51,09	(18,45)
Durée de séjour totale	5,09	(4,97)
Durée de séjour moyenne	4,22	(3,90)
Nombre total d'hospitalisations	1,16	(0,43)

TAB. 3.10: Caractéristiques des hospitalisations à la Maternité Régionale

Localisations	n	%
Organes génitaux féminins (hors sein)	65	85,5
Sein	9	11,8
Appareil digestif (hors système hépato-biliaire)	1	1,3
Os, Tissu conjonctif et autres tissus mous	1	1,3
Total	76	100,0

TAB. 3.11: Nombre de patientes distinctes prises en charge par localisation à la Maternité Régionale

3.3 Prise en charge partagée

3.3.1 Flux inter-établissements

Nous avons construit le treillis de Galois de la relation qui lie les patients aux établissements dans lesquels ils ont séjourné pour le traitement d'un cancer. Cela permet de représenter l'ensemble des combinaisons possibles de fréquentation. Pour n établissements, ce nombre de combinaisons vaut 2^n. Ce qui donne pour $n = 4$, 16 possibilités.

Le treillis de la figure 3.7 illustre la fréquentation des quatre établissements. Il s'agit d'un treillis d'héritage selon les attributs (ici, les établissements). Sur l'étiquette de chaque concept figurent :

- les établissements de son intension, sans redondance ;
- le nombre de patients qui ont fréquenté ces établissements, en valeur absolue et en pourcentage de la population totale.

En lisant le treillis du haut vers le bas, on retrouve d'abord les 8032 patients sélectionnés. Les nœuds du premier rang montrent la fréquentation de chaque établissement (de gauche à droite) :

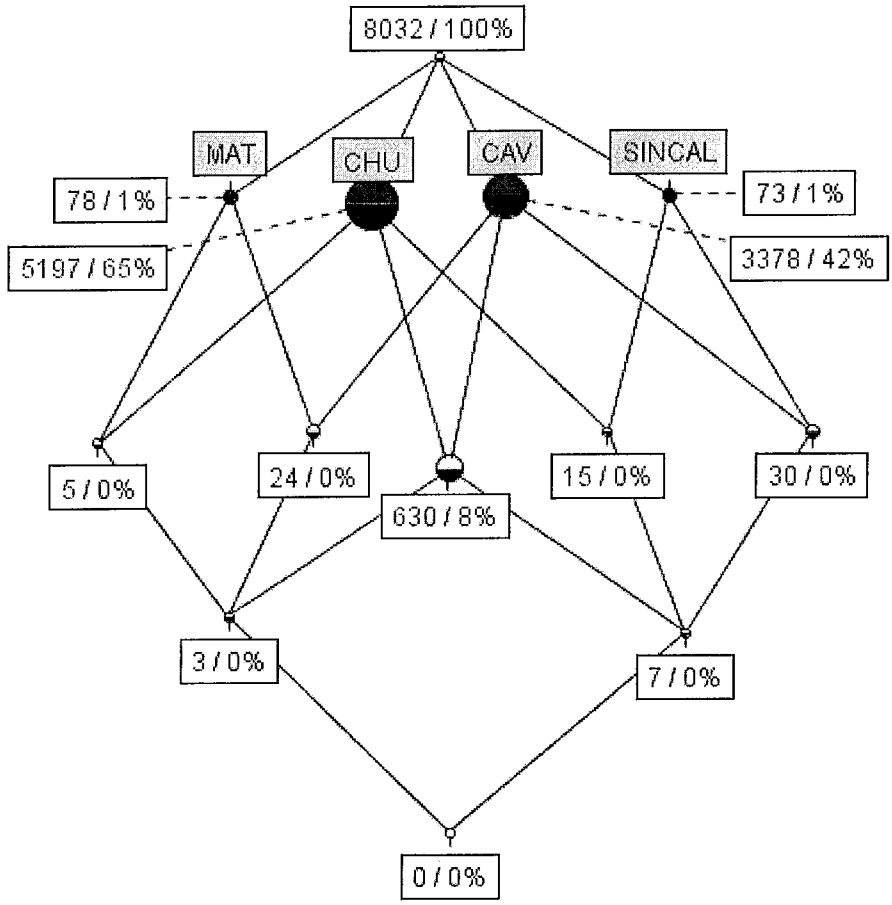


FIG. 3.7: Partage de la prise en charge des cancers

- 78 patients pour la Maternité,
- 5197 patients au CHU,
- 3378 patients au Centre Alexis Vautrin,
- 73 patients au SINCAL.

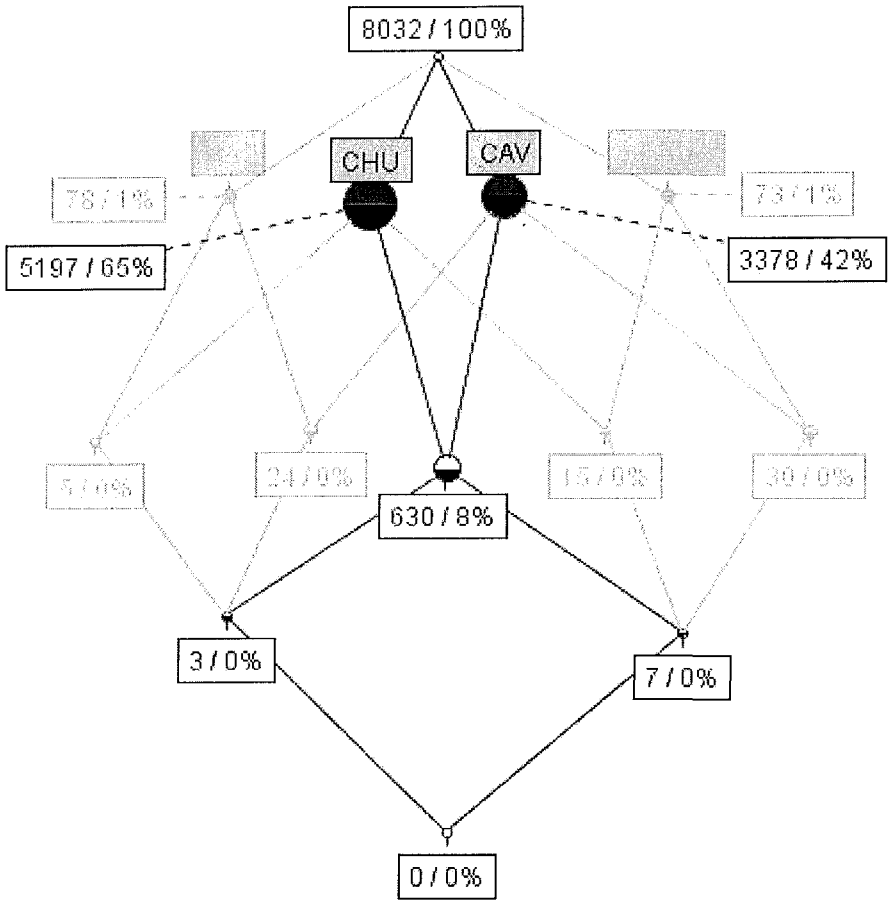


FIG. 3.8: Patients partagés par le CHU et le CAV

Le descendant commun des nœuds représentant le CHU et le CAV montre que 630 patients ont fréquenté ces deux établissements (Figure 3.8). De même, le CHU partage la prise en charge de 5 patients avec la maternité, et 15 avec le SINCAL. Le CAV partage 24 patients avec la

