



HAL
open science

Elaboration d'un cadre méthodologique pour l'analyse de l'information médicale de la tarification à l'activité

Thierry Boudemaghe

► **To cite this version:**

Thierry Boudemaghe. Elaboration d'un cadre méthodologique pour l'analyse de l'information médicale de la tarification à l'activité. Médecine humaine et pathologie. Université Montpellier, 2016. Français. NNT : 2016MONTT022 . tel-01496764

HAL Id: tel-01496764

<https://theses.hal.science/tel-01496764>

Submitted on 27 Mar 2017

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.



THÈSE

Pour obtenir le grade de
Docteur

Délivré par l'**Université de Montpellier**

Préparée au sein de l'école doctorale Information,
Structures, Systèmes
Et de l'unité de recherche Biostatistique, Epidémiologie,
Recherche Clinique

Spécialité : **Biostatistique**

Présentée par **Thierry Boudemaghe**

**Elaboration d'un cadre méthodologique
pour l'analyse de l'information médicale de
la tarification à l'activité**

Soutenue le **9 décembre 2016** devant le jury composé de

Mme Béatrice TROMBERT-PAVIOT, Professeur des Universités, Université de St Etienne	Rapporteur
M Laurent MOLINIER, Professeur des Universités, Université de Toulouse	Rapporteur
M Jean-Pierre DAURES, Professeur des Universités, Université de Montpellier	Examineur
M Paul LANDAIS, Professeur des Universités, Université de Montpellier	Directeur



Elaboration d'un cadre méthodologique pour l'analyse de l'information médicale de la tarification à l'activité

Résumé

En dépit de la production standardisée de millions d'enregistrements décrivant année après année l'activité d'hospitalisation complète en France, aucune étude globale, aucun suivi précis et répété dans le temps, n'a pu être mené à partir de ces données.

Si ces données sont exploitées dans une grande variété d'études spécifiques, elles peinent à trouver une utilisation significative dans l'analyse de l'activité et du recrutement des établissements de santé, ou dans celle des besoins de la population. Tout un pan d'analyse de l'efficience du système de santé reste ainsi inaccessible.

Nous nous fixons comme objectif de contribuer à développer un cadre précis d'exploitation de ces données. Cette démarche se fera en trois temps :

- Définition d'une méthode de consolidation des données ;
- Construction de référentiels d'analyse des données ;
- Elaboration de méthodes d'analyse sur deux thématiques générales : la caractérisation de l'activité des établissements et l'étude de leur recrutement, avec exemple d'application de ces méthodes.

Mots clés

Tarification à l'activité, Information médicale, Entropie, Référentiel

Construction of a methodological framework for the analysis of medical information within the French prospective payment system

Summary

Millions of computerized records describing inpatient hospitalization activity are produced year after year in France.

In spite of the availability of this massive amount of data, no global, iterative and well calibrated related study on health system efficiency has been possible.

There are of course many specific studies partially or totally relying on these data, but their systematic use for assessing hospital activity, catchment areas or population health needs remains to be implemented.

Our work aims to contribute to create a methodological framework for analyzing these data through a three step approach :

- Definition of data consolidation methods ;
- Creation of adequate data repositories ;
- Determination of analysis methods for two general topics : hospital activity characterization and study, with example applications.

Keywords

Prospective payment system, Medical information, Entropy, Data repository

Unité de recherche Biostatistique, Epidémiologie, Recherche Clinique

Institut Universitaire de Recherche Clinique

641, avenue du doyen Gaston GIRAUD 34093 Montpellier cedex 5

Remerciements

Parmi les nombreuses personnes ayant contribué à me permettre de réaliser ce travail, certaines l'ont probablement fait à leur insu, d'autres sans que je ne m'en aperçoive...

Tout a probablement commencé avec mes premiers pas dans le laboratoire de Biomathématiques de la Faculté de Médecine de la Timone, à Marseille, chez le Pr Roux. C'est en particulier grâce à la pédagogie du Pr Bernard Fichet que je dois d'avoir pu aborder les statistiques ; il m'a donné des bases sur lesquelles j'ai toujours pu compter. C'est avec son assentiment, qui valait tous les encouragements, que je me suis inscrit au DEA de Biostatistique de Montpellier.

J'y ai fait la rencontre du Pr Daurès, qui deviendrait 8 ans plus tard mon chef de service. Je ne saurais dire la reconnaissance que j'éprouve à son égard pour tout ce qu'il a pu m'apprendre avec sa hauteur de vue, et la formidable leçon de bienveillance, de curiosité et d'esprit d'initiative que j'ai reçue en travaillant à ses côtés. Je n'aurais probablement pas mené ce travail à son terme sans les encouragements qu'il a su me prodiguer à un moment critique.

Enfin, je tiens ici à remercier mes parents, mon épouse et mes enfants. Vous êtes le sel de la vie sans qui tout ceci n'aurait pas de sens.

A tous, j'adresse toute mon immense gratitude pour tout ce que vous m'avez donné.

Table des matières

Introduction	5
Chapitre 1 Stratégie de consolidation des données	17
1. Intérêt et qualité des données. Consolidations et non consolidation	17
2. Les référentiels de codage	18
2.1. Aspects généraux touchant aux référentiels de codage	18
2.2. Les référentiels du PMSI	21
2.2.1. Codage des diagnostics	22
2.2.2. Codage des Actes	24
2.2.3. L'échelle des GHM et des GHS	26
2.2.4. Autres classifications d'intérêt	28
2.2.5. La pointe de l'iceberg : le numéro FINESS	28
3. Stratégie de gestion des référentiels	30
3.1. Diagnostics et Actes	30
3.2. Les GHM	31
3.3. Les établissements	31
Chapitre 2 Consolidation et exploitation du chaînage	32
1.1. Introduction	32
1.2. Matériel et méthode	33
1.2.1. Les tests	33
1.2.2. Stratégie d'application des tests et algorithme général	37
1.2.3. Données analysées et moyens techniques	40
1.3. Résultats	41
1.4. Applications	41
1.4.1. Sélection d'un échantillon pour l'étude des trajectoires	41
1.4.2. Construction d'une table des trajectoires	45
1.4.3. Etude sur la mortalité hospitalière dans la chirurgie de la hanche	47
1.4.4. Addictologie	47
1.4.5. Etude sur les ré-hospitalisations dans l'insuffisance cardiaque	47
Chapitre 3 Méthode d'évaluation du niveau de référence de l'activité	49
1. Introduction	49
2. Les outils mathématiques : l'entropie et les divergences entre distributions	50
2.1. L'entropie de Shannon	50
2.1.1. Le concept	50
2.1.2. Quelques propriétés de l'entropie de Shannon	51
2.1.3. Décomposition de l'entropie	53

2.2.	Divergence entre distributions.....	59
3.	Analyse de l'entropie de la matrice d'activité.....	61
3.1.	Matrice d'activité et modèle de partition.....	61
3.1.1.	Matrice d'activité.....	61
3.1.2.	Modèle de partition de l'activité.....	62
3.2.	Homogénéité et standardisation des données.....	63
3.2.1.	Etablissements de référence.....	63
3.2.2.	Standardisation des effectifs de classe.....	65
3.2.3.	Standardisation de l'entropie.....	66
3.3.	Résultats.....	67
3.4.	Discussion.....	73
Chapitre 4 Caractérisation de l'activité des établissements.....		74
1.	Introduction.....	74
2.	Méthode.....	74
2.1.	Etablissements sélectionnés.....	74
2.2.	Le découpage de l'activité en classes.....	75
2.3.	Stratégie de classification.....	76
3.	Analyse préalable et détermination des paramètres de la classification.....	77
3.1.	Caractéristiques des données.....	77
3.2.	Choix des paramètres de la classification hiérarchique.....	81
3.2.1.	Hypothèse de travail.....	81
3.2.2.	Distance, critère d'agrégation et indicateur de qualité.....	81
4.	Résultats.....	83
4.1.	Détermination des classifications d'intérêt.....	83
4.2.	Profils pour les classifications de 4 à 6 groupes.....	86
4.3.	Classification retenue.....	91
5.	Discussion.....	92
Conclusion.....		95
Annexes.....		104
I. RSS non groupé format 017.....		105
II. RSS groupé format 117.....		107
III. Format de RSA format 223.....		108
IV. Bibliographie sur le chaînage des données PMSI.....		112
V. Racines de GHM classées par contribution croissante à l'entropie.....		116
VI. Publication - Medico-economic decision making: an international comparison.....		139

VII. Publication – Outcomes After Hip Fracture Surgery Compared With Elective Total Hip Replacement.....	158
VIII. Publication – Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI)	170
IX. Publication – Identifying Life-Threatening Admissions for Drug Dependence or Abuse (ILIADDA): Derivation and Validation of a Model	178

Table des figures

Figure 1. Evolution de la dépense de santé par habitant entre 1975 et 2010. Pourcentage d'évolution au cours des cinq années précédentes pour les années 1975 à 2010.....	6
Figure 2. Représentation simplifiée de la chaîne de production des informations médicales de la T2A.....	14
Figure 3 - Cas particulier de repérage à tort d'une incohérence d'âge	35
Figure 4 - Estimation du délai maximum entre 2 séjours.....	37
Figure 5 - Algorithme général de détermination du niveau de fiabilité de l'identifiant patient	39
Figure 6 - Structure générale de la base - la table des RSA et les tables satellites	40
Figure 7 - Structure de la population.....	42
Figure 8- Affectation de l'identifiant du RSA de référence	46
Figure 9. Accroissement de l'entropie pour la variable aléatoire multinomiale $M(4, 10)$	53
Figure 10 - Représentation schématisée d'une classification hiérarchique	55
Figure 11 - Etapes de sélection des établissements.....	65
Figure 12 - Matrice d'activité et contribution à l'entropie des classes d'activité	67
Figure 13 - Entropie des prises en charge - Plan d'analyse	68
Figure 14 - Contribution cumulée à l'entropie des racines de GHM	71
Figure 15 - Caractérisation de l'activité des établissements - Plan d'analyse	76
Figure 16 - Population des établissements - Caractéristique des classes d'activité	78
Figure 17 - Nombre (proportion) d'établissements en fonction du pourcentage d'activité - Classes 1 à 5	79
Figure 18 - Nombre (proportion) d'établissements en fonction du pourcentage d'activité - Classes 1 à 5	80
Figure 19 - Classification à 4 groupes - Profils moyens	86
Figure 20 - Classification à 5 groupes - Profils moyens	88
Figure 21 - Classification à 5 groupes - Profils moyens des 2 nouveaux groupes	88
Figure 22 - Classification à 6 groupes - Profils moyens	90
Figure 23 - Classification à 6 groupes - Profils moyens des 2 nouveaux groupes	91
Figure 24 - Profils d'activité des CHU de Nîmes, CH Alès et CH Bagnols sur Cèze ..	92
Figure 25 - Profils CHU Nîmes et CH Alès - Patients du bassin alésien.....	93
Figure 26 - Profils du CHU de Nîmes et du CH Bagnols - Patients du bassin bagnolais	93

Introduction

« La tarification à l'activité (T2A), introduite en 2004-2005 pour financer l'activité de court séjour des hôpitaux publics et privés, avait pour but d'améliorer l'efficacité des établissements de santé et du secteur hospitalier. Or le suivi de l'impact de la T2A sur l'évolution de l'activité et de la productivité hospitalière reste à ce jour partiel. »[1]

Le constat effectué dans cette citation d'un document de l'IRDES résume parfaitement la démarche qui est à l'origine de notre travail. En effet, il est remarquable d'observer qu'en dépit de la production standardisée de millions d'enregistrements décrivant année après année l'activité d'hospitalisation complète en France, aucune étude globale, aucun suivi précis et répété dans le temps, n'ait pu être mené à partir de ces données.

Il existe certes de très nombreuses utilisations de ces données, dans une grande variété d'études, sur un spectre diversifié de domaines et d'objectifs, allant de la description épidémiologique de pathologies cancéreuses [2-4] à des études d'activité menées par la tutelle en vue de décision et de planification [5]. Ces données sont même utilisées dans la définition réglementaire de périmètres et de seuils d'activité [6].

Pourtant, elles peinent à trouver une utilisation significative dans l'analyse globale de l'activité des établissements de santé, dans celle des besoins de la population, et encore plus dans celle du recrutement des établissements. Tout un pan de l'analyse de l'efficacité du système de santé reste inaccessible alors que ces données ont été générées par la démarche même de recherche d'efficacité.

Devant cette situation, nous avons souhaité contribuer au développement d'un cadre précis d'exploitation de ces données.

Le principe du système de paiement prospectif de l'activité

Avec l'élévation du niveau de vie et l'augmentation de l'espérance de vie, les pays industrialisés doivent faire face à une augmentation de leurs dépenses de santé.

Quand on se penche par exemple sur la France, l'Allemagne et les Etats-Unis, les chiffres de l'OCDE[7] indiquent une augmentation des dépenses de santé par habitant de l'ordre de 20% (Etats-Unis) à 40% (France et Allemagne) entre 1970 et 1975. Au cours des cinq années suivantes, le rythme se ralentit pour la France et l'Allemagne mais reste du même ordre aux Etats-Unis, voir progresse encore pour atteindre 28% entre 1985 et 1990. Ce n'est que cinq ans plus tard que cette progression retrouve un niveau compatible avec celui de la France et de l'Allemagne.

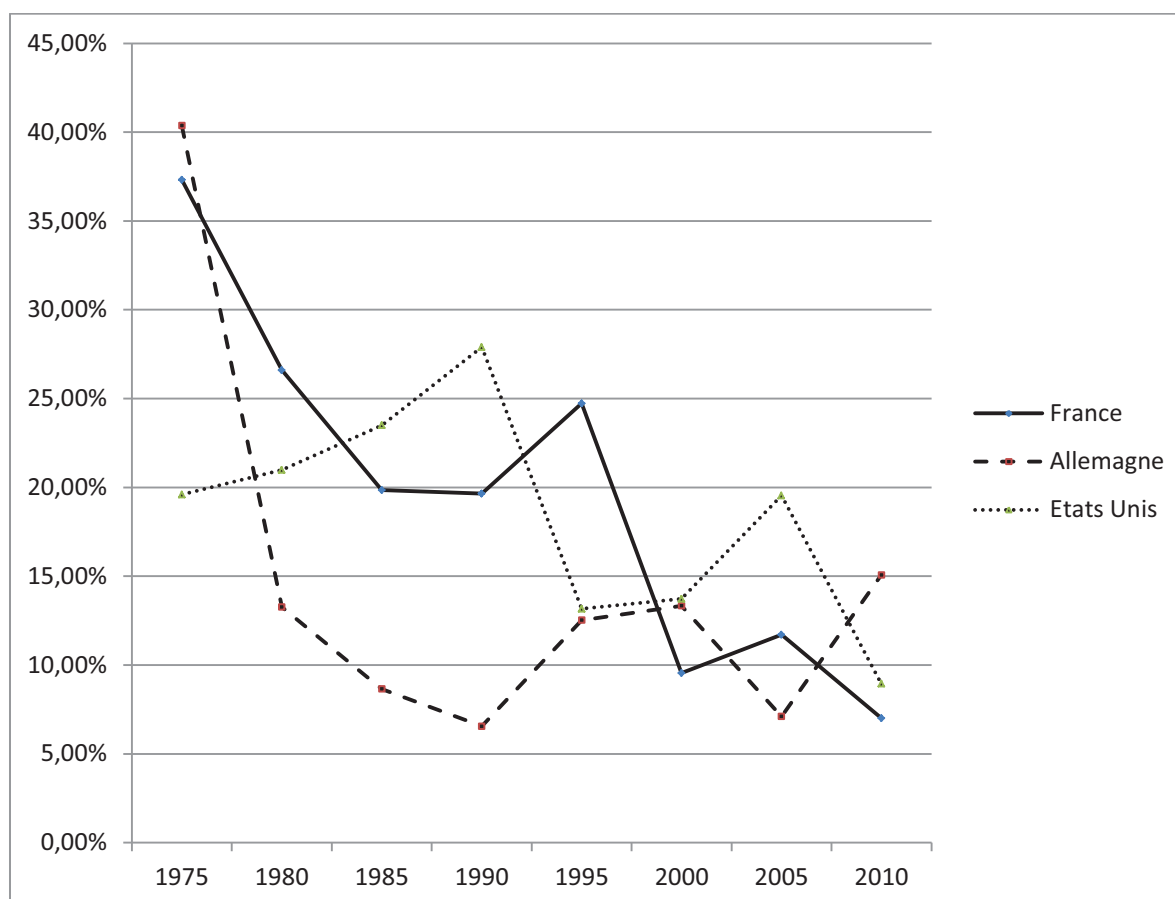


Figure 1. Evolution de la dépense de santé par habitant entre 1975 et 2010. Pourcentage d'évolution au cours des cinq années précédentes pour les années 1975 à 2010.

Conjuguée au choc pétrolier et au contexte économique international, la situation à la fin des années 70 ne pouvait pas être soutenue. D'autre part, des études médico-économiques étaient menées sur l'efficacité des systèmes de santé et leurs coûts. Enfin, facteur méconnu mais important, les capacités de l'informatique à stocker et traiter de gros volumes de données rendaient accessibles la réalisation de calculs et d'études à des échelles encore jamais atteintes jusque-là.

C'est dans ce contexte qu'est réalisée en 1980 l'étude fondatrice de Robert Fetter, en tout cas celle que l'histoire retiendra : « *Case mix definition by diagnosis related groups* »[8]. Avec cette étude, on rapproche pour la première fois, par une analyse statistique menée à grande échelle, le coût de l'hospitalisation avec son contenu médical. Cette démarche vient révolutionner un système dans lequel le principe était jusque-là de rembourser un dollar pour un dollar dépensé.

Afin d'introduire une démarche médicale dans la détermination des tarifs de remboursement, Fetter réalise une analyse de variance portant sur des groupes de séjours prédéterminés selon une logique médicale, les « *Diagnosis Related Groups* » (DRG) et élabore la liste des 467 DRG initiaux.

Il démontre que la part de variance expliquée est suffisamment importante pour permettre d'élaborer des tarifs *a priori* des hospitalisations en fonction de leurs caractéristiques médicales. Ces tarifs sont réputés couvrir, en moyenne, les dépenses engagées lors de la prise

en charge en hospitalisation. C'est le système de paiement prospectif qui voit ainsi le jour. Il sera généralisé et connaîtra ensuite de nombreuses évolutions et déclinaisons (hospitalisation pour rééducation, psychiatrie, activité ambulatoire...). Dès 1982, ce système est adopté aux Etats-Unis.

Le contexte français : de la coupe aux lèvres

• Les réformes du système de santé et de son financement

La démarche de contrôle des dépenses de santé s'intègre dans un contexte plus général d'organisation de la santé et de planification sanitaire. C'est notamment le cas en France où toute une série de réformes ont progressivement structuré le paysage sanitaire[9].

Quelques grandes étapes jalonnent ce parcours :

- La loi portant réforme hospitalière de 1970[10]. Ce texte est véritablement fondateur : il crée la notion de service public hospitalier, avec la notion d'égalité et de continuité d'accès aux soins, mais instaure également la carte sanitaire et prévoit la coopération inter-établissement au plan local et régional (syndicat inter-hospitalier de secteur et régional), définit les types d'établissements court séjour (dits aussi Médecine, Chirurgie, Obstétrique - MCO) et moyen séjour (soins de suite et de réadaptation - SSR) les structures de gouvernance, met en place le régime des autorisations d'activité et d'équipement... Enfin, son article 52 stipule :

« Une réforme de la tarification des soins dispensés dans les établissements devra intervenir dans le délai de un an à dater de la promulgation de la présente loi.

Cette réforme fera notamment apparaître le coût réel des diverses prestations fournies dans chacune des unités d'hospitalisation définie à l'article 4 ainsi que les frais d'acquisition des prothèses et des médicaments coûteux ».

La disposition prévue à la première ligne sera appliquée en 2004 avec l'entrée en vigueur de la tarification à l'activité (TAA, ou T2A). Celle prévue à la seconde n'a toujours pas été appliquée en 2016. Seule l'étude nationale de coûts, portant sur un échantillon limité d'établissements, a été mise en place.

- La loi portant réforme hospitalière de 1991[11]. Ce texte introduit l'évaluation et l'analyse de l'activité des établissements de santé, dans un objectif d'amélioration et d'optimisation de l'offre de soins et de la prise en charge du patient. Il précise également les types d'établissements de santé en termes de court séjour (MCO), moyen séjour (SSR), psychiatrie, long séjour et hospitalisation à domicile (HAD), ainsi que leur missions (soins, enseignement, recherche).

Ce texte instaure un élément important dans la structuration de l'offre de soins : les schémas régionaux d'organisation sanitaire (SROS).

D'un point de vue politique et organisationnel, ils font apparaître la notion de contractualisation entre établissements de santé et représentant de l'Etat et de l'Assurance Maladie, autour d'objectifs et de moyens. Leur mise en place s'accompagne de toute une

structure de gouvernance locale et régionale : comité national et régional d'organisation sanitaire, commission régionale d'évaluation médicale des établissements, conférence sanitaire de secteur.

Au plan technique, les SROS sont censés être élaborés « sur la base d'une mesure des besoins de la population et de leur évolution, compte tenu des données démographiques et des progrès des techniques médicales et après une analyse, quantitative et qualitative, de l'offre de soins existante. » et ont pour objet « de prévoir et de susciter les évolutions nécessaires de l'offre de soins, en vue de satisfaire de manière optimale la demande de santé ». Quelles mesures ? Avec quelles méthodes ? Sur quelles données ? Le texte n'en dit rien, et pour cause.

Enfin, ce texte définit également l'organisation administrative et financière des établissements (conseil d'administration, organes représentatifs, projet d'établissement, budget et rapports avec l'Etat dans sa définition...).

- Les ordonnances dites Juppé de 1996[12-14]. Ces ordonnances sont promulguées pour mettre en œuvre le plan Juppé. Elles constituent un tournant dans l'architecture de notre système de santé. Du point de vue de son financement, qu'elle bouleverse : mise en place de l'Objectif National de Dépenses d'Assurance Maladie (ONDAM), voté par la représentation nationale chaque année à l'occasion de la Loi de Financement de la Sécurité Sociale (LFSS), création de la Caisse d'Amortissement de la Dette Sociale (CADES). Mais également avec la restructuration des acteurs : la création des Agences Régionales d'Hospitalisation (ARH) vient consolider la décentralisation du système de soins. De plus ces structures associent les services de l'Etat et ceux de l'Assurance Maladie (on a pu dire, à cet égard, qu'il s'agissait d'une démarche d'étatisation du système de santé).

- Le plan Hôpital 2007. Cette série de mesures annoncées en Conseil des Ministres fin 2002 se concrétise par l'ordonnance du 4 septembre 2003[15]. Le régime des autorisations est simplifié par la suppression de la carte sanitaire, mais le rôle des SROS est renforcé. Par ailleurs ce texte porte sur les modalités de coopération inter-hospitalière et le mode de financement des investissements (partenariats public-privé).

- Par la suite, l'ordonnance 2005-406 du 2 mai 2005[16] met en place la nouvelle gouvernance : organisation en pôles dans les établissements, instauration de la Tarification A l'Activité (TAA, ou T2A), possibilité de placer sous administration provisoire un établissement en déficit.

- La loi HPST (Hôpital Patient Santé Territoire) de 2009[17]. Dans cette nouvelle étape, les ARH sont transformées en Agences Régionales de Santé (ARS), avec des prérogatives étendues à la mise en œuvre de la politique de santé au plan régional et à la régulation de l'offre de soins. Le texte prévoit également la possibilité de constituer des Communautés Hospitalières de Territoire (CHT). Enfin ce texte substitue à la notion de service public hospitalier celle de mission de service public, élargissant ainsi cette notion au-delà de celle de la prise en charge hospitalière.

Parallèlement à l'évolution de l'organisation sanitaire, le mode de financement des établissements de santé a lui aussi évolué dans le domaine du court séjour (MCO) :

- Le mode de financement est initialement basé sur le prix de journée : les séjours sont payés aux établissements en fonction de prix de journées élaborés en fonction du type de prise en charge (médecine, chirurgie, réanimation...). Ce système inflationniste ne comporte aucune incitation ni à la qualité, ni à l'efficacité dans la consommation de ressources ;

- La dotation générale de fonctionnement (DGF). En 1983, ce mode est remplacé pour les établissements publics par une enveloppe annuelle, la dotation générale de fonctionnement (DGF). La DGF est reconduite chaque année selon un taux directeur ; ce système conduisant à des situations de rente ou au contraire de sous-dotation chronique en fonction de la situation de l'établissement lors de l'entrée en vigueur de ce mode de financement ;

- La Tarification A l'Activité (TAA, ou T2A). A partir de 2005, le financement par T2A prend progressivement le relais de la DGF pour le champ d'activité MCO, pour le remplacer totalement en 2008. Ce système reprend le principe du paiement prospectif mis en place 20 ans plus tôt sur la base des travaux de Fetter. Il présente également des risques inflationnistes ; néanmoins, dans la mesure où il s'inscrit dans le cadre de l'ONDAM, les dérapages peuvent être atténués par une régulation des tarifs à la baisse pour contenir les dépenses dans l'objectif fixé.

Ces réformes sont traversées par deux contradictions, politique et technique, qui s'atténuent dans le temps. Politiquement, il s'agit de la tension entre décentralisation et autonomie des établissements d'une part, et planification et centralisation décisionnelle d'autre part. Il semble que le législateur veut miser sur une plus grande autonomie des acteurs, et leur capacité à s'organiser spontanément de façon optimale pour assurer la mission de prise en charge de l'état de santé des populations. Mais pour s'organiser (avant même de songer à une stabilité, et encore moins une stabilité optimale, de comportement), le système a besoin de règles du jeu, d'un cadre dans lequel il va consommer des ressources pour générer le maintien de la santé des populations. C'est ici que peut apparaître la nécessité, ou la tentation, de multiplier les directives, les règles et les contrôles. Il s'agit alors d'une défiance à l'égard de l'initiative de terrain et d'un libre jeu des acteurs qui pourrait tendre à une activité anarchique, non coordonnées, avec explosion de la consommation de ressources.

Cette attitude nous semble être fondamentalement liée au deuxième axe de contradiction que nous relevons dans les réformes du système de santé : la distorsion technique entre les objectifs opérationnels affichés et les moyens scientifiques et techniques mobilisés pour les atteindre ; et d'une manière générale le manque non seulement de connaissances, mais de moyens de générer de la connaissance (outils, méthodes), sur le système de santé. Certes, il existe d'innombrables rapports, études et publications scientifiques sur le système de santé, son financement, son efficacité, et un grand nombre de ces contributions sont de grande qualité. Il demeure néanmoins une absence de vision globale et unifiée du système, de son évolution, de l'adéquation entre l'offre de soins et les besoins en santé de la population ; le tout dans l'environnement doublement évolutif des besoins de santé et des progrès techniques.

- **Le Programme de Médicalisation des Systèmes d'Information**

Les décideurs et les législateurs ne sont pas ignorants de cette situation ni inactifs. C'est ainsi que, concernant l'hospitalisation, le Programme de Médicalisation des Systèmes d'Information (PMSI) a vu le jour dès 1982. Cette initiative est directement issue de l'observation des travaux de Fetter. Une expérimentation est lancée au deuxième semestre 1982 par une note du directeur des hôpitaux, Jean de Kervasdoué[18], avec un démarrage prévu début 1983. Il s'agit de recueillir, à la sortie du patient, des informations concernant son séjour et de constituer ainsi un Résumé de Séjour, qui deviendra très rapidement un Résumé Standardisé de Séjour (RSS).

Fin 1985, l'informatisation de recueil est instituée[19,20]. La première liste de Groupes Homogènes de Malades (GHM), l'équivalent français des DRG de Fetter, est diffusée en 1986[21], elle comporte 471 GHM. Au cours des années suivantes, d'autres textes préciseront les objectifs et les modalités du recueil des RSS[22], les structures hospitalières dédiées à leur traitement (Départements d'Information Médicale, DIM), et les missions des professionnels chargés de ce traitement[23,24].

Sans dispositif contraignant, sans impact sur le budget, ce système ne repose alors que sur le volontariat et l'implication des différents professionnels dans la production de l'information. Ainsi, au début des années 90, on constate un taux d'exhaustivité dans la production des RSS de l'ordre de 80%[25].

C'est dans ce contexte qu'une expérimentation de mise en adéquation de l'activité et des coûts, menée en Languedoc-Roussillon, est généralisée en 1995[25] et conduit à la mise en place du système des points ISA (Indice Synthétique d'Activité). Il s'agit d'un système affectant chaque GHM d'un nombre de points ISA. L'activité d'un établissement, d'une région, de la France, peut alors être exprimée en points ISA. En les rapportant à la dépense engagée pour les produire, on obtient alors la valeur du point ISA de l'établissement, de la région, de la France. Ceci permet de comparer ces valeurs et de contribuer à définir des stratégies de réallocation de ressources en faveur de régions défavorisées, par exemple.

Au cours des dix années suivantes, le PMSI prendra ainsi progressivement une place de plus en plus importante dans l'analyse de l'activité des établissements, aussi bien dans les rapports avec la tutelle que dans les dialogues internes. Il faut ici évoquer l'activité inlassable du Pr Grémy, membre du conseil scientifique du PMSI, en faveur de ce dispositif et, au-delà, du développement de l'Information Médicale. Il faut également mentionner les travaux du Pr Daurès dans le cadre de l'expérimentation ayant conduit à la création des points ISA, système précurseur de celui de la T2A.

La mise en œuvre progressive de la T2A en 2004, achevée en 2008, ainsi que les premières réglementations s'appuyant sur des mesures d'activités réalisées à partir des données du PMSI, achèveront de donner à l'Information Médicale sa place comme instrument d'aide à la décision, et au-delà, dans la démarche de maîtrise médicalisée des dépenses de santé formalisée par les ordonnances Juppé.

Simultanément, l'Agence Technique de l'Information sur l'Hospitalisation (ATIH)[26] était créée en 2000, un décret venant définir son organisation et son fonctionnement[27]. L'ATIH est une agence de l'Etat, placée sous la tutelle des ministres chargés de la santé, des affaires sociales et de la sécurité sociale. Elle a pour mission de collecter et d'analyser les données, elle gère les aspects techniques des dispositifs de financement et élabore et maintient les nomenclatures de santé. Elle intervient sur tous les champs du PMSI et a vu son domaine d'intervention étendu au domaine médico-social en 2015.

Cette démarche s'est conjuguée, parfois heurtée, à l'évolution de l'informatique. En effet, il faut rappeler qu'en 1983 la micro-informatique n'existe pour ainsi dire pas ; le concept lui-même paraît encore irréaliste à certains. Seuls règnent sans partage les *mainframes*, grosses installations centralisées fonctionnant sous des systèmes d'exploitation de type Unix, auxquels accèdent des terminaux aux fonctionnalités réduites, qui ne sont pas encore ce que l'on appellera plus tard des clients légers. A titre de repère, on mentionnera que Microsoft est créé en 1975, Apple en 1976. IBM présente son premier *Personal Computer* (PC) en 1981. Microsoft lance le système d'exploitation Windows® en 1983, quant au célèbre MacIntosh d'Apple, il apparaît en 1984.

Avec sa logique de recueil, stockage et analyse de données dont le volume ne pouvait qu'être destiné à croître, le PMSI s'est implicitement placé sous l'égide des technologies de l'information et de la communication. Ses concepteurs en avaient-ils conscience ? Toujours est-il que la tendance à la production de données massives étant elle-même toute aussi massive dans les pays industrialisés, les progrès de l'informatique ont permis de répondre aux exigences techniques du PMSI.

Trente ans après les premières expérimentations, la T2A est maintenant installée comme un des instruments incontournables du financement et de l'analyse de l'activité hospitalière.

Du PMSI à la T2A : enrichissement et complexification des données

• La chaîne de production et de transmission des données à l'ATIH

La note de Jean de Kervasdoué citée plus haut peut être considérée comme le lointain prédécesseur de ce qui est aujourd'hui le « Guide méthodologique de production des informations relatives à l'activité médicale et à sa facturation en Médecine, Chirurgie, Obstétrique et Odontologie ». Il s'agit de la première définition du contenu des RSS, accompagnée de quelques consignes de recueil. On y retrouve toute la logique du recueil de données en vue de leur exploitation : liste et définition des items, recours à leur codage, définition de la méthodologie et de la nomenclature de codage.

La note détermine un ensemble de 9 informations ou item à recueillir :

1. Age
2. Sexe
3. Catégorie socioprofessionnelle
4. Date d'entrée dans l'entité juridique
5. Date de sortie de l'entité juridique

6. Diagnostic principal
7. Diagnostics autres
8. Procédures (actes)
9. Mode de sortie

On notera au passage l'anglicisme « procédures » « *procedure* » signifiant acte en anglais.

Les items sont définis :

« Les items 1 à 5 et 9 s'entendent comme définis par la circulaire 554 DH du 3 juillet 1979 (statistique H 80) »

Puis la méthodologie de détermination de leur valeur, son codage, et donc la définition de la nomenclature, ou référentiel, support de ce codage :

« L'item 6 correspond :

- au diagnostic
 - . considéré comme la cause de l'hospitalisation par le médecin,
 - mais
 - . déterminé à la sortie de l'hôpital ;
- le degré de précision minimum est celui de la Classification internationale des maladies (CIM) 9ème révision OMS, à 4 chiffres (une précision supérieure est à fortiori acceptable).»

On notera que le recours au codage n'est pas explicitement mentionné, du moins dans ce document, et semble être considéré comme allant de soi.

Ce modèle de recueil va être affiné en relevant, pendant le séjour à l'hôpital, des informations concernant chaque passage dans une unité médicale distincte dans un Résumé d'Unité Médicale (RUM). Le RSS sera alors constitué de l'ensemble des RUM produits au cours du séjour. Si le patient n'a fréquenté qu'une seule unité médicale, alors son RUM est le RSS.

Le contenu du RSS lui-même va évoluer au fil des ans, intégrant progressivement de nouvelles informations comme le diagnostic relié, le poids du nouveau-né, ou encore l'IGS 2 (indice de gravité simplifié 2, employé en réanimation). En 2016, le RUM comporte 36 items (voir annexe I).

Deux étapes vont se rajouter à la chaîne de production et de transmission des données. Ces étapes consistent en un traitement du fichier de RSS par des logiciels fournis par l'ATIH :

- **Le groupage.** Au cours de cette étape, chaque RUM (ou série de RUM) composant les RSS est analysé par un algorithme qui détermine à quel GHM le RSS est affecté. On obtient en sortie un nouveau fichier, dit de RSS groupés (RSSG, voir format actuel en annexe II).

- **L'anonymisation.** Le fichier de RSSG est traité pour produire un fichier de Résumés Standardisés Anonymisés (RSA, voir format actuel en annexe III). Plusieurs traitements sont effectués :

Agrégation. Là où un RSSG pouvait comporter plusieurs RUM, donc plusieurs enregistrements (plusieurs lignes) dans le fichier d'entrée, le fichier de RSA n'en comporte plus qu'une. On passe donc d'un modèle 1 séjour → 1 à n enregistrements à un modèle 1 séjour → 1 enregistrement

Anonymisation proprement dite. Les variables considérées comme identifiantes sont supprimées ou dégradées : la date de naissance est remplacée par l'âge à l'entrée, le code postal est remplacé par un code géographique pouvant regrouper plusieurs codes postaux, de façon à ce que la population totale atteigne ou dépasse les 1000 habitants, les dates d'entrée et de sortie sont supprimées et remplacées par la durée de séjour et le mois de sortie.

Détermination du code tarifaire de base. Sur la base du GHM essentiellement, complétée dans certains cas par d'autres informations comme la présence d'actes spécifiques, le code définissant le tarif auquel sera payé le séjour à l'établissement est déterminé. Ce code est appelé le Groupe Homogène de Séjour (GHS), à chaque GHS correspond un tarif.

Détermination de modulateurs tarifaires. Au fil des ans, des modulateurs tarifaires ont été ajoutés afin de prendre en compte des séjours dont les caractéristiques s'éloignent trop d'un modèle moyen, faussant la logique d'un paiement prospectif. A titre d'exemple, on peut citer les forfaits dits « extrême haut », somme forfaitaire payée en sus du tarif du GHS pour chaque journée d'hospitalisation au-delà d'une borne de durée de séjour. Ces modulateurs se sont multipliés au fil des ans. Il faut citer parmi les plus importants le forfait réanimation et le forfait soins intensifs, réputés rendre compte du surcoût occasionné par le passage du patient en unité de réanimation ou de soins intensifs.

Enfin, au moment du passage en T2A, l'intrication avec la chaîne de facturation est accrue par l'adjonction d'un fichier comportant des informations sur les droits au remboursement, le rapprochement avec le fichier des RSS se faisant par le numéro de séjour.

Simultanément, d'autres informations sont relevées et transmises en même temps que les RSS, sous la forme de fichiers complémentaires. Il faut citer en particulier le fichier des molécules onéreuses (MolOn) et des dispositifs médicaux implantables (DMI), mais ce ne sont pas les seuls : prélèvement multiples d'organes, dialyses péritonéales... Le corpus des informations transmises c'est considérablement accru au fil des ans, pour affiner et moduler toujours plus le paiement de leur activité aux établissements.

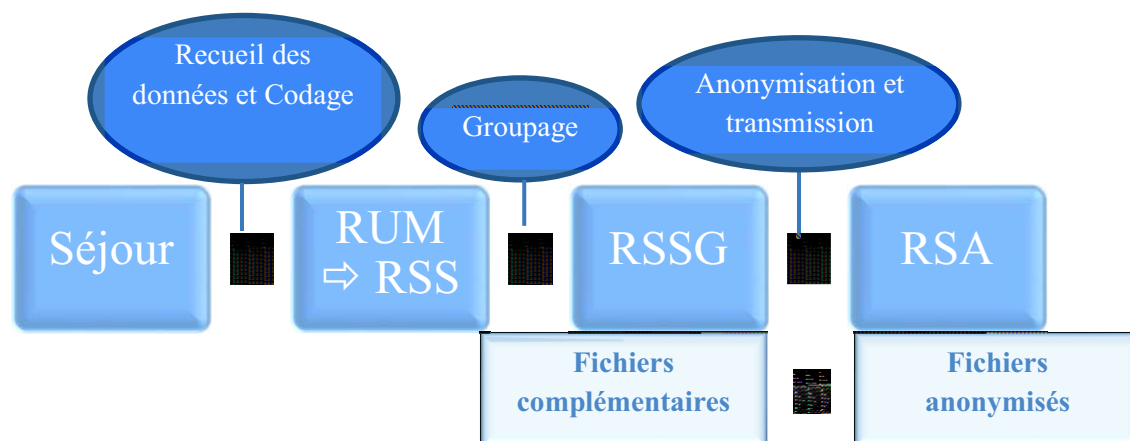


Figure 2. Représentation simplifiée de la chaîne de production des informations médicales de la T2A.

Les établissements transmettent leur activité à l'ATIH tout au long d'un exercice qui couvre l'année civile. L'ATIH constitue ainsi chaque année une base nationale de RSA, dont il était possible jusqu'en 2015 (pour l'année 2014) de faire l'acquisition.

- **Un atout majeur : l'identifiant unique de chaînage**

A partir de l'année 2001, une innovation majeure est réalisée : la mise en place d'un identifiant anonyme de chaînage propre à chaque patient[28], sur la base d'un travail universitaire réalisé au CHU de St Etienne [5].

Il s'agit de réaliser une empreinte numérique non réversible de quelques traits identifiants du patient : numéro de sécurité sociale, sexe, date de naissance. Cette empreinte est réalisée par « hashage » (anglicisme provenant du terme « *hashing* »), un procédé produisant, à partir d'une chaîne de caractères donnée en entrée, une chaîne de caractères en sortie, et présentant les caractères suivants :

- irréversibilité : il est impossible de remonter à la chaîne source à partir de la chaîne finale ;
- taux de collision virtuellement nul : la probabilité pour deux chaînes différentes de générer deux empreintes identiques est de l'ordre de 10^{-48} .

En outre, deux chaînes très proches produiront deux empreintes très différentes.

Afin de rendre une attaque par dictionnaire pratiquement impossible, les variables identifiantes sont combinées à une clef secrète avant d'être « hashées ». Au niveau de l'ATIH, l'empreinte produite est à combinée à une autre clef secrète avant d'être hashée à nouveau.

L'identifiant unique de chaînage a pour intérêt d'être stable dans le temps et d'un établissement à l'autre : les mêmes variables identifiantes donneront les mêmes empreintes quel que soit l'établissement fréquenté.

Naturellement, des erreurs sont possibles (saisie erronée, changement de numéro de sécurité sociale). De plus, le choix du numéro de sécurité sociale implique certaines limitations intrinsèques : indiscernabilité des jumeaux de même sexe, rupture de continuité dans le cas

de changement de numéro de sécurité sociale, ou encore cas des étrangers sans numéro de sécurité sociale. On peut cependant considérer qu'avec l'identifiant de chaînage nous disposons globalement d'un formidable outil de suivi de la trajectoire des patients.

Un monde à explorer

Depuis des années, la France s'est dotée d'un système de recueil pratiquement exhaustif des hospitalisations complètes. Sous l'effet de la T2A, la qualité de ce recueil ne cesse de s'améliorer. Depuis 2009, la stabilisation de la méthodologie de codage, la stabilisation de la structure du référentiel des GHM, la professionnalisation progressive du codage, ont rendu envisageable le suivi dans le temps de l'activité.

Enfin, grâce à l'identifiant anonyme de chaînage, il est possible d'étudier la trajectoire des patients. Tout ceci a contribué à créer une opportunité unique d'une meilleure connaissance de l'environnement sanitaire et de la santé des populations. Il devient enfin possible de parler d'aide à la décision médicalisée.

Il est cependant nécessaire de s'entourer des précautions méthodologiques nécessaires, car l'exploitation fiable et pertinente des données de la T2A passe par la maîtrise de leur environnement méthodologique :

- définition des items ;
- méthodologie de recueil des données, permettant de savoir quoi chercher et comment l'interpréter ;
- référentiels utilisés, avec leur intérêt et leurs limites ;
- l'évolution des trois éléments précédents au cours du temps.

Ainsi, une étude menée sur un recul de plusieurs années devra nécessairement tenir compte de la variation de la méthodologie de recueil et des référentiels, où être conçue de façon à ne pas être affectée par ces variations.

Nous avons souhaité présenter dans ce travail un cadre méthodologique adapté pour l'exploitation de ces données, avec quelques exemples d'application, dans lesquels nous présentons des méthodes d'analyse originales.

Le chapitre 1 présente une approche des stratégies de validation des données. Après une partie consacrée aux aspects fondamentaux et théoriques des référentiels de codage, les principales nomenclatures et référentiels employés sont abordés. Outre le codage des diagnostics et des actes et la classification des GHM, nous abordons plus en détail la question du numéro FINESS des établissements et des questions qu'il soulève. Enfin, ce chapitre se termine par une description de notre stratégie de gestion des référentiels.

Dans le chapitre 2, nous exposons une exploitation du chaînage. Après avoir décrit une méthode de validation de la fiabilité du chaînage, nous présentons les résultats obtenus et donnons quelques exemples d'utilisation des données obtenues.

Le chapitre 3 propose une méthode d'évaluation du niveau de référence de l'activité basée sur l'analyse de l'entropie de la distribution d'activité des établissements en racines de GHM. Sur la base des données consolidées grâce à l'effort fait sur la maîtrise des référentiels, nous pouvons établir une sélection raisonnée d'établissements à inclure dans l'étude. Le niveau de référence des racines de GHM est ensuite analysé sur cette base d'établissements.

Enfin, à partir des résultats du chapitre 3, le chapitre 4 consiste en l'élaboration d'une typologie des établissements basée sur le profil de référence de leur activité. Par l'application d'une méthode de clustering, nous regroupons entre eux des établissements de profils similaires et nous individualisons cinq grands types d'établissements.

En conclusion, nous soulignons les enjeux d'une meilleure maîtrise des référentiels, et au-delà, des modèles de représentation des données. Par ailleurs, l'importance de la question de la visualisation des données est également soulignée. Enfin, la perspective – la nécessité – de l'application des techniques d'ingénierie des connaissances aux données de l'Information Médicale est relevée.

Chapitre 1 Stratégie de consolidation des données

1. Intérêt et qualité des données. Consolidations et non consolidation.

Un RSA peut être vu comme une ligne de facturation pour une prestation de soins hospitaliers. Toute l'originalité de cet objet tient en ce que, au-delà de ce que l'on pourrait appeler « les codes produits » (le GHS et les modulateurs tarifaires), il comporte des informations supplémentaires :

- des informations médicales (DP, DR, DAS, Actes, mais aussi durée de séjour, passage en réanimation...);
- des informations concernant le patient (Age, Sexe, Code géographique).

Dans la mesure où la production de RSA s'intègre dans la chaîne de facturation de l'activité des établissements, elle peut être vue comme quasi-exhaustive. De plus, la définition d'une méthodologie précise de codage et la montée en puissance de la professionnalisation du codage dans les établissements, contribuent à un certain degré de fiabilité de ces données lorsque la qualité de la facturation est en jeu. Il est illusoire en revanche d'attendre un degré de précision trop élevé dans le codage.

Il est important de comprendre que ces données ne sont pas des données recueillies selon les standards de la recherche clinique. Le monitoring, par exemple, ne consisterait ici qu'à s'assurer de l'exhaustivité du recueil. Le data management ne porterait que sur des aspects ciblés, dans une optique de consolidation de la recette : ne pas oublier de comorbidités, s'assurer de la présence d'éléments de preuve pour les comorbidités codées et du respect de la méthodologie de codage.

Enfin, ce tableau doit être modulé par l'hétérogénéité constatée dans la base. On y trouve en effet des établissements de types différents, ne mettant certainement pas tous l'accent sur la qualité du codage de la même manière et sur les mêmes sujets. Tous n'ont pas le même intérêt à, ou les moyens de, consacrer la ressource nécessaire à une qualité optimale du codage et à la production de l'information médicale en général. A titre d'exemple, le financement des centres hospitaliers de proximité (ex-hôpitaux locaux) n'était pas jusqu'à début 2016 lié à la T2A.

Certes la charge incompressible liée à cette activité est faible : il est possible de désigner un clinicien non spécialiste responsable de l'Information Médicale pour satisfaire aux obligations réglementaires, on peut confier le recueil et le codage à du personnel des secrétariats médicaux, faire procéder aux envois de données par un technicien du service informatique... Cependant la qualité du codage ne sera évidemment pas au niveau de celle d'un établissement comportant une équipe de Techniciens de l'Information Médicale (TIM), régulièrement formés et travaillant en équipe.