maternité et 30 avec le SINCAL. 10 patients ont fréquenté trois établissements :

- 3 la maternité, le CHU et le CAV,
- 7 le CHU, le CAV et le SINCAL.

Au bas du treillis, aucun patient n'a fréquenté les quatre établissements.

Une autre façon de visualiser le treillis est de ne faire figurer dans un concept que les objets qu'il ne lègue pas à ses descendants (héritage selon les objet). Chaque patient apparaît alors dans un unique concept. Sur la figure 3.9, on peut voir que pour 4557 patients, la prise en charge s'est déroulée uniquement au CHU. De même, 2704 patients n'ont été hospitalisés qu'au CAV.

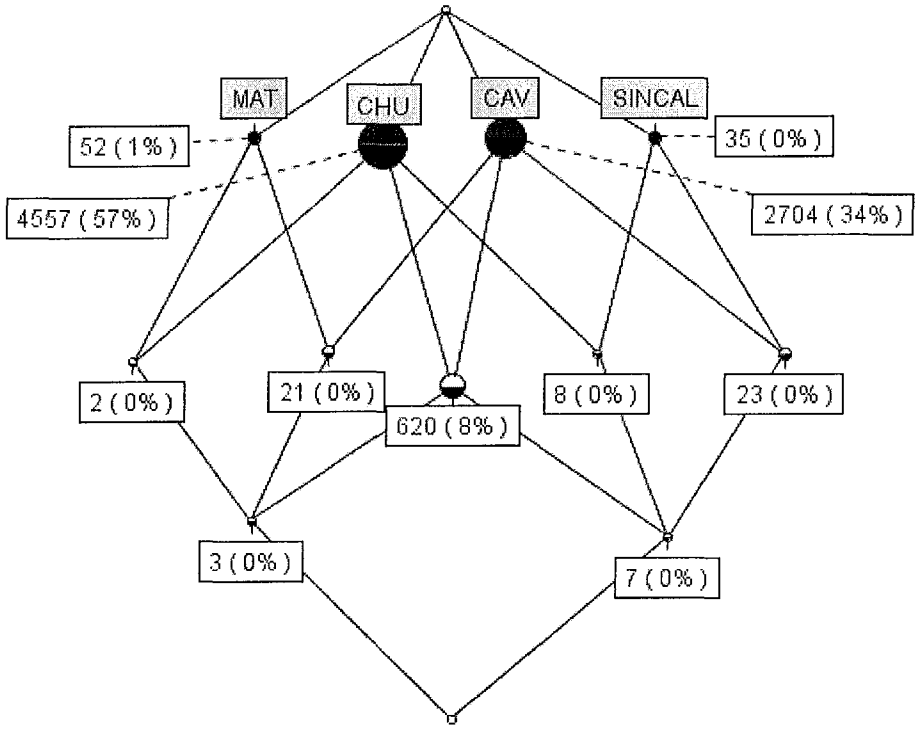


FIG. 3.9: Partage de la prise en charge des cancers : treillis d'héritage

3.3.2 Partage entre le CHU et le Centre Alexis Vautrin

Compte tenu des effectifs, nous nous focaliserons dans cette partie sur l'analyse de la prise en charge des patients communs au CHU et au Centre Alexis Vautrin.

3.3.3 Population

On compte 402 hommes et 228 femmes chez les patients communs au CHU et au CAV. La moyenne d'âge est de 56,1 ans. Les hommes sont plus âgés que les femmes (+ 2,3 ans).

3.3.4 Hospitalisations

Les patients de cette sous-population totalisent 11872 hospitalisations, dont 76% au CAV. Le tableau 3.12 retrace les caractéristiques de ces hospitalisations. D'une manière générale, les hospitalisations au CAV sont plus 3 fois plus fréquentes qu'au CHU, mais 3 fois plus courtes. Les mêmes constatations peuvent être faites quelque soit le sexe.

Il ressort de l'analyse du case-mix de cette sous-population (Figure 3.10) que 63% des hospitalisations concernent des GHM de radiothérapie.

3.3.5 Localisation des tumeurs

Le tableau 3.13 et la figure 3.11 montrent les files actives des patients communs par localisation.

Par ordre de fréquence décroissante, 140 d'entre eux sont traités pour une localisation pleuro-pulmonaire, 120 pour une tumeur maligne des voies aéro-digestives supérieures, 93 pour une tumeur hématologique, et 88 pour une localisation digestive (hors système hépato-biliaire).

Établissement	CHU		CAV		
	Moy	$(E-T)$	Moy	$(E-T)$	p
Hommes					
Durée totale de séjour	20,2	(27,1)	24,0	(22,3)	0,0358
Durée moyenne de séjour du patient	8,2	(11,9)	2,7	(4,5)	$< 0,0001$
Nombre moyen d'hospitalisations du patient	4,2	(5,5)	15,1	(11,5)	$< 0,0001$
femmes					
Durée totale de séjour	19,2	(20,2)	23,1	(20,1)	0,0508
Durée moyenne de séjour du patient	8,4	(12,5)	2,9	(4,2)	$< 0,0001$
Nombre moyen d'hospitalisations du patient	5,3	(7,7)	12,8	(11,6)	$< 0,0001$
Hommes et femmes					
Durée totale de séjour	19,8	(24,8)	23,7	(21,5)	0,0048
Durée moyenne de séjour du patient	8,3	(12,1)	2,8	(4,4)	$< 0,0001$
Nombre moyen d'hospitalisations du patient	4,6	(6,4)	14,3	(11,6)	$< 0,0001$

TAB. 3.12: Comparaison des hospitalisations des patients pris en charge par le CHU et le CAV