Il faut donc à nouveau souligner que cette base ne présente pas en elle-même les caractéristiques et la qualité d'un recueil spécifiquement dédié à l'épidémiologie d'une part, et d'autre part qu'elle n'a pas *a priori* d'objectif de santé publique, même si cet argument a pu être mis en avant dans les débuts du recueil du PMSI. Pour autant, son exhaustivité, son

contenu en information médicale et la robustesse des informations ayant un impact sur la facturation, lui confèrent un intérêt majeur pour des études épidémiologiques.

La seule façon de concilier ces deux aspects sera de procéder à une consolidation des données avant d'exploiter la base, mais cette consolidation devra être ciblée en fonction des objectifs de l'étude.

2. Les référentiels de codage

2.1.Aspects généraux touchant aux référentiels de codage

L'utilisation d'un référentiel pour coder une information, et la nature même d'un référentiel, présentent un caractère d'évidence. Il nous paraît cependant intéressant de s'y arrêter un instant pour considérer cet objet, ses propriétés, les conséquences de son utilisation. Nous emploierons indistinctement les termes de référentiel, nomenclature ou terminologie, même si le terme de terminologie semble assez largement répandu.

L'emploi de registres, de listes, de nomenclatures pour le recueil systématique de données est avéré depuis longtemps dans l'histoire de l'humanité. En Europe, dans le domaine médical, les premiers certificats de décès remontent au XV^{ème} siècle, en Italie du Nord.

Par la suite, l'importance de ce qui n'était au départ qu'un simple outil à façon fut de plus en plus reconnue. Dans le domaine médical en particulier, l'épidémiologiste William Farr (1807 - 1883) écrivait en 1839 dans son rapport annuel sur les naissances, décès et mariages en Angleterre[29] :

"The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has in many instances been denoted by three or four terms, and each term has been applied to as many different diseases; vague, inconvenient names have been employed, or complications have been registered instead of primary diseases. The nomenclature is of as much importance in this department of inquiry, as weights and measures in the physical sciences, and should be settled without delay."

Que l'on peut traduire par :

« Les avantages d'une nomenclature statistique uniforme, pour imparfaite qu'elle soit, sont si évidents, qu'il est surprenant qu'aucune attention n'ait été prêtée à son utilisation systématique dans les certificats de décès. Chaque maladie a, en de nombreuses occasions, été désignée par trois ou quatre appellations, et chaque appellation a elle-même été aussi souvent appliquée à des maladies différentes ; des noms vagues, peu pratiques, ont été employés, ou des complications ont été enregistrées au lieu des maladies initiales. Dans notre domaine d'étude, une nomenclature est aussi importante que le système des poids et mesures en physique, et cette question devrait être traitée sans délai. »

On ne peut aborder la question des terminologies sans évoquer Eugen Wüster (1898 - 1977). Cet ingénieur autrichien se présente de façon atypique : une formation d'ingénieur en

électricité, une passion pour l'Espéranto avec la rédaction d'un dictionnaire Espéranto – Allemand, des travaux essentiellement axés sur le développement de terminologies tout d'abord techniques dans le but de résoudre des problèmes pratiques de communication d'information.

Par la suite, son travail dans le domaine des terminologies prendra une ampleur considérable et il élaborera une théorie générale de la terminologie, diffusée à titre posthume au cours des années 80[30], dans l'objectif de fonder la terminologie en tant que discipline scientifique. Il déploiera dans ce domaine une activité considérable : secrétaire de l'ISO/TC 37¹[31] de 1952 à 1971, impliqué dans la création de l'UNISIST² pour le compte de l'UNESCO, fondateur du centre Infoterm³[32] de Vienne... ne sont que quelques-uns des points les plus marquants de son parcours.

L'œuvre de Wüster, son impact et sa postérité sont si importantes qu'il n'est pas possible d'en rendre compte dans ce chapitre qui a pour objet de mettre en perspective les principaux enjeux méthodologiques liés à l'emploi de terminologie dans le PMSI. Il est néanmoins intéressant de citer quelques éléments clés de sa théorie, et de son cheminement ultérieur.

La théorie générale de la terminologie repose sur cinq principes[33] :

1. La terminologie part du concept, sans s'occuper du langage.
2. Un concept est clairement défini et séparé des autres et on peut lui assigner une place dans un système structuré.
3. Idéalement, un concept est défini en « intension » (par ses caractéristiques)⁴.
4. Un concept est défini de façon univoque et bijective par un terme et un seul : pas de synonyme, pas de polysémie.
5. L'association entre concept et terme est permanente.

Sur la base de ces cinq principes se développeront des évolutions, puis des critiques, de cette théorie. On peut globalement diviser le parcours de cette théorie en trois phases[34,35] :

- dans les années 80, la théorie de Wüster est accueillie de façon bienveillante et fait ses premiers pas académiques ;
- dans les années 90, les premières discordances et critiques se font jour avec la dissémination et la mise en œuvre de la théorie ;

¹ *International Organization for Standardization, Technical Committee "Terminology (principles and coordination)"*

² *Universal System for Information in Science and Technology*

³ *International Information Centre For Terminology*

⁴ Un concept est défini par :

- Son "*intension*" quand sa définition indique le concept superordonné et ses caractères distinctifs (une fracture de la diaphyse du fémur est une fracture d'un os localisée au niveau de la diaphyse du fémur).
- Son "*extension*" quand il est décrit par l'énumération de tous les concepts subordonnés i.e. critère de division (gaz rare : hélium, néon, argon, krypton, xénon, radon).

- dans la première décennie des années 2000 apparaissent les critiques de fond (positives ou négatives) de la théorie en tant que telle.

Chacun de ces principes peut être discuté, amendé, voir remis en question, et c'est ce qui va se produire : on fera remarquer par exemple que la dissociation entre concept et langage est problématique et ne prend pas en compte son évolution, ou encore que l'idée de concepts bien définis et séparés ne correspond pas toujours à la réalité, de même que l'univocité de la désignation et l'exclusion de la synonymie et de la polysémie.

Les travaux de Wüster se placent, et ont été analysés, commentés, contestés, dans le champ de la linguistique et des sciences humaines. Mais la question des terminologies se présente aujourd'hui comme un domaine interdisciplinaire, à la croisée des chemins entre linguistique, sociologie, intelligence artificielle, sciences de l'information, philosophie... Wüster lui-même en a bien conscience dans ses dernières productions[36].

En effet, avec le développement des sciences de l'information, la terminologie rencontre un domaine et un outil très puissant de mise en application systématique. Au minimum, le caractère systématique, standardisé et reproductible, des terminologies en font un outil pragmatique de choix pour l'indexation et l'échange d'information, même si, comme nous allons le voir plus bas, d'autres problèmes apparaissent très rapidement. Au-delà, les moyens technologiques informatiques, dont la puissance n'a cessé de croître, rendent possibles ou tout au moins envisageables des applications de plus en plus sophistiquées.

Il est nécessaire pour compléter cette mise en perspective d'aborder la notion d'ontologie. Le terme lui-même provient de la philosophie et signifie « étude de l'être ». Dans le domaine des technologies de l'information, une ontologie est une formalisation explicite d'un système de concepts (concepts, relations et contraintes dans un domaine donné). Dans la mesure où la démarche « ontologique » implique une étude, ou une méthode d'étude, des caractéristiques et des relations entre les objets (abstraits ou concrets), elle va s'articuler assez spontanément avec la logique des prédicats. Passant dans le domaine de l'intelligence artificielle dans les années 80, le terme acquiert un statut technique, désignant un système constitué d'entités et de leur relation[37].

Les chercheurs en intelligence artificielle cherchent alors à décliner ce concept de façon théorique puis opérationnelle. Sans poursuivre trop avant sur ce vaste sujet connexe, indiquons néanmoins que l'on peut sommairement repérer deux grands domaines d'application des ontologies dans le domaine de l'intelligence artificielle et de l'ingénierie des connaissances.

D'une part, la poursuite du projet d'une intelligence artificielle permettant à un robot de réaliser une tâche simple en utilisant une série de connaissances sur son environnement. Ici, une question théorique récurrente est celle du « *frame problem* », qui ressort, pour simplifier, de la capacité à intégrer le sens commun dans le raisonnement (intégrer spontanément les conséquences d'une action même si elles ne sont pas toutes explicitement décrites)[38,39]. Aujourd'hui, ce projet se poursuit à travers des techniques diversifiées (intelligence collaborative, réseaux neuronaux et « *deep learning* »).

D'autre part, nous concernant plus directement, le stockage, la communication et le requêtage interopérable de bases de connaissances. Dans ce domaine, de nombreuses initiatives ont vu le jour et se sont concrétisées par des formalismes de description et de manipulation de connaissance. Cette démarche a même abouti à l'élaboration d'un standard décrivant une famille de langages de logique destinés à permettre l'échange de connaissances entre systèmes informatiques (ISO 26162) [40,41], ainsi qu'un standard de développement de nomenclature (ISO 25964) [42-44].

Etant donné le développement massif de l'internet durant les années 90, le W3C[45] (World Wide Web Consortium) a bien entendu été amené à travailler sur ses questions. Le W3C est une organisation fondée en 1994 par Tim Berners-Lee. Il rassemble une communauté internationale autour du développement des standards de l'internet, groupés par grands thèmes (« *Web of Telecommunication* », « *Web of Data* », « *Web of Things* »...) traités par des groupes de travail. En 2001 le groupe de travail « *Semantic Web* » était créé, ayant pour objectif de développer des standards permettant l'échanges et la manipulation de connaissances. Concernant les nomenclatures, il faut noter la création de SKOS⁵[46], un standard libre aligné sur l'ISO 25964. Ce groupe a produit de nombreux travaux autour de langages de description et de requêtage d'ontologie (RDF⁶, OWL⁷, SPARQL⁸...). Fin 2013, il était réuni avec d'autres au sein de la thématique plus générale du « *Web of Data* » ; une dénomination moins ambitieuse ne faisant plus mention de la sémantique.

Les ontologies sont aujourd'hui très utilisées. Dans le domaine des sciences de la vie en particulier, le *National Center for Biomedical Ontology* a créé et maintient le site BIOPORTAL[47-50]. Ce site héberge et permet d'interroger ou télécharger des ontologies dans le domaine biomédical. Nul doute que se jouent ici des questions aux lourdes conséquences politiques, sociologiques et culturelles, dans la production, le contrôle et la diffusion de la connaissance au XXIème siècle.

Pour en revenir à la notion de référentiel de codage, on peut considérer que les ontologies en constituent aujourd'hui la forme la plus élaborée, ayant étendu des relations entre concepts initialement purement hiérarchiques (lien d'appartenance exclusive à une classe) à un domaine beaucoup plus étendu de relations à valences sémantiques diverses (appartenance multiple, causalité, association dans un processus...).

2.2. Les référentiels du PMSI

Outre les codages usuels (par exemple pour le sexe), le recours à des référentiels élaborés pour le codage des variables est inscrit dans les gènes du PMSI dès la première expérimentation.

⁵ SKOS : Simple Knowledge Organisation System

⁶ RDF : Resource Description Framework

⁷ OWL : Web Ontology Language

⁸ SPARQL : Acronyme récursif pour SPARQL Protocol and RDF Query Language

2.2.1. Codage des diagnostics

La classification employée pour le codage des diagnostics est la Classification Internationale des Maladies et des problèmes de santé connexes, 10^{ème} révision (CIM 10)[51]. En 2016, c'est une déclinaison adaptée par l'ATIH qui est utilisée pour le recueil des données du PMSI[52].

L'origine de la CIM remonte à Jacques Bertillon⁹, qui présente en 1893, à Chicago, lors du congrès de l'*International Statistics Institute*, la première nomenclature des maladies[53]. Celle-ci va progressivement être adoptée par la communauté internationale, et lors de la création de l'Organisation Mondiale de la Santé (OMS) en 1945, celle-ci se verra confier l'évolution et la mise à jour de la CIM. La 10^{ème} révision de la CIM a été adoptée en 1990 et est entrée en vigueur le 1^{er} janvier 1993. La 11^{ème} révision est actuellement en cours.

Actuellement, le CIM se présente sous la forme d'une classification hiérarchique comportant 22 chapitres :

I	Certaines maladies infectieuses et parasitaires
II	Tumeurs
III	Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire
IV	Maladies endocriniennes, nutritionnelles et métaboliques
V	Troubles mentaux et du comportement
VI	Maladies du système nerveux
VII	Maladies de l'œil et de ses annexes
VIII	Maladies de l'oreille et de l'apophyse mastoïde
IX	Maladies de l'appareil circulatoire
X	Maladies de l'appareil respiratoire
XI	Maladies de l'appareil digestif
XII	Maladies de la peau et du tissu cellulaire sous-cutané
XIII	Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif
XIV	Maladies de l'appareil génito-urinaire
XV	Grossesse, accouchement et puerpéralité
XVI	Certaines affections dont l'origine se situe dans la période périnatale
XVII	Malformations congénitales et anomalies chromosomiques
XVIII	Symptômes, signes et résultats anormaux d'examens cliniques et de laboratoire, non classés ailleurs
XIX	Lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes
XX	Codes d'utilisation particulière
XXI	Causes externes de morbidité et de mortalité
XXII	Facteurs influant sur l'état de santé et motifs de recours aux services de santé

⁹ J.Bertillon est le petit-fils d'Achille Guillard statisticien et botaniste, créateur du terme *démographie* qui en 1853, à l'occasion du congrès de statistique de Bruxelles propose de former une nomenclature uniforme des causes de décès. La tâche sera confiée à Marc D'Espine et à William Farr, un des pères de l'épidémiologie, de la notion de facteurs de risque et initiateur de la surveillance épidémiologique. Le père de Bertillon est statisticien de la ville de Paris, charge dont son fils héritera. Le frère de Jacques, Alphonse Bertillon est le créateur de l'identité judiciaire fondée sur un système de reconnaissance anthropométrique.

On notera la particularité des trois derniers chapitres. Le chapitre XX est destiné à permettre d'utiliser des codes temporaires, ou particulièrement dédiés à une thématique non ou insuffisamment couverte dans les autres chapitres. Actuellement, il comporte entre autres les résistances aux antibiotiques, une problématique qui s'est développée depuis quelques années jusqu'à atteindre aujourd'hui à un niveau préoccupant ; à ce titre il fallait pouvoir en rendre compte dans le codage des pathologies.

Le chapitre XXI décrit toutes les situations pouvant causer une lésion traumatique, et une atteinte à la santé en général : accidents, agression, tentative de suicide, mais aussi situation de guerre, catastrophe naturelle ou encore complications médico-chirurgicales, accident du travail, etc.

Enfin, le chapitre XXII décrit des facteurs associés ou préexistants aux pathologies, voire potentiellement aggravants ou même causaux, ainsi que les circonstances dans lesquelles un patient peut avoir recours aux services de santé. On trouvera dans ce chapitre des codes permettant par exemple de décrire dans une certaine mesure le statut social, familial, marital ou de dépendance du patient. Certaines circonstances particulières de recours aux soins sont aussi décrites, comme les activités de dépistage ou de suivi de prise en charge.

En termes d'arborescence, on trouve à l'intérieur de chaque chapitre des blocs, qui se déclinent en rubriques. Les codes terminaux se trouvent en général à l'intérieur des rubriques. Dans un certain nombre de cas, l'arborescence se poursuit en sous-rubriques. Un code typique sera composé d'une lettre suivi de 3 chiffres, avec un point « . » précédant le dernier chiffre.

Par exemple :

IX Maladies de l'appareil circulatoire

I20-I25 Cardiopathies ischémiques

I20 Angine de poitrine

I20.0 Angine de poitrine instable

I20.1 Angine de poitrine avec spasme coronaire vérifié

I20.8 Autres formes d'angine de poitrine

I20.9 Angine de poitrine, sans précision

Cette structure se prête à des déclinaisons nationales spécifiques, comportant l'ajout de codes fils aux codes terminaux. Ainsi en France a-t-il été décidé, pour les besoins du recueil du PMSI, de rajouter un degré de précision au code I20.0, avec le code :

I20.0+0 Angine de poitrine [angor] instable avec élévation des marqueurs biochimiques [enzymes] myocardiques

Hormis ces trois chapitres, la classification elle-même se fait selon une logique anatomico-fonctionnelle. On peut voir que les chapitres concernent à la fois des processus physiopathologiques (infections, tumeurs) des appareils ou des organes pour 15 chapitres, puis des symptômes et des lésions traumatiques.

On retrouve cette logique mixte, avec parfois un aspect prédominant sur l'autre, à l'intérieur des chapitres : le chapitre sur les maladies de l'appareil respiratoire comporte par exemple un bloc « Grippe et pneumopathie » et un autre « Maladies chroniques des voies respiratoires inférieures », tandis que celui sur les maladies de l'appareil digestif se décline entièrement en blocs basés sur l'anatomie.

La CIM comporte un mécanisme de prise en compte de ses propres limites et de celles du codeur, avec l'utilisation systématique, à tous les niveaux de la hiérarchie, de blocs, de rubriques et de codes comportant dans leur libellé la mention « autre » (les codes dits « en .8 ») ou encore « sans précision » (les codes dits « en .9 »), ainsi que l'on peut le voir dans l'exemple donné plus haut. Ceci permet de rendre compte de formes pathologiques appartenant à un chapitre, un bloc, une rubrique, mais sans libellé correspondant, ou de coder une pathologie dont on ne connaît pas précisément la forme, mais dont on sait qu'elle relève d'une rubrique donnée.

2.2.2. Codage des Actes

Le codage des actes médico-chirurgicaux est l'autre grand chapitre du recueil des données du PMSI, avec un impact majeur sur le groupage et la valorisation des séjours, sans compter la justification d'activité auprès des tutelles (actes de chirurgie cardiaque, par exemple).

Cotation et codage

Jusqu'à l'apparition du PMSI, la question du relevé des actes par nature ne s'était pas posée. Il n'existait qu'un relevé de l'activité liée à la réalisation des actes, défini par la Nomenclature Générale des Actes Professionnels des Médecins, des Chirurgiens-dentistes, des Sages-femmes et des auxiliaires médicaux (NGAP). Cette nomenclature de l'Assurance maladie avait été promulguée par arrêté en 1972[54], et depuis régulièrement remise à jour.

L'objet de la NGAP est de coter, et non coder, l'activité : à chaque acte est associé une lettre-clé et une quantité, à chaque lettre-clé correspond une valeur, le produit de la quantité par la valeur de la lettre-clé donne alors le montant payé au réalisateur de l'acte.

Par exemple, en 2005, la lettre-clé KCC vaut 2,09 euros. Une appendicectomie est cotée KCC 70, elle est donc payée 146,30 euros par l'Assurance Maladie au chirurgien. Sur la base de ce mode de détermination du montant des actes, une série d'autres dispositions venaient moduler cette somme : association d'actes, délai ou condition de réalisation, etc.

Ce système était entièrement tourné vers des modalités de valorisation les plus souples possibles en jouant sur la valeur de la lettre-clé, de la cotation, des conditions complémentaires, etc. Il présentait cependant un défaut majeur, celui de masquer complètement l'activité. En effet, l'appendicectomie est cotée KCC 70, au même titre que l'arthrodèse des trois articulations d'un doigt. Dès lors, sachant qu'il s'agit de la seule information portée par la facture, l'information sur la nature de l'acte est perdue et l'Assurance Maladie paie sans savoir ce qui a été réalisé.

Ne serait-ce que de ce point de vue, la mise en place du PMSI était, au sens littéral du terme, révolutionnaire, puisqu'elle plaçait le recueil de l'activité au centre du processus, sa facturation ne venant qu'ensuite, comme produit d'aval.

Le Catalogue des Actes Médicaux (CDAM)

Lors de l'expérimentation de 1982, aucune nomenclature française des actes n'existe. Jean de Kervasdoué prescrit l'emploi de la classification des actes de l'OMS. Celle-ci avait été publiée en 1978, avant d'être abandonnée en 1989 en raison des difficultés à la maintenir au regard de l'évolution technique[55].

C'est le Catalogue des Actes Médicaux qui l'a remplacée en 1986. Cette classification présentait l'avantage d'être élaborée selon une méthodologie propre reposant sur six principes[56] :

- exhaustivité de l'activité décrite ;
- identification de l'acte par un code ;
- terminologie utilisée par les équipes médicales ;
- mise à jour annuelle ;
- application limitée aux établissements de soins concernés par la dotation hospitalière de financement ;
- liberté laissée aux experts pour le niveau de précision.

Le CDAM est structuré en cinq champs couvrant chacun un domaine des actes :

- Alpha : actes diagnostiques et thérapeutiques
- Bêta : actes d'anesthésiologie
- Gamma : actes d'imagerie médicale
- Mû : actes de radiothérapie
- Oméga : actes de réanimation

Enfin, un index de complexité relative (ICR) est créé, supposé rendre compte de la complexité des actes. Cet indice permet de quantifier de façon standardisée l'intensité de l'activité en termes d'actes médicaux.

La Classification Commune des Actes Médicaux (CCAM)

La CCAM a été pensée pour faire suite au CDAM, et surtout remplacer à la fois le CDAM et la NGAP, donc unifier le codage des actes pour le PMSI et la facturation de l'activité.

D'emblée, des principes forts ont été adoptés dans la perspective d'une exploitation par les systèmes d'information. Les codes alphanumériques structurés comportent une racine de 7

caractères, dont les 4 premiers sont des lettres à valeur sémantique (les deux premiers digits sont consacrés à la topographie, le troisième à l'action, le quatrième à l'accès), déterminant ainsi non seulement un répertoire de codes décrivant les actes existant, mais fixant également des règles pour l'introduction de nouveaux codes. Enfin, l'ICR est conservé.

Les libellés associés aux codes sont également structurés selon des règles syntaxiques permettant une interprétation sans ambiguïté du texte. Des informations complémentaires sont ajoutées : notes concernant les conditions d'usage, notes concernant les conditions de facturation.

Introduite dès 2002 de façon facultative, la CCAM remplace définitivement le CDAM en 2005.

En 2015, l'ATIH met en place la « CCAM descriptive », une para-CCAM qui, au motif de rendre compte plus finement des actes, superpose aux codes de la CCAM diffusée par l'Assurance Maladie, des codes réputés décrire plus finement certains actes. Un exemple :

L'acte CCAM :

AHLA002 Implantation d'électrode de stimulation du nerf phrénique par thoracotomie, avec pose d'un stimulateur externe

est scindé dans la CCAM descriptive en trois actes :

AHLA002-01 Implantation d'électrode de stimulation du nerf phrénique par thoracotomie, avec pose d'un générateur externe

AHLC802-01 Implantation d'électrode de stimulation du nerf phrénique par thoracoscopie, ou par thoracotomie vidéoassistée, avec pose d'un générateur externe

AHLA802-02 Implantation d'électrode de stimulation du nerf phrénique par coelioscopie ou par laparotomie, avec pose d'un générateur externe

Avec la contrainte de recueillir dans les données PMSI les actes descriptifs, mais de transmettre dans la chaîne de facturation les actes Assurance Maladie, c'est une double contrainte qui est imposée aux établissements : à la fois celle de faire évoluer leurs systèmes d'information, dans un contexte où la relation avec les éditeurs de logiciels est toujours très difficile, et celle de veiller à la double maintenance des nomenclatures. Il paraît difficile d'imaginer la pertinence et l'efficacité d'une telle mesure.

2.2.3. L'échelle des GHM et des GHS

Au cœur du dispositif, l'échelle des GHM décrit la variable censée synthétiser les éléments du séjour en une information ultime. Ainsi, la distribution de l'effectif des séjours dans les GHM, appelé l'éventail des cas, ou selon l'anglicisme passé dans le langage courant, le « case-mix » est censé représenter le profil d'activité d'un établissement.

La version 0 des Groupes Homogènes de Malades est diffusée en 1986[21] et comporte 471 GHM. Cette liste va régulièrement s'enrichir de nouveaux GHM. Une étape importante est

l'introduction des GHM décrivant les hospitalisations de jour (CM24 : catégorie majeure 24, pour 24h00), dès la version 1. Par la suite, on relève essentiellement la version 4, publiée fin 1996 pour application en 1997, qui consacre l'utilisation de la CIM 10 et n'accepte plus de codes de la CIM 9. Cette période avait d'ailleurs été l'occasion de travaux sur une méthode de comparaison des case mix entre 1996 et 1997[57].

La version 7 (2002), est la première version intégrant la CCAM. C'est en 2004, avec la version 9, qu'apparaissent les GHS, porteurs du tarif. Cette version voit également le codage des GHM entièrement modifié, avec l'introduction de codes structurés comportant la catégorie majeure diagnostique (CMD) d'appartenance du GHM (système nerveux, système digestif, etc.), le type médical ou chirurgical, et le niveau de sévérité, de trois niveaux à ce moment : sans comorbidité associé, avec comorbidité, avec comorbidité sévère. A ce moment-là, le nombre de GHM est passé de 471 à 693.

Enfin, une évolution majeure se produit en 2009 avec l'introduction de la version 11 : un codage structuré entièrement refondu comportant 4 niveaux de sévérité ainsi que la disparition de la CM 24, dont les GHM sont reventilés dans les différentes CMD d'origine, avec un dernier digit en J permettant de les identifier. Ainsi, le code d'un GHM est constitué d'une racine, les 5 premiers digits, et d'un digit indiquant la sévérité (ou, selon les cas, la prise en charge en ambulatoire ou sur une très courte durée). Les GHM ne comportant pas de niveau de sévérité, dits non segmentés, se terminent par la lettre Z.

La réaffectation des GHM d'hospitalisation de jour à leur CMD d'appartenance est un exemple de l'importance des choix faits en termes de représentation des données. En effet, il était jusque-là très difficile de réaliser simplement des tableaux de bords par spécialité (identifiée à la catégorie majeure) puisque la structure même du code GHM ne permettait pas de déterminer de quel appareil ou organe dépendait un GHM de la CM 24. Il fallait pour cela étudier les libellés et constituer manuellement un tableau d'association. La version 11 règle ce problème.

L'ATIH fournit et maintient une méta-classification multiaxiale des GHM. Un des axes regroupe les GHM en groupe d'activité, groupe de planification et discipline d'activité. Ceci permet d'effectuer des regroupements lors des analyses d'activité.

Un autre axe est la qualification des GHM selon leur Catégorie d'Activité de Soin (CAS). Identifiant initialement l'Activité de Soin (ASO) du GHM en M, C ou O (médecine, chirurgie ou obstétrique), cet axe a été modifié sans autre forme de procès en 2012 lors du passage à la version 11d des GHM. Dans la foulée, la classification en groupes d'activité, groupes de planification et discipline d'activité était entièrement modifiée avec parfois réutilisation des mêmes codes pour désigner des entités différentes...

Sans compatibilité descendante avec les versions antérieures, ceci rendait du même coup inutilisables les données antérieures. Fort heureusement, la classification est restée stable depuis. Il existe aujourd'hui 6 CAS : chirurgie, médecine, obstétrique, interventionnel, nouveau-nés, séances.

2.2.4. Autres classifications d'intérêt

Dans la mesure où toutes les variables qualitatives des RSA sont codées, on pourrait étendre cette liste à toute une série de micro-classifications, parmi lesquelles deux en particulier doivent être citées :

- la classification des modes d'entrée et de sortie, qui présente l'intérêt de porter sur deux variables à la fois. Par exemple, le codage de la variable « mode d'entrée » par la valeur 7 indiquant un transfert en provenance d'un autre établissement, il nécessite de préciser la nature de cet établissement dans la variable « provenance » (1 : établissement de court séjour, 2 : établissement de moyen séjour, etc.).

Le codage du type d'unité médicale fréquentée met en évidence de façon éclatante les questions liées à l'évolution des référentiels. Introduite en 2005, cette variable avait pour objectif de repérer les unités médicales effectuant une activité particulière (Réanimation, Soins Intensifs, Néonatalogie... 16 activités identifiées). Par la suite le panel des activités repérées s'est progressivement étendu, jusqu'à comporter aujourd'hui 45 activités identifiées, décrites par un système de codes entièrement refondu fin 2013.

2.2.5. La pointe de l'iceberg : le numéro FINESS

Le numéro FINESS (Fichier National des Etablissements Sanitaires et Sociaux) est le numéro d'identification des établissements de santé. Il s'agit d'un code numérique à 9 caractères dont les deux premiers (trois pour l'outre-mer) indiquent le département. Le numéro FINESS est une information précieuse qui permet de faire le lien vers le domaine des établissements, qui est décrit par deux bases de données :

La base de données FINESS

Le numéro FINESS est créé en 1979, dans le cadre plus général de la constitution d'un répertoire des établissements. Les modalités de constitution et de fonctionnement de ce répertoire seront ensuite régulièrement actualisées par des textes, dont le plus récent est l'arrêté du 13 novembre 2013 relatif à la mise en place d'un répertoire national des établissements sanitaires et sociaux[58]. Le répertoire FINESS est maintenu par la DREES (Direction des Etudes, des Evaluations et de la Statistique) et peut être interrogé en ligne[59].

Le grand intérêt de cette base de données consultable en ligne mais également téléchargeable sur la plateforme française d'ouverture des données publiques[60], est de répertorier non seulement les établissements mais également leurs types d'activité, leurs types d'autorisations ou encore leurs équipements lourds.

La base de données de la SAE

La Statistique annuelle des établissements de santé (SAE) est une enquête administrative exhaustive obligatoire, réalisée chaque année par la DREES auprès de tous les établissements de santé de France, pour recueillir des informations sur leur activité, leurs capacités, leurs équipements, et leurs personnels médicaux et non-médicaux. L'enquête de SAE a été lancée pour la première fois en 1995, sur les données de 1994 et fait l'objet d'un arrêté[61].

Outre une série d'informations similaires à celle de la base FINESS (autorisations, matériels), cette enquête a pour objectif de recueillir des informations beaucoup plus détaillées sur les moyens et l'activité. Nombre de salles de bloc opératoire, type de réanimation, nombre de séjours réalisés en unité de chirurgie ambulatoire, ou encore nombre d'équivalent temps plein de praticiens hospitaliers, ne sont que quelques-unes des informations disponibles dans cette base qui peut être consultée en ligne[62] et téléchargée.

Malheureusement, aucun contrôle qualité n'est exercé sur cette base, ce qui en rend les données non fiables au-delà de quelques fondamentaux qui restent bien utiles, par exemple présence d'un bloc opératoire ou d'une réanimation. Cette base doit donc être exploitée avec prudence, dans un périmètre bien défini.

Problèmes liés au numéro FINESS

Toutes les informations précédemment décrites sont liées au numéro FINESS des établissements et on comprend donc l'importance de cette donnée. Or, derrière la simplicité et l'évidence de la définition du numéro FINESS se trouve une grande complexité.

En premier lieu, il faut distinguer entre le FINESS dit juridique, identifiant la personne morale, et le FINESS dit établissement ou géographique, qui identifie le ou les sites géographiques où s'effectue l'activité.

A partir de là vont s'entrecroiser une série de problématiques conceptuelles et techniques qui rendent pratiquement inutilisable cette information pourtant indispensable :

- La base données PMSI comporte pour les établissements publics, le numéro FINESS juridique, et pour les établissements privés le numéro FINESS géographique. Il n'est donc pas possible de considérer une entité juridique privée dans son ensemble ;
- Aucune information d'historique n'est disponible concernant les numéros FINESS ou les établissements. Or, le paysage sanitaire est en perpétuel mouvement : des établissements peuvent ouvrir, fermer, fusionner ;
- Aucune vérification de l'intégrité des données n'est réalisée au niveau des données PMSI. A titre d'exemple, à partir de début 2013, l'Institut de Cancérologie de Montpellier, de FINESS juridique 340780493 et de FINESS géographique 340000207, transmet dans ses données son numéro FINESS géographique.

Enfin, nous n'avons pu trouver de texte de référence définissant la notion d'établissement géographique. On trouve sur les sites des ARS le texte suivant sur une page consacrée au répertoire FINESS (voir par exemple sur le site de l'ARS Limousin[63]) :

« La notion d'établissement retenue résulte du croisement de deux critères :

Le critère géographique : est un établissement toute unité géographiquement distincte d'une autre. Ainsi un Centre Hospitalier Régional comprend autant d'établissements que d'implantations géographiques différentes. Un établissement principal et son annexe située à 1 km constituent deux établissements.

Le critère budgétaire : pour une même implantation géographique, on distingue autant d'établissements qu'il y a de budgets annexes (sanitaire) ou distincts (social). Ainsi un centre d'Aide par le Travail et son foyer d'hébergement constituent deux établissements même s'ils sont implantés à la même adresse. »

On trouve également dans une instruction[64] concernant les Etablissements d'Hébergement des Personnes Agées Dépendantes (EHPAD) l'indication suivante :

« 8 - La notion d'établissement secondaire. Cette notion a été initialement utilisée pour permettre l'immatriculation de toutes les implantations géographiques (établissements et ses annexes) sans entraîner dans FINESS une surestimation du nombre d'établissements par rapport à la réalité des structures.

La notion d'établissement secondaire, établissement à part entière doté de son propre n° FINESS, qualifie les établissements ayant une adresse propre mais pas une autonomie de gestion. Ces derniers sont obligatoirement rattachés à un établissement principal et un seul.

Dans le cas où un établissement secondaire est identifié, une vérification sur la véracité du numéro FINESS de l'établissement principal renseigné sera à opérer. »

Ces éléments paraissent clairs, et en tout cas à même de permettre de déterminer si un lieu de délivrance de soins est un établissement géographique ou non. Cependant une difficulté surgit rapidement avec la notion « d'unité géographique » d'une part, avec celle de « distincte d'une autre » d'autre part, qui s'avèrent être assez difficiles à apprécier au simple vu de données disponibles dans une base, sans se rendre sur place pour constater la réalité des choses. Quant au critère budgétaire, on voit qu'il est implicitement lié à la notion d'adresse, qui elle-même n'est pas simple (adresse postale, adresse physique) ni forcément fiable (un même site physique de taille importante et attenant à plusieurs rues différentes peut avoir plusieurs adresses).

A l'usage, tenter de mettre en correspondance les données de la base PMSI avec celle de la SAE s'avère extrêmement complexe et risqué. Le simple rapprochement entre les numéros FINESS des deux bases est lourd, demandant plusieurs opérations suivies d'une vérification manuelle de cas particuliers repérable simplement, rien ne garantissant que d'autres erreurs moins repérables ne passent pas inaperçues.

3. Stratégie de gestion des référentiels

Nous avons mis en place autour des données du PMSI un cadre tentant de maîtriser les principaux aspects des problèmes liés aux nomenclatures, et malheureusement non gérés car mal identifiés par les autorités de tutelle.

3.1. Diagnostics et Actes

Nous maintenons une table des diagnostics et des actes.

La table des diagnostics est issue de l'exploitation des tables de la fonction groupage ; des dates de validité pour les codes sont maintenues.

La table des actes est issue de l'exploitation des fichiers de la CCAM mis à disposition par l'Assurance maladie, avec un suivi si possible des dates de création et de suppression des actes.

De nombreuses tables de référence sont également disponibles selon des thèmes définis : table de référence des actes de cardiologie interventionnelle tels que définis par la circulaire du 12 août 2009[65], table de référence des actes de neurochirurgie (arrêté du 8 juin 2005 sur les objectifs quantifiés d'offre de soins[6]), table de référence des actes d'accouchement (élaborée par nous-même), ou encore diverses tables de référence de diagnostics ou d'actes tirées de l'exploitation des sources de la fonction groupage.

3.2.Les GHM

Nous avons implémenté plusieurs tables autour de la dimension GHM :

- une table des GHM comportant les associations GHM - GHS pour chaque version depuis la version 9, avec les tarifs, bornes basses, bornes hautes, etc. Cette table ne comporte ces informations que pour les établissements publics. Elle est maintenue par l'ajout d'enregistrements à chaque nouvelle version de la classification, après vérification de l'intégrité des libellés. En effet, nous avons eu à plusieurs reprises la mauvaise surprise de constater des variations dans le libellé des GHM (ajout d'une virgule, d'un espace...), ceci créant des doublons et par suite des risques de produit cartésien dans les requêtes. Les données sont issues des publications officielles annuelles (arrêté prestation) et de la mise en ligne d'informations par l'ATIH qui les accompagne ;
- une table des GHS, comportant la liste des GHS avec leurs tarifs et leurs modulateurs tarifaires depuis la version 9. Ces informations sont données pour les établissements publics et privés, avec des dates de début et de fin de validité. Cette table est maintenue en y intégrant chaque année les données publiées par l'Assurance Maladie sur son site[66] ;
- une table de regroupement des GHM, comportant les regroupements de GHM de l'ATIH. Nous avons en 2012 réalisé un transcodage rétrospectif des versions antérieures dans la nouvelle classification, ce qui fait que nous disposons d'une compatibilité descendante ;
- une table comportant le tarif des suppléments par période ;
- quelques tables de référence propres à des activités spécifiques (chirurgie cardiaque, séances...).

3.3.Les établissements

Nous avons constitué une table comportant l'ensemble des établissements présents dans la base nationale PMSI, en associant par un numéro FINESS de rattachement les établissements dont nous avons pu déterminer qu'ils étaient identiques, mais portant des numéros FINESS différents d'une année sur l'autre. Elle comporte également des dates donnant la période d'activité d'un numéro FINESS.

Cette table est maintenue par comparaison avec les données de la nouvelle année à intégrer à la base. Les numéros qui disparaissent comme ceux qui apparaissent sont vérifiés en utilisant la base de la SAE et l'internet. Lorsque cela est nécessaire, des rapprochements sont faits.

Chapitre 2 Consolidation et exploitation du chaînage

1.1.Introduction

Rappel sur le procédé technique actuel

Actuellement, l'identifiant anonyme (IdPat) est créé en 2 étapes :

- Au niveau de l'établissement

Le numéro de sécurité sociale, la date de naissance et le sexe du patient sont combinés à une clé secrète et hashés. Un premier numéro anonyme est ainsi généré.

Lors de cette étape, des contrôles également appliqués sur les données fournies en entrée, ainsi que sur le numéro de séjour et le rapprochement entre les numéros anonymes générés et les RSS.

Enfin, un « numéro de séjour » est généré, qui permet de calculer par la suite les délais inter-séjours. Ce numéro de séjour est élaboré de la façon suivante : pour chaque identifiant, une date de référence est générée. L'intervalle écoulé entre cette date et la date d'entrée du séjour est calculé, c'est ce chiffre qui est retenu comme numéro de séjour.

Quel que soit l'établissement, chaque fois que le hashage des traits identifiants produira cet identifiant, la même date de référence sera choisie. Il ainsi est possible d'ordonner les séjours chronologiquement par ordre croissant de numéro de séjour.

- Au niveau de l'ATIH

Le premier numéro anonyme est combiné avec une seconde clé secrète et l'ensemble est à nouveau hashé. On obtient ainsi un numéro de très haut niveau de sécurité. Le résultat des tests effectués sur les identifiants est conservé. Ainsi, les identifiants sont dotés d'une série de codes retour résultant de ces tests, et permettant d'éliminer les identifiants présentant des erreurs lors de leur utilisation.

L'emploi des données chaînées dans la recherche

Des expériences de chaînage des données PMSI ont été menées avant la mise en place systématisée de l'identifiant de chaînage. Ainsi, en 2003, dans une étude sur un observatoire du cancer colorectal [67], les données PMSI de 1996 à 2006 de quatre établissements étaient chaînées par hashage de traits identifiants des patients. Il s'agissait d'estimer la capacité des données du PMSI à fournir une bonne estimation du nombre de cas incidents de cancer.

Par la suite, les études utilisant l'identifiant anonyme de patient ne vont se développer que vers la fin des années 2000 et prendre plus d'importance dans les années 2010.

Nous avons effectué une recherche bibliographique sur PubMed portant sur les termes :

(PMSI ET (France OU French)) OU (French ET administrative ET database) ;

complétée par

(PMSI ET linkage) OU (French ET claims ET database ET linkage) OU (French ET hospital ET database ET linkage).

Cette recherche résulte en un ensemble de 260 articles

Chaque article a ensuite été analysé pour déterminer si le chaînage des données PMSI était un élément significatif de l'article. Nous obtenons ainsi 46 articles couvrant une période allant de 2003 à 2016.

En découpant la période 2000 à 2016 par tranches de 4 ans, nous observons la répartition suivante :

Année	Nb articles
2000 - 2004	2
2005 - 2008	3
2009 - 2012	14
2013 - 2016	27

Le sujet le plus abordé est de loin le cancer, avec 14 articles sur 41 portant sur ce thème, essentiellement pour des estimations d'incidence en conjugaison avec les données de registres[3,68-70]. Les données chaînées sont assez souvent évaluées par comparaison avec le gold-standard représenté par les registres.

Les analyses de la qualité du chaînage

Nous avons retrouvé un article s'intéressant à la qualité du chaînage, en termes d'analyse des identifiants en erreur[71]. Dans les autres cas, les précautions méthodologiques dans l'emploi des données chaînées consistent à éliminer les identifiants comportant des erreurs. Parmi, la littérature non indexée, nous avons retrouvé une étude de la DREES sur le chaînage et sa qualité [72], et analysant les grandes caractéristiques des séjours chaînés. Là encore, la méthode consiste à éliminer les identifiants en erreurs.

Cette étude de la littérature nous amène à deux constats :

- La qualité des identifiants de chaînage n'a pas été étudiée au-delà des erreurs lors de leur génération.
- La qualité des identifiants de chaînage est confondue avec la qualité du chaînage ; il faudrait plus exactement parler de l'exploitabilité du chaînage.

C'est pourquoi nous proposons dans ce chapitre une méthode d'évaluation de la fiabilité de l'identifiant du chaînage. Une stratégie d'exploitation du chaînage et quelques applications sont ensuite données.

1.2. Matériel et méthode

1.2.1. Les tests

Pour évaluer la fiabilité de l'identifiant de chaînage, nous avons défini des tests de concordance des RSA entre eux.

Il est possible d'utiliser les informations des RSA pour vérifier la cohérence de trois types d'informations :

- cohérence des informations de base : le sexe et l'âge devraient être concordants au fil des hospitalisations ;
- cohérence chronologique : le numéro de séjour (tel que défini plus haut) et la durée de séjour devraient être cohérents avec les mois et années de sortie des séjours ;
- cohérence médicale : si un certain nombre de tests portant sur le contenu du RSS sont effectués au niveau de l'établissement (cohérence entre sexe et acte chirurgical par exemple), il est également possible de vérifier la concordance médicale des données entre séjour.

Nous définissons ainsi 5 tests :

1. Identifiant sans erreur.

2. Cohérence du sexe.

Le sexe doit rester le même sur toute la série de RSA associés à l'identifiant de chaînage.

3. Cohérence de l'âge.

L'âge doit rester cohérent d'un RSA à l'autre : par exemple, les âges de deux RSA successifs au cours de la même année ne devraient pas différer de plus d'un an (anniversaire survenu entre les deux hospitalisations).

Cette règle se formalise de la façon suivante :

Etant donné un séjour S_i , avec l'âge à l'entrée A_i et l'année de sortie Y_i , suivi d'un séjour S_j , avec âge à l'entrée A_j et année de sortie Y_j , la condition :

$$(A_j - A_i) \leq (Y_j - Y_i) + 1$$

devrait être remplie

Dans certains cas très particuliers, on peut observer de façon naturelle une différence d'âge de deux ans entre deux séjours avec la même année de sortie :

- le premier séjour débute l'année N *avant* l'anniversaire (âge à l'entrée = A_i) et se termine l'année suivante $Y_i = N+1$;

- le séjour suivant se termine l'année $Y_j = N+k$, mais il débute *après* l'anniversaire du patient. L'âge à l'entrée est alors $A_j = A_i + (Y_j - Y_i) + 2$ car deux anniversaires se sont écoulés entre les deux entrées.

On aura alors $(A_j - A_i) > (Y_j - Y_i) + 1$, et nous repérerons alors à tort ces séjours comme discordants. La figure 3 illustre ce cas de figure avec l'exemple de deux années consécutives. Compte tenu des conditions très particulières nécessaires à la survenue de ce cas de figure, nous considérons qu'il est préférable de conserver la condition telle que définie plus haut.

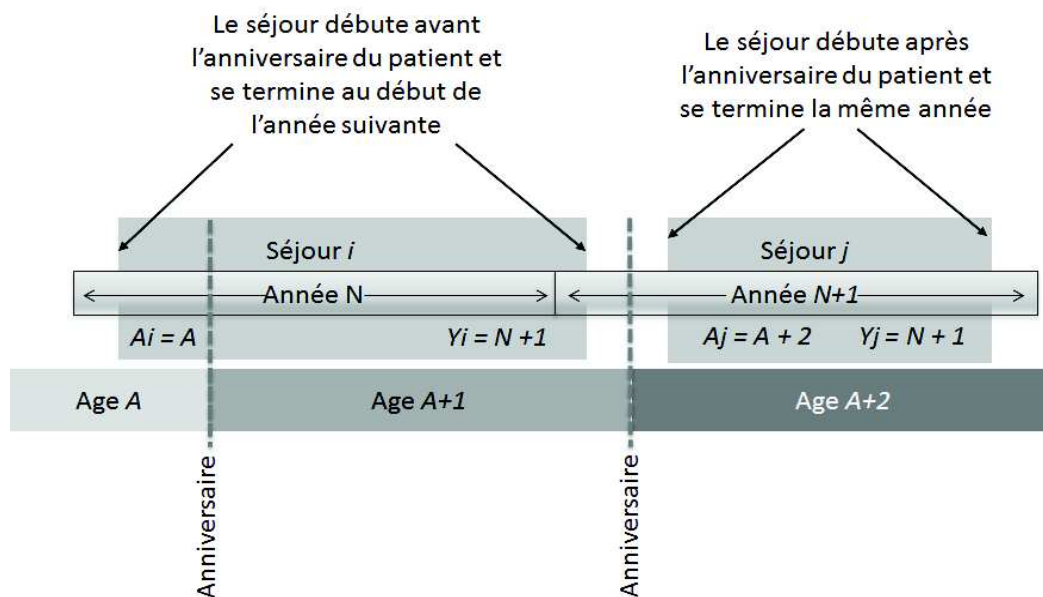


Figure 3 - Cas particulier de repérage à tort d'une incohérence d'âge

4. Cohérence de la chronologie.

Compte tenu des propriétés du numéro de séjour décrites plus haut, on devrait avoir les relations suivantes :

- un séjour postérieur à un autre relativement à son numéro de séjour ne doit pas avoir une année et un mois de sortie antérieurs au séjour précédent ;
- un séjour peut être contenu à l'intérieur d'un autre (prestation inter-établissement avec production de deux séjours différents), mais les deux séjours ne doivent pas se chevaucher : le séjour 2 peut démarrer pendant le séjour 1, mais il doit se terminer au plus au même moment que le séjour 1. De plus on considère que les deux séjours ne peuvent être de la même durée (transfert vers un autre établissement le jour même de l'entrée et retour de cet établissement le jour même de la sortie).