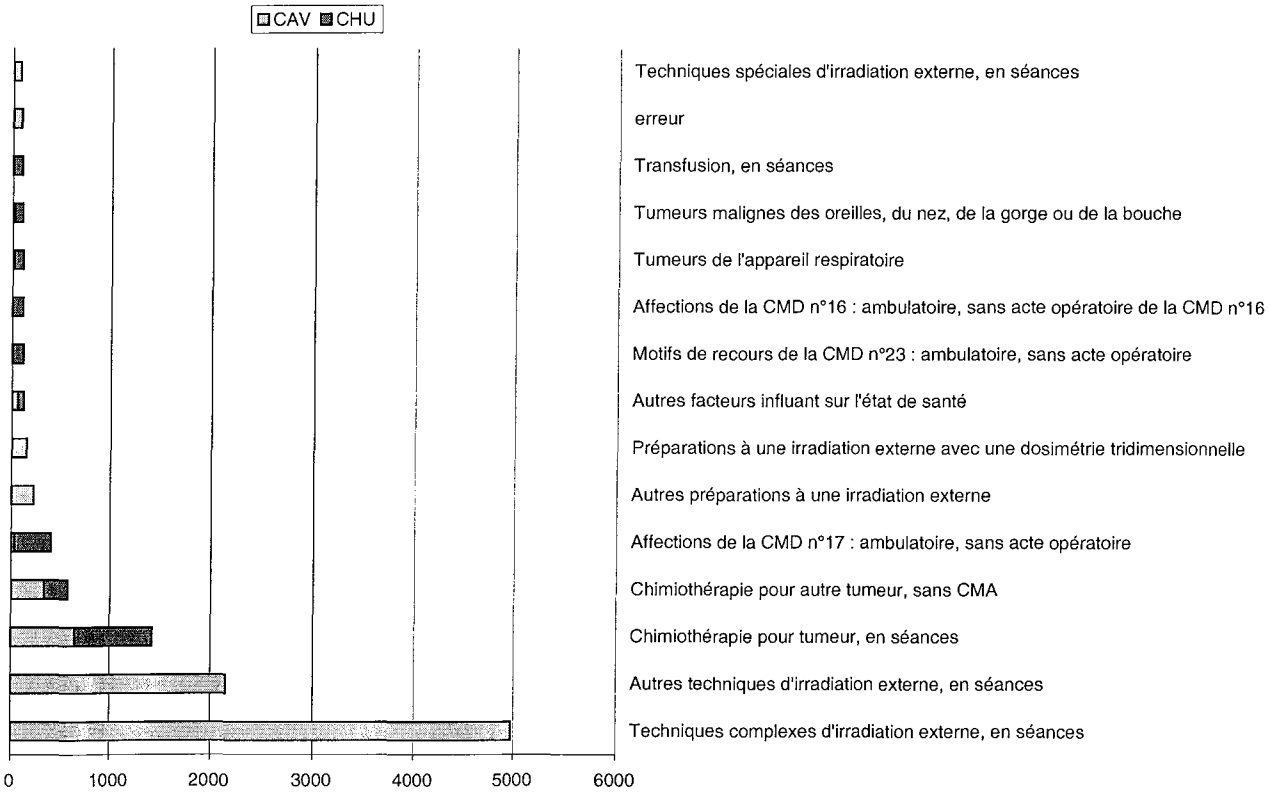


FIG. 3.10: Case-mix des patients communs au CHU et CAV

Localisations	Hommes n	Hommes %	Femmes n	Femmes %	Total n	Total %
Poumon	101	23,5	38	14,3	139	20,0
V.A.D.S.	97	22,6	23	8,7	120	17,3
Sang et des organes hématopoïétiques	46	10,7	47	17,7	93	13,4
Appareil digestif (hors système hépato-biliaire)	55	12,8	33	12,5	88	12,7
Système nerveux central et oeil	25	5,8	22	8,3	47	6,8
Os, Tissu conjonctif et autres tissus mous	19	4,4	22	8,3	41	5,9
Organes génitaux masculins (y compris prostate)	34	7,9	0	0,0	34	4,9
Système urinaire (hors prostate)	20	4,7	9	3,4	29	4,2
Autres	9	2,1	16	6,0	25	3,6
Cutanées	14	3,3	9	3,4	23	3,3
Organes génitaux féminins (hors sein)	0	0,0	22	8,3	22	3,2
Glandes endocrines	9	2,1	7	2,6	16	2,3
Sein	0	0,0	15	5,7	15	2,2
Système hépato-biliaire	0	0,0	2	0,8	2	0,3
Total	429	100,0	265	100,0	694	100,0

TAB. 3.13: Nombre de patients distincts pris en charge par localisation et par sexe au CHU et au CAV