Ces contraintes peuvent se formaliser de la façon suivante :

Soit deux séjours successifs S_i et S_j :

S_i , de numéro de séjour N_i , de durée D_i , de mois et année de sortie M_i et Y_i ;

S_j , de numéro de séjour N_j , de durée D_j , de mois et année de sortie M_j et Y_j .

La condition suivante doit alors être respectée :

$$\left[(N_i + D_i) \leq N_j \text{ ET } N_j \leq \left(N_i + D_i + \left((Y_j - Y_i) \times 12 + (M_j - M_i) \right) \times 31 + (31 - D_j) \right) \right]$$

OU

$$\left[N_i \leq N_j \leq (N_i + D_i) \text{ ET } (N_j + D_j) \leq (N_i + D_i) \text{ ET } D_j < D_i \right]$$

La première partie de la condition décrit la succession de deux séjours, la seconde un séjour en contenant un autre.

Détaillons la première partie :

$(N_i + D_i) \leq N_j$ signifie que le séjour suivant commence au plus tôt le jour de la sortie du précédent.

Cette condition donne une borne inférieure pour N_j , il 'agit ensuite de déterminer une borne supérieure.

$N_i + D_i + ((Y_j - Y_i) \times 12 + (M_j - M_i)) \times 31 + (31 - D_j)$ est une approximation par excès du délai maximal possible entre deux séjours compte tenu de leur année et mois de sortie respectifs :

(1) $((Y_j - Y_i) \times 12 + (M_j - M_i)) \times 31$ représente le nombre de jours (par excès, en affectant 31 jours à chaque mois) entre le début du mois de sortie du premier séjour et le début du mois de sortie du séjour suivant :

Si $Y_j = Y_i$, c'est évident

Si $Y_j > Y_i$ et $M_j > M_i$, idem : $(Y_j - Y_i) \times 12$ représente le nombre de mois d'une année à l'autre, auquel on ajoute la différence entre M_j et M_i .

Par exemple, si on a $Y_i = 2010$, $M_i = 4$ et $Y_j = 2012$, $M_j = 6$, on voit que le nombre de mois est de $(2012 - 2010) \times 12 + (6 - 4) = 24$; c'est-à-dire que 24 mois se sont écoulés entre le début d'avril 2010 et le début de juin 2012.

Si $Y_j > Y_i$ et $M_j < M_i$, le nombre de mois écoulés sera

$$(Y_j - Y_i) \times 12 - (M_i - M_j) = (Y_j - Y_i) \times 12 + (M_j - M_i)$$

Par exemple, si on a $Y_i = 2010$, $M_i = 9$ et $Y_j = 2012$, $M_j = 6$, on voit que le nombre de mois est de $(2012 - 2010) \times 12 + (6 - 9) = 24 - 3 = 21$.

(2) En supposant que la sortie de S_j se soit faite le dernier jour de M_j , l'entrée de S_j se sera donc fait au plus tard au jour $(31 - D_j)$ de M_j .

(3) en supposant que la sortie du premier séjour se soit faite le premier jour de M_i , le nombre maximal de jours (par excès) séparant les deux entrées sera donc :

$$N_i + D_i + ((Y_j - Y_i) \times 12 + (M_j - M_i)) \times 31 + (31 - D_j)$$

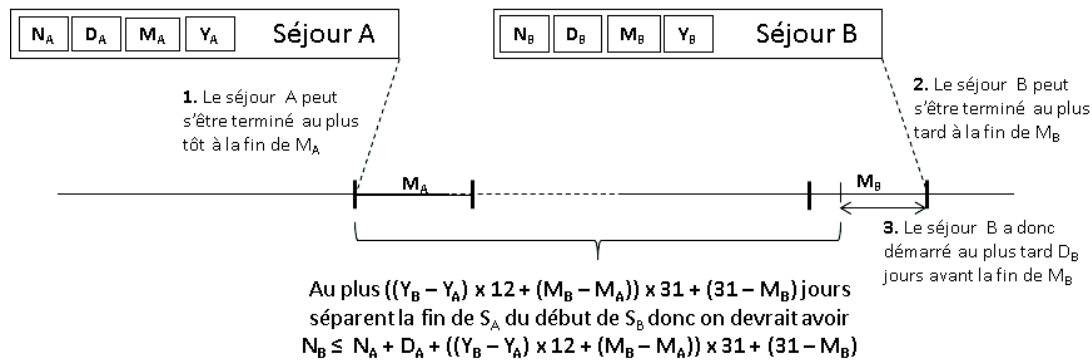


Figure 4 - Estimation du délai maximum entre 2 séjours

La seconde partie de la condition de respect de la cohérence chronologique est plus simple :

$N_i \leq N_j \leq (N_i + D_i)$ indique que le second séjour commence pendant le premier, au plus tôt le même jour, au plus tard le dernier.

$(N_j + D_j) \leq (N_i + D_i)$ indique que le second séjour doit finir au plus tard en même temps que le premier

$D_j < D_i$ indique que le premier doit obligatoirement être plus long que le second.

Le second séjour peut donc commencer ou finir en même temps que le premier, mais pas les deux.

5. Cohérence médicale.

Compte tenu de l'information médicale contenue dans les RSA, nous pouvons établir les contrôles suivants :

- le patient ne peut être indiqué comme étant décédé qu'une seule fois, à la fin de la série ;
- on considère qu'une femme ne peut accoucher plus de deux fois sur une période de 1 an ;
- on établit une limite de 6 au nombre d'interruption de grossesses possibles par an.

D'autres contrôles pourraient être imaginés, par exemple le nombre d'interventions en considérant la parité des organes (par exemple, pas plus de deux interventions pour cataracte dans la série des RSA), mais nous considérons que la qualité du codage ne permet pas encore d'aller jusqu'à ce type de contrôles plus poussés.

1.2.2. Stratégie d'application des tests et algorithme général

L'objectif étant d'apprécier si l'identifiant anonyme associé à une série de RSA est bien propre à un patient et un seul dans cette série, nous appliquons les tests dans l'ordre où ils ont été énoncés.

En effet, lorsqu'un test est positif, c'est-à-dire qu'une incohérence est détectée, il peut s'agir d'un vrai positif : l'erreur détectée reflète bien la réalité, et il ne s'agit effectivement pas du même patient ; ou d'un faux positif : l'erreur ne reflète pas la réalité, elle est due à une erreur humain ou technique, et on ne peut pas dire sur qu'il ne s'agit pas du même patient.

Nous avons donc listé les tests par ordre décroissant du degré de confiance que nous accordons à leur valeur prédictive positive (VPP) compte tenu de notre connaissance des systèmes d'information administratifs et cliniques.

Par exemple, les admissions se font, depuis longtemps déjà et dans la majorité des cas, en utilisant la carte Vitale, avec la plupart du temps une recherche d'antériorité. De plus, le domaine de valeurs est très limité. Du coup, une erreur humaine ou technique sur la saisie du sexe d'un patient paraît assez peu probable, et nous considérons que ce test est probablement celui dont la VPP est la plus élevée.

Vient ensuite le test sur l'âge, pour lequel on sait que les erreurs de saisie de date de naissance peuvent se produire un peu plus fréquemment. Même si cette variable est récupérée sur la carte Vitale, sa modification en une valeur erronée peut se produire dans un certain nombre de cas : recherche d'antériorité et sélection d'un dossier avec une date différente qui va écraser la valeur, patient fréquemment réadmis aux urgences dans un contexte de faible disponibilité de ses pièces d'identité et pièces justificatives d'Assurance Maladie, etc.

La chronologie pourra encore plus facilement être altérée étant donné la problématique de saisie des mouvements, assez répandue dans les établissements, et sachant que de plus une erreur de mouvement sur une date d'entrée ou de sortie a assez peu de chance d'être détectée si elle ne se manifeste pas par un problème de facturation ou de production du PMSI.

Enfin, même si nous avons restreint la cohérence médicale à des tests que nous considérons comme simples et robustes, de nombreuses circonstances peuvent amener à des saisies erronées non détectées. Par exemple, un mode de sortie décès à tort peut être indétectable si il n'engendre pas une atypie détectable ou recherchée.

Ainsi, chaque identifiant est successivement testée pour les 5 tests, et seuls les identifiants passant un test sont soumis au suivant. Ce processus définit ainsi 6 niveaux successifs de fiabilité, de 0 à 5.

Dans la mesure où les tests se basent sur la cohérence des données entre RSA, seuls les IdPat associés à une série d'au moins 2 RSA sont testés. Les IdPat ne correspondant qu'à un RSA sont affectés d'un niveau de fiabilité 5.

Enfin, le niveau de fiabilité des IdPat est initialisé à 5 avant les tests. L'algorithme général est donné sur la figure suivante.

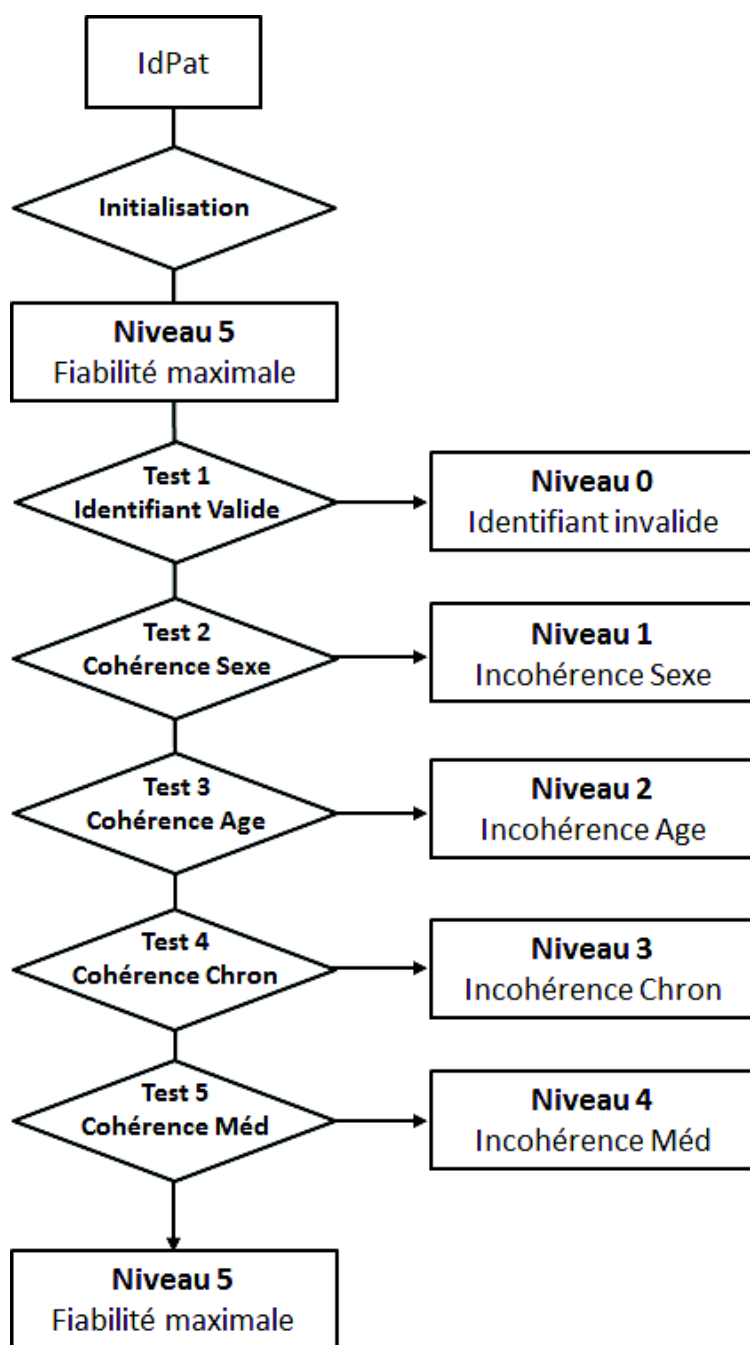


Figure 5 - Algorithme général de détermination du niveau de fiabilité de l'identifiant patient

Avant et durant la réalisation des tests, nous avons effectué les choix suivants :

- les RSA en erreur (GHM débutant par 90) sont éliminés ;
- les RSA séances ne sont pas pris en compte dans le calcul de cohérence chronologique en raison de la possibilité d'avoir des RSA calendaires, c'est-à-dire regroupant sur un enregistrement les séances effectuées sur plusieurs semaines ;
- les IdPat de niveau inférieur à 5 sont enregistrés de façon à pouvoir être étudiés.

1.2.3. Données analysées et moyens techniques

Pour ce travail, nous avons utilisé la base nationale des RSA des années 2009 à 2014, représentant environ 141 millions d'enregistrements pour 42 millions d'identifiants anonymes.

Sur la base du format des RSA, un modèle conceptuel de données est développé, consistant en une table principale comportant les RSA et une série de tables satellites comportant les diagnostics associés significatifs (DAS), les actes (ACT), les Zones d'Unité Médicale (ZUM) ainsi que les enregistrements ANO (IdPat, numéro de séjour et codes retour du traitement de l'IdPat) correspondants.

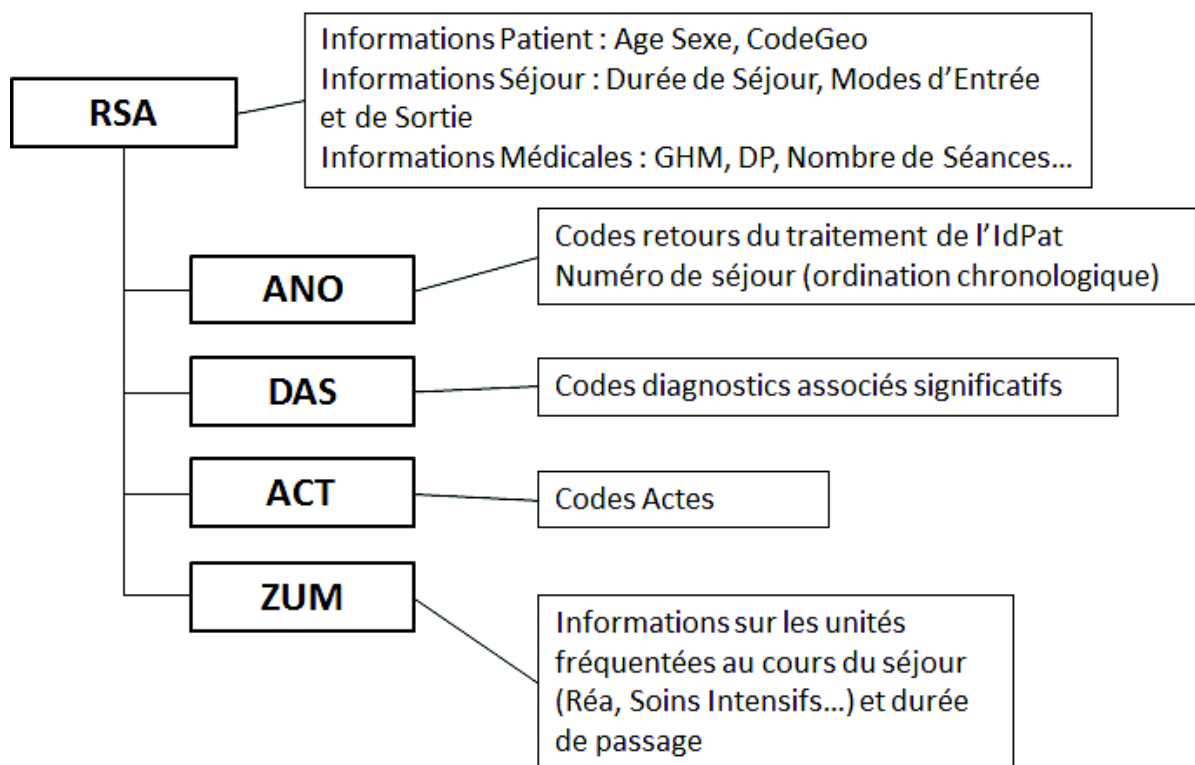


Figure 6 - Structure générale de la base – La table des RSA et les tables satellites

Les données sont chargées dans une base de données Microsoft SQL Server TM hébergée sur un serveur DELL TM PowerEdge TM T610R, équipé de 2 processeurs Intel TM Xeon 5649@2.53 GHz, doté de 32 Go de Ram et de 3 To de disque dur à 15 000 tours/min.

Les tests sont implémentés sous forme de scripts SQL, ce qui permet de tirer parti du traitement parallèle inhérent au SQL : il est possible d'appliquer les tests sur toute la base en même temps.

Le temps total de traitement avoisine les 80 heures.

1.3.Résultats

Nous obtenons les résultats suivants :

Niveau	IdPat	RSA	% IdPat	% RSA
0	281 911	2 377 093	0,67%	1,69%
1	17 134	134 392	0,04%	0,10%
2 RSA et plus	17 134	134 412	0,08%	0,11%
2	17 028	318 655	0,04%	0,23%
2 RSA et plus	17 028	318 655	0,08%	0,27%
3	133 607	1 070 898	0,32%	0,76%
2 RSA et plus	133 607	1 070 898	0,63%	0,91%
4	23 001	577 973	0,05%	0,41%
2 RSA et plus	23 001	577 973	0,11%	0,49%
5	40 667 877	136 234 194	98,88%	96,82%
2 RSA et plus	21 170 480	115 736 797	99,11%	98,22%
Total	42 140 558	140 713 205	100,00%	100,00%
2 RSA et plus *	21 361 250	117 838 715	100,00%	100,00%

Table 1 - Distribution des niveaux de fiabilité - 2009 à 2014

On constate tout premièrement que les séries comportant au moins 2 RSA représentent la moitié des IdPat (21 millions sur 42) mais bien entendu beaucoup plus d'enregistrements (118 millions sur 140).

D'autre part, la forte proportion d'IdPat de niveau 5 parmi les séries à 2 RSA ou plus est très satisfaisante et permet d'envisager une exploitation fiable de la base pour l'étude des trajectoires.

Un examen des données de niveau 1 et 2 montre que des cas de confusion mère / nouveau-né et des cas de néonatalogie sont retrouvés dans 58,6% des cas (respectivement 18 256 et 1 769 cas sur 34 162 IdPat).

L'examen du niveau 3 (incohérence de chronologie) montre des cas de chevauchement dans la majorité des cas : 128 268 (96,00%) cas sur 133 607.

Enfin, la majorité des données de niveau 4 (22 997 sur 23 001, 99,98%) est constituée par des IdPat pour lesquels le patient est indiqué comme décédé plusieurs fois ou avant le dernier RSA de la série.

1.4.Applications

1.4.1. Sélection d'un échantillon pour l'étude des trajectoires

Stratification de la population

Ce traitement nous permet d'aborder la question de la sélection d'une population représentative pour l'étude des trajectoires. Dans cet objectif, nous allons stratifier la population, dont nous excluons les séances. Au vu de ce qui précède, on peut caractériser chaque IdPat selon le nombre de RSA qui lui sont associés et son niveau de fiabilité ; nous

allons appeler « type » d'un IdPat sa caractérisation selon ces deux critères et employer la notation $R[\text{Nombre de RSA}] : L[\text{Niveau de fiabilité}]$.

Le nombre de RSA et la fiabilité pourront être exprimés selon une valeur unique, plusieurs valeurs, une étendue de valeur. L'astérisque représente toute valeur. Par exemple :

- $R^*:L^*$ Type général décrivant tous les IdPat.
- $R2:L5$ IdPat de niveau 5 associés avec 2 RSA.
- $R2,4-6:L3-5$ IdPat de niveau 3 à 5, associés avec 2 ou 4 à 6 RSA.
- $R2^*:L3-5$ IdPat de niveau 3 à 5, associés avec 2 RSA ou plus.
- $R^*5:L3-5$ IdPat de niveau 3 à 5, associés avec jusqu'à 5 RSA.

Enfin on peut omettre la partie "R" ou "L" :

- $L5$ IdPat de niveau 5, indépendamment du nombre de RSA associés.
- $R1$ IdPat associés avec un seul RSA, indépendamment du niveau de fiabilité.

Nous partageons donc la population d'IdPat et de RSA en strates selon les types d'IdPat. Ainsi, n_1 IdPat de type R1 seront associés à n_1 RSA, n_2 IdPat de type R1 seront associés à $2n_2$ RSA, et d'une manière générale n_k IdPat de type Rk seront associés à $k.n_k$ RSA. A l'intérieur d'une strate, des sous-groupes d'IdPat et de leurs RSA associés peuvent être constitués, ainsi par exemple n_{k5} IdPat de type Rk seront associés à $k.n_{k5}$ RSA.

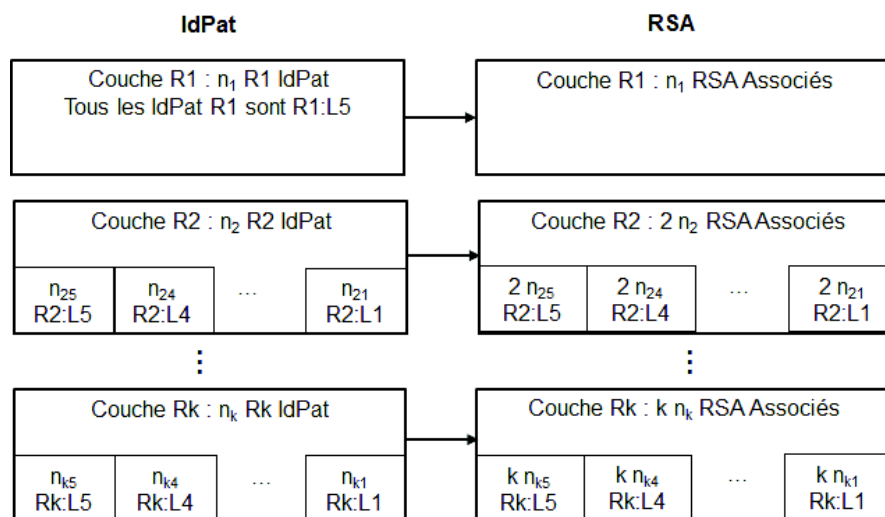


Figure 7 - Structure de la population

Stratégie de sélection des IdPat et des RSA les plus représentatifs

Dans la mesure où nous nous intéressons aux trajectoires, nous étudions les IdPat associés à 2 RSA ou plus ($R2^*$).

Etant donné la façon dont la population est structurée, la sélection des IdPat les plus fiables devra répondre au mieux aux trois critères suivants :

C1 : Le nombre total des IdPat de type L5 doit être représentatif par rapport à la totalité de la population des IdPat.

C2 : Le nombre total des RSA de type L5 doit être représentatif par rapport à la totalité de la population des RSA.

C3 : A l'intérieur de chaque strate, le nombre d'IdPat et de RSA de type L5 doit être représentatif de la totalité de la population de la strate

La table 2 page suivante permet d'évaluer ces trois critères en fonction de la distribution des données parmi les différentes strates et selon les niveaux de fiabilité.

Critère C1 : Proportion cumulée d'IdPat parmi les strates ($R2-n/R2^*$)

Critère C2 : Proportion cumulée de RSA parmi les strates ($R2-n/R2^*$)

Critère C3 : Proportion de données de type L5 dans chaque strate ($Rn:L5/Rn$)

Strate R	Proportion cumulée de données de type L5 : R2-n:L5 / R2		Proportion de données de type L5 dans la strate : Rn:L5 / Rn
	IdPat - C1	RSA - C2	C3
2	42,69%	21,93%	99,33%
3	64,45%	38,69%	99,32%
4	76,62%	51,19%	99,07%
5	83,87%	60,50%	98,99%
6	88,47%	67,59%	98,81%
7	91,51%	73,06%	98,71%
8	93,60%	77,34%	98,50%
9	95,07%	80,74%	98,41%
10	96,14%	83,48%	98,15%
11	96,92%	85,70%	98,04%
12	97,52%	87,54%	97,92%
13	97,97%	89,05%	97,67%
14	98,32%	90,31%	97,58%
15	98,60%	91,38%	97,43%
16	98,82%	92,28%	97,38%
17	99,00%	93,06%	96,98%
18	99,14%	93,73%	96,96%
19	99,26%	94,30%	96,92%
20	99,36%	94,80%	96,78%
21	99,44%	95,24%	96,79%
22	99,51%	95,63%	96,86%
23	99,57%	95,98%	96,37%
24	99,62%	96,28%	96,63%
25	99,66%	96,55%	96,28%
26	99,69%	96,79%	96,37%
27	99,72%	97,00%	95,86%
28	99,75%	97,19%	96,36%
29	99,77%	97,36%	95,66%
30	99,79%	97,51%	95,67%
31	99,81%	97,65%	95,40%
32	99,83%	97,78%	95,93%
33	99,84%	97,89%	95,02%
34	99,85%	98,00%	96,58%
35	99,86%	98,10%	95,57%
36	99,87%	98,19%	95,64%
37	99,88%	98,28%	95,18%
38	99,89%	98,36%	95,30%
39	99,90%	98,43%	95,29%
40	99,90%	98,50%	95,57%
41	99,91%	98,56%	95,84%
42	99,92%	98,63%	95,13%
43	99,92%	98,68%	93,95%
44	99,93%	98,74%	95,51%
45	99,93%	98,79%	94,26%

Table 2 - Distribution des IdPat, des RSA, des IdPat, globale et de type L5, parmi les strates

Au vu de la table 2, on pourra par exemple choisir la population R2-42:L5 comme étant représentative :

- les IdPat R2-42:L5 représentent 99,92% du total des IdPat associés à 2 RSA et plus ;
- les RSA associés représentent 98,63% du nombre total de leurs RSA associés ;
- la proportion de données de niveau de fiabilité 5 dans chaque strate est supérieure à 95%.

Cette stratégie, présentée ici pour la population générale pourra être déclinée sur des sous-populations définies selon les critères voulus, par exemple la sélection de toutes les patients présentant au cours de leur parcours une pathologie déterminée.

1.4.2. Construction d'une table des trajectoires

La détermination du niveau de fiabilité permet de créer dans la base une table des trajectoires pour chaque IdPat, en les enrichissant avec de nouvelles informations élaborées sur la base de celles déjà disponibles. Cette table comporte essentiellement les informations suivantes :

Attribut	Description
NumPat	Identifiant anonyme du patient
IdRSA	Identifiant de RSA
FINESS	Numéro FINESS de l'établissement
NumSej	Numéro de séjour (ordonnancement chronologique)
Ordre_sej	Numéro d'ordre chronologique du séjour
S	Sexe
AgeA	Age en années à l'entrée
Seances	Indicateur de séjour pour séance
DC	Indicateur de décès du patient
ACC	Indicateur de séjour pour accouchement
NbSej	Nombre de séjours de la série
Fiabilité	Niveau de fiabilité de l'identifiant
GHM	GHM du séjour
Eff	Nombre de séances si séjour pour séance, sinon 1
DP	DP du séjour
DurSejT	Durée de séjour
DurSejREA	Durée de séjour en Réanimation
DurSejSI	Durée de séjour en Soins Intensifs
MoisS	Mois de sortie
AnneeS	Année de sortie
Delai_Sej_Prec_HS	Délai depuis le précédent séjour hors séance
Ordre_Sej_Prec_HS	Numéro d'ordre du précédent séjour hors séance
Delai_Sej_Suiv_HS	Délai jusqu'au prochain séjour hors séance
Ordre_Sej_Suiv_HS	Numéro d'ordre du prochain séjour hors séance
IdRSaref	Identifiant du RSA de Référence si séjour intercurrent
DC_Sej_Suiv	Indicateur de décès du patient lors du séjour suivant

ACC : Le groupage des accouchements ayant varié d'une année à l'autre, nous nous sommes basés pour cet indicateur sur la présence dans le RSA d'un acte CCAM d'accouchement.

DurSejREA : Ce champ contient la durée de séjour en réanimation cumulée pour tout le séjour (somme des durées de passage si plusieurs passages).

DurSejSI : Idem pour les soins intensifs.

Delai_Sej_Prec_HS : Cette information est calculée pour les séries de niveau 5. On somme les délais entre séjour pour les séjours hors séance depuis le dernier séjour hors séance. Valeur par défaut : -1. En cas de séjour intercurrent (voir plus bas), valeur = -9.

Ordre_Sej_Prec_HS : Ordre chronologique du séjour hors séance précédent, pour les séries de niveau 5. Valeur par défaut : -1. En cas de séjour intercurrent valeur = -9.

Delai_Sej_Suiv_HS : Délai calculé pour les séries de niveau 5, par sommation des délai entre séjour des séjours hors séance, jusqu'au prochain séjour hors séance. Valeur par défaut : -1. En cas de séjour intercurrent valeur = -9. En cas de séjour pour séance, valeur = -99.

Ordre_Sej_Suiv_HS : Ordre chronologique du séjour hors séance suivant, pour les séries de niveau 5. Valeur par défaut : -1. En cas de séjour intercurrent, valeur = -9. En cas de séjour pour séance, valeur = -99.

IdRSAref : Lorsque le séjour est identifié comme un séjour intercurrent (voir la deuxième sous-condition dans le calcul de la cohérence chronologique), cet attribut prend comme valeur l'identifiant du RSA du séjour principal (voir figure 7 ci-dessous). Les attributs **Delai_Sej_Prec_HS**, **Ordre_Sej_Prec_HS**, **Delai_Sej_Prec_HS**, **Ordre_Sej_Prec_HS**, prennent alors comme valeur -9.

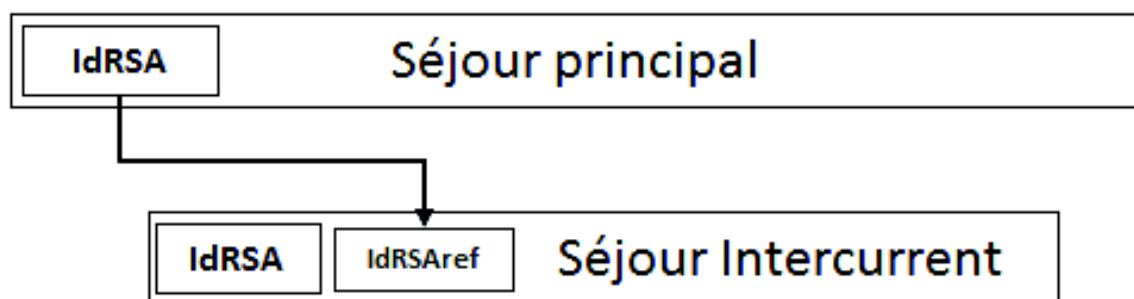


Figure 8- Affectation de l'identifiant du RSA de référence

DC_Sej_Suiv : Indique si le patient décède lors du séjour suivant (séries de niveau 5).

Grâce à la consolidation opérée sur les séjours pour séances et les séjours intercurrents, ces informations supplémentaires permettent une sélection de données beaucoup plus fiables, ainsi qu'un travail sur les délais de ré-hospitalisation.

Nous allons donner ci-dessous trois exemples de travaux conduits à partir d'extractions réalisées en exploitant ces informations.

1.4.3. Etude sur la mortalité hospitalière dans la chirurgie de la hanche

Cette étude portait sur une comparaison de la survenue de complications post-opératoires et de la mortalité hospitalière entre patients opérés pour fracture de hanche et patients opérés pour chirurgie programmée de remplacement de la hanche.

Le modèle de données comportait :

- identification du type de patient (chirurgie sur fracture ou chirurgie de remplacement) ;
- indicateurs de comorbidité ou décès survenu au cours de l'hospitalisation ;
- indicateurs de terrain : comorbidité préexistante ;
- indicateurs temporels liés aux hospitalisations ultérieures : délai de ré-hospitalisation, passage (et durée de passage) en réanimation ou soins intensifs ou décès lors du séjour suivant.

Le calcul des indicateurs de comorbidité et de survenu d'évènement s'est avéré particulièrement complexe. En effet les domaines de définition de ces attributs relèvent de combinaisons de plusieurs attributs des RSA (DP, DR, actes, GHM). De plus, concernant les indicateurs de comorbidité, il s'agissait de rechercher la présence de ces informations dans les séjours antérieurs.

Un jeu de données comportant environ 700 000 séjours a été fourni, et ce travail a donné lieu à publication[73].

1.4.4. Addictologie

L'objectif de cette étude était d'élaborer un modèle prédictif d'évènement intra-hospitalier pour les patients admis pour dépendance ou abus de drogue.

L'extraction requise portait sur les motifs d'admission, et devait également comporter le délai jusqu'au séjour suivant ainsi qu'un indicateur de survenue de l'évènement. Elle comportait environ 66 000 séjours.

Le domaine de définition de cet indicateur se projetait sur un ensemble composite d'attributs des RSA et de la table des trajectoires : passage en soins intensifs, passage en réanimation, ou décès. De plus, plusieurs éléments pouvant coexister (décès survenant à l'issue d'un séjour avec passage en soins intensifs), un ordre a dû être établi : Décès > Réa > Soins Intensifs.

Ce travail a donné lieu à un article actuellement soumis : « Identifying Life-Threatening Admissions for Drug Dependence or Abuse (ILIADDA): Derivation and Validation of a Model », Nguyen L, Boudemaghe T et al.

1.4.5. Etude sur les ré-hospitalisations dans l'insuffisance cardiaque

Pour ce travail dans le cadre d'un mémoire de Master 2, nous avons été sollicités pour fournir des données concernant les ré-hospitalisations suite à une hospitalisation pour insuffisance cardiaque (IC).

Sur la base d'une connaissance experte formalisée par le clinicien, les critères de repérage des hospitalisations pour insuffisance cardiaque ont été formulés.

Nous avons ensuite élaboré un jeu de données de base comportant les trajectoires des patients ayant été hospitalisés au moins une fois pour IC de 2009 à 2014. Partant de ce jeu de données, nous avons produit une extraction comportant les données d'intérêt grâce à des critères d'inclusion et d'exclusion définis par le clinicien :

- inclusion : Hospitalisation pour IC en 2012 ou 2013 (années de sortie 212 ou 2013) ;
- exclusion : Hospitalisation pour IC entre 2009 et 2011.

De plus, pour l'étude des ré-hospitalisations, le clinicien avait défini une liste de GHM d'intérêt. Par exemple, on ne considère pas comme une ré-hospitalisation une hospitalisation pour une opération de la cataracte ou une hospitalisation de jour pour une endoscopie digestive.

Le jeu de données comportait également des informations concernant le terrain des patients, intégrant la notion de chronicité ou d'ancienneté, et d'acuité.

Deux informations en particulier nécessitaient une bonne maîtrise des données de trajectoire :

- le repérage de la première hospitalisation pour IC entre 2012 et 2013. Ici, il s'agissait de bien sélectionner la première hospitalisation en cas de plusieurs occurrences dans la période considérée ;
- le délai de ré-hospitalisation spécifiquement pour IC après la première hospitalisation. Il s'agissait de repérer dans les hospitalisations ultérieures les hospitalisations pour IC, et de plus de sommer les délais des séjours successifs pour obtenir le délai de ré-hospitalisation pour IC. Ce point s'est avéré tout particulièrement délicat à traiter, car nous souhaitions :
 - ne pas prendre en compte les séjours débutant immédiatement après le séjour précédent, que nous rattachons aux séjours intercurrents (on considère qu'ils prolongent la prise en charge du séjour précédent et donc ne doivent pas être pris en compte en tant que tels) ;
 - n'étudier les ré-hospitalisations que pour les séjours non intercurrents.

Enfin, la production des informations sur le terrain des patients était là encore particulièrement délicate. En effet, une même pathologie pouvait être considérée comme ancienne si elle était survenue au cours d'un séjour précédent, ou aiguë si elle survenait au cours du séjour pour insuffisance cardiaque.

Chapitre 3 Méthode d'évaluation du niveau de référence de l'activité

1. Introduction

Depuis longtemps déjà, la question de la qualification de l'activité en fonction de sa complexité ou de sa lourdeur a été posée, avec les rapports qu'elle entretient avec celle de la qualification des établissements via l'analyse de leur profil d'activité. Il s'agit d'un enjeu important d'aide à la décision en planification sanitaire.

Plusieurs travaux ont été menés pour définir des méthodes de classification de l'activité selon une logique qualitative[74,75], ou encore sur l'identification d'un type d'établissement sur la base de son profil d'activité[76].

Plusieurs problèmes se posent lorsque l'on étudie ces questions. Tout d'abord, nous ne disposons d'aucune définition ou critères décrivant précisément ce que serait la notion de « référence » en ce qui concerne l'activité de soins. Ce terme polysémique laisse percevoir plusieurs aspects, tous pertinents :

- complexité des soins. Le terme est sujet à de multiples interprétations : matériel rare et cher, compétences médicales et soignantes rares, ou encore nécessité d'une association de moyens considérés comme simples individuellement ;
- lourdeur des patients. Sans nécessiter de moyens complexes, la pathologie peut engager des moyens lourds et coûteux (infection sévère, état grabataire avec escarres...) ;
- rareté de la maladie. Une pathologie rare peut requérir des compétences médicales spécialisées, sans moyens particulièrement lourds ou coûteux.

Nous conserverons pour la suite de ce travail le terme de « référence », consacré par l'usage. Cependant l'analyse de cette notion, dont le constat est évident mais le contenu difficile à cerner, nous amène à considérer le terme de spécificité de l'activité. En effet, le terme de spécificité désigne le fait pour une chose d'être particulière, singulière, propre à un individu, un objet, une situation. En l'occurrence, il nous apparaît que le trait commun des différents aspects de la référence est le fait qu'il s'agit à chaque fois de situations particulières, propres à peu de patients (par exemple, une greffe cœur-poumons), comparativement à des pathologies ou des situations beaucoup plus répandues (par exemple, une hospitalisation pour endoscopie digestive diagnostique, ou tout simplement le cas de l'accouchement sans complication).

On pourrait décider de réunir des experts pour classer une à une les racines de GHM par exemple. Cependant, au-delà du temps important que cette méthode demanderait, des problèmes se poseraient très probablement dans la comparaison des différentes racines. Pour reprendre l'exemple précédent, comment déterminer ce qui, d'un geste interventionnel simple ou d'un accouchement sans complication, est le plus ou le moins spécifique ? De plus, tout le processus devrait être renouvelé en cas de création ou de suppression de racines. Finalement, cette façon de procéder pourrait ne pas recueillir le consensus nécessaire à son adoption.

C'est pourquoi nous avons cherché à mettre au point une méthode qui se baserait sur les propriétés statistiques de l'activité en termes d'hétérogénéité de distribution, en utilisant l'entropie de Shannon.

2. Les outils mathématiques : l'entropie et les divergences entre distributions

2.1. L'entropie de Shannon

2.1.1. Le concept

Le concept d'entropie nous vient initialement de la thermodynamique, où il fut introduit en 1865 par Rudolf Clausius. En thermodynamique, l'entropie quantifie l'énergie dispersée. Par un léger glissement sémantique, l'idée a été élargie à la notion de dispersion proprement dite. C'est ainsi que le concept a par la suite fait florès, dans la théorie de l'information et la théorie des probabilités (entropie de Rényi, entropie de Shannon, entropie conditionnelle) et en mathématique en général (entropie topologique, entropie métrique), mais aussi en économie ou encore en écologie des populations.

Nous nous intéresserons ici plus particulièrement à sa déclinaison dans le domaine des probabilités, en partant de son application en théorie de l'information, l'entropie de Shannon[77]. Etant donné une variable aléatoire multinomiale de k classes, de probabilité p_1, \dots, p_k , l'entropie de Shannon H_s se définit par :

$$H_s = - \sum_{i=1}^k p_i \log_2(p_i)$$

L'entropie de Shannon a par la suite été étudiée par Rényi[78]. Celui-ci définit ce qui deviendra l'entropie de Rényi, dite aussi entropie d'ordre α , qui dans le cas d'une distribution multinomiale s'écrit :

$$H_\alpha = \frac{1}{(1-\alpha)} \log_2(\sum_i p_i^\alpha), \alpha \neq 1.$$

L'entropie de Shannon correspond alors au cas limite où α tend vers 1.

En ce qui concerne l'entropie de Shannon, nous allons illustrer cette formule à partir d'un exemple.

Supposons que l'on tire au hasard un symbole parmi un jeu de symboles numérotés entre 1 et N , avec $N = 2^n$. Si un interlocuteur doit deviner de quel symbole il s'agit en posant un minimum de questions, il devra procéder par dichotomie, en demandant d'abord si le numéro du symbole est supérieur ou inférieur à $N/2$, puis $N/4$ ou $3N/4$, etc. soit au total $n = \log_2(N)$ questions. On appellera $\log_2(N)$ la quantité d'information contenue dans (apportée par) le symbole. Par extension, on conservera cette formule dans les cas où N n'est pas de la forme 2^n .

Supposons à présent que les symboles soient répartis en k classes, d'effectif n_1, \dots, n_k et de probabilité p_1, \dots, p_k . Si l'on connaît la classe i dans laquelle se trouve le symbole, la quantité

d'information nécessaire pour déterminer de quel symbole il s'agit sera alors $\log_2(n_i)$. La quantité d'information apportée par les classes sera donc $\sum_{i=1}^k p_i \log_2(n_i)$. Il reste cependant à déterminer la quantité d'information nécessaire pour connaître la classe dans laquelle se trouve le symbole.

On constate que l'on peut développer $\log_2(N)$ de la façon suivante :

$$\log_2(N) = \sum_i (p_i \log_2 N) = \sum_i \left(p_i \log_2 \left(\frac{n_i}{p_i} \right) \right) = - \sum_i (p_i \log_2(p_i)) + \sum_i (p_i \log_2(n_i))$$

On voit que $\log_2(N)$ – information nécessaire pour identifier le symbole – peut donc se décomposer en une somme constituée de l'entropie de Shannon d'une part, d'autre part de l'information nécessaire moyenne la classe du symbole étant connue. On considérera pour la suite comme acquis que l'entropie de Shannon correspond à la quantité d'information nécessaire pour déterminer la classe à laquelle appartient le symbole.

On voit bien de ce qui précède le lien étroit entre incertitude et information. Dans son article, Rényi exprime parfaitement ce lien de la façon suivante :

“The entropy of a probability distribution can be interpreted not only as a measure of uncertainty but also as a measure of information. As a matter of fact, the amount of information which we get when we observe the result of an experiment (depending on chance) can be taken numerically equal to the amount of uncertainty concerning the outcome of the experiment before carrying it out.”

Que l'on peut traduire par :

« L'entropie d'une distribution de probabilité peut non seulement s'interpréter comme une mesure de l'incertitude, mais aussi comme une mesure de l'information. En fait, la quantité d'information obtenue lors de l'observation du résultat d'une expérience (reposant sur de l'aléatoire) peut être considérée comme numériquement égale à la quantité d'incertitude concernant le résultat de cette expérience avant de la réaliser. »

2.1.2. Quelques propriétés de l'entropie de Shannon

Soit une variable aléatoire multinomiale à k classes $M(k, n)$. La valeur minimum de l'entropie de Shannon de sa distribution est 0, sa valeur maximum dépend du nombre de classes et est égale à $\log_2(k)$. On peut donc normer l'entropie de $M(k, n)$ à 1 en la divisant par $\log_2(k)$.

L'ensemble des distributions possibles de n individus sur les k classes peut être considéré comme une partition d'une entier n en k classes. Nous appellerons configuration une partition donnée.

En associant chaque configuration avec son entropie, on peut définir une relation d'ordre sur les configurations. On constate cependant que la fonction associant à chaque configuration son entropie n'est pas bijective : en effet il peut y avoir des ex-aequo.

Exemple pour $M(4, 10)$:

Ordre et numérotation et des configurations		$M(4, 10)$				H_s	$H_s/\log_2(k)$
Ordre	Numérotation normée à 1	$C1$	$C2$	$C3$	$C4$		
1	0,04	10	0	0	0	0,0000	0,0000
2	0,09	9	1	0	0	0,4690	0,2345
3	0,13	8	2	0	0	0,7219	0,3610
4	0,17	7	3	0	0	0,8813	0,4406
5	0,22	8	1	1	0	0,9219	0,4610
6	0,26	6	4	0	0	0,9710	0,4855
7	0,30	5	5	0	0	1,0000	0,5000
8	0,35	7	2	1	0	1,1568	0,5784
9	0,39	6	3	1	0	1,2955	0,6477
10	0,43	7	1	1	1	1,3568	0,6784
11	0,48	5	4	1	0	1,3610	0,6805
12	0,52	6	2	2	0	1,3710	0,6855
13	0,57	5	3	2	0	1,4855	0,7427
14	0,61	4	4	2	0	1,5219	0,7610
15	0,65	6	2	1	1	1,5710	0,7855
16	0,70	4	3	3	0	1,5710	0,7855
17	0,74	5	3	1	1	1,6855	0,8427
18	0,78	4	4	1	1	1,7219	0,8610
19	0,83	5	2	2	1	1,7610	0,8805
20	0,87	4	3	2	1	1,8464	0,9232
21	0,91	3	3	3	1	1,8955	0,9477
22	0,96	4	2	2	2	1,9219	0,9610
23	1,00	3	3	2	2	1,9710	0,9855

On voit que les configurations 15 et 16 ont la même entropie

La courbe d'accroissement de l'entropie en fonction des configurations ordonnées selon leur numérotation normée a la forme suivante :

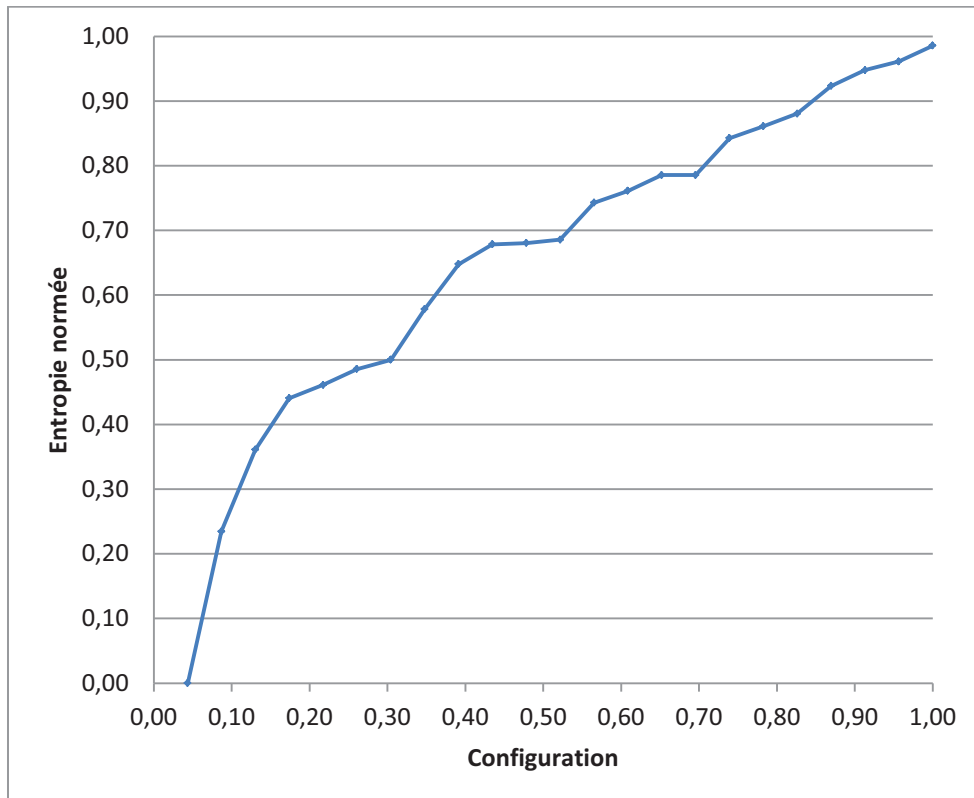


Figure 9. Accroissement de l'entropie pour la variable aléatoire multinomiale $M(4, 10)$

2.1.3. Décomposition de l'entropie

Nous commencerons par nous doter de la définition d'un objet que nous appellerons une classification hiérarchique mono-axiale. On appelle classification hiérarchique mono-axiale à r niveaux, la donnée de n individus et $r+1$ variables aléatoires multinomiales, dites de niveau 0, 2, ..., r :

$$M_0(1, n)$$

$$M_1(k_1, n)$$

$$M_2(k_2, n)$$

...

$$M_r(k_r, n)$$

Présentant les propriétés suivantes :

- $1 < k_1 \leq k_2 \leq \dots \leq k_r$. On adoptera l'usage consistant à désigner le niveau $v+1$ comme inférieur au niveau v . Cet usage rend compte de l'image d'une ramification progressive de la classification.
- Chaque classe d'un niveau donné est partitionnée par une ou plusieurs classes du niveau immédiatement inférieur.
- On note que $M_0(1, n)$ est un cas particulier à une classe de probabilité 1, contenant tous les individus, on l'appelle la classe chapeau.

On notera

- $n_{v+1,j}$ l'effectif de la classe C_j de la variable M_{v+1}
- $p_{v+1,j}$ la probabilité associée à la classe C_j de la variable M_{v+1}

D'après la définition dont nous nous sommes dotés, C_j appartient à la partition d'une des classes C_i de la variable de niveau supérieur M_v

Afin de représenter cette relation entre C_j et C_i on notera :

- $k_{v+1,v,i}$ le nombre de classes de la variable M_{v+1} composant la partition de la classe C_i de la variable de niveau v M_v . On note que l'on a

$$\sum_{i=1}^{k_v} k_{v+1,v,i} = k_{v+1}$$

- $n_{v+1,j} = n_{(v+1,j),(v,i)}$ l'effectif de la classe C_j de la variable M_{v+1} , membre de la partition de la classe C_i de la variable de niveau supérieur M_v . On peut donc écrire

$$n_{v,i} = \sum_{j=1}^{k_{v+1,v,i}} n_{v+1,j}$$

- $p_{(v+1,j),(v,i)}$ la probabilité associée à la classe C_j de la variable M_{v+1} au sein de la partition de la classe C_i de la variable de niveau supérieur M_v . On a donc

$$p_{(v+1,j),(v,i)} = n_{v+1,j}/n_{v,i} = n_{(v+1,j),(v,i)}/n_{v,i} \text{ et } \sum_{j=1}^{k_{v+1,v,i}} p_{(v+1,j),(v,i)} = 1$$

- On notera bien que $p_{(v+1,j),(v,i)}$ ne représente pas la même chose que $p_{v+1,j}$.

En effet, on a :

$$p_{v+1,j} = n_{v+1,j}/n$$

Alors que

$$p_{(v+1,j),(v,i)} = n_{v+1,j}/n_{v,i}$$

Enfin on notera que

$$p_{(v,i)}p_{(v+1,j),(v,i)} = n_{v,i}/n \cdot n_{v+1,j}/n_{v,i} = n_{v+1,j}/n = p_{(v+1,j)}$$

Cette structure peut être illustrée par le schéma suivant :

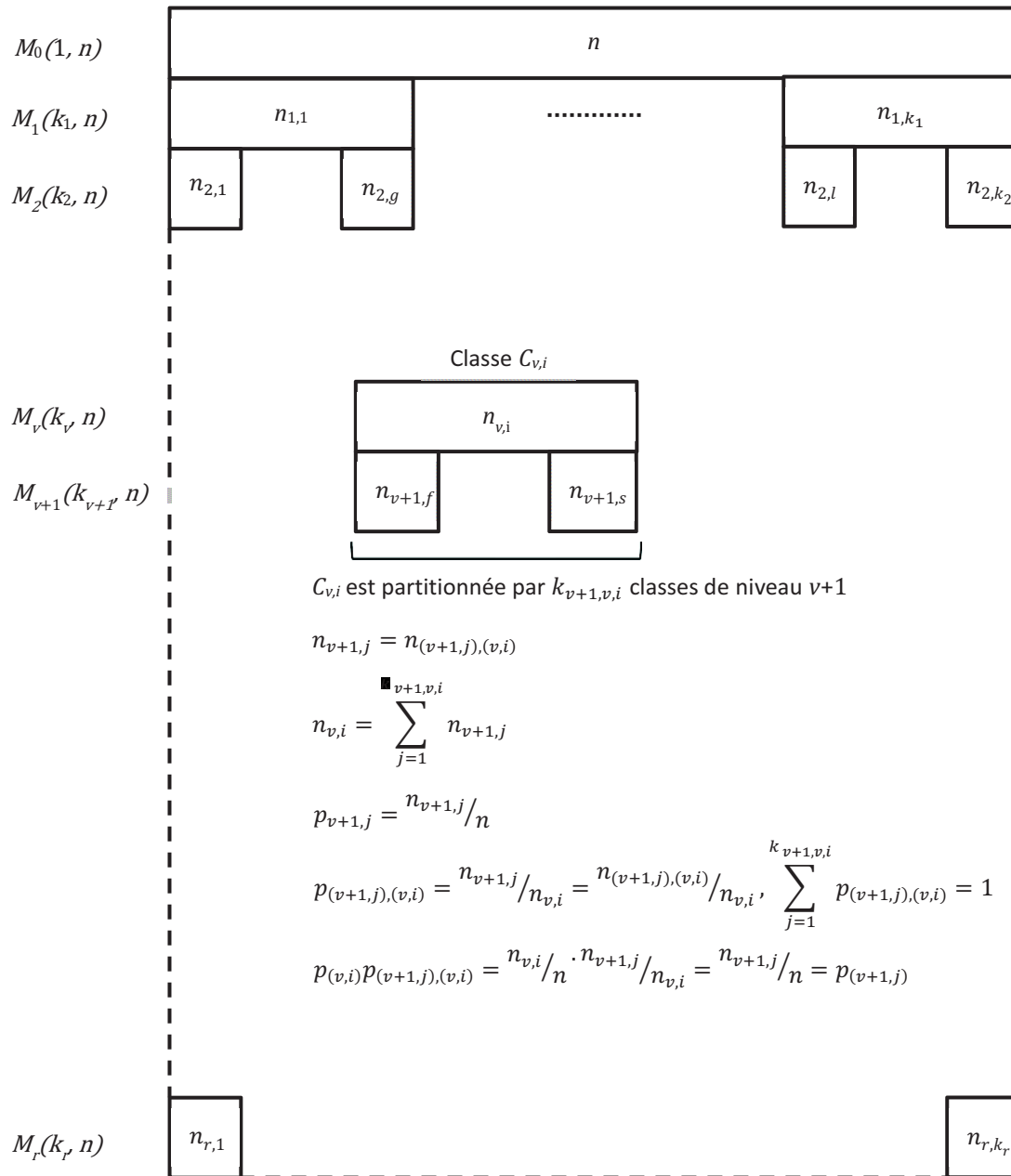


Figure 10 - Représentation schématisée d'une classification hiérarchique

Nous allons à présent nous intéresser à la décomposition de l'entropie de la classification.

1. L'entropie totale de la classification est donnée par :

$$H_S(M_r) = H_S T = - \sum_{z=1}^{k_r} p_{r,z} \log_2(p_{r,z})$$

2. Si l'on étudie à présent l'entropie de M_1 on a :

$$H_S(M_1) = - \sum_{i=1}^{k_1} p_{1,i} \log_2(p_{1,i}) \quad (1)$$

En remarquant que $p_{1,i} = n_{1,i}/n$ et donc que $1/p_{1,i} = n/n_{1,i}$, on peut réécrire (1) de la façon suivante :

$$H_S(M_1) = \sum_{i=1}^{k_1} p_{1,i} \log_2(n/n_{1,i}) = \sum_{i=1}^{k_1} p_{1,i} \log_2(n) - \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) \quad (2)$$

Soit

$$H_S(M_1) = \log_2(n) - \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) \quad (2)$$

3. Dans cette expression, le terme

$$- \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) \quad (2)$$

peut être décomposé de la façon suivante :

en étudiant $\log_2(n_{1,i})$, et sachant que l'on a

$$\sum_{j=1}^{k_{2,1,i}} p_{(2,j),(1,i)} = 1$$

on peut écrire

$$\log_2(n_{1,i}) = \sum_{j=1}^{k_{2,1,i}} (p_{(2,j),(1,i)} \log_2(n_{1,i}))$$

de plus, pour tout $j \in (1, \dots, k_{2,1,i})$, on peut écrire

$$n_{1,i} = n_{(2,j),(1,i)} / p_{(2,j),(1,i)}$$

on peut donc écrire

$$\begin{aligned}\log_2(n_{1,i}) &= \sum_{j=1}^{k_{2,1,i}} \left(p_{(2,j),(1,i)} \log_2 \left(n_{(2,j),(1,i)} / p_{(2,j),(1,i)} \right) \right) \\ &= \sum_{j=1}^{k_{2,1,i}} p_{(2,j),(1,i)} \log_2(n_{(2,j),(1,i)}) - \sum_{j=1}^{k_{2,1,i}} p_{(2,j),(1,i)} \log_2(p_{(2,j),(1,i)}) \quad (3)\end{aligned}$$

On reconnaît dans l'expression

$$- \sum_{j=1}^{k_{2,1,i}} p_{(2,j),(1,i)} \log_2(p_{(2,j),(1,i)})$$

$H_S(C_i)$ l'entropie de Shannon calculée pour la partition de la classe C_i de niveau 1 en classes de niveau 2 : on appellera cette entropie **l'entropie interne de la classe C_i** .

4. On a donc

$$- \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) = - \sum_{i=1}^{k_1} p_{1,i} \left[\sum_{j=1}^{k_{2,1,i}} p_{(2,j),(1,i)} \log_2(n_{(2,j),(1,i)}) + H_S(C_i) \right]$$

soit

$$- \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) = - \sum_{i=1}^{k_1} \left[\sum_{j=1}^{k_{2,1,i}} p_{1,i} p_{(2,j),(1,i)} \log_2(n_{(2,j),(1,i)}) \right] - \sum_{i=1}^{k_1} p_{1,i} H_S(C_i) \quad (4)$$

Sachant que

$$p_{1,i} p_{(2,j),(1,i)} = p_{(2,j)}$$

et que

$$n_{(2,j),(1,i)} = n_{2,j}$$

on a donc

$$- \sum_{i=1}^{k_1} p_{1,i} \log_2(n_{1,i}) = - \sum_{j=1}^{k_2} p_{2,j} \log_2(n_{2,j}) - \sum_{i=1}^{k_1} p_{1,i} H_S(C_i) \quad (4)$$

On remarque que le terme

$$\sum_{i=1}^{k_1} p_{1,i} H_S(C_i)$$

représente la somme des entropies internes des classes de niveau 1, pondérées par leur probabilité. On notera $H_{w,1}$ cette quantité, on l'appellera **entropie pondérée de niveau 1**.

On peut alors écrire

$$H_S(M_1) = \log_2(n) - \sum_{j=1}^{k_2} p_{2,j} \log_2(n_{2,j}) - H_{w,1} \quad (4)$$

5. Le terme

$$- \sum_{j=1}^{k_2} p_{2,j} \log_2(n_{2,j})$$

Peut être à nouveau développé en

$$- \sum_{l=1}^{k_3} p_{3,l} \log_2(n_{3,l}) - \sum_{j=1}^{k_2} p_{2,j} H_S(C_j)$$

avec

$$H_{w,2} = \sum_{j=1}^{k_2} p_{2,j} H_S(C_j)$$

entropie pondérée de niveau 2.

L'entropie de M_1 devient alors :

$$H_S(M_1) = \log_2(n) - \sum_{l=1}^{k_3} p_{3,l} \log_2(n_{3,l}) - H_{w,1} - H_{w,2}$$

Par récurrence, on peut développer jusqu'à atteindre le niveau r .

On aura alors :

$$H_S(M_1) = \log_2(n) - \sum_{z=1}^{k_r} p_{r,z} \log_2(n_{r,z}) - \sum_{i=1}^{r-1} H_{w,i}$$

6. En étudiant le terme $-\sum_{z=1}^{k_r} p_{r,z} \log_2(n_{r,z})$, et sachant que $n_{r,z} = n \cdot p_{r,z}$, on peut écrire

$$- \sum_{z=1}^{k_r} p_{r,z} \log_2(n_{r,z}) = - \sum_{z=1}^{k_r} p_{r,z} \log_2(n) - \sum_{z=1}^{k_r} p_{r,z} \log_2(p_{r,z})$$

Or on a vu plus haut que

$$H_{ST} = - \sum_{z=1}^{k_r} p_{r,z} \log_2(p_{r,z})$$

On a donc

$$H_S(M_1) = H_{ST} - \sum_{i=1}^{r-1} H_{w,i}$$

Soit

$$H_{ST} = H_S(M_1) + \sum_{i=1}^{r-1} H_{w,i}$$

7. $H_S(M_1)$ étant également l'entropie pondérée de niveau 0 (une seule classe de probabilité 1), la décomposition de l'entropie d'une classification hiérarchique mono-axiale à r niveaux peut donc s'écrire :

$$H_{ST} = \sum_{i=0}^{r-1} H_{w,i}$$

avec

$$H_{w,i} = \sum_{j=1}^{k_i} p_{i,j} H_S(C_j)$$

somme de l'entropie interne des classes de niveau i , pondérée par la probabilité associée à ces classes.

Cas particuliers

- **Classification à 1 niveau**

On n'a qu'une seule distribution, et $H_{ST} = - \sum_{i=1}^k p_i \log_2(p_i)$. C'est l'entropie de Shannon dans sa forme originale.

- **Classification à 2 niveaux**

$$H_{ST} = - \sum_{i=1}^{k_1} p_{1,i} \log_2(p_{1,i}) + \sum_{i=1}^{k_1} p_{1,i} H_S(C_i) = \sum_{i=1}^{k_1} p_{1,i} \left[\log_2 \left(\frac{1}{p_{1,i}} \right) + H_S(C_i) \right]$$

La décomposition de l'entropie à deux niveaux a déjà été décrite dans des travaux en économie[79] et employée dans des travaux d'économie de la santé sur l'impact sur la qualité des soins de la T2A[1].

On appellera la quantité

$$CH_i = p_{1,i} \left[\log_2 \left(\frac{1}{p_{1,i}} \right) + H_S(C_i) \right]$$

la contribution de la classe C_i à l'entropie totale.

2.2. Divergence entre distributions

La notion d'entropie d'une distribution est proche du thème plus général de l'étude des divergences entre distributions : si l'on peut caractériser la dispersion d'une distribution, alors on peut se poser la question de caractériser la dispersion de cette distribution relativement à une autre distribution donnée, ou de plusieurs distributions entre elles. Ceci conduit à définir des mesures de similarité entre distributions, présentant des propriétés mathématiques différentes selon les objectifs recherchés. Une propriété particulièrement importante, mais pas toujours recherchée, sera le fait de constituer une distance au sens mathématique du terme, c'est-à-dire vérifiant les propriétés suivantes :

Positivité (ou nullité) : $D(A, B) \geq 0$

Symétrie : $D(A, B) = D(B, A)$

Séparation : $D(A, B) = 0$ si et seulement si $A = B$

Inégalité triangulaire : $D(A,B) + D(B, C) \leq D(A, C)$

Cette problématique est cruciale le domaine de la classification des données : classification proprement dite, analyse discriminante, reconnaissance de motifs. Le champ d'application est extrêmement vaste puisque touchant à toutes les situations impliquant une différenciation, une classification, une reconnaissance de données : bio-informatique (génomique), apprentissage machine (« *machine learning* »), traitement du signal, cryptographie...

Les applications dans le domaine de la génomique, en particulier, n'ont pas manqué d'attirer notre attention par leur similitude avec notre problématique (notamment la notion de traitement d'un grand nombre de données disposées en matrice).

L'étude exhaustive de ce vaste champ de recherche n'entre pas dans le cadre de notre travail. Néanmoins, afin de nous assurer de la pertinence et de la validité de nos choix méthodologiques, nous avons conduit une revue bibliographique de base pour identifier les principaux points de repère dans le développement de ce domaine.

La distance de Mahalanobis[80] constitue une première étape fondamentale. Cet indicateur a été conçu par le chercheur du même nom dans le cadre d'études anthropologiques, portant sur les caractéristiques du crâne humain dans différents groupes ethniques. Il s'agissait de pallier à un défaut du « *Coefficient of Racial Likeness* » de Karl Pearson, qui constituait un test de divergence entre un individu statistique et une population d'appartenance supposée, et non sa mesure. Mahalanobis crée alors la statistique D^2 , qui deviendra la distance de Mahalanobis.

La distance de Mahalanobis entre un individu et une population (représentée par son vecteur moyen) peut être considérée comme une distance euclidienne normée par l'écart type.

Soit une population X de vecteur moyen $\bar{X} = (\bar{x}_1, \dots, \bar{x}_n)^T$ et de matrice de variance-covariance S .

Soit $y = (y_1, \dots, y_n)^T$ un individu observé.

La distance Mahalanobis entre y et \bar{X} est donnée par :

$$d_M(y, \bar{X}) = \sqrt{(\bar{X} - y)^T S^{-1} (\bar{X} - y)}$$

Cette expression présente les critères mathématiques d'une distance.

Par la suite, cette distance est raffinée par Bhattacharyya[81,82]. Bien que l'usage semble avoir consacré le terme de distance de Bhattacharyya, il s'agit en réalité d'une mesure de similarité qui n'offre plus les caractères d'une distance. En revanche, elle permet de prendre en compte les différences de variance entre les deux distributions comparées. Ainsi, dans le cadre d'une classification, deux classes ayant des moyennes proches mais des écarts types

très différents auront une distance de Bhattacharya élevée, alors que leur distance de Mahalanobis sera très proche de 0.

Dans le cas de deux distributions multinomiales p et q comportant k classes, la distance de Bhattacharya est donnée par :

$$D_b(p, q) = -\ln(BC(p, q))$$

avec

$$BC(p, q) = \sum_{i=1}^k \sqrt{p_i q_i},$$

BC étant appelé le coefficient de Bhattacharya.

Enfin, l'étude des distances, et plus largement des similarités, entre distributions va se développer sous l'appellation plus générale de divergences. Les divergences sont classées en plusieurs familles, dont la principale est celle des f-divergences[83-85].

Pour la suite, nous avons retenu la distance de Hellinger[86]. Dans le cas de deux distributions multinomiales p et q comportant k classes, la distance de Hellinger est donnée par :

$$D_H(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}$$

Lorsque les deux distributions sont totalement identiques, $D_H(p, q)$ prend la valeur 0, la valeur 1 lorsqu'elles sont entièrement différentes ($\forall i, p_i = 0$ ou $q_i = 0$).

Outre l'intérêt d'être une distance mathématique, la distance de Hellinger offre l'avantage d'être comprise entre 0 et 1 et de permettre de prendre en compte des valeurs nulles pour les deux distributions.

3. Analyse de l'entropie de la matrice d'activité

3.1. Matrice d'activité et modèle de partition

3.1.1. Matrice d'activité

L'activité d'hospitalisation, prise en son sens quantitatif de nombre de séjours, peut être vue comme un espace découpé selon un axe établissement d'une part, nature d'activité de l'autre (l'activité entendue dans son sens quantitatif). Pour la suite, nous désignerons par le seul terme d' « activité » l'activité entendue dans son sens quantitatif.

Le découpage en établissements est immédiat et relativement fixé (mais il est toujours possible de regrouper des établissements). Le découpage de l'activité par nature, en revanche, peut se faire selon plusieurs logiques, ou modèles : GHM, regroupements de GHM, caractéristiques de séjours en diagnostics, actes, durée de séjour ou autre indépendamment de la notion de GHM... Il existe une multitude de découpages possibles

tant qu'ils constituent une partition de l'activité : un ensemble de parties non vides, deux à deux disjointes, dont l'union recouvre l'activité. Nous appellerons ces parties des classes d'activité, ou encore des prises en charges. Ainsi, la partition de l'activité selon un modèle donné résultera en un ensemble de classes d'activité.

On peut donc considérer l'activité de soins décrite par la base nationale des RSA comme une matrice combinant les établissements au découpage choisi de l'activité. Pour la suite, nous considérerons les établissements comme les lignes de la matrice, et les classes d'activité résultant du modèle de partition comme ses colonnes. Chaque cellule de la matrice comporte alors le nombre de séjours effectué dans un établissement donné pour une classe donnée d'activité. Nous appellerons cette matrice la matrice d'activité. Cette matrice peut également être vue comme une classification hiérarchique à 2 niveaux : un premier niveau d'activité et un second correspondant aux établissements.

On perçoit immédiatement les parallèles qui peuvent être faits avec l'utilisation de l'entropie dans l'analyse de cette matrice. Il s'agit de confirmer ou d'infirmer, de façon formelle et avec une méthode définie et reproductible, des caractéristiques de cette activité communément admises, intuitivement ou à dire d'expert. Nous allons à présent étudier la matrice d'activité pour l'année 2014.

3.1.2. Modèle de partition de l'activité

Un modèle de partition est défini par des caractéristiques choisies des séjours. Nous avons retenu trois caractéristiques que nous considérons comme pertinentes pour définir un type de prise en charge :

- le GHM, portant l'information médicale principale décrivant le séjour ;
- le GHS portant l'information tarifaire. Certains GHM peuvent être associés à un GHS et donc à un tarif différent en fonction de la réalisation d'une technique particulière, dont le GHS rend compte du coût. Il est donc également porteur d'une information médicale sur le séjour, ce qui rend pertinente son association avec le GHM ;
- la durée de séjour. On considère que cette information peut également rendre compte d'une certaine complexité ou lourdeur du séjour, en raison de l'état du patient ou de la complexité des soins. La durée de séjour est répartie en 3 classes : 0 jour, 1 jour, 2 jours et plus. Le choix des classes de durée est basé sur une segmentation classique dans les analyses, distinguant les séjours ambulatoires (au sens générique du terme d'une hospitalisation sans nuitée), les séjours ne comportant qu'une nuit, considérés comme relevant de situations légères ou programmées, et toutes les autres situations, au-delà d'une nuit.

La combinaison entre GHM, GHS et classe de durée de séjour constitue le niveau de granularité le plus fin, c'est celui que nous avons retenu. Il est ensuite possible de réduire le niveau de granularité, notamment en combinant les GHM selon la classification de l'ATIH en groupes d'activité, groupes de planification, disciplines d'activité. D'autres axes définis par la classification de l'ATIH peuvent également être utilisés (activité de soin, catégorie d'activité de soins).

En l'occurrence, nous avons choisi d'étudier les racines de GHM, c'est ce qui constituera le modèle de partition pour la suite.

Nous excluons de cette étude la fraction d'activité représentée par les séances, pour les raisons suivantes :

- déséquilibre des effectifs. A établissements comparables, les volumes importants accompagnant toujours l'activité de séance créeront un biais dans l'analyse de la distribution de l'activité ;
- mode de recueil. Les séances de radiothérapie ne sont pas recueillies par les établissements privés, il y aurait donc un biais important si celles-ci étaient prises en compte.

On notera P la partition de l'activité, ensemble des k classes obtenues par application du schéma de partition. $P = \{C_1, \dots, C_k\}$.

3.2. Homogénéité et standardisation des données

3.2.1. Etablissements de référence

L'objectif de l'analyse de l'entropie est de repérer l'hétérogénéité des prises en charge en observant leur distribution sur les établissements. Il est donc nécessaire de travailler sur une base d'établissements comparables au vu de leurs caractéristiques. Pour cela, nous nous appuyons sur la base de la SAE.

Le terme comparable fait référence au fait que les établissements considérés disposent tous au moins des mêmes grands types d'installation, et, qu'au moins en théorie, ils pourraient tous réaliser toutes les activités possibles.

Le choix des critères de sélection des établissements de référence est critique : trop faibles, ils sélectionneraient des établissements trop dissemblables donc non comparables. Trop exigeants, ils ne sélectionneront que peu d'établissements, de grande taille. Ces établissements réaliseront certes toute la palette des activités, mais risquent d'être non plus seulement comparables, mais similaires.

Ceci affaiblirait le pouvoir discriminant de l'analyse. De plus, définir des critères trop fins nous exposerait aux risques de qualité de la base de la SAE qui ont été mis en évidence dans la partie consacrée à la consolidation des données et à la construction de référentiels.

Prenant en compte ces éléments, nous avons retenu pour l'année 2014, dans la base de la SAE, les établissements (Etb SAE) répondant aux critères suivants :

- en activité toute l'année, avec transmission des données d'activité pour toute l'année ;
- présence d'un service d'urgences ;
- présence d'un bloc opératoire ;
- présence d'une réanimation ;

- présence d'une maternité.

On obtient une liste de 379 Etb SAE, identifiés par leur numéro FINESS d'entité juridique (FINESS_EJ, ou FI_EJ), que nous associons ensuite à 424 établissements (sur un total de 1 974) issus de la base nationale (« Etb BN »), identifiés par le numéro FINESS (FINESS_BN, ou FI_BN), reporté dans les RSA, qui peut correspondre soit au numéro FINESS d'entité juridique, soit au numéro FINESS d'établissement (au sens d'un site géographique distinct).

Nous avons constaté que dans certains cas, plusieurs Etb BN peuvent être associés à un même Etb SAE (Etb SAE multi-Etb BN) : on dénombre ainsi 26 Etb SAE recouvrant 71 Etb BN. Nous n'avons en revanche pas trouvé de cas où plusieurs Etb SAE peuvent être associés à un même Etb BN.

L'examen des Etb SAE retenus montre qu'ils couvrent une large palette en termes de nombre de séjours. Afin d'éviter des biais liés à une faible capacité, donc de petits effectifs, nous décidons de ne retenir que les établissements représentant 99% du nombre total de séjours hors séance (NbSejHS) de la liste complète. Pour l'année 2014, le seuil est de 6 463 séjours hors séance.

L'application de ce filtre aboutit à retenir 359 Etb SAE, représentant 404 Etb BN. Les 359 Etb SAE retenus incluent les 26 Etb SAE multi-Etb BN. Le processus de sélection des établissements est résumé dans la figure suivante.

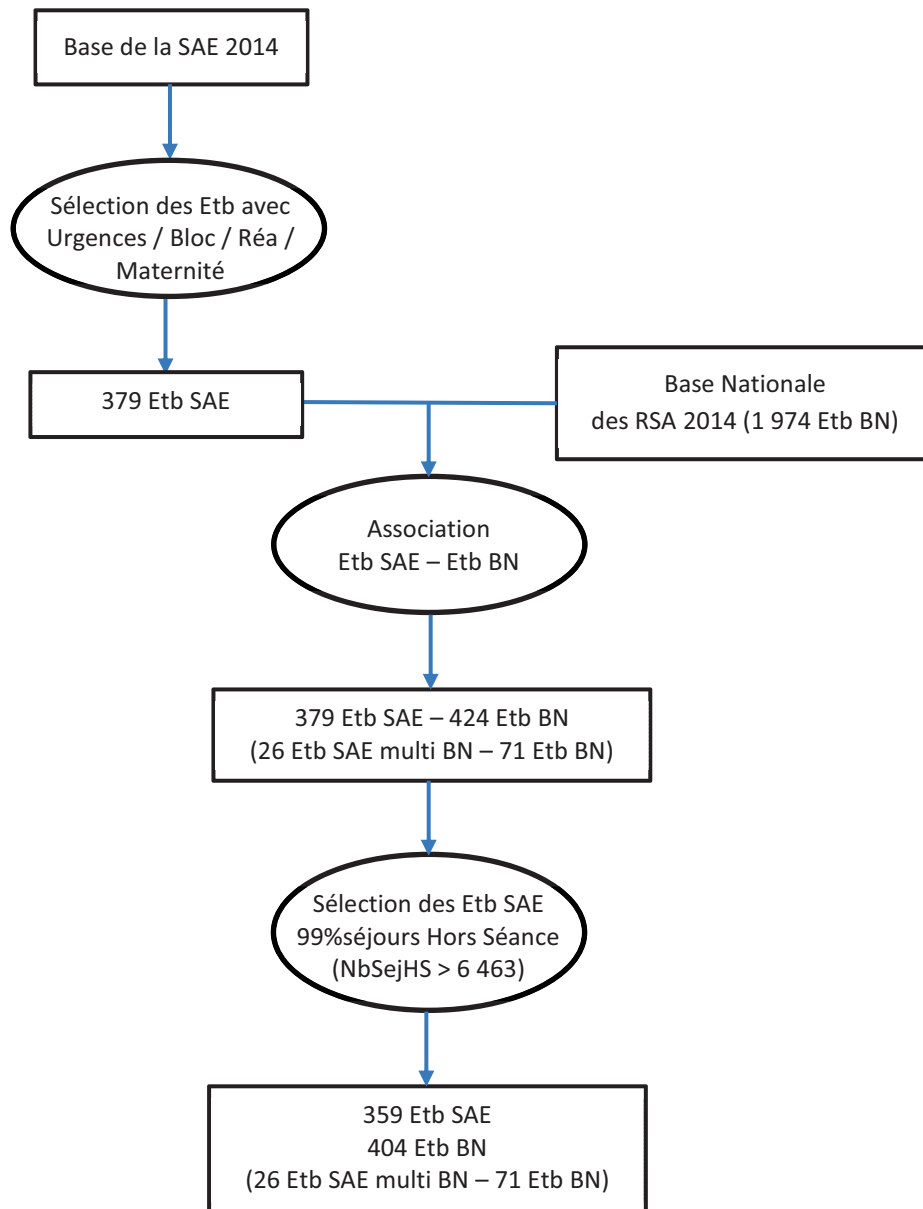


Figure 11 - Etapes de sélection des établissements

3.2.2. Standardisation des effectifs de classe

La différence de taille des établissements constitue un biais dans l'analyse de la distribution des modalités d'activité et doit donc être neutralisée. Une première solution consiste à standardiser les effectifs sur la base d'un nombre total de séjours de référence arbitrairement choisi, par exemple 10 000.

Ainsi, pour un établissement E , avec un nombre total de séjours hors séance N , si l'on observe pour une prise en charge C_i un effectif n_i , son effectif standardisé nt_i sera égal à $n_i \cdot 10\,000 / N$.

Néanmoins, on peut considérer que cette standardisation est insatisfaisante, diluant une activité qui ne serait pas particulièrement spécifique en tant que telle, mais simplement sous-

représentée dans l'établissement en raison des caractéristiques de l'établissement (capacité en lits, capacités du plateau technique) et non de l'activité.

Pour pallier cet effet, on peut considérer qu'une prise en charge donnée est issue, au sein de l'établissement, d'une fraction spécifique de l'activité. Par exemple, si la prise en charge considérée est chirurgicale, on peut alors considérer qu'elle est plus spécifiquement issue de la fraction des séjours chirurgicaux.

On serait alors amené à standardiser l'effectif non pas sur la base d'un nombre total de séjours de 10 000 dans l'établissement, mais d'un nombre de séjours chirurgicaux de 10 000 au sein de l'établissement. Cette logique peut se décliner selon la discipline, la classe de durée de séjour... On la formalise de la façon suivante :

Etant donné $P = \{C_1, \dots, C_k\}$ la partition des classes d'activité obtenue par application du schéma de partition, on définit un ensemble G de m ($m \leq k$) regroupements R de classes d'activité de P , ces regroupement étant disjoints et recouvrant la totalité des k classes de P :

$$G = \{R_1, \dots, R_m\}, m \leq k, \text{ avec } \forall i, j R_i \cap R_j = \emptyset \text{ et } \bigcup_{i=1}^m R_i = P$$

Reprenant pour l'établissement E d'effectif N la prise en charge C_i d'effectif n_i , et sachant que $C_i \in R_g$ d'effectif n_g au sein de l'établissement, on aura alors comme effectif standardisé

$$ntm_i = n_i \cdot 10\,000/n_g.$$

Sachant que $n_g < N$, on aura donc toujours $nt_i > ntm_i$, nous appellerons cet effet la concentration des effectifs dans l'établissement.

En limite, $m = k$; la concentration des effectifs dans les établissements est maximale avec un effectif standardisé constant à 10 000 pour tous les établissements. On aboutit alors à une distribution uniforme, donc une entropie maximale.

Dans un premier temps, nous avons choisi de standardiser en dilution maximale (sur l'effectif total de l'établissement).

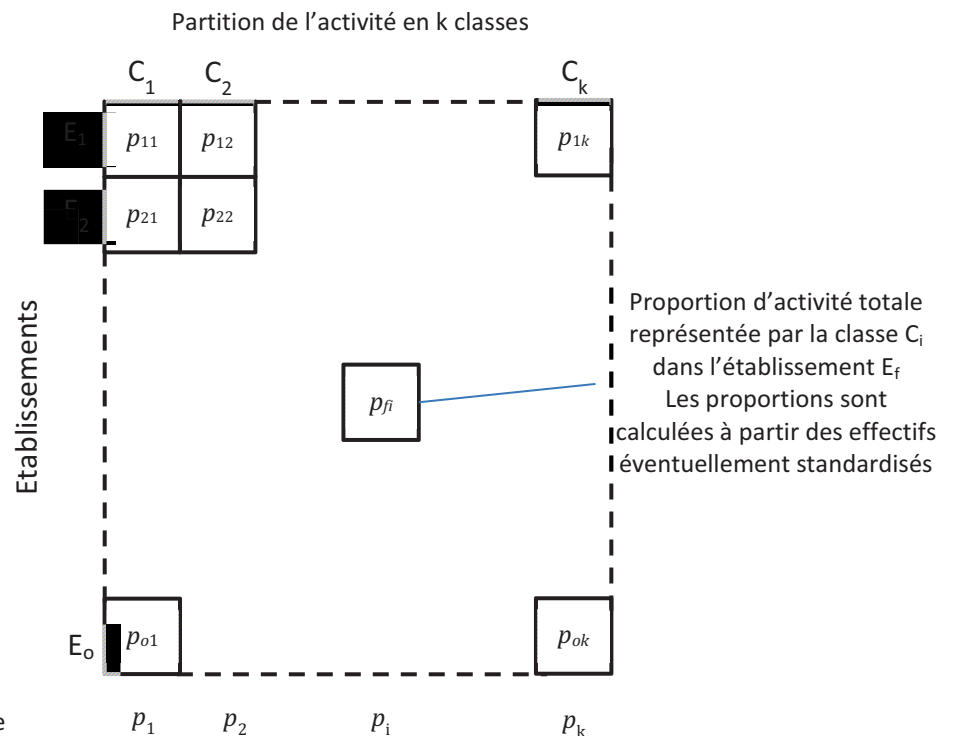
3.2.3. Standardisation de l'entropie

Nous avons vu que la matrice d'activité est une classification hiérarchique à 2 niveaux, dont l'entropie est donnée par $H_{ST} = \sum_{i=1}^{k_1} p_{1,i} [\log_2(p_{1,i}) + H_S(C_i)]$, avec $H_S(C_i)$ entropie interne de la classe C_i .

Si l'on prend comme niveau 1 de la classification le partitionnement en classes, on voit alors que les termes $p_{1,i} [\log_2(p_{1,i}) + H_S(C_i)]$ représentent la contribution de chaque classe d'activité, ou prise en charge, à l'entropie totale.

C'est cette quantité que nous retiendrons dans notre analyse, et que nous appellerons, pour des raisons de commodité, l'entropie de la prise en charge.

En effet, ceci permet de prendre en compte la fréquence de la prise en charge dans l'activité et pourrait, éventuellement, compenser une faible entropie interne due à la concentration sur peu d'établissements d'une prise en charge par ailleurs très fréquente. C'est pourquoi nous considérons cette démarche comme une « standardisation » de l'entropie.



Entropie interne de C_i

$$H_S(C_i) = - \sum_{g=1}^o p_{g,i} \log_2(p_{g,i})$$

Contribution à l'entropie totale de C_i

$$CH_i = p_i [\log_2(1/p_i) + H_S(C_i)]$$

Entropie totale de la matrice d'activité

$$H_{ST} = \sum_{i=1}^k CH_i$$

Proportion de contribution à l'entropie de la matrice d'activité

$$pCH_i = CH_i / H_{ST}$$

Figure 12 - Matrice d'activité et contribution à l'entropie des classes d'activité

3.3. Résultats

Nous avons calculé la contribution à l'entropie, en valeur absolue et relative, des racines de GHM pour l'année 2014 (645 racines de GHM hors séance).

La figure suivante reprend les différentes étapes de l'analyse.

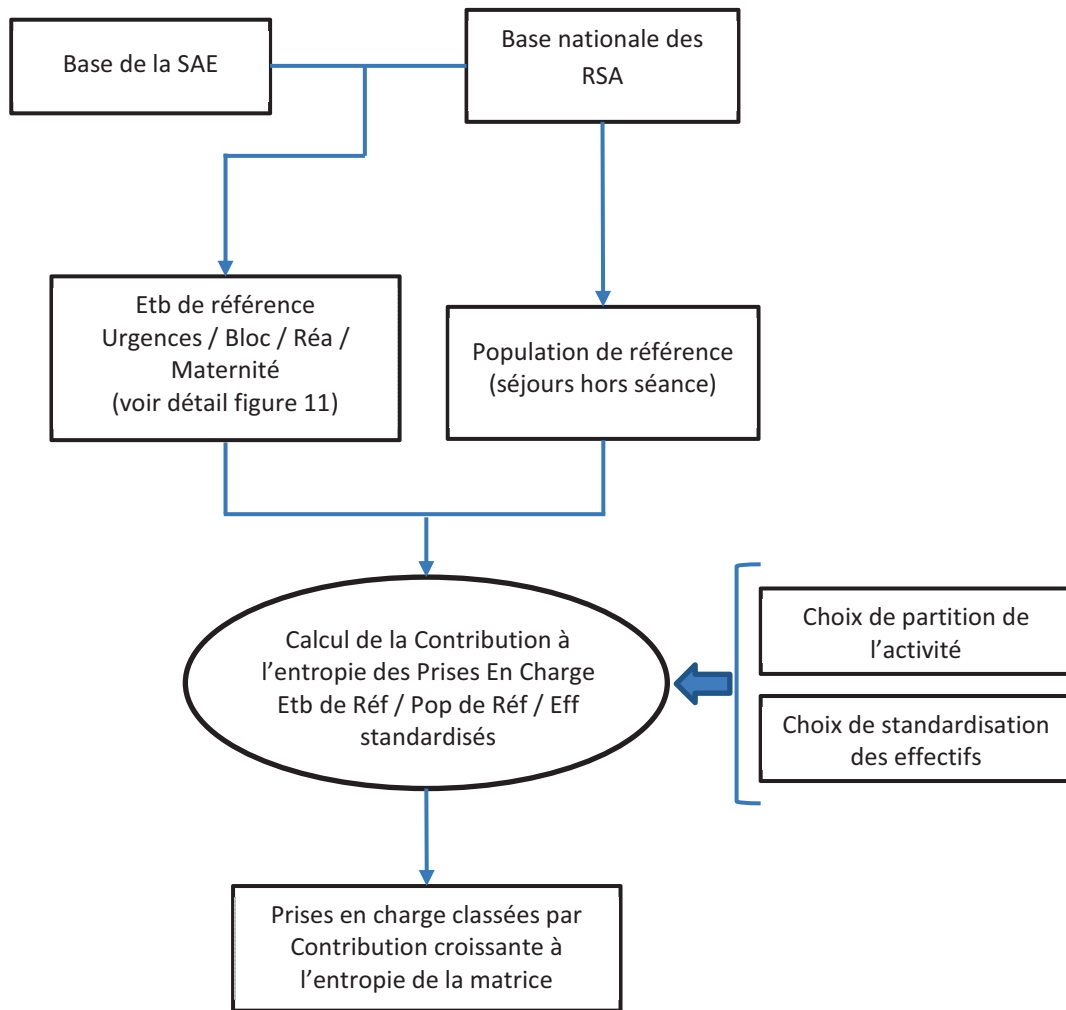


Figure 13 - Entropie des prises en charge - Plan d'analyse

L'entropie totale de la matrice d'activité est de 15,61.

Nous présentons dans le tableau suivant les 25 premières racines classées par ordre croissant de contribution à l'entropie totale. La liste complète est donnée en annexe V.

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale
25Z03	Maladies dues au VIH, âge inférieur à 13 ans	1,88E-06	1,88E-06
06M15	Suivi de greffes de l'appareil digestif	2,10E-06	3,98E-06
05K18	Varices avec acte autre que ligature et éveinage, en ambulatoire	3,85E-06	7,83E-06
27C03	Transplantations pancréatiques	5,43E-06	1,33E-05
07K05	Séjours comprenant une endoscopie biliaire thérapeutique ou diagnostique sans anesthésie, en ambulatoire	9,17E-06	2,24E-05
05C09	Autres interventions cardiothoraciques, âge inférieur à 2 ans, sans circulation extracorporelle	1,39E-05	3,63E-05
13M08	Assistance médicale à la procréation	1,57E-05	5,21E-05
15C03	Interventions majeures sur l'appareil cardiovasculaire, groupes nouveau-nés 1 à 7	1,87E-05	7,08E-05
15C05	Interventions chirurgicales, groupes nouveau-nés 8 à 9	2,05E-05	9,13E-05
27C04	Transplantations pulmonaires	2,12E-05	1,12E-04
27Z04	Greffes de cellules souches hématopoïétiques, en ambulatoire	2,44E-05	1,37E-04
12K03	Séjours de la CMD 12 comprenant une endoscopie génito-urinaire sans anesthésie : séjours de moins de deux jours	2,68E-05	1,64E-04
12C09	Autres interventions pour tumeurs malignes de l'appareil génital masculin	2,68E-05	1,91E-04
25Z02	Maladies dues au VIH, avec décès	2,70E-05	2,18E-04
12M09	Symptômes et autres recours aux soins de la CMD 12	2,84E-05	2,46E-04
09M15	Symptômes et autres recours aux soins concernant les affections des seins	2,84E-05	2,74E-04
22Z02	Brûlures étendues	2,84E-05	3,03E-04
16M18	Autres affections hématologiques concernant majoritairement la petite enfance	2,92E-05	3,32E-04
15C06	Interventions chirurgicales, groupe nouveau-nés 10	2,93E-05	3,61E-04
27C05	Transplantations cardiaques	2,94E-05	3,91E-04
11M20	Autres affections uronéphrologiques concernant majoritairement la petite enfance	3,17E-05	4,22E-04
22C03	Brûlures non étendues avec parages de plaie ou autres interventions chirurgicales	3,41E-05	4,57E-04
01M38	Autres affections neurologiques concernant majoritairement la petite enfance	3,94E-05	4,96E-04
14C06	Césariennes avec naissance d'un mort-né	4,16E-05	5,38E-04
07C07	Interventions diagnostiques sur le système hépato-biliaire et pancréatique pour affections non malignes	4,19E-05	5,80E-04

Tableau des 25 dernières racines classées par ordre croissant de contribution à l'entropie totale :

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale
05M08	Arythmies et troubles de la conduction cardiaque	0,0960	6,15E-03
05C17	Ligatures de veines et éveinages	0,0971	6,22E-03
16M11	Autres troubles de la lignée érythrocytaire, âge supérieur à 17 ans	0,1006	6,45E-03
14M03	Affections de l'ante partum sans intervention chirurgicale	0,1022	6,55E-03
23M06	Autres facteurs influant sur l'état de santé	0,1039	6,65E-03
06C12	Interventions réparatrices pour hernies inguinales et crurales, âge supérieur à 17 ans	0,1041	6,67E-03
01C15	Libérations du médian au canal carpien	0,1105	7,08E-03
08C44	Autres interventions sur la main	0,1119	7,17E-03
23K02	Explorations nocturnes et apparentées : séjours de moins de 2 jours	0,1149	7,36E-03
05K06	Endoprothèses vasculaires sans infarctus du myocarde	0,1161	7,44E-03
04M05	Pneumonies et pleurésies banales, âge supérieur à 17 ans	0,1223	7,83E-03
14C08	Césariennes pour grossesse unique	0,1310	8,39E-03
06M03	Autres gastroentérites et maladies diverses du tube digestif, âge supérieur à 17 ans	0,1314	8,42E-03
17M06	Chimiothérapie pour autre tumeur	0,1347	8,63E-03
05K10	Actes diagnostiques par voie vasculaire	0,1492	9,56E-03
14Z08	Interruptions volontaires de grossesse : séjours de moins de 3 jours	0,1535	9,83E-03
05M09	Insuffisances cardiaques et états de choc circulatoire	0,1706	1,09E-02
23M20	Autres symptômes et motifs de recours aux soins de la CMD 23	0,2070	1,33E-02
14Z13	Accouchements uniques par voie basse chez une primipare	0,2126	1,36E-02
03K02	Affections de la bouche et des dents avec certaines extractions, réparations et prothèses dentaires	0,2307	1,48E-02
14Z14	Accouchements uniques par voie basse chez une multipare	0,2853	1,83E-02
06K02	Endoscopies digestives thérapeutiques et anesthésie : séjours de moins de 2 jours	0,3301	2,11E-02
02C05	Interventions sur le cristallin avec ou sans vitrectomie	0,5133	3,29E-02
15M05	Nouveau-nés de 3 300 g et âge gestationnel de 40 SA et assimilés (groupe nouveau-nés 1)	0,5178	3,32E-02
06K04	Endoscopie digestive diagnostique et anesthésie, en ambulatoire	0,6417	4,11E-02

Analyse des résultats

Les racines classées en premier correspondent dans leur majorité à des prises en charge que l'on classerait intuitivement parmi les plus spécifiques, les racines classées en dernier à des prises en charge que l'on classerait intuitivement parmi les moins spécifiques

Certaines racines qui ne semblent pas particulièrement spécifiques apparaissent parmi les 25 premières :

La racine 05K18 - Varices avec acte autre que ligature et éveinage, en ambulatoire, ne semble pas particulièrement spécifique et apparaît pourtant en 3ème position. Après analyse, on constate que cette racine disparaît à compter de mars 2014. Son effectif pour l'année 2014 ne se monte alors qu'à 50 séjours, ce qui explique son classement. Elle peut donc être éliminée de la liste.