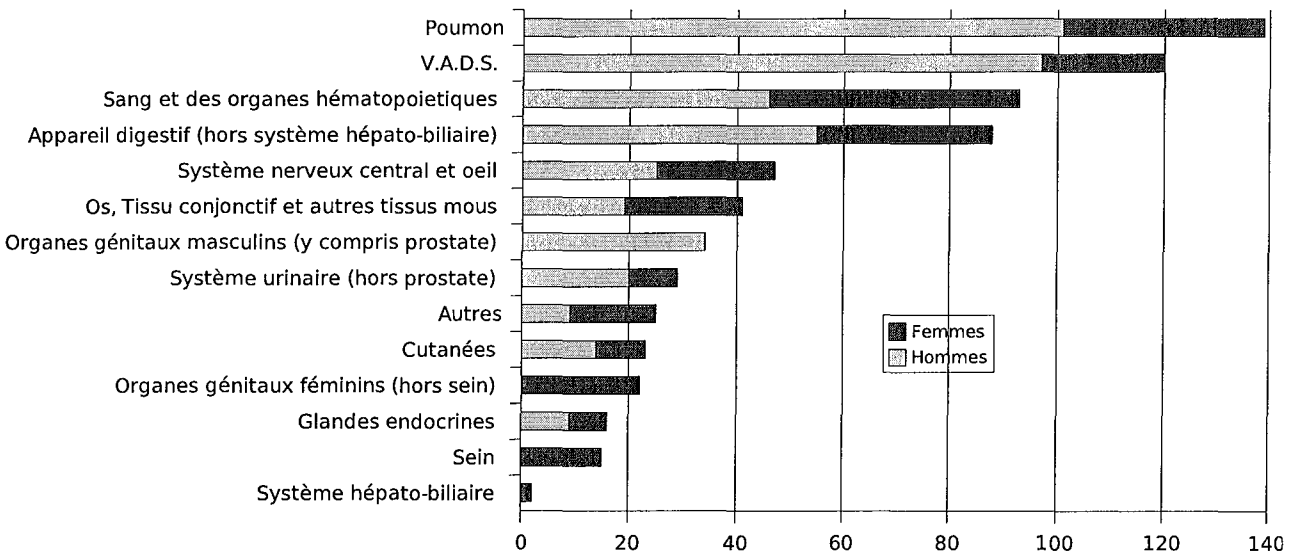
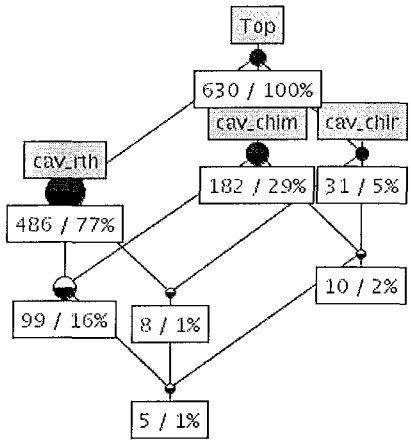
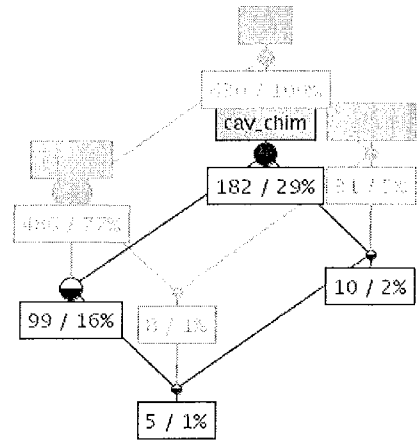


FIG. 3.11: Localisation des cancers des patients traités au CHU et CAV

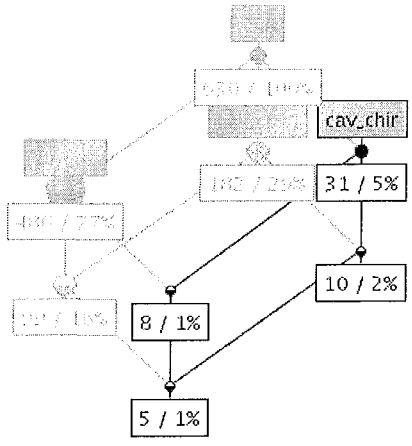
3.3.6 Nature de la coopération CHU-CAV

Comme le montre l'analyse du case-mix des patients communs, un grand nombre d'entre eux ont bénéficié d'une radiothérapie. Or, le CHU ne dispose pas d'équipements de radiothérapie. Pour étudier la véritable nature de la coopération entre ces deux établissements, nous avons détaillé la prise en charge de ces patients par le CAV. Le treillis de la figure 3.12 illustre cette prise en charge. 486 des patients communs ont bénéficié au CAV d'une radiothérapie. 182 y ont reçu une chimiothérapie, parmi lesquels 99 ont également eu des cures de radiothérapie. On compte 31 patients ayant bénéficié d'une intervention chirurgicale au CAV, dont 8 ont aussi reçu de la radiothérapie. La coopération entre les deux établissements s'est limitée à la radiothérapie pour 384 (65%) patients. À l'inverse, elle n'inclut pas de radiothérapie pour 144 (24%) d'entre eux.

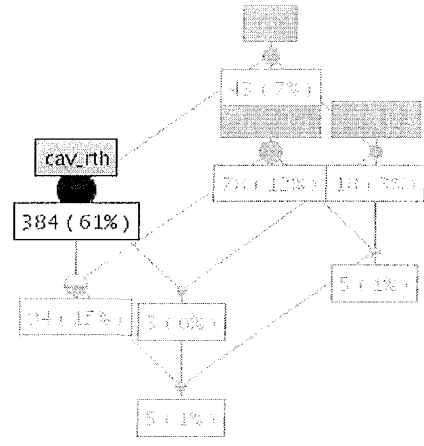




182 patients ont reçu de la chimiothérapie au CAV



31 patients y ont subi une intervention chirurgicale



384 patients y ont séjourné exclusivement pour de la radiothérapie

FIG. 3.12: Prise en charge par le CAV des patients communs au CHU et au CAV

3.4 Mise en œuvre des motifs séquentiels fréquents

3.4.1 Aspects pratiques

Nous avons étudié le potentiel de la recherche de motifs séquentiels fréquents pour caractériser la trajectoire de soins. Pour simplifier l'approche, nous avons travaillé sur la sous-population des patients ayant présenté une localisation primitive pleuro-pulmonaire.

Construction des trajectoires individuelles

Pour un patient, les diagnostics principaux de tous ses séjours en unité médicale sont isolés, de même que la date et le nom de l'établissement dans lequel ils ont été posés. En les classant par ordre chronologique, nous reconstituons ainsi sa trajectoire de soins.

Les diagnostics sont recodés de manière différente selon leur nature :

- `chim` pour le code `Z511=Chimiothérapie pour tumeur`,
- `rth` pour les code `Z5100` et `Z5101` de radiothérapie,
- réduit à leur trois premiers caractères pour les autres.

Ce choix permet de garder la trace des cures de chimiothérapies et radiothérapie. et il conserve une information suffisamment précise sur la nature des séjours ne comportant pas les diagnostics principaux précédents. Enfin, lorsqu'une hospitalisation est classée dans un GHM chirurgical, elle est repérée et marquée du code `chir_N°GHM`

Exemple 3.1

Un patient hospitalisé au CHU pour une intervention chirurgicale ciblée sur sa tumeur, ayant bénéficié d'une cure de chimiothérapie au CHU et enfin de deux séances de radiothérapie au CAV aura la trajectoire suivante : ⟨`(chu_chir)(chu_chim)(cav_rth)(cav_rth)`⟩

L'étape de fouille de données

L'analyse porte sur 624 patients porteurs d'une tumeur maligne primitive de localisation pleuro-pulmonaire. La base des trajectoires de ces patients est fouillée avec un seuil de 10% (soit 62 patients).

3.4.2 Résultats

L'algorithme retrouve 38 motifs séquentiels fréquents contenant jusqu'à 18 items (Tableau 3.14). Les sept motifs de taille 1 figurent dans le tableau 3.15. 334 patients ont été hospitalisés au moins une fois CHU avec un diagnostic principal de Tumeur maligne des bronches et du poumon. 194 (31%) on bénéficié d'une radiothérapie au CAV. 163 (26%) ont reçu une chimiothérapie au CHU et 102 (16,3%) au CAV. Le motif ⟨(chu_Z08)⟩, supporté par 137 patients, correspond au diagnostic `Examen de contrôle après traitement d'une tumeur maligne`, et ⟨(chu_Z45)⟩, supporté par 63 (10%) patients à `Ajustement et entretien d'un dispositif interne`. Enfin, 110 patients supportent le motif ⟨(chu_Chir113)⟩ qui indique la survenue d'une hospitalisation au CHU classée dans le GHM `Interventions majeures sur le thorax`. Notons qu'un patient peut supporter plusieurs motifs.

Une propriété fondamentale des motifs séquentiels stipule que « *Si une séquence n'est pas fréquente, alors toutes ses super-séquences ne sont pas fréquentes* ». Par conséquent, tous les motifs extraits dans cette étape de fouille sont des super-motifs des motifs présents dans le tableau 3.15.

Citons ici à titre d'exemple quelques uns des motifs fréquents de taille 2 ou plus :

- 64 (10%) patients comptent jusqu'à 18 hospitalisations successives au CAV pour de la radiothérapie ;
- 65 (10%) ont bénéficié de 5 hospitalisations au moins pour chimiothérapie au CAV. On retrouve le même motif pour 76 patients au CHU ;
- les motifs indiquant une alternance de chimiothérapie/radiothérapie, ou chimiothérapie

au CHU / chimiothérapie au CAV n'apparaissent pas. On peut en déduire que moins de 10% des patients sont concernés par ces situations.

Taille	n
1	7
2	7
3	5
4	3
5	3
>6	13
Total	38

TAB. 3.14: Nombre de motifs séquentiels fréquents extraits en fonction de leur taille

Motif	Nombre de patients	Support(%)
⟨(chu_C34)⟩	334	53,5
⟨(cav_rth)⟩	194	31,0
⟨(chu_chim)⟩	163	26,1
⟨(chu_Z08)⟩	137	21,9
⟨(chu_Chir113)⟩	110	17,6
⟨(cav_chim)⟩	102	16,3
⟨(chu_Z45)⟩	63	10,0

TAB. 3.15: Les motifs séquentiels fréquents de taille 1

4 Discussion

4.1 Utilisation du PMSI dans l'étude de la trajectoire de soins

Le PMSI constitue une source considérable d'informations. Celles-ci sont d'ordre médical avec des diagnostics et des procédures médico-chirurgicales, mais aussi démographique (âge, sexe, code postal) ou médico-économique avec la valorisation des hospitalisations classées en GHM. C'est le seul système d'information médicalisé qui collecte des données à l'échelle nationale, en routine et de manière standardisée, aussi bien dans les secteur privé que public. Il est de ce fait un dénominateur commun pour la description de l'activité médicale de court-séjour.

Bien qu'il ait été conçu au départ avec le séjour pour unité d'analyse, l'introduction du chaînage anonymisé permet aujourd'hui de reconstituer la trajectoire de soins des patients hospitalisés en secteur MCO. Notre travail montre que grâce au chaînage et à l'étude de la trajectoire de soins, le PMSI permet de mieux décrire l'activité hospitalière autour des pathologies chroniques qu'il ne le faisait auparavant.

C'est le cas pour la cancérologie. L'analyse par séjour et la faible diversité des GHM de cancérologie rend difficilement compte de la diversité des cas de cancer, de leur localisation et du nombre de patients concernés. Notre étude a permis de constituer des files actives de patients par appareil anatomique, pris en charge dans les établissements de soins publics de Nancy. Nous avons décrit une partie des mécanismes de coopération entre ces établissements. Les résultats

montrent qu'une petite partie seulement des patients (10%) bénéficie d'une prise en charge partagée par plusieurs établissements. Cette prise en charge partagée s'effectue généralement entre le CHU et Le Centre Alexis Vautrin. La majeure partie de la coopération entre ces deux établissements s'articule autour de la radiothérapie, le CHU ne disposant pas de tels équipements. Néanmoins, dans 35% des cas, elle ne se limite pas qu'aux séances d'irradiation. En ce qui concerne les deux autres établissements de l'étude, la Maternité Régionale et le SINCAL, leur participation reste globalement marginale dans l'activité de cancérologie. Mais, lorsqu'elle existe, la prise en charge est partagée avec le CAV ou le CHU dans près de la moitié des cas.

Le PMSI présente des limites plus ou moins fortes en fonction du but recherché par l'analyse de la trajectoire de soins. Dans le cas de la cancérologie, il ne permet pas de développer un suivi par tumeur car les codes de diagnostics CIM10 ne comportent pas d'information sur l'histologie des lésions. En outre, leur précision topographique est parfois limitée et son codage par les praticiens peut s'avérer relativement instable au fil des hospitalisations. Le PMSI trouve une autre limite dans le champ des activités médicales qu'il recouvre. Il ne s'applique qu'aux hospitalisations dans le secteur MCO ou SSR, et ne traite ni de l'activité de consultation, ni de la médecine de ville. Or, pour les praticiens du CHU, les consultations représentent une part importante de leur activité de cancérologie.

4.2 Apports de l'extraction de connaissances dans les bases de données

La trajectoire de soins est un objet complexe. Nous avons mis en oeuvre des méthodes d'extraction des connaissances dans les bases de données pour l'analyser. Elles justifient ici leur intérêt dans le traitement des données volumineuses et complexes issues du PMSI. Nous montrons qu'elles peuvent constituer un point de départ dans la construction d'une typologie

des trajectoires de soins, domaine dans lequel les connaissances sont encore limitées.

Les treillis de Galois nous ont permis de caractériser les modes de coopération entre les quatre établissements nancéens du site hautement spécialisé. Ils représentent un outil de classification non supervisée à partir de données qualitatives. Malgré un aspect théorique qui peut paraître abrupt aux néophytes, nous pensons qu'ils peuvent être manipulés et interprétés par les experts du domaine médical. Ils présentent en effet l'avantage de supporter une interface de visualisation et de représentation graphique relativement intuitive. L'aide qu'ils nous ont apportée dans l'exploration automatique et interactive des données s'inscrit dans le processus d'ECBD.