Quand on attribue aux racines de GHM ordonnées un numéro d'ordre normé à 1 et qu'on représente la proportion cumulée de contribution à l'entropie en fonction des racines de GHM, on obtient une courbe d'aspect exponentiel. De fait, on peut linéariser les données par une transformation en $-1/\log(x)$ de la proportion de contribution avant de réaliser le cumul de ces valeurs transformées, normées à 1.

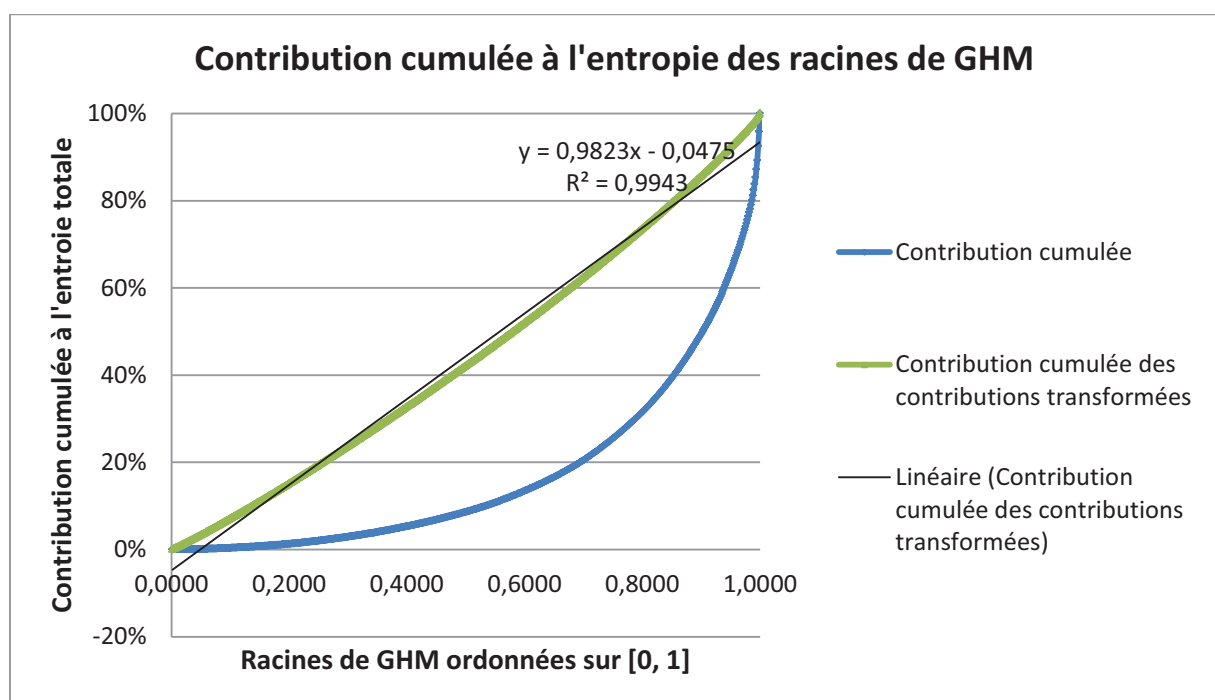


Figure 14 - Contribution cumulée à l'entropie des racines de GHM

Stabilité des résultats

Nous avons vu qu'il est possible de standardiser les effectifs de classes en faisant varier ce que nous avons appelé la dilution des effectifs, en fonction de choix de regroupement de classes. On peut donc examiner la stabilité des résultats lorsque la dilution change.

Nous avons donc recalculé la contribution des racines de GHM en adoptant deux autres critères de standardisations des effectifs de classe : le type de séjours (durée de séjour

répartie en 3 classes : 0 jour, 1 jour, 2 jours et plus) et la discipline d'activité (DA° selon le regroupement de GHM de l'ATIH.

On observe une très grande concordance des résultats avec une standardisation selon le type de séjours, relativement moins bonne avec une standardisation selon les DA.

On peut représenter les discordances en valeur absolue entre classement par les tableaux suivants :

Discordance entre classement initial et classement avec standardisation sur le type de séjour

Ecart (valeur absolue)	Nombre de racines	Pourcentage	Pourcentage cumulé
0	340	52,71%	52,71%
1	218	33,80%	86,51%
2	59	9,15%	95,66%
3	14	2,17%	97,83%
4	7	1,09%	98,91%
5	4	0,62%	99,53%
6	2	0,31%	99,84%
12	1	0,16%	100,00%

Discordance entre classement initial et classement avec standardisation sur la DA

Ecart (valeur absolue)	Nombre de racines	Pourcentage	Pourcentage cumulé
0	204	31,63%	31,63%
1	220	34,11%	65,74%
2	121	18,76%	84,50%
3	45	6,98%	91,47%
4	19	2,95%	94,42%
5	11	1,71%	96,12%
6	8	1,24%	97,36%
7	5	0,78%	98,14%
8	3	0,47%	98,60%
9	3	0,47%	99,07%
10	2	0,31%	99,38%
11	1	0,16%	99,53%
13	1	0,16%	99,69%
14	2	0,31%	100,00%

On voit que dans le premier cas plus de 50% des classements sont concordants (52,71%), et dans 95% des cas (95,66%) ne divergent au plus que de 2 rangs (sur 645 possibles).

Dans le second cas, les classements concordants ne représentent plus que 32% des cas (31,63%), et que l'on atteint 95% des cas (96,12%) pour une divergence d'au plus 5 rangs.

3.4.Discussion

Intérêt

Nous considérons que même si dans le cas de la standardisation sur les DA le degré de concordance est moindre, il reste tout de même satisfaisant et vient conforter, avec le celui en standardisation sur les types de séjour, le classement initial.

Cette méthode présente donc l'intérêt d'être stable, standardisée et maîtrisée, reproductible chaque année afin de prendre en compte l'évolution des pratiques et la banalisation des techniques.

Elle permet de faire varier le niveau de description des prises en charges. Dans la mesure où la partition de l'activité remplit les bonnes propriétés (classes disjointes, leur union recouvrant toute l'activité étudiée), on peut envisager d'affiner la description des prises en charge, voire d'utiliser une description non basée sur le GHM, en associant par exemple les diagnostics principaux. Ceci pourrait permettre de prendre en compte des racines de GHM recouvrant une diversité importante de pathologies.

Limites

Le degré de fiabilité et de précision des données disponibles dans la SAE ne nous permet malheureusement pas d'affiner la standardisation des données. Il serait intéressant, par exemple, de pouvoir standardiser les effectifs sur les capacités en lits les concernant.

De plus, cette méthode reposant sur les caractéristiques statistiques des classes de prise en charge, elle est nécessairement dépendante des choix opérés en termes de partition de l'activité. Un soin particulier devra être apporté à cette phase, qui doit s'appuyer sur une vision experte de l'activité. D'autre part, le choix de la partition devra également se faire en tenant compte des effectifs. En effet, l'expérience nous a montré qu'une partition trop fine, conduisant à un grand nombre de classes de faibles effectifs, produit des résultats difficilement exploitables.

La lecture des résultats devra également se faire en s'appuyant sur une ressource experte afin d'apporter les précisions, voire les modifications, nécessaires à la pertinence de leur utilisation.

Chapitre 4 Caractérisation de l'activité des établissements

1. Introduction

Le classement de l'activité selon son niveau de référence peut utilement être appliqué à la caractérisation du profil d'activité des établissements. Cette question est étroitement liée à celle du niveau de référence de l'activité. Elle constitue un thème d'étude récurrent [76,87] car elle est essentielle dans la visibilité que le décideur ou le régulateur peut avoir du paysage sanitaire et de son évolution.

En première approche, cette question peut assez rapidement amener à un raisonnement auto-référent : caractériser l'activité en fonction de la nature des établissements où elle est réalisée (par exemple, dans les CHU), puis s'appuyer sur les résultats obtenus pour caractériser les établissements.

Nous avons au chapitre précédent bâti une méthode permettant de caractériser l'activité indépendamment de la nature des établissements. Nous allons donc à présent utiliser celle-ci pour caractériser l'activité des établissements.

2. Méthode

Dans un premier temps, nous sélectionnons les établissements susceptibles d'être analysés. Nous comparons ensuite la distribution de l'activité par racines de GHM entre établissements et nous appliquons une méthode de classification pour regrouper les établissements en groupes présentant des caractéristiques similaires.

2.1. Etablissements sélectionnés

Nous avons retenu pour l'année 2014, dans la base de la SAE, les établissements (Etb SAE) répondant aux critères suivants :

- en activité toute l'année, avec transmission des données d'activité pour toute l'année ;
- présence d'un bloc opératoire ;
- présence d'une réanimation.

On obtient une liste de 689 Etb SAE, identifiés par leur numéro FINESS d'entité juridique, que nous associons ensuite à 779 Etb BN (sur un total de 1 974) issus de la base. 52 Etb SAE sont associés à plusieurs Etb BN, représentant un total de 142 Etb BN.

Nous conservons les établissements représentant 99% du nombre total de séjours hors séance (NbSejHS) de la liste complète. Pour l'année 2014, le seuil est de 5 568 séjours hors séance.

L'application de ce filtre aboutit à retenir 653 Etb SAE, représentant 743 Etb BN. Les 653 Etb SAE retenus incluent les 52 Etb SAE multi-Etb BN.

2.2. Le découpage de l'activité en classes

L'activité, représentée en racines de GHM, peut être ordonnée selon sa contribution à l'entropie totale. Nous pouvons alors découper l'activité en classes équivalentes en termes de contribution à l'entropie.

L'activité est représentée par 645 racines de GHM. De plus on constate un ordre de grandeur entre contributions aux deux extrémités du spectre d'activité de l'ordre de 10^5 :

- contribution de la première racine de GHM : $2,48 \cdot 10^{-5}$ soit $1,86 \cdot 10^{-3}\%$ de l'entropie totale ;
- contribution de la dernière racine de GHM : 0,55, soit 4,12% de l'entropie totale.

Le découpage de l'activité en classes devra satisfaire aux contraintes suivantes :

- éviter un émiettement de l'activité des établissements sur un trop grand nombre de classes ;
- prendre en compte l'ordre de grandeur de la contribution des dernières classes par rapport à celle des premières ;
- comporter un nombre de classes suffisamment important pour faire ressortir les différences entre établissements.

Nous choisissons ainsi un découpage en 10 classes d'activité équi-entropiques, dans lesquelles les racines se répartissent de la façon suivante :

Classe	Nombre de Racines
1	344
2	104
3	60
4	42
5	32
6	23
7	18
8	12
9	7
10	3

A titre d'illustration, les classes 9 et 10 contiennent les racines suivantes :

Classe 9

- 14Z08 Interruptions volontaires de grossesse : séjours de moins de 3 jours
- 05M09 Insuffisances cardiaques et états de choc circulatoire
- 23M20 Autres symptômes et motifs de recours aux soins de la CMD 23
- 14Z13 Accouchements uniques par voie basse chez une primipare
- 03K02 Affections de la bouche et des dents avec certaines extractions, réparations et prothèses dentaires
- 14Z14 Accouchements uniques par voie basse chez une multipare
- 06K02 Endoscopies digestives thérapeutiques et anesthésie : séjours de moins de 2 jours

Classe 10

02C05	Interventions sur le cristallin avec ou sans vitrectomie
15M05	Nouveau-nés de 3 300 g et âge gestationnel de 40 SA et assimilés (groupe n-nés 1)
06K04	Endoscopie digestive diagnostique et anesthésie, en ambulatoire

2.3.Stratégie de classification

Sur la base des établissements retenus et du découpage de l'activité en classes, nous constituons une matrice comportant en ligne les établissements et en colonne leur proportion d'activité dans les différentes classes.

La nature de la variable étudiée est particulière : plutôt que des coordonnées dans un espace multidimensionnel, les proportions d'activité dans les classes représentent d'avantage plusieurs aspects, volontairement segmentés, d'un même attribut, à savoir le profil d'activité de l'établissement.

Dès lors, la classification que nous cherchons à opérer repose moins sur une proximité quantitative que qualitative, portant sur une similarité dans la nature des établissements, repérée par la comparaison du profil d'activité exprimé sur les dix classes d'entropie croissante. C'est pourquoi nous choisissons de procéder à une classification hiérarchique ascendante. Cette classification est réalisée en utilisant le logiciel R.

La figure 15 reprend les différentes étapes de l'analyse.

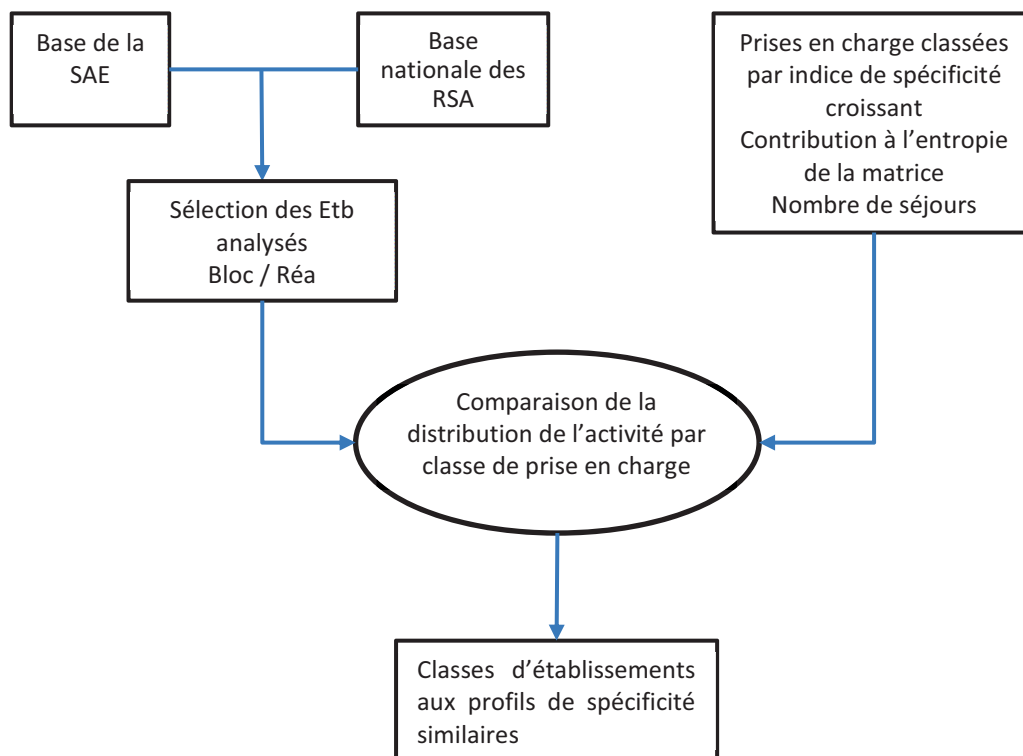


Figure 15 - Caractérisation de l'activité des établissements - Plan d'analyse

3. Analyse préalable et détermination des paramètres de la classification

La réalisation d'une classification hiérarchique nécessite de faire trois choix :

- distance employée ;
- critère d'agrégation des groupes ;
- nombre de groupes retenu.

Afin d'effectuer ces choix, nous étudions tout d'abord les données.

3.1. Caractéristiques des données

Les paramètres descriptifs des classes d'activité sont les suivants :

	Classes									
	1	2	3	4	5	6	7	8	9	10
Minimum	2,16	2,81	3,01	2,19	2,04	0,48	1,64	1,14	0,03	0
1er Quartile	4,88	6,32	7,34	7,74	7,72	6,68	9,3	7,15	9,46	8,26
Médiane	6,37	7,51	8,67	9,18	9,9	9,57	11,06	8,8	11,71	12,73
Moyenne	6,98	7,94	8,59	9,07	9,72	9,42	11,34	9,67	11,95	15,33
Ecart-type	3,23	2,59	2,09	2,14	3,01	3,41	3,66	4,51	4,28	9,02
3ème Quartile	7,87	8,92	9,78	10,19	11,31	12,09	13,13	11,18	14,18	21,92
Maximum	31,45	23,21	22,44	25,42	28,99	23,57	41,07	46,78	33,4	45,63
Coeff Varia	0,46	0,33	0,24	0,24	0,31	0,36	0,32	0,47	0,36	0,59

Ce tableau se lit, par exemple pour la classe 1 de la façon suivante : parmi les établissements retenus pour l'analyse, le pourcentage minimum d'activité dans la classe 1 est de 2,16%, avec une médiane de 6,37%, une moyenne est de 6,98% et un écart-type de 3,23% ; le maximum est de 31,45% et l'écart inter-quartile entre le premier et le troisième quartile est de $(7,87 - 4,88) = 2,99$.

Les données du tableau sont représentées figure 16.

On peut constater :

- une proportion d'activité croissante, avec un pic local dans la classe 7 ;
- une baisse progressive du coefficient de variation jusqu'aux classes 6 - 7, suivie d'une augmentation très importante sur les trois dernières classes ;
- un minimum à 0 ou presque pour les classes 9 et 10, qui sont celles de l'activité la moins spécifique, tandis que le minimum de la classe 1 est de 2,16%. Cette apparente asymétrie est en réalité inversée : alors que les classes 9 et 10 comportent seulement 10 racines (dont 4 relevant de l'obstétrique et de la néonatalogie, et deux portant sur les endoscopies digestives), la classe 1 en comporte 344, couvrant toutes les disciplines. De plus, le premier quartile de la classe 1 est pratiquement inférieur de moitié à celui de la classe 10.

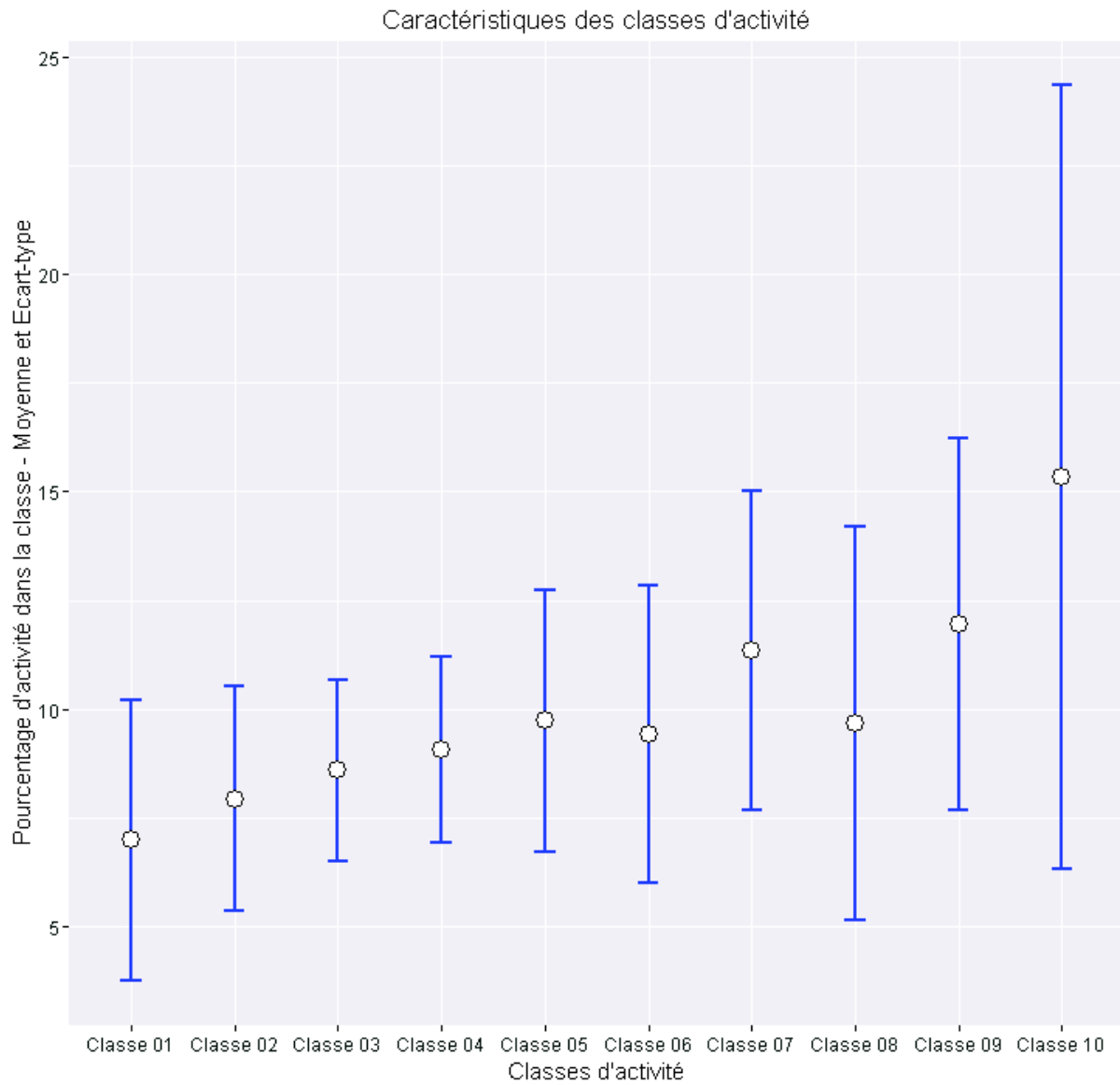


Figure 16 - Population des établissements - Caractéristique des classes d'activité

Nous étudions à présent la distribution du nombre d'établissement en fonction du pourcentage d'activité dans chaque classe :

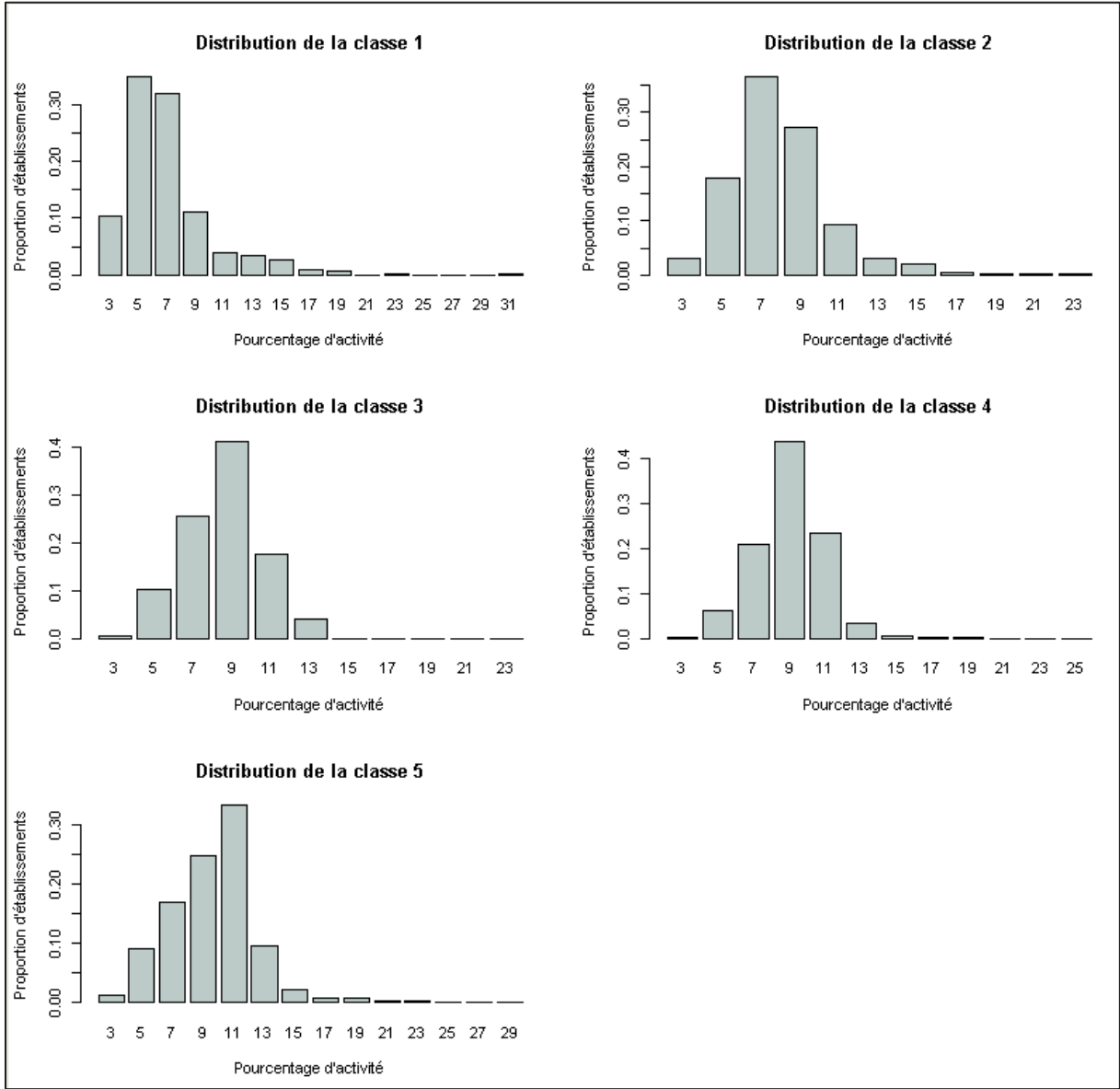


Figure 17 - Nombre (proportion) d'établissements en fonction du pourcentage d'activité - Classes 1 à 5

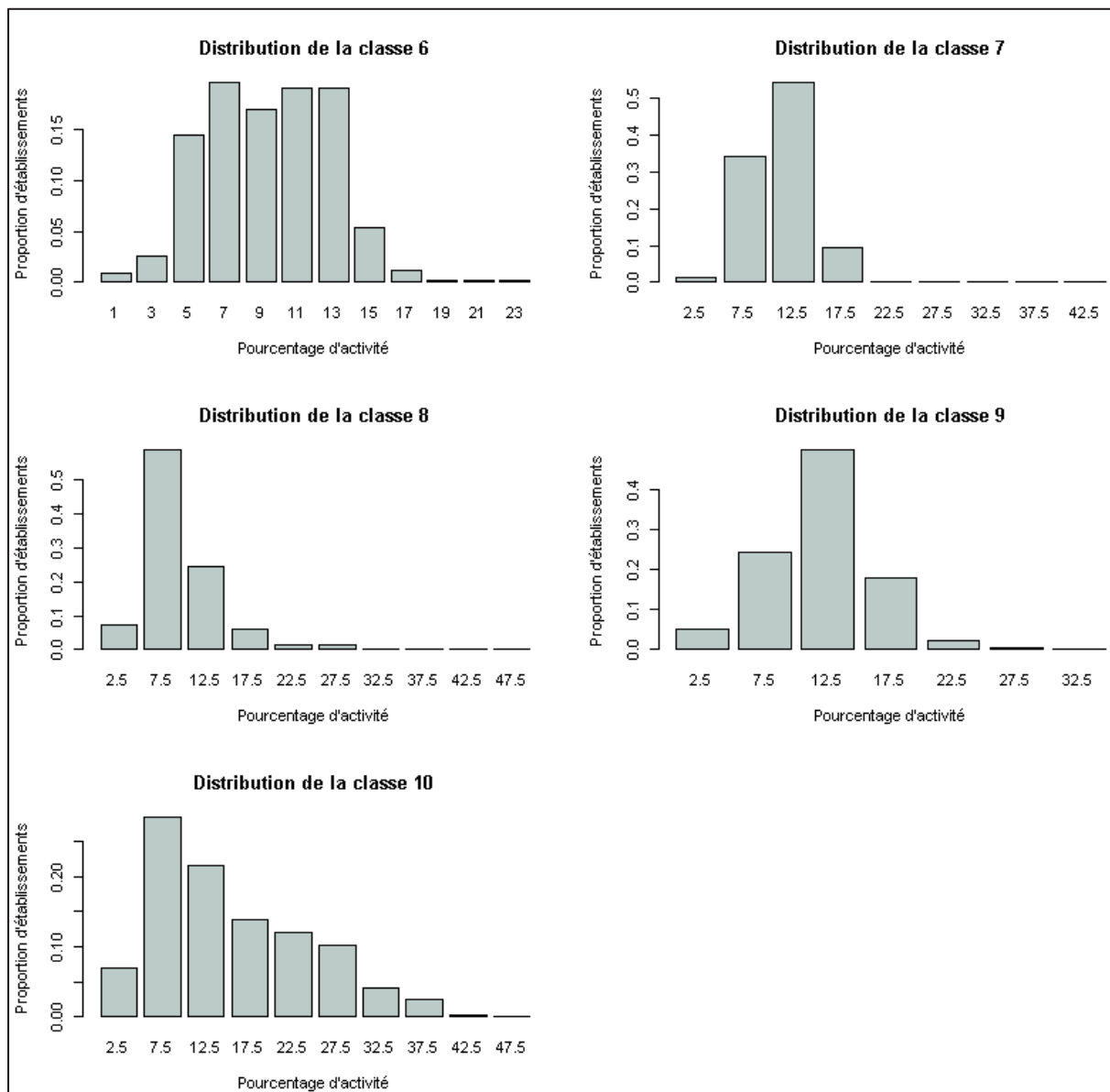


Figure 18 - Nombre (proportion) d'établissements en fonction du pourcentage d'activité - Classes 1 à 5

Ici on notera essentiellement :

- la distribution très étroite autour de petites valeurs de la classe 1, qui semble logique compte tenu de la spécificité plus importante de cette classe d'activité ;
- l'aspect de la classe 6, qui pourrait presque évoquer une bimodalité malgré son faible coefficient de variation ;
- l'aspect symétrique des classes 5 et 10 : croissance régulière puis décroissance brutale pour la classe 5, croissance rapide puis décroissance régulière pour la classe 10

3.2.Choix des paramètres de la classification hiérarchique

3.2.1. Hypothèse de travail

Les établissements présentent une diversité de taille et de types (public, privés, Centre de Lutte Contre le Cancer, Centre Hospitalier Universitaire). De plus, on observe dans le paysage sanitaire une diversité de besoins de santé et de stratégies de réponse à ces besoins. Ces stratégies peuvent être régulées (par exemple via les Schéma Régionaux d'Organisation Sanitaire) ou non (choix d'activité des opérateurs).

D'après ces observations et notre connaissance empirique du paysage sanitaire, nous formulons l'hypothèse d'un axe de spécialisation croisé avec un axe d'étendue du spectre d'activité, les deux axes étant indépendants : il est par exemple possible qu'un établissement couvre un large spectre d'activité, avec une spécialisation (un pic de fréquence) dans un ou plusieurs segments particuliers.

Enfin, compte tenu de la diversité des établissements et des situations, nous pouvons également considérer que l'on s'attend à trouver des groupes d'effectifs relativement différents : l'hypothèse d'une différence de profil d'activité intègre celle d'au moins un groupe d'un petit nombre d'établissements réalisant de l'activité très spécifique en proportion plus élevée.

Au vu de ces éléments, nous posons l'hypothèse de quatre à six groupes d'établissements distincts.

3.2.2. Distance, critère d'agrégation et indicateur de qualité

Distance

Compte tenu de la nature des données (distribution multinomiale), nous avons choisi en première intention la distance de Hellinger. Nous avons ensuite utilisé la distance euclidienne pour comparer les résultats obtenus et observer leur stabilité.

Méthodes d'agrégation

Nous avons retenu le critère de Ward [88]. Ce critère est basé sur la minimisation de la variance intra-groupe à chaque étape d'agrégation.

Critères de qualité de la classification

Il existe de nombreux critères d'évaluation d'une classification hiérarchique. Ils combinent de façons variées trois types de quantités :

- les variances intra et inter-groupe ;
- les distances intra et inter-groupe (minimum, maximum, moyenne, au barycentre...);
- des indices qualitatifs basés sur les distances.

Nous avons retenu les deux indices suivants :

- **Indice de Dunn**[89]. Cet indice est basé sur un rapport entre un indicateur de distance inter-groupe et un indicateur de distance intra-groupe :

La distance inter- groupe est représentée par la plus petite des distances inter-groupe minimales (distance minimale entre deux points i et j de groupes différents k et k') :

$$d_{min} = \min_{k,k'} D_{k,k'} \text{ avec } D_{k,k'} = \min_{i \in G_k, j \in G_{k'}} D_{i,j}$$

La distance intra-groupe est représentée par la plus grande des distances intra-groupe maximales (distance maximale entre deux points i et j de groupes différents k et k') :

$$d_{max} = \max_k d_k \text{ avec } d_k = \max_{i,j \in G_k, i \neq j} d_{i,j}$$

L'indice de Dunn est alors défini par

$$I = \frac{d_{min}}{d_{max}}$$

- **Indice Gamma**[90]. Ici, il s'agit de rendre compte du bon ordonnancement des points dans les groupes :

- On compte tous les cas où une paire de points appartenant au même groupe est séparée par une distance inférieure à celle d'une paire de points n'appartenant pas au même groupe. On note cette quantité s^+ .
- On compte tous les cas où une paire de points appartenant au même groupe est séparée par une distance supérieure à celle d'une paire de points n'appartenant pas au même groupe. On note cette quantité s^- .
- Les ex-aequo sont éliminés.

L'indice Gamma est alors défini par

$$I = \frac{s^+ - s^-}{s^+ + s^-}$$

Nous avons choisi ces deux indices car ils nous paraissent être les moins à même d'être perturbés par la variabilité des données, notamment l'indice Gamma. L'indice de Dunn sera en revanche affecté par les valeurs extrêmes. De plus, son calcul repose sur l'utilisation de la distance euclidienne, ce qui perturbera également les résultats.

Critère de comparaison entre classifications

Nous comparons les résultats obtenus entre distance de Hellinger et distance euclidienne avec l'indice de Rand.

Cet indice fait le rapport entre les paires de points classées de façon concordante par les deux méthodes.

Si 'on note yy le nombre de paires de points classées dans le même groupe par les deux méthodes, et nn le nombre de paires de points non classées dans le même groupe par les deux méthodes, et Nt le nombre total de paires, l'indice de Rand est alors défini par

$$I = \frac{yy + nn}{Nt}$$

Cet indice varie entre 0 (aucune concordance) et 1 (concordance maximale).

4. Résultats

4.1. Détermination des classifications d'intérêt

Nous avons successivement testé des classifications allant de 4 à 10 groupes, et calculé à chaque fois les indices de Dunn, Gamma et Rand correspondants (en gras, les maximums et les maximums locaux) :

Nb Groupes	Distance de Hellinger		Distance euclidienne		Rand
	Dunn	Gamma	Dunn	Gamma	
4	0,0732	0,7089	0,0732	0,7516	0,8672
5	0,0836	0,7187	0,0732	0,7594	0,9189
6	0,0761	0,7575	0,0719	0,8005	0,8990
7	0,0859	0,8062	0,0864	0,8130	0,9036
8	0,0578	0,6818	0,0581	0,7049	0,8855
9	0,0578	0,7140	0,0581	0,7419	0,8940
10	0,0578	0,7347	0,0581	0,7504	0,9011

Lorsque l'on examine les effectifs par groupes pour les classifications allant de 4 à 7 groupes, on observe les résultats suivants :

Nb Groupes	Effectif par groupe						
	1	2	3	4	5	6	7
4	102	199	296	56	0	0	0
5	102	199	296	21	35	0	0
6	102	123	296	76	21	35	0
7	102	123	288	76	21	35	8

Avec les transitions suivantes :

4 à 5 groupes

Groupe Ini \ Fin	1	2	3	4	5
1	102				
2		199			
3			296		
4				21	35

5 à 6 groupes

Groupe Ini \ Fin	1	2	3	4	5	6
1	102					
2		123		76		
3			296			
4					21	
5						35

6 à 7 groupes

Groupe Ini \ Fin	1	2	3	4	5	6	7
1	102						
2		123					
3			288				8
4				76			
5					21		
6						35	

A l'analyse des transitions, nous constatons les éléments suivants :

- le groupe 1 (102 établissements), formé dès la première classification, reste stable tout au long des transitions ;
- le groupe 4 (56 établissements) se scinde immédiatement en 2 groupes qui ne changent plus ensuite ;
- dans la transition de 5 à 6 groupes, le nouveau groupe (76 établissements) se forme par séparation du groupe 2 qui en comportait précédemment 199, ce qui représente 38% de son effectif ;
- le passage de 6 à 7 groupes s'opère par le détachement de 8 établissements seulement à partir du groupe 3, présent et stable depuis la classification en 4 groupes.

Ces 8 établissements sont les suivants :

FI_EJ	FI_BN	Etablissement	Statut
210000030	210780136	CLINIQUE MEDICO CHIR DE CHENOVE	Privé
310788799	310780150	CLINIQUE DU COURS DILLON	Privé
420787061	420010050	CLINIQUE MUTUALISTE	Public (Dotation globale)
440000941	440000412	CLINIQUE BRETECHE VIAUD	Privé
590781621	590781621	CENTRE HOSPITALIER DE LE CATEAU	Public
620003814	620000026	ETAB HOPALE CTRE CALOT/HELIO	Public
710000464	710781824	CENTRE ORTHOPEDIQUE MEDICO-CHIR	Privé
750001067	750301160	CLINIQUE DES MAUSSINS	Privé

L'examen du case-mix ordonné en classes d'entropie croissante de ces établissements montre des profils très particuliers qui concentrent entre 60 et 80% de leur activité sur les classes 3 à 7. Cette activité est elle-même centrée sur quelques segments d'activité uniquement (chirurgie orthopédique, ORL, gynécologie...)

Le regroupement de ces 8 établissements, s'il améliore l'indice de qualité, ne nous paraît pas pertinent dans l'optique d'une classification globale des établissements. Nous allons donc écarter les classifications en 7 groupes et au-delà et étudier les classifications en 4 à 6 groupes.

4.2.Profils pour les classifications de 4 à 6 groupes

Les profils moyens d'activité pour les classifications de 4 à 6 groupes sont les suivants :

4 groupes

Groupe	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10
1	6,23	8,04	7,49	8,23	9,18	6,02	12,05	14,10	11,75	16,93
2	5,08	6,17	7,47	8,37	7,54	7,90	11,30	6,49	13,60	26,11
3	7,17	8,14	9,42	9,80	10,99	11,82	11,54	9,41	11,96	9,75
4	14,10	12,94	10,25	9,28	11,76	8,32	9,15	14,26	6,38	3,54

Nombre, type d'établissements et capacité

Nombre et Type d'établissements

Groupe	NbEtb	Privé	Public
1	102	94	8
2	199	179	20
3	296	18	278
4	56	5	51

Nombre Moyen de Lits

Global	Privé	Public
199,38	193,33	270,50
125,02	122,12	150,90
303,75	128,00	315,13
1 014,46	161,60	1 098,08

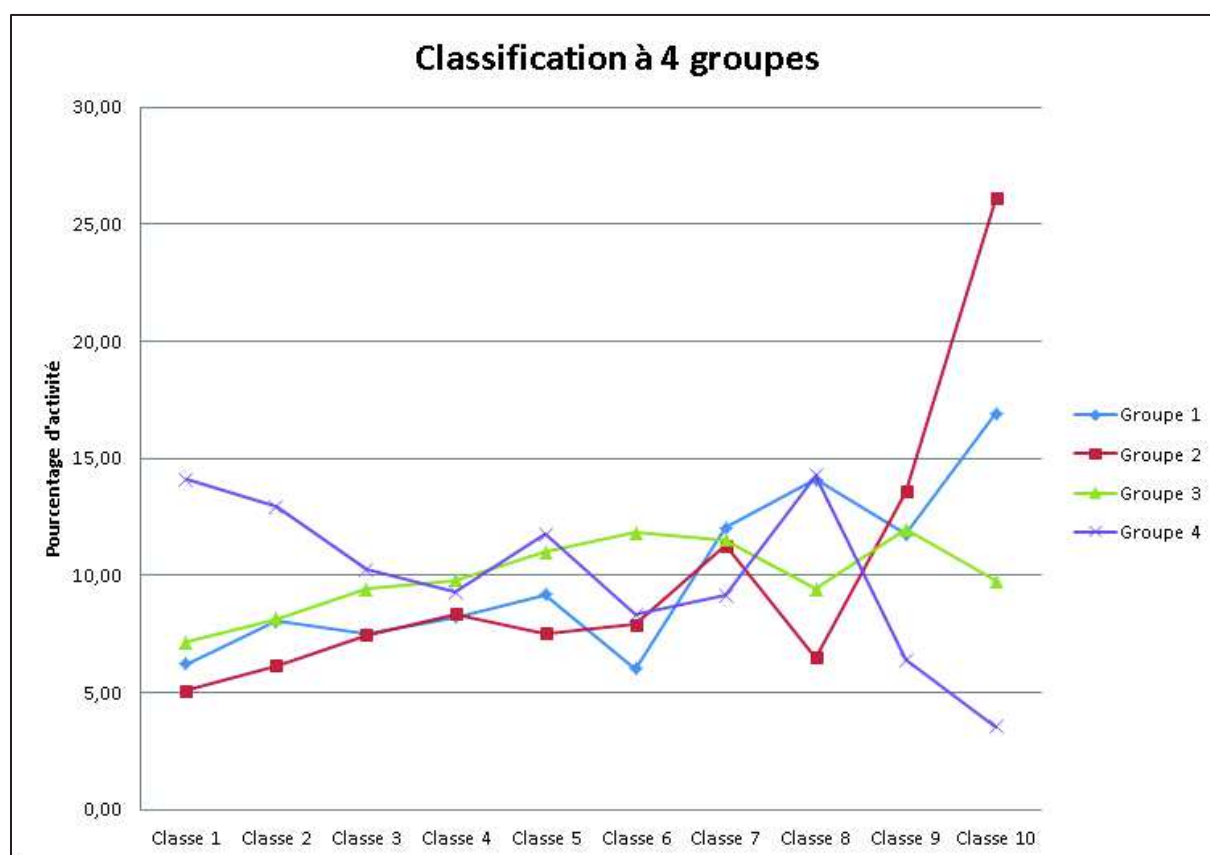


Figure 19 - Classification à 4 groupes - Profils moyens.

On peut immédiatement constater que les établissements publics et privés sont relativement séparés par la classification, avec deux groupes par type d'établissement. Les établissements privés sont essentiellement concentrés dans les groupes 1 et 2 :

- le groupe 1 présente un profil d'activité de proximité, couvrant surtout les dernières classes d'activité, avec des capacités de moyenne importance ;
- le groupe 2 dénote une forte spécialisation sur la dernière classe d'activité, on y trouve des établissements de taille plus faible.

Les établissements publics sont plutôt regroupés dans les groupes 3 et 4 :

- un groupe 3 polyvalent, couvrant un large spectre d'activité, sans véritable pic de spécialisation. Outre tous les Hôpitaux Inter Armées (HIA) sauf un, ce groupe comporte de grands CH et les CHU d'Orléans, Metz et Fort de France.

Enfin, un groupe 4 comportant une forte proportion d'activité très spécifique mais aussi un pic en classe 8, avec une part d'activité de proximité relativement plus faible. Ce groupe contient en particulier les Centre de Lutte Contre le Cancer (CLCC) et tous les autres CHU.

5 groupes

Groupe	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10
1	6,23	8,04	7,49	8,23	9,18	6,02	12,05	14,10	11,75	16,93
2	5,08	6,17	7,47	8,37	7,54	7,90	11,30	6,49	13,60	26,11
3	7,17	8,14	9,42	9,80	10,99	11,82	11,54	9,41	11,96	9,75
4	14,58	12,73	7,86	7,65	13,28	4,80	10,57	22,99	4,30	1,24
5	13,82	13,06	11,69	10,26	10,85	10,44	8,30	9,02	7,63	4,92

Nombre, type d'établissements et capacité

Nombre et type d'établissements

Groupe	NbEtb	OQN	STC
1	102	94	8
2	199	179	20
3	296	18	278
4	21	4	17
5	35	1	34

Nombre Moyen de Lits

Global	OQN	STC
199,38	193,33	270,50
125,02	122,12	150,90
303,75	128,00	315,13
166,10	153,25	169,12
1 523,49	195,00	1 562,56

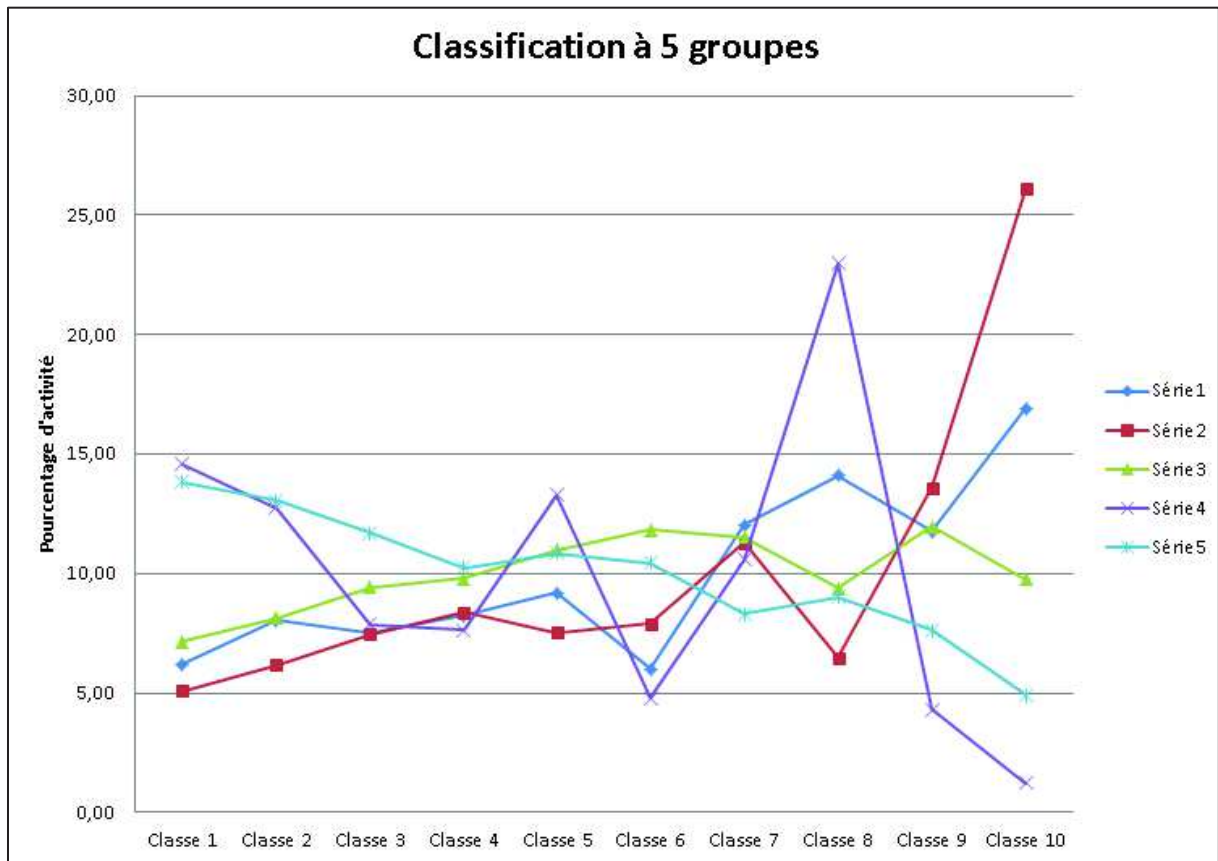


Figure 20 - Classification à 5 groupes - Profils moyens.

Le groupe 4 a généré le groupe 5 par séparation de 35 établissements :

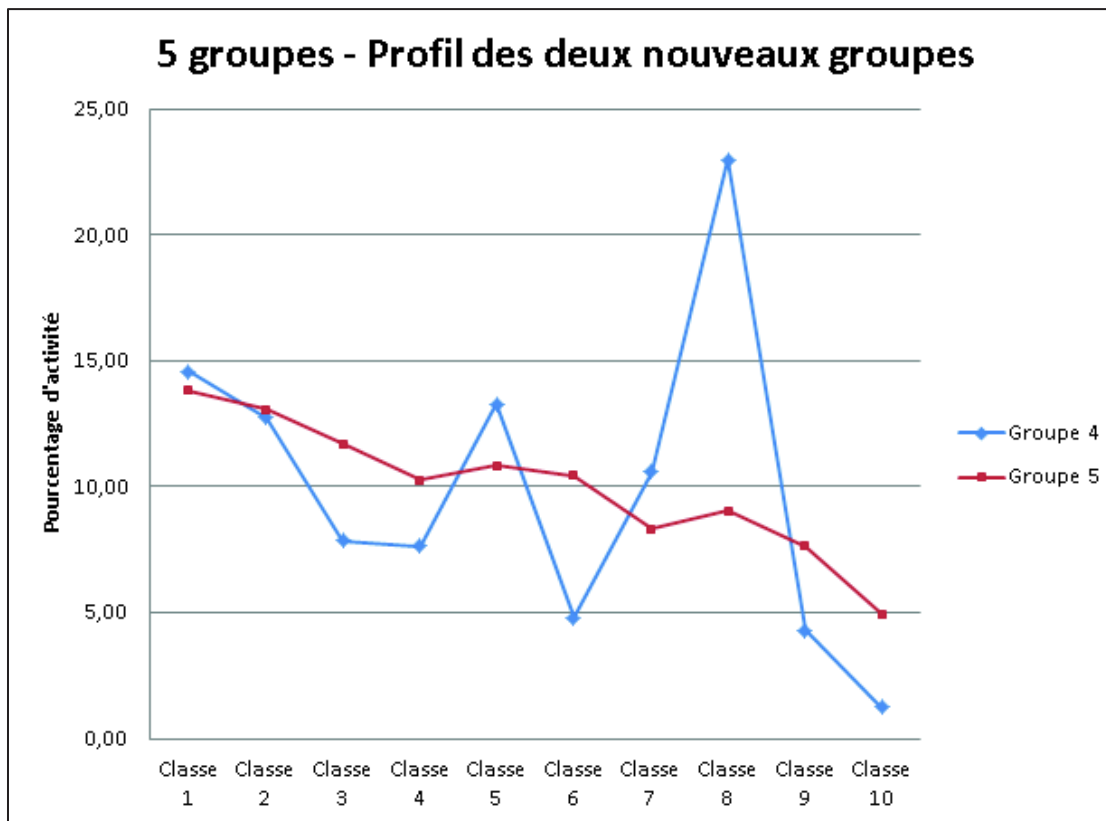


Figure 21 - Classification à 5 groupes - Profils moyens des 2 nouveaux groupes.

Sur 35 établissements, le nouveau groupe comporte 28 CHU, l'HIA Ste Anne (Toulon) et le CLCC Jean Godinot.

On y trouve également l'Hôpital Lentral, entièrement dédié aux enfants. Cet établissement présente le profil suivant :

Classe	Pourcentage d'activité
1	19,57%
2	16,37%
3	18,26%
4	8,83%
5	8,91%
6	17,71%
7	3,91%
8	2,51%
9	3,14%
10	0,80%

L'examen des racines de GHM de la classe 1 par exemple montre une activité sur toutes les racines avec âge inférieur à 18 ans, les pathologies congénitales, etc.

De la même manière, l'examen du case-mix du seul établissement privé présent dans ce groupe montre une pluri-spécialisation dans des activités très spécifiques : chirurgie de l'obésité, chirurgie cardiaque, médecine interventionnelle dans le traitement de la douleur...

Le groupe 4 conserve notamment tous les autres CLCC. La classe 8, pour laquelle on observe un pic pour les établissements de ce groupe, comporte en effet la racine 17M06 Chimiothérapie pour autre tumeur, ce qui explique cette caractéristique pour ces établissements.

6 groupes

Groupe	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9	Classe 10
1	6,23	8,04	7,49	8,23	9,18	6,02	12,05	14,10	11,75	16,93
2	5,17	5,95	6,91	8,36	7,41	8,08	9,60	6,30	16,29	25,92
3	7,17	8,14	9,42	9,80	10,99	11,82	11,54	9,41	11,96	9,75
4	4,94	6,53	8,36	8,37	7,73	7,60	14,04	6,78	9,24	26,41
5	14,58	12,73	7,86	7,65	13,28	4,80	10,57	22,99	4,30	1,24
6	13,82	13,06	11,69	10,26	10,85	10,44	8,30	9,02	7,63	4,92

Nombre, type d'établissements et capacité

Nombre et type d'établissements

NumClust	NbEtb	OQN	STC
1	102	94	8
2	123	111	12
3	296	18	278
4	76	68	8
5	21	4	17
6	35	1	34

Nombre Moyen de Lits

Global	OQN	STC
199,38	193,33	270,50
131,39	128,77	155,58
303,75	128,00	315,13
114,70	111,26	143,88
166,10	153,25	169,12
1 523,49	195,00	1 562,56

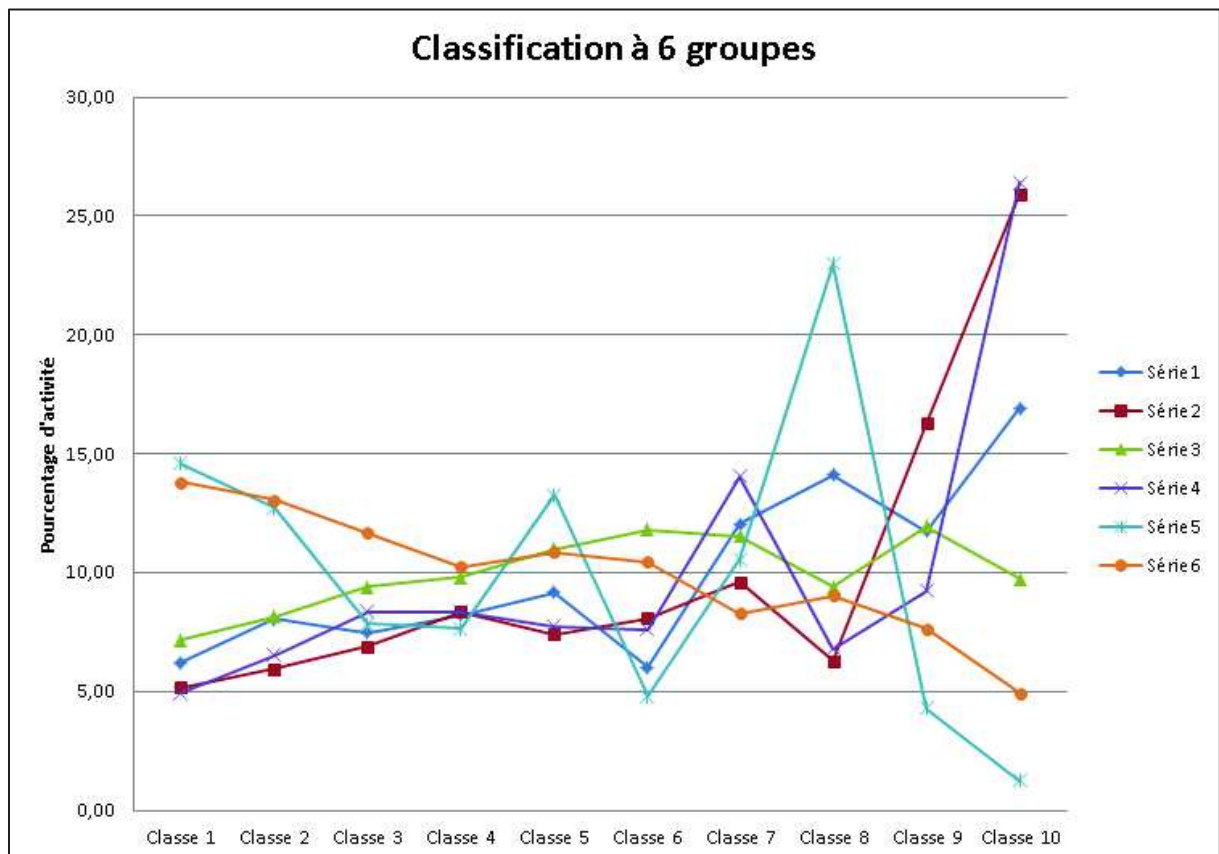


Figure 22 - Classification à 6 groupes - Profils moyens.

Le groupe 2 génère le groupe 4 par séparation de 76 établissements (les groupes 4 et 5 sont renumérotés 5 et 6) :

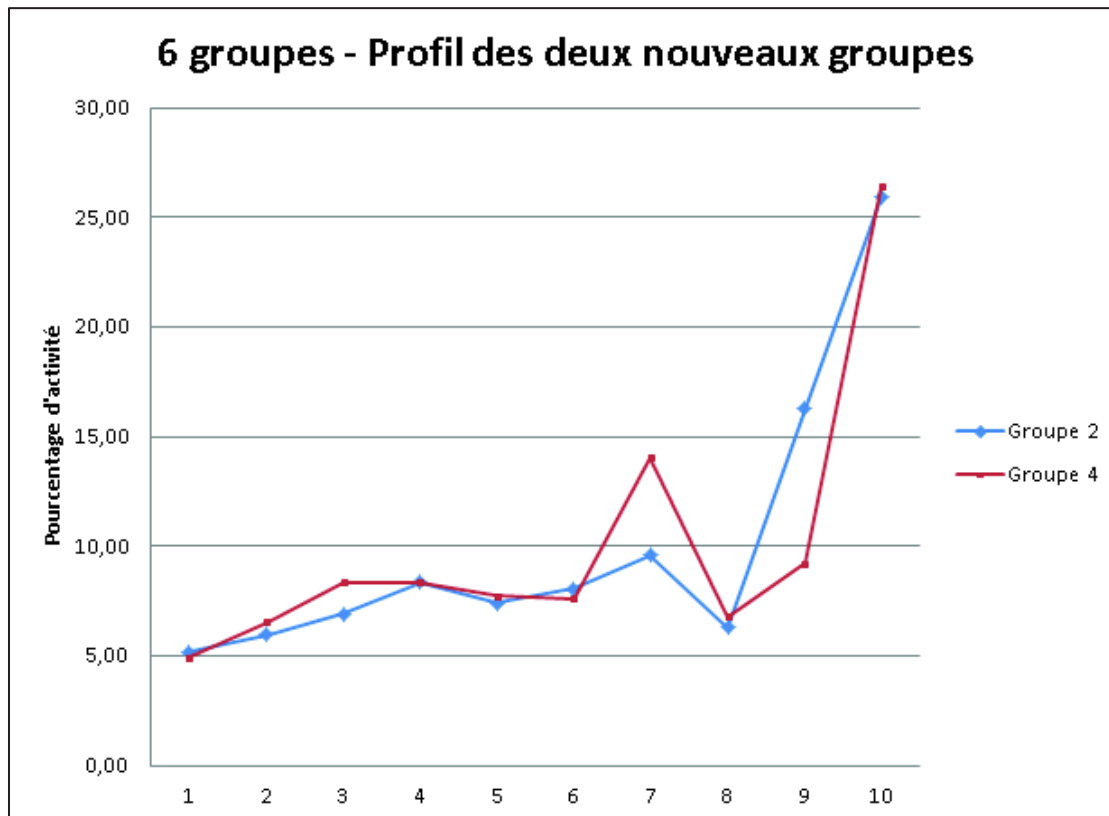


Figure 23 - Classification à 6 groupes – Profils moyens des 2 nouveaux groupes.

Ici cette séparation ne nous apparaît pas pertinente au vu de la similitude des profils moyens des deux groupes. En effet la seule différence réside dans un pic plus important d'activité moyennement spécifique (classe 7), et plus faible dans la première classe d'activité peu spécifique (classe 9).

4.3. Classification retenue

Au vu de ces données, nous retenons la classification en 5 groupes, qui nous paraît mettre en évidence la typologie globale suivante :

Groupe 1 : établissements généralistes de proximité

Groupe 2 : établissements spécialisés de proximité

Groupe 3 : établissements généralistes de recours

Groupe 4 : établissements spécialisés de recours

Groupe 5 : établissements généralistes de référence

5. Discussion

Intérêt

Nous avons élaboré une méthode de classification des établissements, permettant d'identifier 5 grands types. Au-delà d'une typologie statique, cette méthode permet une lecture du paysage sanitaire et de l'activité des établissements. De plus, le caractère distinct des profils semble valider rétrospectivement la classification d'activité élaborée au chapitre 2.

Enfin, cette méthode peut également être utilisée pour évaluer les stratégies de coopération inter-établissement. Considérons la situation locale, associant au sein d'un même territoire le CHU de Nîmes, de type 5, et les CH d'Alès et de Bagnols sur Cèze, tous deux de type 3. La distribution de leur activité est la suivante :

	CHU NIMES	CH ALES	CH BAGNOLS SUR CEZE
Classe 1	10,91%	6,52%	7,31%
Classe 2	10,78%	7,38%	7,32%
Classe 3	10,79%	7,82%	11,79%
Classe 4	10,04%	8,60%	7,69%
Classe 5	12,95%	11,48%	12,81%
Classe 6	13,05%	12,26%	12,89%
Classe 7	8,15%	8,51%	11,55%
Classe 8	10,25%	9,79%	5,82%
Classe 9	7,52%	13,36%	11,23%
Classe 10	5,55%	14,28%	11,58%

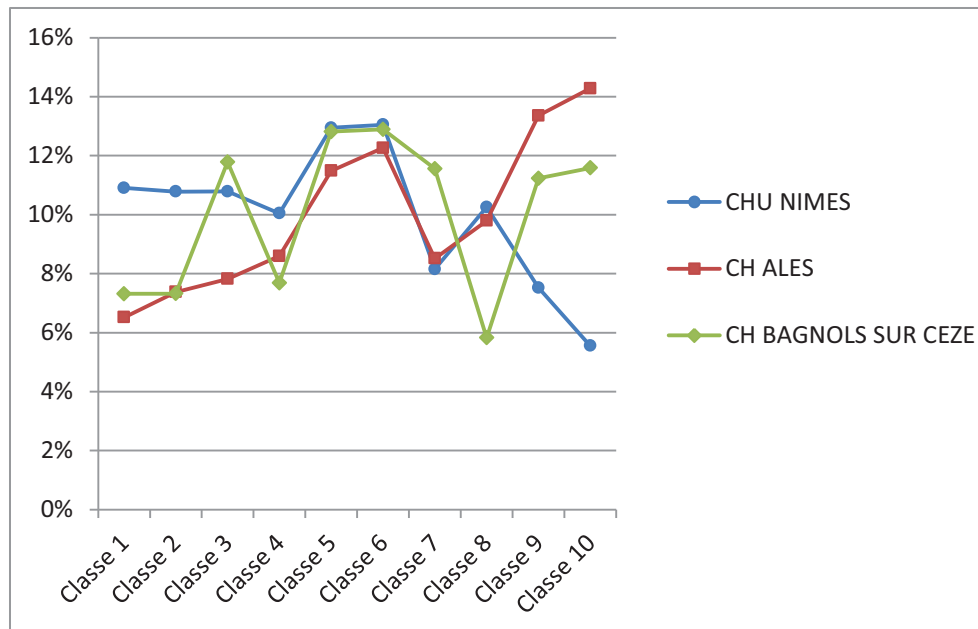


Figure 24 - Profils d'activité des CHU de Nîmes, CH Alès et CH Bagnols sur Cèze.

Du simple fait de leur type, ces établissements sont déjà complémentaires. La figure 24 montre bien un pourcentage d'activité important dans les premières classes pour le CHU de Nîmes, tandis que cette tendance s'inverse dans les dernières au profit des deux CH.

On peut néanmoins pousser l'analyse plus loin et s'interroger spécifiquement sur la prise en charge des patients des bassins alésien et bagnolais. Nous allons donc examiner le profil d'activité du CHU de Nîmes pour les patients alésiens puis les patients bagnolais, et le comparer aux profils de ces deux établissements pour les mêmes patients :

	CHU NIMES / Patients Alès	CH ALES / Patients Alès		CHU NIMES / Patients Bagnols	CH BAGNOLS SUR CEZE / Patients Bagnols
Classe 1	11,54%	5,66%	Classe 1	13,42%	6,93%
Classe 2	12,11%	6,51%	Classe 2	14,37%	7,36%
Classe 3	9,80%	6,77%	Classe 3	10,23%	11,80%
Classe 4	9,88%	7,64%	Classe 4	9,86%	7,56%
Classe 5	14,16%	10,20%	Classe 5	12,01%	12,62%
Classe 6	13,89%	10,93%	Classe 6	10,12%	12,49%
Classe 7	8,05%	7,57%	Classe 7	7,87%	11,72%
Classe 8	12,03%	8,62%	Classe 8	14,00%	6,01%
Classe 9	4,13%	12,22%	Classe 9	4,14%	11,28%
Classe 10	4,43%	13,33%	Classe 10	3,99%	12,23%

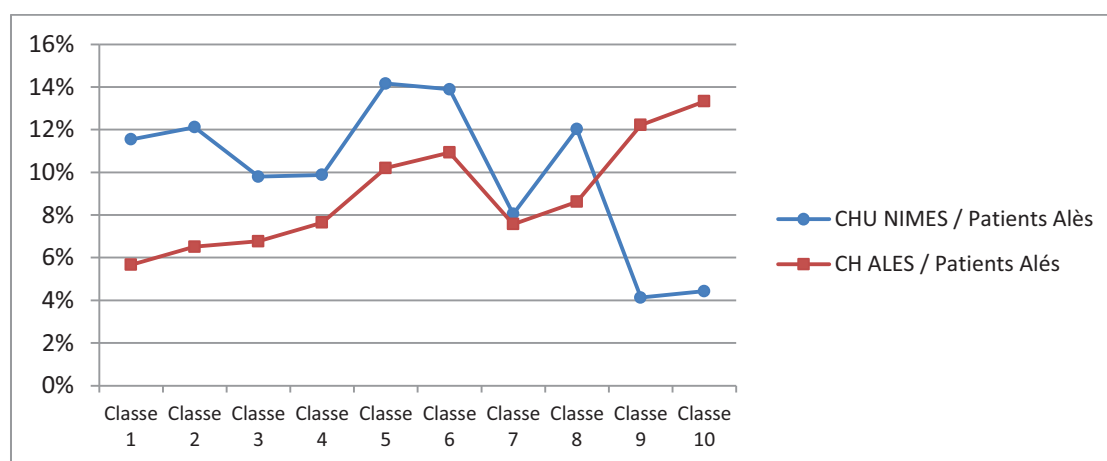


Figure 25 - Profils CHU Nîmes et CH Alès - Patients du bassin alésien.

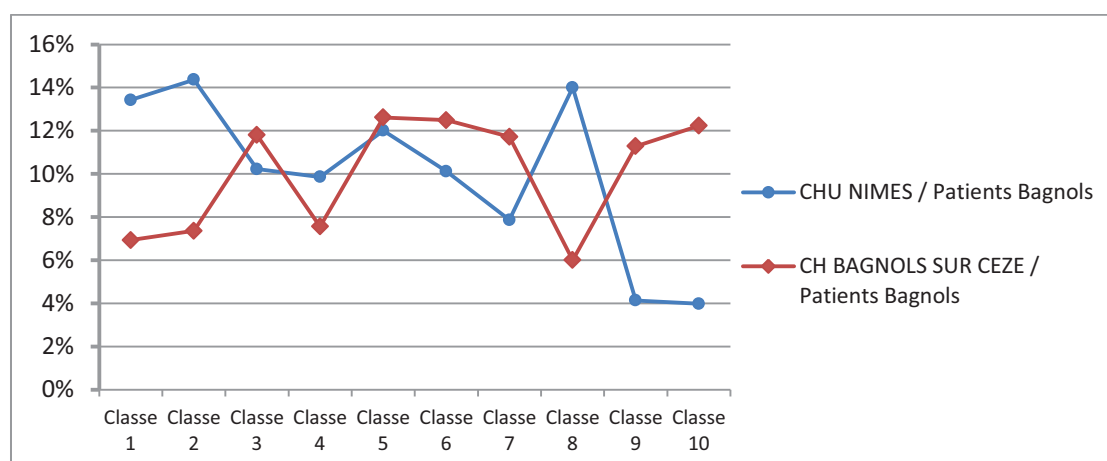


Figure 26 - Profils CHU de Nîmes et CH Bagnols - Patients du bassin bagnolais.

Ces figures mettent en évidence une réalisation des activités de plus forte spécificité par le CHU de Nîmes, tandis que les activités de proximité sont réalisées dans les CH. Il y a donc bien une répartition graduée des prises en charge entre les établissements.

Limites

La mise en œuvre de cette méthode nécessite un important travail de préparation des données. La mise en correspondance des données de la base nationale avec celles de la SAE en particulier, est particulièrement lourde. Le décalage entre les modèles de données, les référentiels ou encore les périodes de référence, l'incertitude ou l'inexactitude des données déclaratives de la SAE, empêchent une exploitation simple et systématique des deux bases. Il est donc nécessaire de revenir fréquemment au détail des données pour s'assurer qu'un résultat surprenant n'est pas causé par une erreur au niveau de ces dernières.

Les résultats sont actuellement peut maniables. La distribution de l'activité en classe, combinée à la classification ATIH en regroupement de GHM, permet un examen assez rapide de l'activité d'un établissement et le repérage de ses principaux segments. Cependant la manipulation des données en tableau est peu pratique et freine leur utilisation. De plus, la mise en perspective géographique fait défaut pour une lecture clarifiée des données.

Au plan statistique, l'emploi du critère de Ward sur des données dont les distances sont mesurées avec la distance de Hellinger pose question. En effet, le critère de Ward a pour objet de minimiser l'augmentation de la variance intra-groupe à chaque étape d'agrégation des données : cette méthode fait donc appel au carré de la distance euclidienne. On peut donc s'interroger sur l'impact de ces mesures différentes sur la classification et son efficacité.

Perspectives

Cette méthode trouve une articulation naturelle avec l'analyse du recrutement ainsi que nous en avons donné un bref exemple ci-dessus. Au-delà du recrutement, une mise en relation avec une étude de l'attractivité des établissements semble pertinente. Ceci nous paraît nécessairement impliquer une dimension géographique et cartographique.

Les techniques statistiques pourront probablement être améliorées, en particulier par le développement sous R d'une version du critère de Ward basée sur la distance de Hellinger ; il sera intéressant d'étudier le résultat obtenu et de le comparer au résultat actuel.

Enfin, l'intégration de la dimension temporelle, combinée à la géographie, apporterait une vision dynamique du paysage sanitaire, permettant de suivre l'évolution qualitative de l'offre, en la mettant en rapport par exemple avec l'évolution démographique et les déterminants socio-économiques.

Conclusion

Si l'intérêt de l'exploitation des données PMSI semble être admis de tous, les conditions de leur exploitation systématique et fructueuse ne sont pas forcément réunies. Par ce travail, nous pensons avoir contribué à la construction d'un cadre méthodologique de leur exploitation ; le terme d'exploration serait peut-être même plus approprié, tant nous avons la sensation d'un véritable continent digital à explorer.

De nombreux aspects restent à développer, notamment la question d'une méthode standardisée, reproductible et systématisable d'analyse de l'attractivité des établissements. Ceci implique de disposer d'un référentiel stabilisé des sites des établissements de santé, ainsi que d'une base des distances (en temps si possible) entre les unités géographiques du PMSI, à savoir les codes géographiques, et les sites. L'attractivité devra probablement être abordée segment d'activité par segment d'activité, en tenant compte d'une série de facteurs susceptibles de perturber l'analyse (effet tourisme, effet urgence – où le patient ne choisit pas, ou moins, l'établissement d'accueil...), en opérant les choix les plus pertinents possibles en termes de standardisation, etc. Toutes problématiques auxquelles nous nous sommes confrontées au cours de ce travail.

Il faut également prendre acte de l'intrication croissante entre données, méthode d'analyse et méthode de présentation pour en extraire l'information pertinente. La réalisation de ce travail nous a permis d'identifier plusieurs chantiers qui doivent être simultanément menés :

- un important travail de structuration, de consolidation, de maintenance et de diffusion maniable et agile des référentiels : intégration standardisée de la notion temporelle avec la définition des conditions de validité, représentation en XML des nomenclatures, accès par web-services, outils de gestion de thesaurus... Il est évident que cela doit faire l'objet d'un véritable projet qui ne peut être, au moins, que d'envergure nationale ;

- une action de consolidation et d'homogénéisation des modèles de représentation des données. Notre travail d'exploitation conjointe des données du PMSI et de la SAE nous amène à constater qu'il ne semble pas y avoir pas, à ce jour, de définition précise de ce qu'est un établissement. Partant de là, le modèle de représentation, et donc le modèle de données, sont nécessairement défaillants. Ceci entraîne une déperdition importante de qualité et d'efficacité dans l'exploitation des données, ce qui en retour pose la question de l'intérêt du recueil pour les producteurs de données. A quoi bon faire un effort et dépenser de l'énergie dans le recueil de données mal définies et difficilement exploitables ? Il faut inverser le sens de l'intérêt à faire pour les producteurs de données ;

- un développement des techniques de présentation des données, qui devraient être intégrées dans les analyses dès leur conception. La question « comment vais-je présenter les données ? » devrait être posée dès le début, à égale importance avec « comment vais-je répondre à la question ? ». Les outils de présentation graphique, et notamment de cartographie, devraient être banalisés et accessibles (ergonomie, simplicité d'utilisation, prix), ce qui implique, en particulier dans le cas de la cartographie, la diffusion de bases cartographiques propres au PMSI, par exemple le fichier des contours des codes

géographiques PMSI, actualisé à chaque changement de leur contenu, avec maintenance des périodes de validité, ou encore le fichier de géolocalisation des établissements de santé (mais encore faut-il avoir clairement défini cette notion).

Une exploitation systématisée, fluidifiée, des données de l'Information Médicale est nécessaire. En raccourcissant le cycle hypothèse - analyse - conclusion - nouvelle hypothèse, elle permettrait d'avancer beaucoup plus rapidement dans la connaissance de l'offre de santé et de sa configuration, contribuant à fournir des instruments de planification bien informée. Elle faciliterait leur mise en rapport avec d'autres données démographiques, épidémiologiques, socio-économiques, et permettrait ainsi de contribuer à une meilleure connaissance de l'état de santé des populations et de son évolution. Les moyens techniques dont nous disposons aujourd'hui en termes de capacité de stockage et de traitement de l'information permettent d'envisager ce saut qualitatif.

Enfin, une des directions incontournables nous paraît être celle de l'ingénierie des connaissances. Depuis une cinquantaine d'années, dans le sillage du développement de la micro-informatique et de l'internet, portées par d'extraordinaires progrès technologiques et d'énormes enjeux scientifiques, politiques et économiques, les techniques de représentation des données se sont développées. Elles ont donné naissance, il y a maintenant près de vingt ans, aux premiers langages de description des faits, puis des connaissances. Les acronymes de RDF, OWL, ou encore SPARQL, ne peuvent désormais plus rester inconnus et inusités dans le domaine de l'Information Médicale et il appartiendra à notre génération et aux suivantes de mettre en œuvre les premières bases de connaissances PMSI.

Les révolutions techniques nous emportent loin de nos bases ; elles nous bousculent, c'est pour cela qu'elles nous ouvrent de nouveaux horizons. C'est le cas dans le domaine de l'Information Médicale comme ailleurs. Nous avons tous les outils pour les explorer, il faut les mettre en œuvre avec ordre et méthode, au service des valeurs fondatrices de notre système de santé : le progrès médical, l'égalité d'accès aux soins, la solidarité.

Bibliographie

- 1 Or Z, Bonastre J, Journeau F, *et al.* Activité, productivité et qualité des soins des hôpitaux avant et après la T2A - DT 56 - IRDES. 2013.[www.irdes.fr/ EspaceRecherche/ DocumentsDeTravail/ DT56SoinsHospitaliersT2A.pdf](http://www.irdes.fr/EspaceRecherche/DocumentsDeTravail/DT56SoinsHospitaliersT2A.pdf)
- 2 Beyeme-Ondoua J-P. Evaluation of the quality of surveillance data for colorectal cancer from the national PMSI database in 2003. *Santé Publique Vandoeuivre-Lès-Nancy Fr* 2007;**19**:471–80.
- 3 Remontet L, Mitton N, Couris CM, *et al.* Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol* 2008;**23**:681–8. doi:10.1007/s10654-008-9282-y
- 4 Colonna M, Mitton N, Schott A-M, *et al.* Joint use of epidemiological and hospital medico-administrative data to estimate prevalence. Application to French data on breast cancer. *Cancer Epidemiol* 2012;**36**:116–21. doi:10.1016/j.canep.2011.12.001
- 5 Trombert-Paviot B, Martin C, Quantin C, *et al.* Du chaînage des résumés d'hospitalisation en oncologie à la connaissance épidémiologique des cancers pour la planification. *Rev Médicale Assur Mal* 2003;**32**:235–41.
- 6 Arrêté du 8 juin 2005 pris en application des articles L. 6121-2, L. 6114-2 et L. 6122-8 du code de la santé publique et du décret n° 2005-76 du 31 janvier 2005 relatif aux objectifs quantifiés de l'offre de soins prévus à l'article L. 6121-2 du code de la santé publique. 2005.
- 7 Health policies and data - OECD. <https://www.oecd.org/health/health-systems/> (accessed 3 May2016).
- 8 Fetter RB, Shin Y, Freeman JL, *et al.* Case mix definition by diagnosis-related groups. *Med Care* 1980;**18**:iii, 1-53.
- 9 Landais P, Boudemaghe T, Suehs C, *et al.* Computerized Medico-Economic Decision Making: An International Comparison. In: Venot A, Burgun A, Quantin C, eds. *Medical Informatics, e-Health*. Springer Paris 2014. 191–220.http://link.springer.com/chapter/10.1007/978-2-8178-0478-1_9 (accessed 6 Oct2016).
- 10 Loi n° 70-1318 du 31 décembre 1970 portant réforme hospitalière. 1970. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000874228> (accessed 5 May2016).
- 11 Loi n° 91-748 du 31 juillet 1991 portant réforme hospitalière. 1991. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000720668&dateTexte=20160505> (accessed 5 May2016).
- 12 Ordonnance n° 96-344 du 24 avril 1996 portant mesures relatives à l'organisation de la sécurité sociale. 1996. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000742093> (accessed 5 May2016).

- 13 Ordonnance n° 96-345 du 24 avril 1996 relative à la maîtrise médicalisée des dépenses de soins. 1996.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000192992&dateTexte=> (accessed 5 May2016).
- 14 Ordonnance n° 96-346 du 24 avril 1996 portant réforme de l'hospitalisation publique et privée. 1996.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000742206&categorieLien=id> (accessed 5 May2016).
- 15 Ordonnance n° 2003-850 du 4 septembre 2003 portant simplification de l'organisation et du fonctionnement du système de santé ainsi que des procédures de création d'établissements ou de services sociaux ou médico-sociaux soumis à autorisation. 2003.
https://www.legifrance.gouv.fr/jo_pdf.do?numJO=0&dateJO=20030906&numTexte=26&pageDebut=15391&pageFin=15404
- 16 Ordonnance n° 2005-406 du 2 mai 2005 simplifiant le régime juridique des établissements de santé. 2005.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006051597>
- 17 Loi n° 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires. 2009.
<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020879475&categorieLien=id> (accessed 5 May2016).
- 18 de Kervasdoué J. Appel des candidatures des établissements hospitaliers à l'expérience de résumés de sorties du projet de médicalisation des systèmes d'information (PMSI). 1982.
- 19 Arrêté du 3 octobre 1985 autorisant l'informatisation des résumés de sortie standardisés dans les établissements d'hospitalisation publics et privés participant au service public hospitalier. 1985.
http://www.atih.sante.fr/sites/default/files/public/content/990/Arrete_1985.pdf
- 20 Circulaire n° 119 du 4 octobre 1985 relative à la mise en place dans les établissements hospitaliers des résumés de sortie standardisés (R.S.S.). 1985.
http://www.atih.sante.fr/sites/default/files/public/content/990/Cir_4-10-85.pdf.pdf
- 21 Circulaire n° 160 du 5 août 1986 relative à la diffusion de la classification des groupes homogènes de malades (G.H.M.). 1986.
http://www.atih.sante.fr/sites/default/files/public/content/990/1986_Cir_V0_GHM.pdf
- 22 Circulaire 178 du 11 mars 1987 relative à la collecte et à l'exploitation des R.S.S. 1987.
- 23 Circulaire DH/P.M.S.I. n° 303 du 24 juillet 1989 relative à la généralisation du programme de médicalisation des systèmes d'information (P.M.S.I.) et à l'organisation de l'information médicale dans les hôpitaux publics. 1989.
- 24 Circulaire 325 du 12 février 1990 relative aux modalités de mise en place des structures de gestion de l'information médicale dans les établissements hospitaliers publics et privés participant au service public. 1990.

- 25 Circulaire DH/DSS/95/23 du 10 mai 1995 relative à la généralisation du dispositif technique, expérimenté en Languedoc-Roussillon, de mesure de l'activité et des coûts des établissements hospitaliers sous compétence tarifaire de l'état, à l'aide des informations provenant du système d'information médicalisé. 1995.
- 26 ATIH : Agence technique de l'information sur l'hospitalisation. <http://www.atih.sante.fr/> (accessed 6 May2016).
- 27 Décret n° 2002-1284 du 22 octobre 2002 relatif à l'organisation et au fonctionnement de l'Agence technique de l'information sur l'hospitalisation, modifiant le code de la santé publique (deuxième partie : Décrets en Conseil d'Etat). 2002. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000410136&fastPos=3&fastReqId=1579157361&categorieLien=id&oldAction=rechTexte> (accessed 6 May2016).
- 28 Circulaire n° 106/DHOS-PMSI du 22 février 2001 relative au chaînage des séjours en établissements hospitalier. 2001.
- 29 Farr W. First annual report of the Registrar General. *Lond Her Majesty's Stat Off* 1839.
- 30 Wüster E, Bauer L, Wüster E, et al. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Springer 1979.
- 31 ISO - Comités techniques - ISO/TC 37 - Terminologie et autres ressources langagières et ressources de contenu. http://www.iso.org/iso/fr/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48104 (accessed 19 May2016).
- 32 International Information Centre For Terminology. <http://www.infoterm.info/> (accessed 27 May2016).
- 33 Temmerman R. *Towards new ways of terminology description: the sociocognitive-approach*. John Benjamins Publishing 2000. https://books.google.fr/books?hl=en&lr=&id=q9gUJ_ykkVAC&oi=fnd&pg=PR11&dq=Towards+new+ways+of+terminology+description.+The+sociocognitive+approach&ots=h1YOcnh3by&sig=82rkeN8J1QmYeZLKOwJUhaJ2Lk8 (accessed 27 May2016).
- 34 Campo A. The reception of Eugen Wüster's work and the development of terminology. Published Online First: 2013. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/9198> (accessed 27 May2016).
- 35 Terminology Theory in Easy-to-Swallow Pills. My Own Terms. 2014. <http://inmyownterms.com/terminology-theory-easy-swallow-pills/> (accessed 27 May2016).
- 36 WÜSTER E de ME. La théorie générale de la terminologie-un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets. Published Online First: 1976. <http://docpatient.net/onto/doc/wuster.pdf> (accessed 30 May2016).

- 37 Gruber TR. Ontology (Computer Science) - definition in Encyclopedia of Database Systems. <http://web.dfc.unibo.it/buzzetti/IUcorso2007-08/mdidattici/ontology-definition-2007.htm> (accessed 27 May2016).
- 38 Shanahan M. The Frame Problem. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy*. 2016. <http://plato.stanford.edu/archives/spr2016/entries/frame-problem/> (accessed 31 May2016).
- 39 Dennett DC. Cognitive Wheels : The Frame Problem of AI. *Philos Psychol Contemp Read* 2006;**433**.https://books.google.fr/books?hl=en&lr=&id=sa9_AgAAQBAJ&oi=fnd&pg=PA433&dq=the+frame+problem&ots=c81LqEXGt1&sig=dkOdKxDEevC9IMahJ4SqzizA7dU (accessed 30 May2016).
- 40 ISO/IEC 24707:2007 - Information technology -- Common Logic (CL): a framework for a family of logic-based languages. ISO. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39175 (accessed 31 May2016).
- 41 Hayes P, Menzel C, Sowa J, *et al*. Common Logic (CL): a framework for a family of logic-based languages. *ISOIEC IS 2007*;**24707**.
- 42 ISO 25964-1:2011(en), Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. <https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en> (accessed 25 Aug2016).
- 43 ISO 25964-2:2013(en), Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies. <https://www.iso.org/obp/ui/#iso:std:iso:25964:-2:ed-1:v1:en> (accessed 25 Aug2016).
- 44 ISO 25964 Thesaurus Schemas - National Information Standards Organization. <http://www.niso.org/schemas/iso25964/#part1> (accessed 25 Aug2016).
- 45 World Wide Web Consortium (W3C). <https://www.w3.org/> (accessed 31 May2016).
- 46 SKOS Simple Knowledge Organization System - home page. <https://www.w3.org/2004/02/skos/> (accessed 4 Oct2016).
- 47 Musen MA, Noy NF, Shah NH, *et al*. The national center for biomedical ontology. *J Am Med Inform Assoc* 2012;**19**:190–195.
- 48 Whetzel PL, Noy NF, Shah NH, *et al*. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res* 2011;**39**:W541–W545.
- 49 Salvadores M, Horridge M, Alexander PR, *et al*. Using SPARQL to query bioportal ontologies and metadata. In: *The Semantic Web–ISWC 2012*. Springer 2012. 180–195.http://link.springer.com/chapter/10.1007/978-3-642-35173-0_12 (accessed 31 May2016).
- 50 NCBO BioPortal. <http://bioportal.bioontology.org/> (accessed 31 May2016).

- 51 WHO | International Classification of Diseases (ICD). WHO. <http://www.who.int/classifications/icd/en/> (accessed 18 May2016).
- 52 Classification Statistique Internationale des Maladies et des Problèmes de Santé Connexes - CIM10-FR à usage PMSI - Volume 1. *Bull Off Ministère Aff Soc Santé - Fasc Spéc* 2016;**2016/9 bis**.http://social-sante.gouv.fr/IMG/pdf/sts_20160009_0001_p000.pdf
- 53 Moriyama IM, Loy RM, Robb-Smith AHT, *et al.* *History of the statistical classification of diseases and causes of death*. 2011.
- 54 Arrêté du 27 mars 1972 relatif à la nomenclature générale des actes professionnels des médecins, des chirurgiens-dentistes, des sages-femmes et des auxiliaires médicaux. 1972.
- 55 WHO | International Classification of Health Interventions (ICHI). WHO. <http://www.who.int/classifications/ichi/en/> (accessed 1 Jun2016).
- 56 MOURGUES F. Actes médicaux: méthodologie d'élaboration. *Inf Hosp* 1986;:45–8.
- 57 Boudernaghe T, Durif L. Methode de comparaison du case-mix entre 1996 et 1997. *J Econ Medicale* 2000;**18**:487–494.
- 58 Arrêté du 13 novembre 2013 relatif à la mise en place d'un répertoire national des établissements sanitaires et sociaux.
- 59 FINESS - Accueil. <http://finess.sante.gouv.fr/finess/jsp/index.jsp> (accessed 17 Sep2013).
- 60 Recherche - Data.gouv.fr. <https://www.data.gouv.fr/fr/search/?q=Extraction+FINESS> (accessed 4 Oct2016).
- 61 Arrêté du 9 septembre 2010 relatif à la collecte et à la transmission des informations nécessaires à l'établissement de la statistique annuelle des établissements de santé. 2010. http://social-sante.gouv.fr/fichiers/bo/2010/10-09/ste_20100009_0100_0045.pdf
- 62 SAE - Accueil. <http://www.sae-diffusion.sante.gouv.fr/> (accessed 17 Sep2013).
- 63 ARS - Agences Régionales de Santé: Etablissements sanitaires et sociaux (FINESS). <http://www.ars.limousin.sante.fr/Etablissements-sanitaires-et-s.110451.0.html> (accessed 4 Oct2016).
- 64 Instruction relative à la refonte du cadre des établissements hébergeant des personnes âgées dans le fichier national des établissements sanitaires et sociaux (FINESS) et précisant les modalités de création d'une catégorie EHPAD (établissements d'hébergement pour personnes âgées dépendantes) dans FINESS et à la création du portail internet personnes âgées dont la gestion sera confiée à la CNSA. 2014. http://circulaire.legifrance.gouv.fr/pdf/2014/07/cir_38525.pdf
- 65 Circulaire n° DHOS/04/2009/279 du 12 août 2009 relative aux activités interventionnelles sous imagerie médicale, par voie endovasculaire, en cardiologie. 2009.
- 66 ameli.fr - T2A, CCAM et LPP. <http://www.ameli.fr/l-assurance-maladie/documentation-technique/t2a-ccam-et-lpp/index.php> (accessed 4 Oct2016).

- 67 Paviot BT, Martin C, Clavel L, *et al.* From DRG databases to an epidemiological observatory for colorectal cancer in a French small area oncology network. *Stud Health Technol Inform* 2003;**95**:829–33.
- 68 Olive F, Gomez F, Schott A-M, *et al.* [Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: a longitudinal approach becomes possible]. *Rev Dépidémiologie Santé Publique* 2011;**59**:53–8. doi:10.1016/j.respe.2010.09.001
- 69 Quantin C, Benzenine E, Hägi M, *et al.* Estimation of national colorectal-cancer incidence using claims databases. *J Cancer Epidemiol* 2012;**2012**:298369. doi:10.1155/2012/298369
- 70 Doat S, Samson S, Fagot-Campagna A, *et al.* Estimation of breast, prostate, and colorectal cancer incidence using a French administrative database (general sample of health insurance beneficiaries). *Rev Dépidémiologie Santé Publique* 2016;**64**:145–52. doi:10.1016/j.respe.2015.12.020
- 71 Bocquier A, Thomas N, Zitouni J, *et al.* [Evaluation of hospital stays linkage quality to study health spatial variation. a feasibility study in three French regions]. *Rev Dépidémiologie Santé Publique* 2011;**59**:243–9. doi:10.1016/j.respe.2011.02.099
- 72 Tardif L. Etude méthodologique du chaînage des séjours - Base PMSI MCO 2004 - Document de travail n°116 - Série statistiques. 2007.<http://www.drees.sante.gouv.fr/IMG/pdf/seriestat116.pdf>
- 73 Le Manach Y, Collins G, Bhandari M, *et al.* Outcomes after hip fracture surgery compared with elective total hip replacement. *JAMA* 2015;**314**:1159–66. doi:10.1001/jama.2015.10842
- 74 Couray Targe S, Jaisson Hot I, Fauconnier J, *et al.* REFREC : Une méthode pour décrire les activités de référence et de recours à l'aide du PMSI. *J Econ MEDICALE* 2003;**21**:43–51.
- 75 Duclos A, Couray Targe S, Groupe de travail des Dim des Chu de Rhône-Alpes. Lyon. FRA. Actualisation de l'outil REFREC en version 9 de la classification des GHM. *9ème Sémin Etude Natl Coûts* 2006;**24**:191–6.
- 76 Mendelsohn M. L'activité des CHU dans la PMSI : peut-on isoler l'activité spécifique des CHU ? 2009.http://www.atih.sante.fr/sites/default/files/public/content/2137/activit__CHU_2009-V2bis.pdf
- 77 Shannon CE. A Mathematical Theory of Communication. *Bell Syst Tech J* 1948;**27**:379–423.
- 78 Rényi A. On Measures of Entropy and Information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Statistical Laboratory of the University of California, Berkeley: 1960. 547–61.http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.bsmmsp/1200512181

- 79 Jacquemin A. Une mesure entropique de la diversification des entreprises. *Rev Économique* 1975;**26**:834–8. doi:10.2307/3500966
- 80 Mahalanobis PC. On the generalized distance in statistics. *Proc Natl Inst Sci Calcutta* 1936;**2**:49–55.
- 81 Bhattacharyya AK. On a Measure of Divergence between Two Multinomial Populations. *Sankhyā Indian J Stat 1933-1960* 1946;**7**:401–6.
- 82 Bhattacharyya AK. On a Measure of Divergence between Two Statistical Populations Defined by their Probability Distributions. *Bull Calcutta Math Soc* 1943;**35**:99–109.
- 83 Csisz I, others. Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der ergodizität von Markoffschen Ketten. *Publ Math Inst Hung Acad* 1963;**8**:95–108.
- 84 Morimoto T. Markov Processes and the H -Theorem. *J Phys Soc Jpn* 1963;**18**:328–31. doi:10.1143/JPSJ.18.328
- 85 Ali SM, Silvey SD. A General Class of Coefficients of Divergence of One Distribution from Another. *J Roy Stat Soc Ser B Methodol* 1966;**28**:131–42.
- 86 Hellinger E. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J Für Reine Angew Math Crelles J* 1909;**1909**. doi:10.1515/crll.1909.136.210
- 87 Kerleau M, Le Vaillant M, Or Z. La diversification de l'activité de court séjour des établissements de santé : l'éclairage par l'indicateur d'entropie. *Doss Solidar Santé* 2005;**2005**:35–47.
- 88 Ward Jr JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;**58**:236–244.
- 89 Dunn† JC. Well-separated clusters and optimal fuzzy partitions. *J Cybern* 1974;**4**:95–104.
- 90 Baker FB, Hubert LJ. Measuring the power of hierarchical cluster analysis. *J Am Stat Assoc* 1975;**70**:31–38.