Notre travail montre la faisabilité de la recherche de motifs séquentiels fréquents avec les données du PMSI. Il dévoile le potentiel de cet outil de fouille de données qui permet d'extraire des « profils » de trajectoires de soins. Ces profils tiennent compte de l'aspect *séquentiel* et *ordonné* dans le temps des épisodes de soins. Ce modèle repose sur le postulat que l'ordre d'apparition des traitements ou des phénomènes morbides chez un patient caractérise des trajectoires de soins différentes. C'est notamment le cas en cancérologie où, par exemple, la prise en charge d'un cancer du sein d'emblée métastatique n'est pas la même que celle d'une tumeur localisée, échappant secondairement au traitement.

4.3 Limites

4.3.1 Limites liées aux données

Une des principales limites de notre travail concerne la fenêtre d'observation des patients. Il existe un phénomène de censure de l'information à la fois dans le temps et dans l'espace. Notre visibilité des trajectoires est bornée par le premier et le dernier jour de l'année 2003. De nombreuses trajectoires peuvent nous apparaître tronquées. C'est particulièrement le cas pour des patients dont la prise en charge s'achève au début de l'année 2003, ou commence à la fin de l'année 2003. Ce biais de censure explique sans doute une part importante des 1000

patients chez lesquels nous n'avons pas retrouvé de localisation primitive. De plus, notre étude s'est limitée aux établissement de soins du secteur non lucratif de Nancy. Or, en France les établissements de soins privés occupent une large place dans le domaine de la cancérologie [17]. D'autre part, notre analyse ne tient compte ni de l'activité de consultation, ni du rôle de la médecine de ville dans la prise en charge du cancer. En outre, bien que nous n'ayons pas exploré le versant moyen séjour de l'activité de cancérologie, le PMSI offre la possibilité technique de chaîner l'information des secteurs SSR et MCO.

Ne disposant pas dans notre étude des actes médicaux présents dans le PMSI, nous avons mesuré l'activité chirurgicale par le biais du classement des hospitalisations par GHM. Ce choix résulte d'une première approche volontairement simplifiée de la trajectoire de soins.

4.3.2 Limites liées à la méthode

Choix du modèle de trajectoire

Bien que centré sur le patient, notre travail reflète encore une vision institutionnelle de la trajectoires de soins. Le suivi des patients se déroule au travers d'une fenêtre spatiale et temporelle dont les bornes sont fixées arbitrairement :

- le 1er janvier 2003 et le 31 décembre 2003 pour la période d'observation,
- l'espace de soins délimité par les quatre établissements de l'étude.

Au delà de considérations liées à la disponibilité des données, on peut se demander quel point de vue adopter sur la trajectoire de soins. Deux modèles s'opposent. Celui que nous avons utilisé adopte le point de vue d'un ou plusieurs prestataires de soins qui voient défiler une cohorte de patients sur une période d'observation fixée. Il relève l'utilisation de l'offre de soins par cette cohorte, indépendamment du début et de la fin d'un processus pathologique. On peut le qualifier d'« absolu » par opposition à un modèle « relatif » résolument centré sur le patient et sa maladie, dans lequel l'observation coïncide avec les limites décelables de la maladie. Tous deux présentent des avantages et des inconvénients mais ne partagent pas obligatoirement les

mêmes finalités. Le modèle absolu parait plus adapté à une mesure de la consommation de soins d'une population, ayant pour unité d'analyse le patient. Il offrirait, par exemple, une approche plus pragmatique que le modèle relatif dans un dispositif d'allocation des ressources basé sur le principe d'une prestation « patient-année », à l'instar des Adjusted Clinical Groups. Le modèle relatif peut quant à lui servir de base dans le domaine de l'évaluation des pratiques, où la trajectoire de soins observée pourrait être comparée à une trajectoire optimale.

Choix des outils

L'extraction des connaissances à partir des treillis de Galois et surtout leur visualisation est limitée par la taille des treillis, c'est-à-dire leur nombre de concepts. Cette taille tend à croître de façon exponentielle avec le nombre d'objets et d'attributs manipulés. À titre d'exemple, imaginons qu'une population de plusieurs milliers de patients fréquente une dizaine d'établissements. Le nombre de combinaisons possibles de prises en charge s'élève à $2^{10} = 1024$!. Malgré la faible probabilité que toutes soient réelles, la représentation de cette fréquentation par un treillis n'est pas visualisable en l'état. Il existe cependant des méthodes d'élagage des treillis permettant de synthétiser l'information qu'ils contiennent. Il peuvent en outre être associés à des algorithmes d'extraction de motifs fréquents pour ne faire figurer que les associations d'attributs les plus fréquentes.

L'extraction de motifs séquentiels fréquents présentent le même inconvénient. Une fouille de la base de séquences construite dans notre travail (cf. 3.4) avec un seuil de 2% retrouve 7910 motifs. Il est alors difficile de déterminer les motifs les plus intéressants sans une étape de post-traitement automatisée. Nous avons développé dans cette optique un outil de navigation qui facilite l'interprétation des résultats de la fouille. Le choix du seuil peut égaler poser problème dans la définition d'une trajectoire fréquente. À partir de quand un motif séquentiel peut-il être considéré comme un profil type ? La réponse est sans doute variable selon le domaine étudié. De plus, la fréquence n'est peut être pas le seul argument pour définir des classes de trajectoires.

Certaines pourraient être isolées à partir d'autres critères (leur coût par exemple).

Les résultats issus de l'ECBD sont étroitement liés à la forme des données cibles. Celle-ci résulte des choix opérés par l'expert qui manipule le système d'extraction. Ainsi, la précision des codes diagnostics CIM10 doit apporter un compromis entre une information trop fractionnée limitant la probabilité d'extraire des profils de trajectoire suffisamment généraux, et une information diluée dans laquelle aucun profil n'est individualisable. Le choix des nomenclatures et de leur granularité prend ici toute son importance.

4.4 Perspectives

Ce travail représente une première étape dans la compréhension de la trajectoire de soins. Nous envisageons à l'avenir plusieurs axes de développements.

Nous espérons travailler sur une période d'observation plus longue, pour bénéficier d'un historique plus complet de la prise en charge des patients. Nous pourrions ainsi tester un modèle de trajectoire orienté « épisode » de soins. La littérature montre que ce dernier type de modèle pose le problèmes de définition des bornes de l'épisode. A priori, le PMSI ne permet pas toujours de marquer avec exactitude le début d'un processus morbide. Il serait intéressant de voir quelles réponses peuvent apporter les outils de l'ECBD à ce problème.