Annexes

I. RSS non groupé format 017

Libellé	Taille	Début	Fin	Obligatoire[1]	Type[2]	Remarques
Numéro FINESS d'inscription ePMSI	9	1	9	O	A	
Version du format du RUM	3	10	12	O	N	017
N° de RSS	20	13	32	O	A	
N° Administratif local de séjour	20	33	52	O	A	
N° de RUM	10	53	62	O	A	
Date de naissance	8	63	70	O	N	JJMMAAAA
Sexe	1	71	71	O	N	1 : homme, 2 : femme
Numéro de l'unité médicale	4	72	75	O	A	
Type d'autorisation du lit dédié	2	76	77	F	N	
Date d'entrée dans l'unité médicale	8	78	85	O	N	JJMMAAAA
Mode d'entrée dans l'unité médicale	1	86	86	O	N	
Provenance (si mode d'entrée est mutation, transfert ou domicile)	1	87	87	F	N	
Date de sortie de l'unité médicale	8	88	95	O	N	JJMMAAAA
Mode de sortie de l'unité médicale	1	96	96	O	N	
Destination (si mode de sortie est mutation, transfert ou domicile)	1	97	97	F	N	
Code postal de résidence (ou 99 suivi du code Insee du pays pour les patients résidant hors de France)	5	98	102	O	N	
Poids du nouveau-né à l'entrée de l'unité médicale (en grammes)	4	103	106	F	N	
Age gestationnel	2	107	108	F	N	
Date des dernières règles	8	109	116	F	N	JJMMAAAA
Nombre de séances	2	117	118	F	N	
Nombre de diagnostics associés (nDA) dans ce RUM	2	119	120	O	N	
Nombre de données à visée documentaire (nDAD) dans ce RUM	2	121	122	O	N	
Nombre de zone d'actes (nZA) dans ce RUM	3	123	125	O	N	
Diagnostic principal (DP)	8	126	133	O	A	
Diagnostic relié (DR)	8	134	141	F	A	
IGS 2	3	142	144	F	N	
Confirmation du codage du RSS	1	145	145	F	N	1 : oui, 2 : non (cf. Note 1)
Type de machine en radiothérapie	1	146	146	F	N	cf. Note 2
Type de dosimétrie	1	147	147	F	N	cf. Note 3
Numéro d'innovation	15	148	162	F	A	
Nombre d'IVG antérieures	2	163	164	F	N	
Année de l'IVG précédente	4	165	168	F	N	
Filler	4	169	172	F	A	
Nombres de naissances vivantes antérieures	2	173	174	F	N	
Zone réservée	3	175	177	F	N	
DA n° 1	8	178	185	F	A	
.....	*8			F	A	
DA n° nDA	8			F	A	

Libellé		Taille	Début	Fin	Obligatoire[1]	Type[2]	Remarques
DAD n° 1		8			F	A	
.....		*8			F	...	
DAD n° nDAD		8			F	A	
Zone d'acte n° 1	Date de réalisation	8			F	N	JJMMAAAA, Signalement si non renseigné
	Code CCAM	7			O	A	
	Extension PMSI	3			F	A	
	Phase	1			O	N	
	Activité	1			O	N	
	Extension documentaire	1			F	A	
	Modificateurs	4			F	A	
	Remboursement exceptionnel	1			F	A	
	Association non prévue	1			F	A	
Nombre de réalisations de l'acte n° nZA pendant le séjour	2			O	N		
...		*29	
Zone d'acte n° nZa	Date de réalisation	8			F	N	JJMMAAAA, Signalement si non renseigné
	Code CCAM	7			O	A	
	Extension PMSI	3			F	A	
	Phase	1			O	N	
	Activité	1			O	N	
	Extension documentaire	1			F	A	
	Modificateurs	4			F	A	
	Remboursement exceptionnel	1			F	A	
	Association non prévue	1			F	A	
Nombre de réalisations de l'acte n° nZA pendant le séjour	2			O	N		

[1] O : obligatoire F : facultatif (la variable peut être laissée à vide)

[2] A : alphanumérique (pas de distinction entre majuscules et minuscules), N : numérique

[3] NA signifie "non applicable"

Note 1 : La fonction groupage suspecte parfois des anomalies de codage pour lesquelles le contenu du RSS doit être modifié ou confirmé.

Cette zone permet la confirmation. Elle doit être faite dans le dernier RUM. Ces anomalies portent soit sur une durée de séjour trop courte soit sur un accouchement classé en dehors de la CMD 14.

Note 2 :

1 : machine dédiée de type "Tomographie" ou "Novalis" ou "Cyberknife" ou "Protonthérapie"

2 : machine équipée d'imagerie portale, de collimateur multilame et de système d'enregistrement et de contrôle et d'un module de repositionnement du malade à distance

3 : machine équipée d'imagerie portale, de collimateur multilame et de système d'enregistrement et de contrôle sans module de repositionnement du malade à distance

4 : machine non équipée d'imagerie portale, de collimateur multilame ou de système d'enregistrement et de contrôle

Note 3 :

1 : dosimétrie de RCMI

2 : tridimensionnelle avec HDV sur volume cible et organe à risque

3 : tridimensionnelle sans HDV sur volume cible et organe à risque

4 : autre type de dosimétrie

II. RSS groupé format 117

Le format de RSS groupé reprend le format de RSS, auquel sont ajoutés les champs suivants :

Libellé	Taille	Début	Fin	Obligatoire	Type	Remarques	
Groupage : version de la classification	2	1	2	O	N		
Groupage : n° de GHM	N° CMD	2	3	4	O	N	
	N° GHM	4	5	8	O	A	
Filler	1	9	9	F	A		
N° de version du format de RSS	3	10	12	O	A	117	

III. Format de RSA format 223

Libellé		Taille	Début	Fin	Remarques
Numéro FINESS e-PMSI		9	1	9	
Numéro de version du format du RSA		3	10	12	223
N° d'index du RSA		10	13	22	
Numéro de version du format du "RSS-groupé"		3	23	25	
Numéro séquentiel de tarifs		3	26	28	
Groupage lu :Version de la classification		2	29	30	
Groupage lu : GHM	CMD	2	31	32	
	Type	1	33	33	
	Numéro	2	34	35	
	Complexité	1	36	36	
Groupage lu : code-retour		3	37	39	
Groupage GENRSA :Version de la classification		2	40	41	
Groupage GENRSA : GHM	CMD	2	42	43	
	Type	1	44	44	
	Numéro	2	45	46	
	Complexité	1	47	47	
Groupage GENRSA : code-retour		3	48	50	
Nombre de RUM composant le RSS d'origine (NbRUM)		2	51	52	
Age en années		3	53	55	
Age en jours		3	58	60	
Sexe		1	59	59	
Mode d'entrée dans le champ du PMSI-MCO		1	60	60	
Provenance		1	61	61	
Mois de sortie		2	62	63	
Année de sortie		4	64	67	
Mode de sortie du champ PMSI-MCO		1	68	68	
Destination		1	69	69	
Type de séjour		1	70	70	
Durée totale du séjour dans le champ du PMSI		4	71	74	
Code géographique de résidence		5	75	79	
Poids d'entrée (en grammes)		4	80	83	
Age gestationnel retenu pour le groupage		2	84	85	
Délai de la date des dernières règles par rapport à la date d'entrée		3	86	88	
Nombre de séances		2	89	90	
Numéro de GHS (du GHM GENRSA)		4	91	94	
Nombre de journées au-delà de la borne extrême haute		4	95	98	
Type de séjour inférieur à la borne extrême basse		1	99	99	
Nb journées EXB		3	100	102	
Top Admission Maison de Naissance		1	103	103	0 : non, 1 : oui
Top Radiation partielle Avastin		1	104	104	0 : non, 1 : oui
Top Radiation partielle Alimta		1	105	105	0 : non, 1 : oui
Top UHCD		1	106	106	
Top GHS minoré		1	107	107	0 : non, 1 : oui
Confirmation du codage du séjour		1	108	108	
Nombre d'autorisations d'unités médicales à portée globale valides (Nb_AutPGV)		1	109	109	
Nombre de suppléments pour hémodialyse hors séances		3	110	112	

Nombre de suppléments pour les entraînements à la dialyse péritonéale automatisée hors séances	3	113	115	
Nombre de suppléments pour les entraînements à la dialyse péritonéale continue ambulatoire hors séances	3	116	118	
Nombre de suppléments pour les entraînements à l'hémodialyse hors séances	3	119	121	
Nombre d'actes menant dans le GHS 9615	3	122	124	9615(aphérèse sanguine)
Nombre de suppléments radiothérapie pédiatrique	3	125	127	
Nombre de suppléments antepartum	3	128	130	
Nombre de zones de suppléments de radiothérapie (Nb_Rdth)	1	131	131	
Top valves aortiques percutanées	1	132	132	1 si date sortie >=15/08/2012 + type autorisation UM 60 + acte pose, blanc sinon
Type de génération automatique du RSA	1	133	133	0 : pas de génération automatique, 1 : FFM, ATU, SE 2 : médecin salariés, 3 : dialyse
GHS si non prise en compte de l'innovation	4	134	137	
Filler	8	138	145	
Nombre de suppléments caisson hyperbare	3	146	148	
Type de prestation de prélèvement d'organe	1	149	149	1, 2 ou 3
Nombre de suppléments pour REA (réanimation)	3	150	152	
Nombre de suppléments soins intensifs provenant de la réanimation	3	153	155	
Nombre de suppléments pour STF (soins intensifs)	3	156	158	
Nombre de suppléments pour SRC (surveillance continue)	3	159	161	
Nombre de suppléments NN1	3	162	164	
Nombre de suppléments NN2	3	165	167	
Nombre de suppléments NN3	3			
Nombre de suppléments REP (réanimation pédiatrique)	3	171	173	
Passage dans un lit dédié de soins palliatifs	1	174	174	
Type de machine en radiothérapie	1	175	175	
Type de dosimétrie	1	176	176	
Numéro d'innovation	15	177	191	
Nombre d'IVG antérieures	2	192	193	
Année de l'IVG précédente	4	194	197	
Nombres de naissances vivantes antérieures	2	198	199	
Filler	1	200	200	
N° séquentiel du RUM ayant fourni le DP	2	201	202	
Diagnostic principal (DP)	6	203	208	
Diagnostic relié (DR)	6	209	214	
Nombre de diagnostics associés dans ce RSA	4	215	218	
Nombre de zones d'actes dans ce RSA	5	219	223	
Filler	15	224	238	
Type d'autorisation à portée globale valide n° 1	2	239	240	
...	...			
Type d'autorisation à portée globale valide n° Nb_AutPGV	2			
Zone de suppléments de radiothérapie n° 1	Code du supplément (n° GHS)	4		
	Nombre de suppléments	3		
...
Zone de	Code du supplément (n° GHS)	4		

suppléments de radiothérapie n° Nb_Rdth	Nombre de suppléments	3			
Informations RUM n°1	N° séquentiel de RUM	2			
	Filler	3			
	Numéro FINESS géographique	9			
	DP	6			
	DR	6			
	IGS2	3			
	Age gestationnel du RUM	2			
	Nombre de diagnostics associés du RUM (Nb_DA_R_1)	2			
	Nombre de zones d'actes du RUM (Nb_Act_R_1)	3			
	Durée séjour partielle	4			
	Premier type d'autorisation d'UM	4			
	Nature du supplément pour le premier type	2			
	Nombre de supplément pour le premier type	4			
	Deuxième type d'autorisation d'UM	4			
	Nature du supplément pour le deuxième type	2			
	Nombre de supplément pour le deuxième type	4			
...
Informations RUM n° NbRUM	N° séquentiel de RUM	2			
	Filler	3			
	Numéro FINESS géographique	9			
	DP	6			
	DR	6			
	IGS2	3			
	Age gestationnel du RUM	2			
	Nombre de diagnostics associés du RUM (Nb_DA_R_NbRUM)	2			
	Nombre de zones d'actes du RUM (Nb_Act_R_NbRUM)	3			
	Durée séjour partielle	4			
	Premier type d'autorisation d'UM	4			
	Nature du supplément pour le premier type	2			
	Nombre de supplément pour le premier type	4			
	Deuxième type d'autorisation d'UM	4			
	Nature du supplément pour le deuxième type	2			
	Nombre de supplément pour le deuxième type	4			
DA n° 1 du RUM n° 1		6			
...	
DA n° Nb_DA_R_1 du RUM n° 1		6			
...	
DA n° 1 du RUM n° NbRUM		6			
...	
DA n° Nb_DA_R_NbRUM du RUM n° NbRUM		6			
Zone acte n° 1 du RUM n° 1	Délai depuis la date d'entrée	3			
	Code CCAM	7			
	Extension PMSI	2			

	Phase	1			
	Activité	1			
	Extension documentaire	1			
	Modificateurs	4			
	Remboursement exceptionnel	1			
	Association non prévue	1			
	Nombre de réalisations de l'acte	2			
	Dates de validité de l'acte compatibles avec les dates du RUM	1			1 : oui, 2 : non
...
Zone acte n° Nb_Act_R_1 du RUM n° 1	Délai depuis la date d'entrée	3			
	Code CCAM	7			
	Extension PMSI	2			
	Phase	1			
	Activité	1			
	Extension documentaire	1			
	Modificateurs	4			
	Remboursement exceptionnel	1			
	Association non prévue	1			
	Nombre de réalisations de l'acte	2			
	Dates de validité de l'acte compatibles avec les dates du RUM	1			1 : oui, 2 : non

Zone acte n° 1 du RUM n° NbRUM	Délai depuis la date d'entrée	3			
	Code CCAM	7			
	Extension PMSI	2			
	Phase	1			
	Activité	1			
	Extension documentaire	1			
	Modificateurs	4			
	Remboursement exceptionnel	1			
	Association non prévue	1			
	Nombre de réalisations de l'acte	2			
	Dates de validité de l'acte compatibles avec les dates du RUM	1			1 : oui, 2 : non

Zone acte n° Nb_Act_R_NbRUM du RUM n° NbRUM	Délai depuis la date d'entrée	3			
	Code CCAM	7			
	Extension PMSI	2			
	Phase	1			
	Activité	1			
	Extension documentaire	1			
	Modificateurs	4			
	Remboursement exceptionnel	1			
	Association non prévue	1			
	Nombre de réalisations de l'acte	2			
	Dates de validité de l'acte compatibles avec les dates du RUM	1			1 : oui, 2 : non

IV. Bibliographie sur le chaînage des données PMSI

Titre	Auteurs	Revue
Patient's permanent identification and PMSI chaining system. Complementary or competing methods?	Herengt G, Toussaint E, Musse JP, Kohler F.	Stud Health Technol Inform. 2002
From DRG databases to an epidemiological observatory for colorectal cancer in a French small area oncology network.	Paviot BT, Martin C, Clavel L, De Laroche G, Rodrigues JM.	Stud Health Technol Inform. 2003
[Analysis of the medical activity related to cancer in a network of multidisciplinary hospitals using claims databases, the reseau Concorde Oncology Network].	Schott AM, Hajri T, Gelas-Dore B, Couris CM, Couray-Targe S, Trillet-Lenoir V, Dumeril B, Grandjean JP, Lledo G, Poncet JL, Colin C, Cautela N, Gilly FN.	Bull Cancer. 2005
[Evaluation of the quality of surveillance data for colorectal cancer from the national PMSI database in 2003].	Beyeme-Ondoua JP.	Sante Publique. 2007
Is it possible to estimate the incidence of breast cancer from medico-administrative databases?	Remontet L, Mitton N, Couris CM, Iwaz J, Gomez F, Olive F, Polazzi S, Schott AM, Trombert B, Bossard N, Colonna M.	Eur J Epidemiol. 2008
Breast cancer incidence using administrative data: correction with sensitivity and specificity.	Couris CM, Polazzi S, Olive F, Remontet L, Bossard N, Gomez F, Schott AM, Mitton N, Colonna M, Trombert B.	J Clin Epidemiol. 2009
[Estimating emergency hospital admissions to gauge short-term effects of air pollution: evaluation of health data quality].	Bois de Fer B, Host S, Chardon B, Chatignoux E, Beaujouan L, Brun-Ney D, Grémy I.	Sante Publique. 2009
Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus.	Weill A, Païta M, Tuppin P, Fagot JP, Neumann A, Simon D, Ricordeau P, Montastruc JL, Allemand H.	Pharmacoepidemiol Drug Saf. 2010
[Evaluation of hospital stays linkage quality to study health spatial variation. a feasibility study in three French regions].	Bocquier A, Thomas N, Zitouni J, Lewandowski E, Cortaredona S, Jardin M, Favier O, Finkel S, Champion F, Bernardy A, Trugeon A, Verger P.	Rev Epidemiol Sante Publique. 2011
[Preventable readmissions of community-acquired pneumonia": Usefulness and reliability of an indicator of the quality of care of patients' care pathways]."	Dely C, Sellier P, Dozol A, Segouin C, Moret L, Lombrail P.	Presse Med. 2012
[Study of hospitalizations for pneumococcal pneumoniae in Centre region, 2004-2008].	Grammatico-Guillon L, Thiercelin N, Mariani S, Lecuyer AI, Goudeau A, Bernard L, Rusch E.	Rev Epidemiol Sante Publique. 2012
Burden of breast cancer with brain metastasis: a French national hospital database analysis.	Benjamin L, Cotté FE, Mercier F, Vainchtock A, Vidal-Trécan G, Durand-Zaleski I.	J Med Econ. 2012
Estimate of deaths due to valvular	Fournier A, Zureik M.	Pharmacoepidemiol

Titre	Auteurs	Revue
insufficiency attributable to the use of benfluorex in France.		Drug Saf. 2012
[Acute and post-acute hospitalizations for stroke in France: recent improvements (2007-2009)].	de Peretti C, Nicolau J, Tuppin P, Schnitzler A, Woimant F.	Presse Med. 2012
Pioglitazone and risk of bladder cancer among diabetic patients in France: a population-based cohort study.	Neumann A, Weill A, Ricordeau P, Fagot JP, Alla F, Allemand H.	Diabetologia. 2012
[Precarious parturients and readmission: pilot study at the Lariboisière-Fernand-Widal Hospital Group in Paris].	Atete-Leblanc R, Bréchat PH, Morel O, Thoury A, Frati A, Barranger E.	Gynecol Obstet Fertil. 2012
[Performance evaluation of hospital claims database for the identification of incident central nervous system tumors compared with a cancer registry in Gironde, France, 2004].	Coureau G, Baldi I, Savès M, Jaffré A, Barat C, Gruber A, Loiseau H, Kostrzewa A, Servas V, Salamon R, Gilleron V.	Rev Epidemiol Sante Publique. 2012
Estimation of national colorectal-cancer incidence using claims databases.	Quantin C, Benzenine E, Håggi M, Auverlot B, Abrahamowicz M, Cottenet J, Fournier E, Binquet C, Compain D, Monnet E, Bouvier AM, Danzon A.	J Cancer Epidemiol. 2012
[Role of French hospital claims databases from care units in epidemiological studies: the example of the Cohorte "Enfant Scanner" study]."	Bernier MO, Mezzarobba M, Maupu E, Caër-Lorho S, Brisse HJ, Laurier D, Brunelle F, Chatellier G.	Rev Epidemiol Sante Publique. 2012
Does insulin glargine increase the risk of cancer compared with other basal insulins?: A French nationwide cohort study based on national administrative databases.	Fagot JP, Blotière PO, Ricordeau P, Weill A, Alla F, Allemand H.	Diabetes Care. 2013
[The role of a hospital-based registry of a regional cancer center for a population-based cancer registry].	Colombani F, Pereira E, Bettaieb J, Gobin L, Cowppli-Bony A, Hoppe S, Coureau G, Picat MQ, Salamon R, Monnereau A, Savès M.	Rev Epidemiol Sante Publique. 2013
Outcomes after acute myocardial infarction in HIV-infected patients: analysis of data from a French nationwide hospital medical information database.	Lorgis L, Cottenet J, Molins G, Benzenine E, Zeller M, Aube H, Touzery C, Hamblin J, Gudjoncik A, Cottin Y, Quantin C.	Circulation. 2013
Role of comorbid conditions in asthma hospitalizations in the south of France.	Aubas C, Bourdin A, Aubas P, Gamez AS, Halimi L, Vachier I, Malafaye N, Chanez P, Molinari N.	Allergy. 2013
Perforations and haemorrhages after colonoscopy in 2010: a study based on comprehensive French health insurance data (SNIIRAM).	Blotière PO, Weill A, Ricordeau P, Alla F, Allemand H.	Clin Res Hepatol Gastroenterol. 2014
Comparative effectiveness of rosuvastatin versus simvastatin in primary prevention	Neumann A, Maura G, Weill A, Ricordeau P, Alla F, Allemand H.	Pharmacoepidemiol Drug Saf. 2014

Titre	Auteurs	Revue
among new users: a cohort study in the French national health insurance database.		
[Drug in pregnancy: studies in the French database EFEMERIS].	Damase-Michel C, Lacroix I, Hurault-Delarue C, Beau AB, Montastruc JL; les partenaires d'EFEMERIS.	Therapie. 2014
[Evaluation of medical information systems as a mean of identification of incident cases of colorectal cancer].	Quantin C, Benzenine E, Hägi M, Auverlot B, Abrahamowicz M, Cottenet J, Fournier E, Binquet C, Compain D, Monnet E, Bouvier AM, Danzon A.	Sante Publique. 2014
Prevalence and incidence of systemic lupus erythematosus in France: a 2010 nation-wide population-based study.	Arnaud L, Fagot JP, Mathian A, Paita M, Fagot-Campagna A, Amoura Z.	Autoimmun Rev. 2014
Allogeneic Stem Cell Transplant With Graft Versus Host Disease In France In 2012 Based On National Hospital Databases (Pmsi).	Thiebaut A, Lafuma A, Bureau I, Boyaval G, Bensoussan C, Godard C.	Value Health. 2014
Comparison of rehospitalization rates in France and the United States.	Gusmano M, Rodwin V, Weisz D, Cottenet J, Quantin C.	J Health Serv Res Policy. 2015
The cost of thromboembolic events in hospitalized patients with breast or prostate cancer in France.	Scotte F, Martelli N, Vainchtock A, Borget I.	Adv Ther. 2015
[Analysis from the French DRG-based information system (PMSI) of conservative surgical treatment for abnormal uterine bleeding in 2008-2010].	Fernandez H, Villefranque V, Panel P.	J Gynecol Obstet Biol Reprod (Paris). 2015
Compliance with pregnancy prevention plan recommendations in 8672 French women of childbearing potential exposed to acitretin.	Raguideau F, Mezzarobba M, Zureik M, Weill A, Ricordeau P, Alla F.	Pharmacoepidemiol Drug Saf. 2015
EpiHosp: A web-based visualization tool enabling the exploratory analysis of complications of implantable medical devices from a nationwide hospital database.	Ficheur G, Ferreira Careira L, Beuscart R, Chazard E.	Stud Health Technol Inform. 2015
[Analysis of reimbursement of dialysis patients' transport expenses in Lorraine].	Allenbach D, Pereira O.	Sante Publique. 2015
[Use in France of Baclofen for Alcohol Dependence from 2007 to 2013: Cohort Study Based on the Databases SNIIRAM and PMSI].	Chaignot C, Weill A, Ricordeau P, Alla F.	Therapie. 2015
An Assessment of Annual Costs of Patients Hospitalized for Spinal Tumors in France: Analysis Using the Pmsi Database.	de Léotoing L, Fernandes J, Tournier C, Jouaneton B, Vainchtock A.	Value Health. 2015
Cost-effectiveness model comparing dual-mobility to fixed-bearing designs for total	Epinette JA, Lafuma A, Robert J, Doz M.	Orthop Traumatol Surg Res. 2016

Titre	Auteurs	Revue
hip replacement in France.		
Early cardiovascular events in women with a history of gestational diabetes mellitus.	Goueslard K, Cottenet J, Mariet AS, Giroud M, Cottin Y, Petit JM, Quantin C.	Cardiovasc Diabetol. 2016
Conditions of prescription of anti-TNF agents in newly treated patients with inflammatory bowel disease in France (2011-2013).	Blotière PO, Rudant J, Barré A, Racine A, Weill A, Peyrin-Biroulet L, Carbonnel F, Alla F.	Dig Liver Dis. 2016
Predictive ability of HAS-BLED, HEMORR2HAGES, and ATRIA bleeding risk scores in patients with atrial fibrillation. A French nationwide cross-sectional study.	Fauchier L, Chaize G, Gaudin AF, Vainchtock A, Rushton-Smith SK, Cotté FE.	Int J Cardiol. 2016
Access to and characteristics of palliative care-related hospitalization in the management of end-stage renal disease patients on renal replacement therapy in France.	Couchoud C, Bello AD, Lobbedez T, Blanchard S, Chantrel F, Maurizi-Balzan J, Moranne O; REIN registry.	Nephrology (Carlton). 2016
[Clinical burden caused by hospitalization for febrile neutropenia in France in 2010-2011: An analysis of the PMSI database].	Freyer G, Scotte F, Borget I, Bruyas A, Vainchtock A, Chouaid C.	Bull Cancer. 2016
Estimation of breast, prostate, and colorectal cancer incidence using a French administrative database (general sample of health insurance beneficiaries).	Doat S, Samson S, Fagot-Campagna A, Tuppin P, Menegaux F.	Rev Epidemiol Sante Publique. 2016
Natural History of Adjustable Gastric Banding: Lifespan and Revisional Rate: A Nationwide Study on Administrative Data on 53,000 Patients.	Lazzati A, De Antonio M, Paolino L, Martini F, Azoulay D, Iannelli A, Katsahian S.	Ann Surg. 2016
[The opening of the French national health database: Opportunities and difficulties. The experience of the Gazel and Constances cohorts].	Goldberg M, Carton M, Gourmelen J, Genreau M, Montourcy M, Le Got S, Zins M.	Rev Epidemiol Sante Publique. 2016

V. Racines de GHM classées par contribution croissante à l'entropie

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
25Z03	Maladies dues au VIH, âge inférieur à 13 ans	2,48E-05	1,86E-06	1,86E-06
06M15	Suivi de greffes de l'appareil digestif	2,80E-05	2,10E-06	3,96E-06
05K18	Varices avec acte autre que ligature et éveinage, en ambulatoire	5,12E-05	3,84E-06	7,80E-06
27C03	Transplantations pancréatiques	7,24E-05	5,43E-06	1,32E-05
07K05	Séjours comprenant une endoscopie biliaire thérapeutique ou diagnostique sans anesthésie, en ambulatoire	1,22E-04	9,13E-06	2,24E-05
05C09	Autres interventions cardiothoraciques, âge inférieur à 2 ans, sans circulation extracorporelle	1,86E-04	1,40E-05	3,63E-05
13M08	Assistance médicale à la procréation	2,09E-04	1,57E-05	5,20E-05
15C03	Interventions majeures sur l'appareil cardiovasculaire, groupes nouveau-nés 1 à 7	2,51E-04	1,88E-05	7,08E-05
15C05	Interventions chirurgicales, groupes nouveau-nés 8 à 9	2,73E-04	2,05E-05	9,13E-05
27C04	Transplantations pulmonaires	2,83E-04	2,12E-05	1,12E-04
27Z04	Greffes de cellules souches hématopoïétiques, en ambulatoire	3,26E-04	2,44E-05	1,37E-04
12C09	Autres interventions pour tumeurs malignes de l'appareil génital masculin	3,56E-04	2,67E-05	1,64E-04
12K03	Séjours de la CMD 12 comprenant une endoscopie génito-urinaire sans anesthésie : séjours de moins de deux jours	3,57E-04	2,67E-05	1,90E-04
25Z02	Maladies dues au VIH, avec décès	3,59E-04	2,69E-05	2,17E-04
12M09	Symptômes et autres recours aux soins de la CMD 12	3,74E-04	2,81E-05	2,45E-04
09M15	Symptômes et autres recours aux soins concernant les affections des seins	3,78E-04	2,84E-05	2,74E-04
22Z02	Brûlures étendues	3,78E-04	2,84E-05	3,02E-04
16M18	Autres affections hématologiques concernant majoritairement la petite enfance	3,89E-04	2,91E-05	3,31E-04
15C06	Interventions chirurgicales, groupe nouveau-nés 10	3,92E-04	2,94E-05	3,61E-04
27C05	Transplantations cardiaques	3,92E-04	2,94E-05	3,90E-04
11M20	Autres affections uronéphrologiques concernant majoritairement la petite enfance	4,20E-04	3,15E-05	4,21E-04
22C03	Brûlures non étendues avec parages de plaie ou autres interventions chirurgicales	4,54E-04	3,40E-05	4,56E-04
01M38	Autres affections neurologiques concernant majoritairement la petite enfance	5,23E-04	3,92E-05	4,95E-04
14C06	Césariennes avec naissance d'un mort-né	5,52E-04	4,14E-05	5,36E-04
07C07	Interventions diagnostiques sur le système hépatobiliaire et pancréatique pour affections non malignes	5,55E-04	4,16E-05	5,78E-04
15C02	Interventions majeures sur l'appareil digestif, groupes nouveau-nés 1 à 7	5,67E-04	4,25E-05	6,20E-04

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
07C08	Autres interventions sur le système hépato-biliaire et pancréatique	6,32E-04	4,74E-05	6,68E-04
14C03	Accouchements par voie basse avec autres interventions	6,44E-04	4,83E-05	7,16E-04
06M10	Ulcères gastroduodénaux compliqués	6,59E-04	4,94E-05	7,65E-04
25C02	Interventions pour maladie due au VIH	6,92E-04	5,19E-05	8,17E-04
07C06	Interventions diagnostiques sur le système hépato-biliaire et pancréatique pour affections malignes	7,14E-04	5,35E-05	8,71E-04
22Z03	Brûlures avec transfert vers un autre établissement MCO, séjours de moins de 2 jours	7,23E-04	5,42E-05	9,25E-04
16M07	Donneurs de moelle	7,40E-04	5,55E-05	9,80E-04
01C11	Craniotomies pour tumeurs, âge inférieur à 18 ans	7,41E-04	5,55E-05	1,04E-03
05C20	Remplacements ou ablations chirurgicale d'électrodes ou repositionnements de boîtier de stimulation cardiaque permanente	7,42E-04	5,56E-05	1,09E-03
18M15	Autres maladies infectieuses concernant majoritairement la petite enfance	7,96E-04	5,97E-05	1,15E-03
05C07	Autres interventions cardiothoraciques, âge inférieur à 2 ans, avec circulation extracorporelle	8,45E-04	6,34E-05	1,21E-03
11C06	Interventions sur l'urètre, âge inférieur à 18 ans	8,48E-04	6,36E-05	1,28E-03
01M39	Troubles de la régulation thermique du nouveau-né et du nourrisson	8,55E-04	6,41E-05	1,34E-03
11M11	Signes et symptômes concernant les reins et les voies urinaires, âge inférieur à 18 ans	8,72E-04	6,54E-05	1,41E-03
17K05	Curiéthérapies de la prostate	8,73E-04	6,54E-05	1,47E-03
13M10	Autres symptômes et recours aux soins de la CMD 13	9,10E-04	6,82E-05	1,54E-03
18M13	Affections de la CMD 18 avec décès : séjours de moins de 2 jours	9,19E-04	6,89E-05	1,61E-03
07M16	Ictères du nouveau-né	9,48E-04	7,11E-05	1,68E-03
14C10	Affections de l'ante partum avec intervention chirurgicale	9,81E-04	7,36E-05	1,75E-03
12M08	Explorations et surveillance des affections de l'appareil génital masculin	1,03E-03	7,72E-05	1,83E-03
27C02	Transplantations hépatiques	1,04E-03	7,78E-05	1,91E-03
19C02	Interventions chirurgicales avec un diagnostic principal de maladie mentale	1,07E-03	8,02E-05	1,99E-03
17M08	Leucémies aigües, âge inférieur à 18 ans	1,13E-03	8,50E-05	2,08E-03
01M33	Troubles du sommeil	1,15E-03	8,61E-05	2,16E-03
16C02	Interventions sur la rate	1,17E-03	8,74E-05	2,25E-03
02M10	Symptômes et autres recours aux soins de la CMD 02	1,17E-03	8,74E-05	2,34E-03
01C09	Pose d'un stimulateur cérébral	1,17E-03	8,81E-05	2,42E-03
10M10	Maladies métaboliques congénitales sévères	1,19E-03	8,93E-05	2,51E-03
05K12	Actes thérapeutiques par voie vasculaire sauf endoprothèses, âge inférieur à 18 ans	1,21E-03	9,04E-05	2,60E-03

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
08C06	Amputations pour affections de l'appareil musculosquelettique et du tissu conjonctif	1,22E-03	9,11E-05	2,69E-03
02C06	Interventions primaires sur l'iris	1,24E-03	9,29E-05	2,79E-03
03C18	Pose d'implants cochléaires	1,26E-03	9,42E-05	2,88E-03
06M21	Autres affections digestives concernant majoritairement la petite enfance	1,28E-03	9,61E-05	2,98E-03
09M13	Explorations et surveillance des affections des seins	1,28E-03	9,61E-05	3,07E-03
08K03	Tractions continues et réductions progressives : autres que hanche et fémur	1,28E-03	9,62E-05	3,17E-03
21M13	Maltraitance	1,30E-03	9,74E-05	3,27E-03
01K07	Embolisations intracrâniennes et médullaires pour hémorragie	1,30E-03	9,77E-05	3,37E-03
10C02	Interventions sur l'hypophyse	1,31E-03	9,85E-05	3,46E-03
12C10	Autres interventions pour affections non malignes de l'appareil génital masculin	1,32E-03	9,87E-05	3,56E-03
16M08	Déficits immunitaires	1,33E-03	9,96E-05	3,66E-03
15C04	Autres interventions chirurgicales, groupes nouveau-nés 1 à 7	1,33E-03	9,99E-05	3,76E-03
08C02	Interventions majeures multiples sur les genoux et/ou les hanches	1,36E-03	1,02E-04	3,86E-03
18M06	Septicémies, âge inférieur à 18 ans	1,37E-03	1,03E-04	3,97E-03
12K02	Endoscopies génito-urinaires et anesthésie : séjours de la CMD 12 et de moins de deux jours	1,37E-03	1,03E-04	4,07E-03
23M13	Désensibilisation nécessitant une hospitalisation	1,39E-03	1,04E-04	4,17E-03
08C12	Biopsies ostéoarticulaires	1,40E-03	1,05E-04	4,28E-03
11K04	Séjours de la CMD 11 comprenant une endoscopie génito-urinaire thérapeutique sans anesthésie : séjours de moins de 2 jours	1,45E-03	1,09E-04	4,39E-03
01C12	Craniotomies pour affections non tumorales, âge inférieur à 18 ans	1,46E-03	1,09E-04	4,50E-03
06C23	Certaines interventions pour stomies	1,47E-03	1,10E-04	4,61E-03
11K07	Séjours de la CMD 11 comprenant la mise en place de certains accès vasculaires, en ambulatoire	1,48E-03	1,11E-04	4,72E-03
10C03	Interventions sur les glandes surrénales	1,53E-03	1,15E-04	4,83E-03
07C11	Dérivations biliaires	1,54E-03	1,15E-04	4,95E-03
07K06	Actes thérapeutiques par voie vasculaire pour des affections malignes du système hépatobiliaire	1,54E-03	1,16E-04	5,06E-03
10C07	Interventions sur le tractus thyroïdienne	1,55E-03	1,16E-04	5,18E-03
13M09	Explorations et surveillance gynécologiques	1,59E-03	1,19E-04	5,30E-03
05M21	Infarctus aigu du myocarde avec décès : séjours de moins de 2 jours	1,60E-03	1,20E-04	5,42E-03
06C05	Interventions sur l'oesophage, l'estomac et le duodénum, âge inférieur à 18 ans	1,61E-03	1,20E-04	5,54E-03
27Z02	Allogreffes de cellules souches hématopoïétiques	1,65E-03	1,24E-04	5,66E-03
08C29	Interventions sur le tissu mou pour tumeurs malignes	1,66E-03	1,25E-04	5,79E-03

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
08M08	Entorses et luxations de la hanche et du bassin	1,68E-03	1,26E-04	5,91E-03
04M27	Autres affections respiratoires concernant majoritairement la petite enfance	1,68E-03	1,26E-04	6,04E-03
01M29	Anévrysmes cérébraux	1,69E-03	1,26E-04	6,17E-03
19M15	Autres psychoses, âge supérieur à 79 ans	1,69E-03	1,27E-04	6,29E-03
03C06	Interventions sur les sinus et l'apophyse mastoïde, âge inférieur à 18 ans	1,72E-03	1,29E-04	6,42E-03
10M19	Autres affections de la CMD 10 concernant majoritairement la petite enfance	1,73E-03	1,30E-04	6,55E-03
15M14	Nouveau-nés de 800 g et âge gestationnel de 28SA et assimilés (groupe nouveau-nés 10)	1,75E-03	1,31E-04	6,68E-03
03C29	Autres interventions sur l'oreille, le nez ou la gorge pour tumeurs malignes	1,75E-03	1,31E-04	6,81E-03
23M07	Autres motifs de recours pour infection à VIH, en ambulatoire	1,81E-03	1,36E-04	6,95E-03
05M10	Cardiopathies congénitales et valvulopathies, âge inférieur à 18 ans	1,82E-03	1,37E-04	7,09E-03
08C55	Interventions sur la cheville et l'arrière-pied pour fractures	1,83E-03	1,38E-04	7,22E-03
16M06	Affections de la rate	1,84E-03	1,38E-04	7,36E-03
04M06	Infections et inflammations respiratoires, âge inférieur à 18 ans	1,84E-03	1,38E-04	7,50E-03
08M24	Tumeurs malignes primitives des os	1,85E-03	1,38E-04	7,64E-03
15M03	Décès précoces et tardifs de nouveau-nés	1,85E-03	1,38E-04	7,78E-03
08K04	Tractions continues et réductions progressives : hanche et fémur	1,91E-03	1,43E-04	7,92E-03
01K03	Autres actes thérapeutiques par voie vasculaire du système nerveux	1,93E-03	1,45E-04	8,06E-03
03C05	Réparations de fissures labiale et palatine	1,94E-03	1,46E-04	8,21E-03
10C08	Autres interventions pour troubles endocriniens, métaboliques ou nutritionnels	1,97E-03	1,48E-04	8,36E-03
22C02	Brûlures non étendues avec greffe cutanée	1,97E-03	1,48E-04	8,51E-03
21M06	Traumatismes imprécis, âge inférieur à 18 ans	1,98E-03	1,48E-04	8,65E-03
02M02	Hyphéma	2,01E-03	1,50E-04	8,80E-03
14C04	Affections du post-partum ou du post abortum avec intervention chirurgicale	2,05E-03	1,53E-04	8,96E-03
08M23	Fractures, entorses, luxations et dislocations du pied	2,06E-03	1,54E-04	9,11E-03
13K05	Endoscopies génito-urinaires diagnostiques sans anesthésie : séjours de la CMD 13 et de moins de 2 jours	2,12E-03	1,59E-04	9,27E-03
04M26	Fibroses kystiques avec manifestations pulmonaires	2,13E-03	1,60E-04	9,43E-03
06M14	Invaginations intestinales aiguës	2,16E-03	1,62E-04	9,59E-03
18M14	Symptômes et autres recours aux soins de la CMD 18	2,17E-03	1,62E-04	9,76E-03

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
03M12	Infections aiguës sévères des voies aériennes supérieures, âge inférieur à 18 ans	2,20E-03	1,65E-04	9,92E-03
22K02	Brûlures sans acte opératoire, avec anesthésie, en ambulatoire	2,21E-03	1,66E-04	0,0101
13M06	Autres infections de l'appareil génital féminin	2,23E-03	1,67E-04	0,0103
08C28	Interventions maxillofaciales	2,33E-03	1,74E-04	0,0104
08C20	Greffes de peau pour maladie de l'appareil musculosquelettique ou du tissu conjonctif	2,34E-03	1,76E-04	0,0106
03C30	Interventions sur l'oreille externe	2,37E-03	1,78E-04	0,0108
03M14	Explorations et surveillance pour affections ORL	2,38E-03	1,78E-04	0,0110
21K02	Traumatismes, allergies et empoisonnements sans acte opératoire, avec anesthésie, en ambulatoire	2,42E-03	1,81E-04	0,0111
15M13	Nouveau-nés de 1 100 g et âge gestationnel de 30 SA et assimilés (groupe nouveau-nés 9)	2,44E-03	1,83E-04	0,0113
05K22	Actes thérapeutiques par voie vasculaire sur les orifices du cœur	2,49E-03	1,87E-04	0,0115
08C57	Libérations articulaires du membre inférieur à l'exception de la hanche et du pied	2,50E-03	1,87E-04	0,0117
16M17	Autres troubles de la lignée érythrocytaire, âge inférieur à 18 ans	2,51E-03	1,88E-04	0,0119
27C06	Transplantations rénales	2,51E-03	1,88E-04	0,0121
14Z11	Accouchements multiples par voie basse chez une primipare	2,52E-03	1,89E-04	0,0123
18M09	Paludisme	2,56E-03	1,92E-04	0,0125
08M38	Entorses et luxations du rachis	2,57E-03	1,92E-04	0,0126
18M10	Maladies infectieuses sévères	2,57E-03	1,93E-04	0,0128
27Z03	Autogreffes de cellules souches hématopoïétiques	2,62E-03	1,97E-04	0,0130
17M13	Polyglobulies	2,64E-03	1,98E-04	0,0132
12C05	Interventions sur les testicules pour tumeurs malignes	2,69E-03	2,02E-04	0,0134
01K02	Autres embolisations intracrâniennes et médullaires	2,75E-03	2,06E-04	0,0136
17C04	Interventions majeures pour affections myéloprolifératives ou tumeurs de siège imprécis ou diffus	2,78E-03	2,08E-04	0,0138
06C20	Interventions sur l'oesophage, l'estomac et le duodénum pour ulcères, âge supérieur à 17 ans	2,82E-03	2,12E-04	0,0141
08M06	Fractures, entorses, luxations et dislocations de la jambe, âge inférieur à 18 ans	2,83E-03	2,12E-04	0,0143
08M33	Ablation de matériel sans acte classant	2,89E-03	2,16E-04	0,0145
12M07	Autres infections et inflammations de l'appareil génital masculin	2,89E-03	2,17E-04	0,0147
12C13	Stérilisation et vasoplastie	2,91E-03	2,18E-04	0,0149
05K23	Ablations, repositionnements et poses de sondes cardiaques supplémentaires par voie vasculaire	2,91E-03	2,18E-04	0,0151
08M05	Fractures de la diaphyse, de l'épiphyse ou d'une	2,94E-03	2,21E-04	0,0154

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	partie non précisée du fémur			
02C10	Autres interventions intraoculaires pour affections sévères	2,96E-03	2,22E-04	0,0156
07M15	Affections hépatiques sévères à l'exception des tumeurs malignes, des cirrhoses et des hépatites alcooliques	2,99E-03	2,24E-04	0,0158
01M37	Autres affections de la CMD 01 avec décès : séjours de moins de 2 jours	3,07E-03	2,30E-04	0,0160
18C02	Interventions pour maladies infectieuses ou parasitaires	3,07E-03	2,30E-04	0,0163
05C02	Chirurgie de remplacement valvulaire avec circulation extracorporelle et avec cathétérisme cardiaque ou coronarographie	3,08E-03	2,31E-04	0,0165
21M11	Effets toxiques des autres substances chimiques	3,09E-03	2,32E-04	0,0167
14Z09	Accouchements hors de l'établissement	3,10E-03	2,32E-04	0,0170
01M09	Affections et lésions du rachis et de la moelle	3,10E-03	2,33E-04	0,0172
07M09	Hépatites chroniques	3,11E-03	2,33E-04	0,0174
08C04	Interventions sur la hanche et le fémur, âge inférieur à 18 ans	3,11E-03	2,34E-04	0,0177
21M15	Rejets de greffe	3,12E-03	2,34E-04	0,0179
05C04	Pontages aortocoronariens avec cathétérisme cardiaque ou coronarographie	3,13E-03	2,35E-04	0,0181
07M12	Suivis de greffe de foie et de pancréas	3,15E-03	2,36E-04	0,0184
02M03	Infections oculaires aiguës sévères	3,16E-03	2,37E-04	0,0186
11K06	Séjours de la CMD 11 comprenant une endoscopie génito-urinaire diagnostique sans anesthésie : séjours de moins de 2 jours	3,24E-03	2,43E-04	0,0188
05M14	Arrêt cardiaque	3,26E-03	2,44E-04	0,0191
09C11	Reconstructions des seins	3,30E-03	2,47E-04	0,0193
20Z03	Abus de drogues non éthyliques sans dépendance	3,40E-03	2,55E-04	0,0196
13C05	Interventions sur le système utéroannexiel pour tumeurs malignes	3,40E-03	2,55E-04	0,0199
05K21	Poses de bioprothèses de valves cardiaques par voie vasculaire	3,46E-03	2,59E-04	0,0201
05K15	Surveillances de greffes de coeur avec acte diagnostique par voie vasculaire	3,47E-03	2,60E-04	0,0204
23M18	Tests allergologiques nécessitant une hospitalisation	3,49E-03	2,61E-04	0,0206
11C10	Interventions pour incontinence urinaire en dehors des interventions transurétrales	3,51E-03	2,63E-04	0,0209
04C03	Autres interventions chirurgicales sur le système respiratoire	3,52E-03	2,64E-04	0,0212
05M19	Surveillances de greffes de coeur sans acte diagnostique par voie vasculaire	3,54E-03	2,65E-04	0,0214
08M32	Ostéomyélites chroniques	3,56E-03	2,67E-04	0,0217
01M28	Hydrocéphalies	3,56E-03	2,67E-04	0,0220