Un autre axe de développement porte sur une amélioration et une meilleure utilisation des méthodes de fouille de données. Nous avons déjà initié ce travail [33] en combinant treillis de Galois et motifs séquentiels fréquents pour classer et visualiser les trajectoires. Nous envisageons également de classer les trajectoires en fonction d'une variable cible, par exemple leur coût, ou de déterminer dans quelle mesure des trajectoires sont semblables en définissant une distance inter-trajectoire.

Notre objectif à terme est de définir une classification des trajectoires de soins. Cette approche est d'un intérêt particulier dans la compréhension et l'évaluation de l'organisation du

système de santé. La disponibilité d'un outil d'analyse et de comparaison de la prise en charge des pathologies au long cours, par plusieurs institutions et professionnels de santé répondrait à des demandes diverses. La tarification à l'activité (T2A) mise en place en 2004, instaure un mode de financement des hôpitaux et cliniques centré sur l'hospitalisation. On peut se demander si le choix de cette unité de production de soins convient à aux pathologies chroniques comme le cancer, de surcroît lorsqu'elles sont prises en charge au sein de réseaux. La typologie et l'organisation de l'offre de soins étant extrêmement variables d'une région à l'autre, quelle sera l'influence de l'arrivée de la T2A sur l'organisation et la coopération entre établissements ?

Conclusion

Plusieurs études, ayant des méthodologies et des points de vue différents, ont été menées en France afin de décrire l'activité du système de soins au travers des trajectoires de patients. Leurs objectifs sont multiples :

- comprendre les mécanismes qui génèrent des flux de patients entre plusieurs prestataires de soins, ou à l'intérieur d'un même établissement ;
- évaluer les modalités médicales de prise en charge de certaines pathologies ;
- évaluer le coût de la prise en charge globale de certaines pathologies ;
- tester des outils de mesure de l'activité médicale ambulatoire.

Les connaissances disponibles sur la trajectoire de soins, notamment hospitalière, restent encore parcellaires. Avec l'introduction du chaînage des hospitalisations par patient, le PMSI apparaît comme un outil médicalisé de suivi longitudinal du parcours de soins hospitalier. En outre, il reste, à l'heure actuelle, le seul système d'information autorisant ce type de démarche à grande échelle. Notre étude montre l'intérêt que le PMSI présente dans l'analyse de la trajectoire des patients en cancérologie. Nous avons pu expliciter la nature de la prise en charge des patients par les quatre établissements du site hautement spécialisé de Nancy, au cours de l'année 2003. Nos résultats décrivent notamment :

- le nombre total de patients pris en charge par localisation tumorale et par établissement ;
- lorsqu'elle est partagée, l'organisation de cette prise en charge entre les différents établissements et le nombre de patients concernés.

L'analyse de la trajectoire de soins s'avère relativement complexe. Celle-ci peut être constituée d'informations d'ordre médical, administratif, économique, démographique, spatial et surtout temporel. Nous avons eu recours à des méthodes empruntées à l'Extraction de Connaissances dans les Bases de Données et au Data Mining. Ces technologies, en plein essor dans le domaine de la recherche bio-médicale, s'adaptent bien à l'étude de la trajectoire de soins. Elles pourraient contribuer à la construction d'une classification des trajectoires à partir des données.

La maîtrise des trajectoires de soins est une des clés de l'organisation du système de santé. Que ce soit dans le domaine de l'évaluation des pratiques professionnelles, de l'adéquation des ressources aux besoins de santé, de la planification ou encore de la régulation des dépenses, les trajectoires de soins offrent la vision panoramique indispensable à la compréhension des relations qui se tissent entre les professionnels de santé et les patients.

Bibliographie

[1] Y BOURGUEIL, M BRÉMOND, A DEVELAY, M GRIGNON, F MIDY, M NAIDITCH et D POLTON : L'évaluation des réseaux de soins - enjeux et recommandations. Rapport technique, Groupe Image-Ensp, CREDES, May 2001.

[2] L PERRIER : *Modes de financement du service public hospitalier et trajectoire optimale du patient en cancérologie pédiatrique.* Thèse de doctorat, Université Lumière Lyon II, dec 2001.

[3] M NAIDITCH : Modélisation des trajectoires : problèmes méthodologiques. *ITBM-RBM*, 21(5):307–12, 2000.

[4] F RIOU et P JARNO : Représentation et modélisation des trajectoires de soins. *ITBM-RBM*, 21(5):313–17, 2000.

[5] J WEINER, B STARFIELD, D STEINWACHS et L MUMFORD : Development and application of a population-oriented measure of ambulatory care case mix. *Medical Care,* ; 29(5) : 452-72.(5):452–72, 1991.

[6] JM RODRIGUES, JF BRULET, B TROMBERT-PAVIOT et F MENNERA : Une approche casemix adaptée a l'épisode de soins. *In L'Informatisation du Cabinet Médical du Futur*, volume 11 de *Informatique et Santé*, pages 241–53. Springer-Verlag France, 1999.

[7] P Le Fur, V Paris, H Picard et D Polton : Les trajectoires des patients en Île-de-france : l'accès aux plateaux techniques. Rapport technique, CREDES, 2000.

[8] T Dart, Y Cui, G Chatellier et P Degoulet : Analysis of patient flows using data-mining. *In* R.Baud et al., éditeur : *The New Navigators : from Professionals to Patients. Proceedings of Medical Informatics Europe 2003*, pages 263–268, 2003.

[9] B Blanchon et al : Prise en charge des cancers bronchiques en ile-de-france. *Rev Mal Respir*, 14(4):839–46., Oct 2000.

[10] R Launois, M Giroud, AC Megnigbeto, K Le Lay, G Presente, MH Mahagne, I Durand et AF Gaudin : Estimating the cost-effectiveness of stroke units in france compared with conventional care. *Stroke*, 35(3):770–5, mar 2004.

[11] L Borella et A Paraponaris : Reconstitution macro-économique des dépenses, suivi micro-économique des coûts du cancer du sein. *Rev Epidemiol Sante Publique*, 50(6):581–92, Dec 2002.

[12] F Le Duff, A Happe, A Burgun, S Levionnois, M Bremond et P Le Beux : Sharing medical data for patient path analysis with data mining method. *Medinfo*, 10(2):1364–8, 2001.

[13] S Bayat Makoei : Étude des trajectoires de soins des patients de 75 ans et plus atteints d'ischémie cardiaque dans le département des deux-sèvres. Mémoire, DESS Traitement Information Médicale Hospitalière, 2003.

[14] G Williams, R Baxter, C Kelman, C Rainsford, H He, L Gu, D Vickers et S Hawkins : Estimating episodes of care using linked medical claims data. *LNCS*, 2557:660 – 671, 2002.

[15] U FAYYAD, G PIATETSKY-SHAPIRO et P SMYTH : The kdd process for extracting useful knowledge from volumes of data. *In Communication of the ACM*, volume 29, pages 27–34, Nov 1996.

[16] IARC : *Cancer incidence in five continents. Volume VIII.* Numéro 155. IARC Sci Publ, 2002.

[17] L BORELLA, P PEUVREL, M SAUVAGE, D MARANINCHI et T PHILIP : Un essai d'exploitation de la base pmsi nationale pour évaluer le volume et les modes de prise en charge du cancer, en secteur hospitalier non lucratif. *Rev Epidemiol Sante Publique*, 48(1):53–70, Jan 2000.

[18] L PERRIER, L BORELLA, P MAREC-BERARD, N CRAPEAU, M BENSADON et T PHILIP : Un essai d'exploitation de la base nationale du pmsi pour évaluer la spécificité et le coût de la prise en charge hospitalière des cancers pédiatriques en france. *Bull Cancer*, 89 (11):975–83, Nov 2002.

[19] L BORELLA, S FINKEL, N CRAPEAU, P PEUVREL, M SAUVAGE, L PERRIER, E LEPAGE, J VILLEMINOT et B GARRIGUES : Volume et coût de la prise en charge hospitalière du cancer en france en 1999. *Bull Cancer*, 89(9):809–21, Sep 2002.

[20] CM COURIS, C FORÊT-DODELIN, M RABILLOUD, C COLIN, J-Y BOBIN, D DARGENT, D RAUDRANT et A-M SCHOTT : Sensibilité et specificité de deux methodes d'identification des cancers du sein incidents dans les services specialisés à partir des données médico-administratives. *Rev Epidemiol Sante Publique*, 52(2):151–60, Apr 2004.

[21] BT PAVIOT, C MARTIN, L CLAVEL, G DE LAROCHE et JM RODRIGUES : From drg databases to an epidemiological observatory for colorectal cancer in a french small area oncology network. *Stud Health Technol Inform*, 95:829–33., 2003.

[22] ATIH : Guide méthodologique de production des résumés de séjour du pmsi en médecine, chirurgie et obstétrique. Rapport technique, ATIH, 2004. URL http ://www.atih.fr.

[23] RB FETTER, Y SHIN, JL FREEMAN, RF AVERILL et JDTHOMPSON : Case mix definition by diagnosis-related groups. *Med Care*, 18(2):1–53, Feb 1980.

[24] PERNNS : Manuel des groupes homogènes de malades. Rapport technique, Ministère de la santé, de la famille et des personnes handicapées., jan 2004.

[25] F VOLLAIRE et L BRUNEL : Pmsi et cancérologie : l'outil dac © ou comment décrire l'activité de cancérologie, 2001. URL http ://dime.ap-hop-paris.fr/Mco/presentation3_dac.php.

[26] M BARBUT et B MONJARDET : *Ordre et Classification*, volume II. Hachette, 1970.

[27] R WILLE : Restructuring lattice theory : an approach based on hierarchies of concepts. *In* I. RIVAL, éditeur : *Ordered Sets*. Reidel, 1982.

[28] R GODIN, G MINEAU, R MISSAOUI et H MILI : Méthodes de classification conceptuelle basées sur les treillis de galois et applications. *Revue d'Intelligence Artificielle*, pages 105–137, 1995.

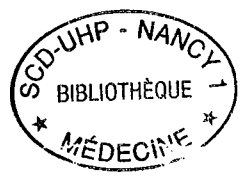
[29] R AGRAWAL et R SRIKANT : Mining sequential patterns. *In proceedings of the Int. Conf. Data Engineering(ICDE'95)*, Taipei, Taiwan, Mar 1995.

[30] SA YEVTUSHENKO : System of data analysis "concept explorer". *In Proceedings of the 7th national conference on Artificial Intelligence KII-2000*, pages 127–134, Russia, 2000. (In Russian).

[31] M JANSSEN : *SIMuLLDA. A Multilingual Lexical Database Application using a Structured Interlingua.* Thèse de doctorat, Universiteit Utrecht, 2002. URL http ://maarten.janssenweb.net/jalaba/JaLaBA.pl.

[32] M SENO et G KARYPIS : Slpminer : An algorithm for finding frequent sequential patterns using length decreasing support constraint. Rapport technique 02-023, Department of Computer Science, University of Minnesota, 2002. URL citeseer.ist.psu.edu/seno02slpminer.html.

[33] N JAY : Recherche et interprétation de motifs fréquents séquentiels dans une base de données médicales. Mémoire, DEA d'informatique. Université Henri Poincaré Nancy 1, 2003.



SCD-UHP - NANCY 1
BIBLIOTHÈQUE
MÉDECINE

VU

NANCY, le **2 juin 2004**
Le Président de Thèse

Professeur **F. KOHLER**

NANCY, le **22 septembre 2004**
Le Doyen de la Faculté de Médecine
par délégation

Professeur **H. COUDANE**

AUTORISE À SOUTENIR ET À IMPRIMER LA THÈSE

NANCY, le **27 septembre 2004**

LE PRÉSIDENT DE L'UNIVERSITÉ DE NANCY 1

Professeur **J.P. FINANCE**

Analyse de la trajectoire de soins en cancérologie

Fouille de données et Extraction des Connaissances à partir du PMSI

Le concept de trajectoire de soins remet le patient au centre de l'analyse de l'activité médicale. Il donne une vision panoramique des modalités d'accès du patient au système de santé, de l'organisation et de la coordination des soins qu'il reçoit. En France, peu de systèmes d'information autorisent ce type de démarche à grande échelle. Avec l'introduction du chaînage des données par patient, le PMSI devient un outil de suivi longitudinal du parcours de soins hospitalier. Nous montrons l'intérêt qu'il présente dans l'analyse de la trajectoire des patients en cancérologie. Par ailleurs nous proposons des méthodes d'exploration de cette trajectoire basées sur l'extraction des connaissances dans les bases de données. Ce travail est une première étape vers l'élaboration de « profils » types de trajectoires et d'une classification de la prise en charge des pathologies chroniques.

Patient pathway analysis in cancerology

Data mining and Knwoledge Discovery in the French DRG Data

THÈSE DE MÉDECINE SPÉCIALISÉE - ANNÉE 2004

Mots Clefs :

PMSI - trajectoire de soins - data mining - cancer

Faculté de Médecine de Nancy

9, avenue de la Forêt de Haye

54505 VANDOEUVRE LES NANCY Cedex