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
03M08	Autres diagnostics portant sur les oreilles, le nez, la gorge ou la bouche, âge inférieur à 18 ans	3,57E-03	2,68E-04	0,0222
11M08	Autres affections des reins et des voies urinaires, âge inférieur à 18 ans	3,59E-03	2,69E-04	0,0225
08C36	Interventions sur le pied, âge inférieur à 18 ans	3,59E-03	2,69E-04	0,0228
03C13	Interventions sur les amygdales et les végétations adénoïdes autres que les amygdalectomies et/ou les adénoïdectomies isolées, âge supérieur à 17 ans	3,62E-03	2,71E-04	0,0230
17M11	Autres leucémies	3,63E-03	2,72E-04	0,0233
08M34	Algoneurodystrophie	3,65E-03	2,73E-04	0,0236
14Z12	Accouchements multiples par voie basse chez une multipare	3,65E-03	2,74E-04	0,0239
13M07	Autres tumeurs de l'appareil génital féminin	3,66E-03	2,74E-04	0,0241
05C08	Autres interventions cardiothoraciques, âge supérieur à 1 an, ou vasculaires quel que soit l'âge, sans circulation extracorporelle	3,66E-03	2,74E-04	0,0244
04M21	Suivis de greffe pulmonaire	3,72E-03	2,79E-04	0,0247
01M36	Accidents vasculaires cérébraux non transitoires avec décès : séjours de moins de 2 jours	3,78E-03	2,83E-04	0,0250
23M11	Autres motifs concernant majoritairement la petite enfance	3,83E-03	2,87E-04	0,0253
07C10	Interventions sur le foie, le pancréas et les veines porte ou cave pour affections non malignes	3,88E-03	2,91E-04	0,0255
11C08	Autres interventions sur les reins et les voies urinaires	3,90E-03	2,93E-04	0,0258
10M20	Problèmes alimentaires du nouveau-né et du nourrisson	3,92E-03	2,94E-04	0,0261
21M04	Réactions allergiques non classées ailleurs, âge inférieur à 18 ans	3,93E-03	2,95E-04	0,0264
02M05	Autres affections oculaires, âge inférieur à 18 ans	3,96E-03	2,97E-04	0,0267
01C03	Craniotomies pour traumatisme, âge supérieur à 17 ans	4,03E-03	3,02E-04	0,0270
13C13	Autres interventions sur l'appareil génital féminin	4,04E-03	3,03E-04	0,0273
11C07	Interventions sur l'urètre, âge supérieur à 17 ans	4,05E-03	3,04E-04	0,0276
12M04	Hypertrophie prostatique bénigne	4,06E-03	3,05E-04	0,0279
01C10	Pose d'un stimulateur médullaire	4,07E-03	3,05E-04	0,0282
03C25	Interventions majeures sur la tête et le cou	4,08E-03	3,06E-04	0,0285
09M14	Symptômes et autres recours aux soins concernant les affections de la peau	4,09E-03	3,07E-04	0,0289
26M02	Traumatismes multiples graves	4,10E-03	3,08E-04	0,0292
10C09	Gastroplasties pour obésité	4,14E-03	3,10E-04	0,0295
19M13	Autres troubles de la personnalité et du comportement avec réactions impulsives	4,16E-03	3,12E-04	0,0298
09C02	Greffes de peau et/ou parages de plaie pour ulcère cutané ou cellulite	4,18E-03	3,13E-04	0,0301
03C26	Autres interventions sur la tête et le cou	4,24E-03	3,18E-04	0,0304

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
11M10	Rétrécissement urétral	4,25E-03	3,18E-04	0,0307
09C13	Interventions pour condylomes anogénitaux	4,26E-03	3,19E-04	0,0311
09C14	Certains curages lymphonodaux pour des affections de la peau, des tissus sous-cutanés ou des seins	4,30E-03	3,22E-04	0,0314
23M16	Convalescences et autres motifs sociaux	4,30E-03	3,22E-04	0,0317
19M12	Anorexie mentale et boulimie	4,34E-03	3,26E-04	0,0320
04M19	Tuberculoses	4,38E-03	3,28E-04	0,0323
19M10	Névroses autres que les névroses dépressives	4,38E-03	3,29E-04	0,0327
07C12	Autres interventions sur les voies biliaires sauf cholécystectomies isolées	4,40E-03	3,30E-04	0,0330
11C12	Injections de toxine botulique dans l'appareil urinaire	4,43E-03	3,32E-04	0,0333
05M18	Endocardites aiguës et subaiguës	4,44E-03	3,33E-04	0,0337
23Z03	Interventions de confort et autres interventions non prises en charge par l'assurance maladie obligatoire	4,46E-03	3,34E-04	0,0340
18M03	Maladies virales, âge supérieur à 17 ans	4,49E-03	3,37E-04	0,0343
13M05	Infections de l'utérus et de ses annexes	4,50E-03	3,37E-04	0,0347
06C21	Autres interventions sur le tube digestif par laparotomie	4,51E-03	3,38E-04	0,0350
10M17	Troubles nutritionnels divers, âge inférieur à 18 ans	4,52E-03	3,39E-04	0,0354
05C06	Autres interventions cardiothoraciques, âge supérieur à 1 an, ou vasculaires quel que soit l'âge, avec circulation extracorporelle	4,53E-03	3,40E-04	0,0357
11K05	Endoscopies génito-urinaires diagnostiques et anesthésie : séjours de la CMD 11 et de moins de 2 jours	4,54E-03	3,41E-04	0,0360
15M12	Nouveau-nés de 1 300 g et âge gestationnel de 32 SA et assimilés (groupe nouveau-nés 8)	4,57E-03	3,43E-04	0,0364
15M11	Nouveau-nés de 1 500 g et âge gestationnel de 33 SA et assimilés (groupe nouveau-nés 7)	4,59E-03	3,44E-04	0,0367
08C13	Résections osseuses localisées et/ou ablation de matériel de fixation interne au niveau de la hanche et du fémur	4,65E-03	3,49E-04	0,0371
06C13	Libérations d'adhérences péritonéales	4,69E-03	3,52E-04	0,0374
02C03	Interventions sur l'orbite	4,69E-03	3,52E-04	0,0378
01M27	Autres tumeurs du système nerveux	4,70E-03	3,52E-04	0,0381
06C16	Interventions sur l'oesophage, l'estomac et le duodénum pour tumeurs malignes, âge supérieur à 17 ans	4,73E-03	3,55E-04	0,0385
02C12	Interventions sur le cristallin avec trabéculéctomie	4,88E-03	3,66E-04	0,0389
11C04	Autres interventions sur la vessie à l'exception des interventions transurétrales	4,92E-03	3,69E-04	0,0392
13C19	Interventions pour stérilité ou motifs de soins liés à la reproduction	4,93E-03	3,69E-04	0,0396

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
09M09	Affections non malignes des seins	4,94E-03	3,71E-04	0,0400
13C06	Interruptions tubaires	5,01E-03	3,76E-04	0,0403
10M15	Troubles métaboliques, âge inférieur à 18 ans	5,01E-03	3,76E-04	0,0407
08M30	Rhumatismes et raideurs articulaires	5,04E-03	3,78E-04	0,0411
22M02	Brûlures et gelures non étendues sans intervention chirurgicale	5,05E-03	3,79E-04	0,0415
26C02	Interventions pour traumatismes multiples graves	5,15E-03	3,86E-04	0,0419
13M03	Tumeurs malignes de l'appareil génital féminin	5,20E-03	3,90E-04	0,0422
08C50	Interventions sur la hanche et le fémur sauf traumatismes récents, âge supérieur à 17 ans	5,20E-03	3,90E-04	0,0426
10M11	Autres maladies métaboliques congénitales	5,22E-03	3,91E-04	0,0430
08C31	Interventions sur la jambe, âge inférieur à 18 ans	5,26E-03	3,94E-04	0,0434
02M07	Autres affections oculaires d'origine diabétique, âge supérieur à 17 ans	5,27E-03	3,95E-04	0,0438
17C05	Autres interventions au cours d'affections myéloprolifératives ou de tumeurs de siège imprécis ou diffus	5,32E-03	3,99E-04	0,0442
05C12	Amputations du membre inférieur, sauf des orteils, pour troubles circulatoires	5,33E-03	4,00E-04	0,0446
19M19	Troubles de l'humeur	5,38E-03	4,03E-04	0,0450
05C13	Amputations pour troubles circulatoires portant sur le membre supérieur ou les orteils	5,38E-03	4,04E-04	0,0454
02M09	Explorations et surveillance pour affections de l'oeil	5,41E-03	4,06E-04	0,0458
01M10	Autres affections cérébrovasculaires	5,43E-03	4,07E-04	0,0462
14Z06	Menaces d'avortement	5,45E-03	4,09E-04	0,0466
08M22	Fractures, entorses, luxations et dislocations de la main	5,47E-03	4,10E-04	0,0471
15M10	Nouveau-nés de 1 700 g et âge gestationnel de 35 SA et assimilés (groupe nouveau-nés 6)	5,47E-03	4,10E-04	0,0475
12M03	Tumeurs malignes de l'appareil génital masculin	5,52E-03	4,14E-04	0,0479
07K02	Endoscopies biliaires thérapeutiques et anesthésie : séjours de moins de 2 jours	5,52E-03	4,14E-04	0,0483
14Z15	Grossesses ectopiques sans intervention chirurgicale	5,55E-03	4,16E-04	0,0487
17C02	Interventions majeures au cours de lymphomes ou de leucémies	5,60E-03	4,20E-04	0,0491
19M17	Maladies et troubles du développement psychologiques de l'enfance	5,61E-03	4,20E-04	0,0495
03M15	Symptômes et autres recours aux soins de la CMD 03	5,62E-03	4,22E-04	0,0500
05K24	Dilatations coronaires et autres actes thérapeutiques sur le coeur par voie vasculaire	5,63E-03	4,22E-04	0,0504
03M13	Infections aiguës sévères des voies aériennes supérieures, âge supérieur à 17 ans	5,64E-03	4,23E-04	0,0508
13C18	Myomectomies de l'utérus	5,65E-03	4,24E-04	0,0512
01K05	Séjours pour douleurs chroniques rebelles	5,65E-03	4,24E-04	0,0517

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	comprenant un bloc ou une infiltration, en ambulatoire			
19M18	Autres maladies et troubles mentaux de l'enfance	5,67E-03	4,25E-04	0,0521
14Z10	Accouchements par voie basse avec naissance d'un mort-né	5,70E-03	4,27E-04	0,0525
21M07	Traumatismes imprécis, âge supérieur à 17 ans	5,72E-03	4,29E-04	0,0529
12C12	Interventions pelviennes majeures chez l'homme pour affections non malignes	5,73E-03	4,29E-04	0,0534
09M10	Tumeurs malignes des seins	5,80E-03	4,35E-04	0,0538
21M12	Autres effets toxiques	5,81E-03	4,36E-04	0,0542
17K07	Affections myéloprolifératives et tumeurs de siège imprécis sans acte opératoire, avec anesthésie, en ambulatoire	5,82E-03	4,36E-04	0,0547
03C12	Interventions sur les amygdales et les végétations adénoïdes autres que les amygdalectomies et/ou les adénoïdectomies isolées, âge inférieur à 18 ans	5,91E-03	4,43E-04	0,0551
16C03	Autres interventions pour affections du sang et des organes hématopoïétiques	5,96E-03	4,47E-04	0,0556
17M09	Leucémies aigües, âge supérieur à 17 ans	5,97E-03	4,47E-04	0,0560
09M11	Tumeurs de la peau	5,98E-03	4,49E-04	0,0565
20Z06	Troubles mentaux organiques induits par l'alcool ou d'autres substances	6,01E-03	4,51E-04	0,0569
08M09	Arthropathies non spécifiques	6,05E-03	4,53E-04	0,0574
01M04	Méningites virales	6,09E-03	4,56E-04	0,0578
21M14	Autres traumatismes et effets nocifs autres que les intoxications	6,11E-03	4,58E-04	0,0583
13K06	Affections de l'appareil génital féminin sans acte opératoire de la CMD 13, avec anesthésie, en ambulatoire	6,17E-03	4,63E-04	0,0587
04M15	Autres diagnostics portant sur le système respiratoire	6,20E-03	4,65E-04	0,0592
01M18	Lésions traumatiques intracrâniennes sévères	6,21E-03	4,65E-04	0,0597
05C18	Autres interventions sur le système circulatoire	6,21E-03	4,65E-04	0,0601
13C20	Exérèses ou destructions de lésions du col de l'utérus sauf conisations	6,26E-03	4,70E-04	0,0606
03K04	Séjours comprenant certains actes non opératoires de la CMD 03, en ambulatoire	6,32E-03	4,74E-04	0,0611
13K03	Séjours de la CMD 13 comprenant une endoscopie génito-urinaire thérapeutique sans anesthésie : séjours de moins de 2 jours	6,34E-03	4,76E-04	0,0616
06M17	Soins de stomies digestives	6,35E-03	4,76E-04	0,0620
04M25	Grippes	6,35E-03	4,76E-04	0,0625
10C05	Interventions sur les parathyroïdes	6,37E-03	4,77E-04	0,0630
23M02	Rééducation	6,41E-03	4,81E-04	0,0635
17K04	Autres irradiations	6,43E-03	4,82E-04	0,0640
17K06	Autres curiethérapies et irradiations internes	6,45E-03	4,84E-04	0,0644

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
02C13	Interventions sur les muscles oculomoteurs	6,47E-03	4,85E-04	0,0649
01K06	Affections du système nerveux sans acte opératoire avec anesthésie, en ambulatoire	6,53E-03	4,90E-04	0,0654
15Z10	Mort-nés	6,57E-03	4,93E-04	0,0659
18M11	Autres maladies infectieuses ou parasitaires	6,64E-03	4,98E-04	0,0664
06M11	Ulcères gastroduodénaux non compliqués	6,65E-03	4,99E-04	0,0669
03C15	Drains transtympaniques, âge supérieur à 17 ans	6,78E-03	5,08E-04	0,0674
17M05	Chimiothérapie pour leucémie aigüe	6,78E-03	5,09E-04	0,0679
25M02	Autres maladies dues au VIH	6,86E-03	5,14E-04	0,0684
07M08	Autres cirrhoses et fibrose hépatique	6,88E-03	5,16E-04	0,0689
05C14	Poses d'un stimulateur cardiaque permanent avec infarctus aigu du myocarde ou insuffisance cardiaque congestive ou état de choc	7,01E-03	5,26E-04	0,0695
11M15	Autres affections des reins et des voies urinaires d'origine diabétique, âge supérieur à 17 ans	7,02E-03	5,26E-04	0,0700
13C09	Laparoscopies ou coelioscopies diagnostiques	7,02E-03	5,26E-04	0,0705
06K03	Séjours comprenant une endoscopie digestive thérapeutique sans anesthésie, en ambulatoire	7,03E-03	5,27E-04	0,0711
02C09	Allogreffes de cornée	7,08E-03	5,31E-04	0,0716
04M23	Autres symptômes et recours aux soins de la CMD 04	7,13E-03	5,35E-04	0,0721
09M04	Lésions, infections et inflammations de la peau et des tissus sous-cutanés, âge inférieur à 18 ans	7,20E-03	5,40E-04	0,0727
14C07	Césariennes pour grossesse multiple	7,21E-03	5,41E-04	0,0732
14C09	Grossesses ectopiques avec intervention chirurgicale	7,22E-03	5,41E-04	0,0737
06K06	Affections digestives sans acte opératoire de la CMD 06, avec anesthésie, en ambulatoire	7,22E-03	5,41E-04	0,0743
07C09	Interventions sur le foie, le pancréas et les veines porte ou cave pour tumeurs malignes	7,22E-03	5,42E-04	0,0748
07M11	Autres affections non malignes du pancréas	7,28E-03	5,46E-04	0,0754
10M12	Tumeurs des glandes endocrines	7,34E-03	5,50E-04	0,0759
21C04	Interventions sur la main ou le poignet à la suite de blessures	7,36E-03	5,52E-04	0,0765
13C11	Dilatations et curetages, conisations pour tumeurs malignes	7,36E-03	5,52E-04	0,0770
05M11	Cardiopathies congénitales et valvulopathies, âge supérieur à 17 ans	7,37E-03	5,53E-04	0,0776
10C11	Interventions sur la thyroïde pour tumeurs malignes	7,38E-03	5,53E-04	0,0781
02C07	Autres interventions extraoculaires, âge inférieur à 18 ans	7,44E-03	5,58E-04	0,0787
08M18	Suites de traitement après une affection de l'appareil musculosquelettique ou du tissu conjonctif	7,48E-03	5,61E-04	0,0792
23M08	Autres motifs de recours chez un patient	7,55E-03	5,66E-04	0,0798

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	diabétique, en ambulatoire			
16M12	Purpuras	7,59E-03	5,69E-04	0,0804
12M05	Autres affections de l'appareil génital masculin	7,60E-03	5,70E-04	0,0810
09C15	Interventions sur la peau, les tissus sous-cutanés ou les seins pour lésions traumatiques	7,62E-03	5,72E-04	0,0815
19M14	Troubles bipolaires et syndromes dépressifs sévères	7,63E-03	5,72E-04	0,0821
02M04	Affections oculaires d'origine neurologique	7,78E-03	5,83E-04	0,0827
03C19	Ostéotomies de la face	7,79E-03	5,84E-04	0,0833
19M20	Autres troubles mentaux	7,83E-03	5,87E-04	0,0839
06C15	Autres interventions sur le tube digestif en dehors des laparotomies	7,94E-03	5,95E-04	0,0844
16M16	Troubles sévères de la lignée érythrocytaire, âge inférieur à 18 ans	7,95E-03	5,96E-04	0,0850
11K03	Endoscopies génito-urinaires thérapeutiques et anesthésie : séjours de la CMD 11 et de moins de 2 jours	7,96E-03	5,97E-04	0,0856
16M14	Explorations et surveillance pour affections du sang et des organes hématopoïétiques	7,97E-03	5,97E-04	0,0862
06C22	Interventions sur l'oesophage, l'estomac et le duodénum pour affections autres que malignes ou ulcères, âge supérieur à 17 ans	7,98E-03	5,99E-04	0,0868
21C06	Greffes de peau ou parages de plaies pour lésions autres que des brûlures	7,99E-03	5,99E-04	0,0874
13C15	Exentérations pelviennes, hystérectomies élargies ou vulvectomies pour affections non malignes	8,00E-03	6,00E-04	0,0880
20Z02	Toxicomanies non éthyliques avec dépendance	8,04E-03	6,03E-04	0,0886
08M21	Entorses, luxations et dislocations du bras et de l'avant-bras, âge supérieur à 17 ans	8,11E-03	6,08E-04	0,0892
05C21	Créations et réfections de fistules artérioveineuses pour affections de la CMD 05	8,16E-03	6,12E-04	0,0899
03C24	Interventions sur les glandes salivaires	8,23E-03	6,17E-04	0,0905
09C07	Biopsies et excisions locales pour des affections non malignes du sein	8,24E-03	6,18E-04	0,0911
15M09	Nouveau-nés de 1 800 g et âge gestationnel de 36 SA et assimilés (groupe nouveau-nés 5)	8,35E-03	6,26E-04	0,0917
21M05	Réactions allergiques non classées ailleurs, âge supérieur à 17 ans	8,38E-03	6,28E-04	0,0923
05C05	Pontages aortocoronariens sans cathétérisme cardiaque, ni coronarographie	8,42E-03	6,31E-04	0,0930
01M05	Infections du système nerveux à l'exception des méningites virales	8,52E-03	6,39E-04	0,0936
11C03	Interventions sur les reins et les uretères et chirurgie majeure de la vessie pour une affection non tumorale	8,53E-03	6,39E-04	0,0943
17C03	Autres interventions au cours de lymphomes ou de leucémies	8,70E-03	6,53E-04	0,0949

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
12C03	Interventions sur le pénis	8,84E-03	6,63E-04	0,0956
08C33	Interventions sur la cheville et l'arrière-pied à l'exception des fractures	8,85E-03	6,63E-04	0,0962
01M13	Troubles de la conscience et comas d'origine non traumatique	8,91E-03	6,68E-04	0,0969
04C04	Interventions sous thoracoscopie	9,01E-03	6,75E-04	0,0976
05K19	Traitements majeurs de troubles du rythme par voie vasculaire	9,05E-03	6,78E-04	0,0983
01M23	Convulsions hyperthermiques	9,09E-03	6,81E-04	0,0989
08C51	Interventions majeures sur le rachis pour fractures, cyphoses et scolioses	9,12E-03	6,84E-04	0,0996
08C53	Interventions sur le genou pour traumatismes	9,28E-03	6,96E-04	0,1003
23M19	Explorations et surveillance pour autres motifs de recours aux soins	9,30E-03	6,97E-04	0,1010
08C54	Interventions sur le genou pour des affections autres que traumatiques	9,50E-03	7,12E-04	0,1017
10M14	Symptômes et autres recours aux soins de la CMD 10	9,56E-03	7,17E-04	0,1024
03C11	Amygdalectomies et/ou adénoïdectomies isolées, âge supérieur à 17 ans	9,59E-03	7,19E-04	0,1032
04M24	Affections de la CMD 04 avec décès : séjours de moins de 2 jours	9,62E-03	7,21E-04	0,1039
11K02	Insuffisance rénale, avec dialyse	9,73E-03	7,30E-04	0,1046
14M02	Affections médicales du post-partum ou du post-abortum	9,84E-03	7,38E-04	0,1053
05K26	Actes thérapeutiques sur les accès vasculaires ou les veines par voie vasculaire	9,87E-03	7,40E-04	0,1061
08M31	Ostéomyélites aiguës (y compris vertébrales) et arthrites septiques	9,88E-03	7,41E-04	0,1068
23C02	Interventions chirurgicales avec autres motifs de recours aux services de santé	9,92E-03	7,44E-04	0,1076
05M23	Symptômes et autres recours aux soins de la CMD 05	0,0100	7,53E-04	0,1083
06M13	Tumeurs malignes de l'oesophage et de l'estomac	0,0100	7,53E-04	0,1091
03M06	Epistaxis	0,0102	7,64E-04	0,1098
08C21	Autres interventions portant sur l'appareil musculosquelettique et le tissu conjonctif	0,0102	7,68E-04	0,1106
11M07	Tumeurs des reins et des voies urinaires	0,0103	7,74E-04	0,1114
05M22	Autres affections de la CMD 05 avec décès : séjours de moins de 2 jours	0,0106	7,91E-04	0,1122
10M03	Diabète, âge inférieur à 36 ans	0,0106	7,94E-04	0,1130
13C10	Ligatures tubaires par laparoscopie ou coelioscopie	0,0106	7,98E-04	0,1138
07M13	Explorations et surveillance des affections du système hépatobiliaire et du pancréas	0,0107	7,99E-04	0,1146
10C10	Autres interventions pour obésité	0,0107	8,00E-04	0,1154
13C14	Exentérations pelviennes, hystérectomies élargies	0,0107	8,02E-04	0,1162

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	ou vulvectomies pour tumeurs malignes			
06M08	Autres affections digestives, âge inférieur à 18 ans	0,0107	8,02E-04	0,1170
21M02	Effets toxiques des médicaments et substances biologiques, âge inférieur à 18 ans	0,0107	8,03E-04	0,1178
05C19	Poses d'un défibrillateur cardiaque	0,0107	8,05E-04	0,1186
15M08	Nouveau-nés de 2 000 g et âge gestationnel de 37 SA et assimilés (groupe nouveau-nés 4)	0,0109	8,19E-04	0,1194
15M02	Transferts précoces de nouveau-nés vers un autre établissement MCO	0,0110	8,25E-04	0,1202
06C07	Interventions mineures sur l'intestin grêle et le côlon	0,0110	8,25E-04	0,1211
05C11	Autres interventions de chirurgie vasculaire	0,0110	8,26E-04	0,1219
03M02	Traumatismes et déformations du nez	0,0112	8,39E-04	0,1227
16M13	Autres troubles de la coagulation	0,0113	8,50E-04	0,1236
05C03	Chirurgie de remplacement valvulaire avec circulation extracorporelle, sans cathétérisme cardiaque, ni coronarographie	0,0114	8,52E-04	0,1244
03M04	Otites moyennes et autres infections des voies aériennes supérieures, âge supérieur à 17 ans	0,0115	8,61E-04	0,1253
01C05	Interventions sur le rachis et la moelle pour des affections neurologiques	0,0115	8,65E-04	0,1261
17M07	Autres affections myéloprolifératives et tumeurs de siège imprécis ou diffus	0,0115	8,65E-04	0,1270
19M16	Autres psychoses, âge inférieur à 80 ans	0,0115	8,65E-04	0,1279
01M07	Maladies dégénératives du système nerveux, âge supérieur à 79 ans	0,0116	8,68E-04	0,1287
09M08	Affections dermatologiques sévères	0,0116	8,72E-04	0,1296
04M16	Traumatismes thoraciques	0,0118	8,84E-04	0,1305
11C09	Créations et réfections de fistules artérioveineuses pour affections de la CMD 11	0,0118	8,86E-04	0,1314
03C27	Interventions sur les amygdales	0,0119	8,90E-04	0,1323
08C59	Ténosynovectomies du poignet	0,0121	9,05E-04	0,1332
21M16	Autres complications iatrogéniques non classées ailleurs	0,0121	9,08E-04	0,1341
01M31	Autres accidents vasculaires cérébraux non transitoires	0,0121	9,09E-04	0,1350
03C21	Interventions pour oreilles décollées	0,0122	9,12E-04	0,1359
16M09	Autres affections du système réticuloendothélial ou immunitaire	0,0122	9,14E-04	0,1368
19M22	Symptômes et autres recours aux soins de la CMD 19	0,0123	9,19E-04	0,1377
06M19	Affections sévères du tube digestif	0,0123	9,25E-04	0,1387
04M04	Pneumonies et pleurésies banales, âge inférieur à 18 ans	0,0124	9,29E-04	0,1396
01C08	Interventions sur les nerfs crâniens ou périphériques et autres interventions sur le	0,0124	9,31E-04	0,1405

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	système nerveux			
06C10	Interventions réparatrices pour hernies et éventrations, âge inférieur à 18 ans	0,0125	9,37E-04	0,1415
08M20	Fractures, entorses, luxations et dislocations du bras et de l'avant-bras, âge inférieur à 18 ans	0,0125	9,38E-04	0,1424
08C25	Prothèses d'épaule	0,0125	9,40E-04	0,1433
09M02	Traumatismes de la peau et des tissus sous-cutanés, âge inférieur à 18 ans	0,0126	9,43E-04	0,1443
04M14	Maladies pulmonaires interstitielles	0,0126	9,46E-04	0,1452
14Z04	Avortements sans aspiration, ni curetage, ni hystérotomie	0,0127	9,49E-04	0,1462
09M12	Explorations et surveillance des affections de la peau	0,0127	9,52E-04	0,1471
08M07	Fractures, entorses, luxations et dislocations de la jambe, âge supérieur à 17 ans	0,0127	9,53E-04	0,1481
23M15	Actes non effectués en raison d'une contre-indication	0,0128	9,56E-04	0,1490
05M07	Thrombophlébites veineuses profondes	0,0130	9,72E-04	0,1500
03M09	Autres diagnostics portant sur les oreilles, le nez, la gorge ou la bouche, âge supérieur à 17 ans	0,0130	9,74E-04	0,1510
05C22	Remplacements de défibrillateurs ou de stimulateurs cardiaques permanents	0,0131	9,85E-04	0,1520
10M08	Acidocétose et coma diabétique	0,0133	9,95E-04	0,1530
08C40	Arthroscopies d'autres localisations	0,0134	1,00E-03	0,1540
01M26	Tumeurs malignes du système nerveux	0,0134	1,01E-03	0,1550
01M17	Sclérose en plaques et ataxie cérébelleuse	0,0135	1,01E-03	0,1560
05K25	Actes thérapeutiques sur les artères par voie vasculaire	0,0135	1,01E-03	0,1570
01M15	Accidents ischémiques transitoires et occlusions des artères précérébrales, âge supérieur à 79 ans	0,0135	1,01E-03	0,1580
05M16	Athérosclérose coronarienne	0,0136	1,02E-03	0,1590
01C06	Interventions sur le système vasculaire précérébral	0,0138	1,03E-03	0,1601
06M20	Tumeurs bénignes de l'appareil digestif	0,0139	1,04E-03	0,1611
15M07	Nouveau-nés de 2 200 g et âge gestationnel de 37 SA et assimilés (groupe nouveau-nés 3)	0,0139	1,04E-03	0,1621
21C05	Autres interventions pour blessures ou complications d'acte	0,0139	1,05E-03	0,1632
01M12	Autres affections du système nerveux	0,0142	1,07E-03	0,1643
04M12	Pneumothorax	0,0144	1,08E-03	0,1653
01M35	Symptômes et autres recours aux soins de la CMD 01	0,0144	1,08E-03	0,1664
11C02	Interventions sur les reins et les uretères et chirurgie majeure de la vessie pour une affection tumorale	0,0146	1,09E-03	0,1675
02M08	Autres affections oculaires d'origine non diabétique, âge supérieur à 17 ans	0,0147	1,10E-03	0,1686

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
02C11	Autres interventions intraoculaires en dehors des affections sévères	0,0148	1,11E-03	0,1697
06C03	Résections rectales	0,0149	1,12E-03	0,1708
01M19	Autres lésions traumatiques intracrâniennes, sauf commotions	0,0150	1,12E-03	0,1720
03M07	Tumeurs malignes des oreilles, du nez, de la gorge ou de la bouche	0,0151	1,13E-03	0,1731
06M07	Maladies inflammatoires de l'intestin	0,0152	1,14E-03	0,1742
12C11	Interventions pelviennes majeures chez l'homme pour tumeurs malignes	0,0152	1,14E-03	0,1754
07K04	Endoscopie biliaire diagnostique et anesthésie, en ambulatoire	0,0153	1,15E-03	0,1765
17M14	Explorations et surveillance pour affections myéloprolifératives et tumeurs de siège imprécis ou diffus	0,0154	1,15E-03	0,1777
08M04	Fractures de la hanche et du bassin	0,0157	1,17E-03	0,1788
11M18	Explorations et surveillance pour affections du rein et des voies urinaires	0,0157	1,17E-03	0,1800
18M12	Explorations et surveillance pour maladies infectieuses ou parasitaires	0,0157	1,18E-03	0,1812
01M24	Epilepsie, âge inférieur à 18 ans	0,0158	1,18E-03	0,1824
03C17	Interventions sur la bouche	0,0158	1,19E-03	0,1836
03K03	Séjours comprenant une endoscopie oto-rhino-laryngologique, en ambulatoire	0,0159	1,19E-03	0,1848
05K20	Autres traitements de troubles du rythme par voie vasculaire	0,0160	1,20E-03	0,1860
01C04	Craniotomies en dehors de tout traumatisme, âge supérieur à 17 ans	0,0161	1,21E-03	0,1872
08C60	Interventions sur le poignet autres que les ténosynovectomies	0,0164	1,23E-03	0,1884
12K06	Séjours comprenant une biopsie prostatique, en ambulatoire	0,0164	1,23E-03	0,1896
04C02	Interventions majeures sur le thorax	0,0165	1,24E-03	0,1909
03M10	Affections de la bouche et des dents sans certaines extractions, réparations ou prothèses dentaires, âge inférieur à 18 ans	0,0165	1,24E-03	0,1921
11M17	Surveillances de greffes de rein	0,0166	1,24E-03	0,1933
10M09	Obésité	0,0167	1,25E-03	0,1946
09C04	Mastectomies totales pour tumeur maligne	0,0168	1,26E-03	0,1959
08C56	Interventions pour infections ostéoarticulaires	0,0168	1,26E-03	0,1971
08M35	Explorations et surveillance de l'appareil musculosquelettique et du tissu conjonctif	0,0175	1,31E-03	0,1984
05K17	Affections cardiovasculaires sans acte opératoire de la CMD 05, avec anesthésie, en ambulatoire	0,0176	1,32E-03	0,1997
07M07	Cirrhoses alcooliques	0,0177	1,33E-03	0,2011
13K04	Endoscopies génito-urinaires diagnostiques et	0,0177	1,33E-03	0,2024

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	anesthésie : séjours de la CMD 13 et de moins de 2 jours			
09M06	Ulcères cutanés	0,0180	1,35E-03	0,2038
12C07	Interventions sur les testicules pour affections non malignes, âge supérieur à 17 ans	0,0181	1,36E-03	0,2051
19M21	Explorations et surveillance pour maladies et troubles mentaux	0,0185	1,38E-03	0,2065
08M37	Fractures du bras et de l'avant-bras, âge supérieur à 17 ans	0,0186	1,39E-03	0,2079
08M27	Sciatiques et autres radiculopathies	0,0187	1,40E-03	0,2093
07M04	Autres affections hépatiques	0,0187	1,40E-03	0,2107
11M16	Autres affections des reins et des voies urinaires, à l'exception de celles d'origine diabétique, âge supérieur à 17 ans	0,0188	1,41E-03	0,2121
12C06	Interventions sur les testicules pour affections non malignes, âge inférieur à 18 ans	0,0190	1,43E-03	0,2135
01C14	Libérations de nerfs superficiels à l'exception du médian au canal carpien	0,0198	1,49E-03	0,2150
13M04	Autres affections de l'appareil génital féminin	0,0198	1,49E-03	0,2165
03C16	Autres interventions chirurgicales portant sur les oreilles, le nez, la gorge ou le cou	0,0201	1,51E-03	0,2180
08C22	Interventions pour reprise de prothèses articulaires	0,0201	1,51E-03	0,2195
08M36	Symptômes et autres recours aux soins de la CMD 08	0,0202	1,52E-03	0,2210
08M15	Tendinites, myosites et bursites	0,0203	1,52E-03	0,2226
04M08	Bronchopneumopathies chroniques	0,0204	1,53E-03	0,2241
06M16	Explorations et surveillance pour affections de l'appareil digestif	0,0204	1,53E-03	0,2256
18M07	Septicémies, âge supérieur à 17 ans	0,0206	1,54E-03	0,2272
18M02	Maladies virales et fièvres d'étiologie indéterminée, âge inférieur 18 ans	0,0207	1,55E-03	0,2287
04M17	Epanchements pleuraux	0,0208	1,56E-03	0,2303
23M21	Désensibilisations et tests allergologiques nécessitant une hospitalisation	0,0209	1,57E-03	0,2318
08C52	Autres interventions majeures sur le rachis	0,0210	1,57E-03	0,2334
09C09	Interventions plastiques en dehors de la chirurgie esthétique	0,0210	1,57E-03	0,2350
17M12	Lymphomes et autres affections malignes hématopoiétiques	0,0211	1,59E-03	0,2366
06K05	Séjours comprenant une endoscopie digestive diagnostique sans anesthésie, en ambulatoire	0,0213	1,60E-03	0,2382
08C38	Autres arthroscopies du genou	0,0214	1,60E-03	0,2398
16M10	Troubles sévères de la lignée érythrocytaire, âge supérieur à 17 ans	0,0215	1,61E-03	0,2414
08M28	Autres rachialgies	0,0215	1,61E-03	0,2430
10M07	Autres troubles endocriniens	0,0218	1,63E-03	0,2446

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
08M19	Autres pathologies de l'appareil musculosquelettique et du tissu conjonctif	0,0219	1,64E-03	0,2463
13C04	Interventions réparatrices sur l'appareil génital féminin	0,0220	1,65E-03	0,2479
03C20	Interventions de reconstruction de l'oreille moyenne	0,0224	1,68E-03	0,2496
13C17	Cervicocystopexie	0,0226	1,69E-03	0,2513
10M18	Troubles nutritionnels divers, âge supérieur à 17 ans	0,0226	1,70E-03	0,2530
11M03	Infections des reins et des voies urinaires, âge inférieur à 18 ans	0,0227	1,70E-03	0,2547
12M06	Prostatites aiguës et orchites	0,0229	1,72E-03	0,2564
03C14	Drains transtympaniques, âge inférieur à 18 ans	0,0232	1,74E-03	0,2581
05M06	Angine de poitrine	0,0234	1,76E-03	0,2599
05C10	Chirurgie majeure de revascularisation	0,0235	1,76E-03	0,2617
16M15	Symptômes et autres recours aux soins de la CMD 16	0,0235	1,76E-03	0,2634
06C08	Appendicectomies compliquées	0,0236	1,77E-03	0,2652
01M08	Maladies dégénératives du système nerveux, âge inférieur à 80 ans	0,0236	1,77E-03	0,2670
18M04	Fièvres d'étiologie indéterminée, âge supérieur à 17 ans	0,0237	1,78E-03	0,2687
05M04	Infarctus aigu du myocarde	0,0237	1,78E-03	0,2705
01M16	Accidents ischémiques transitoires et occlusions des artères précérébrales, âge inférieur à 80 ans	0,0238	1,78E-03	0,2723
04K02	Séjours comprenant une endoscopie bronchique, en ambulatoire	0,0244	1,83E-03	0,2741
08C42	Interventions non mineures sur les tissus mous	0,0245	1,83E-03	0,2760
08C43	Interventions non mineures sur la main	0,0246	1,84E-03	0,2778
08M26	Fractures du rachis	0,0249	1,87E-03	0,2797
06M05	Autres tumeurs malignes du tube digestif	0,0250	1,87E-03	0,2815
06C24	Cures d'événements postopératoires	0,0251	1,88E-03	0,2834
09M07	Autres affections dermatologiques	0,0252	1,89E-03	0,2853
03M03	Otitis moyennes et autres infections des voies aériennes supérieures, âge inférieur à 18 ans	0,0254	1,91E-03	0,2872
05M20	Explorations et surveillance pour affections de l'appareil circulatoire	0,0255	1,91E-03	0,2891
11K08	Lithotritie extracorporelle de l'appareil urinaire, en ambulatoire	0,0255	1,91E-03	0,2910
13C08	Interventions sur la vulve, le vagin ou le col utérin	0,0256	1,92E-03	0,2930
03C10	Amygdalectomies et/ou adénoïdectomies isolées, âge inférieur à 18 ans	0,0257	1,93E-03	0,2949
08M25	Fractures pathologiques et autres tumeurs malignes de l'appareil musculosquelettique et du tissu conjonctif	0,0258	1,94E-03	0,2968
11M19	Autres symptômes et recours aux soins de la CMD	0,0259	1,94E-03	0,2988

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	11			
23M10	Soins de contrôle chirurgicaux	0,0263	1,97E-03	0,3007
03M11	Affections de la bouche et des dents sans certaines extractions, réparations ou prothèses dentaires, âge supérieur à 17 ans	0,0263	1,97E-03	0,3027
13C12	Dilatations et curetages, conisations pour affections non malignes	0,0267	2,00E-03	0,3047
03C09	Rhinoplasties	0,0267	2,00E-03	0,3067
07C13	Cholécystectomies sans exploration de la voie biliaire principale pour affections aiguës	0,0270	2,02E-03	0,3087
06C19	Hémorroïdectomies	0,0274	2,05E-03	0,3108
08M10	Maladies osseuses et arthropathies spécifiques	0,0275	2,06E-03	0,3129
23M09	Chimiothérapie pour affections non tumorales	0,0276	2,07E-03	0,3149
07M10	Pancréatites aiguës	0,0281	2,10E-03	0,3170
10C12	Interventions sur la thyroïde pour affections non malignes	0,0283	2,12E-03	0,3192
08K02	Affections de l'appareil musculosquelettique sans acte opératoire de la CMD 08, avec anesthésie, en ambulatoire	0,0288	2,16E-03	0,3213
08C47	Prothèses de hanche pour traumatismes récents	0,0289	2,16E-03	0,3235
05M15	Hypertension artérielle	0,0289	2,17E-03	0,3256
08M14	Affections du tissu conjonctif	0,0291	2,18E-03	0,3278
09C12	Interventions pour kystes, granulomes et interventions sur les ongles	0,0293	2,20E-03	0,3300
19M11	Névroses dépressives	0,0296	2,22E-03	0,3322
03M05	Troubles de l'équilibre	0,0298	2,24E-03	0,3345
05K05	Endoprothèses vasculaires avec infarctus du myocarde	0,0299	2,24E-03	0,3367
19M07	Troubles mentaux d'origine organique et retards mentaux, âge inférieur à 80 ans	0,0309	2,32E-03	0,3390
09C08	Interventions sur la région anale et périanale	0,0311	2,33E-03	0,3414
03C07	Interventions sur les sinus et l'apophyse mastoïde, âge supérieur à 17 ans	0,0313	2,35E-03	0,3437
11M06	Insuffisance rénale, sans dialyse	0,0313	2,35E-03	0,3461
10C13	Interventions digestives autres que les gastroplasties, pour obésité	0,0315	2,36E-03	0,3484
19M02	Troubles aigus de l'adaptation et du fonctionnement psychosocial	0,0320	2,40E-03	0,3508
04M10	Embolies pulmonaires	0,0322	2,41E-03	0,3532
05C15	Poses d'un stimulateur cardiaque permanent sans infarctus aigu du myocarde, ni insuffisance cardiaque congestive, ni état de choc	0,0322	2,41E-03	0,3557
23M14	Traitements prophylactiques	0,0322	2,41E-03	0,3581
02C02	Interventions sur la rétine	0,0323	2,42E-03	0,3605
01M11	Affections des nerfs crâniens et rachidiens	0,0327	2,45E-03	0,3629

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
08C34	Interventions sur les ligaments croisés sous arthroscopie	0,0331	2,49E-03	0,3654
04M07	Infections et inflammations respiratoires, âge supérieur à 17 ans	0,0339	2,54E-03	0,3680
04M22	Explorations et surveillance pour affections de l'appareil respiratoire	0,0341	2,56E-03	0,3705
11M12	Signes et symptômes concernant les reins et les voies urinaires, âge supérieur à 17 ans	0,0344	2,58E-03	0,3731
04M18	Bronchiolites	0,0351	2,63E-03	0,3757
23K03	Motifs de recours de la CMD 23 sans acte opératoire, avec anesthésie, en ambulatoire	0,0357	2,68E-03	0,3784
09C06	Interventions sur le sein pour des affections non malignes autres que les actes de biopsie et d'excision locale	0,0357	2,68E-03	0,3811
13C07	Interventions sur le système utéroannexiel pour des affections non malignes, autres que les interruptions tubaires	0,0360	2,70E-03	0,3838
07M06	Affections malignes du système hépato-biliaire ou du pancréas	0,0362	2,72E-03	0,3865
13C03	Hystérectomies	0,0362	2,72E-03	0,3892
01M34	Anomalies de la démarche d'origine neurologique	0,0363	2,72E-03	0,3920
06C09	Appendicectomies non compliquées	0,0364	2,73E-03	0,3947
06C14	Interventions sur le rectum et l'anus autres que les résections rectales	0,0367	2,75E-03	0,3974
09Z02	Chirurgie esthétique	0,0370	2,78E-03	0,4002
04M02	Bronchites et asthme, âge inférieur à 18 ans	0,0373	2,80E-03	0,4030
01M22	Migraines et céphalées	0,0375	2,81E-03	0,4058
04M09	Tumeurs de l'appareil respiratoire	0,0378	2,83E-03	0,4087
08C35	Interventions sur le bras, coude et épaule	0,0380	2,85E-03	0,4115
09K02	Affections de la peau, des tissus sous-cutanés et des seins sans acte opératoire de la CMD 09, avec anesthésie, en ambulatoire	0,0397	2,97E-03	0,4145
09C05	Mastectomies subtotaux pour tumeur maligne	0,0398	2,99E-03	0,4175
05M12	Troubles vasculaires périphériques	0,0398	2,99E-03	0,4205
04M13	Oedème pulmonaire et détresse respiratoire	0,0398	2,99E-03	0,4234
06C25	Interventions réparatrices pour hernies à l'exception des hernies inguinales, crurales	0,0401	3,01E-03	0,4264
13C16	Prélèvements d'ovocytes, en ambulatoire	0,0407	3,05E-03	0,4295
04M11	Signes et symptômes respiratoires	0,0408	3,06E-03	0,4326
08C32	Interventions sur la jambe, âge supérieur à 17 ans	0,0409	3,07E-03	0,4356
01M32	Explorations et surveillance pour affections du système nerveux	0,0414	3,11E-03	0,4387
06M06	Occlusions intestinales non dues à une hernie	0,0417	3,13E-03	0,4419
15M06	Nouveau-nés de 2 400 g et âge gestationnel de 38 SA et assimilés (groupe nouveau-nés 2)	0,0427	3,20E-03	0,4451
09M05	Lésions, infections et inflammations de la peau et	0,0431	3,23E-03	0,4483

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
	des tissus sous-cutanés, âge supérieur à 17 ans			
04M03	Bronchites et asthme, âge supérieur à 17 ans	0,0437	3,28E-03	0,4516
06M09	Autres affections digestives, âge supérieur à 17 ans	0,0438	3,28E-03	0,4549
05M17	Autres affections de l'appareil circulatoire	0,0445	3,34E-03	0,4582
10M16	Troubles métaboliques, âge supérieur à 17 ans	0,0445	3,34E-03	0,4615
08C27	Autres interventions sur le rachis	0,0448	3,36E-03	0,4649
08C49	Interventions sur la hanche et le fémur pour traumatismes récents, âge supérieur à 17 ans	0,0450	3,37E-03	0,4683
01M21	Douleurs chroniques rebelles	0,0451	3,38E-03	0,4716
08M29	Autres pathologies rachidiennes relevant d'un traitement médical	0,0454	3,40E-03	0,4750
14C05	Avortements avec aspiration ou curetage ou hystérotomie	0,0454	3,40E-03	0,4785
07M14	Symptômes et autres recours aux soins de la CMD 07	0,0455	3,42E-03	0,4819
08C46	Autres interventions sur les tissus mous	0,0458	3,43E-03	0,4853
11M02	Lithiases urinaires	0,0459	3,44E-03	0,4887
08C58	Arthroscopies de l'épaule	0,0463	3,47E-03	0,4922
06M04	Hémorragies digestives	0,0466	3,49E-03	0,4957
12C04	Prostatectomies transurétrales	0,0471	3,53E-03	0,4992
10M02	Diabète, âge supérieur à 35 ans	0,0485	3,63E-03	0,5029
05M05	Syncopes et lipothymies	0,0486	3,64E-03	0,5065
19M06	Troubles mentaux d'origine organique et retards mentaux, âge supérieur à 79 ans	0,0494	3,70E-03	0,5102
02C08	Autres interventions extraoculaires, âge supérieur à 17 ans	0,0502	3,76E-03	0,5140
06M18	Symptômes et autres recours aux soins de la CMD 06	0,0506	3,80E-03	0,5178
09M03	Traumatismes de la peau et des tissus sous-cutanés, âge supérieur à 17 ans	0,0510	3,82E-03	0,5216
01K04	Injections de toxine botulique, en ambulatoire	0,0510	3,82E-03	0,5254
01M25	Epilepsie, âge supérieur à 17 ans	0,0513	3,85E-03	0,5293
03C28	Interventions sur les végétations adénoïdes	0,0525	3,94E-03	0,5332
08C39	Interventions sur l'avant-bras	0,0527	3,95E-03	0,5372
06C04	Interventions majeures sur l'intestin grêle et le côlon	0,0541	4,06E-03	0,5412
13K02	Endoscopies génito-urinaires thérapeutiques et anesthésie : séjours de la CMD 13 et de moins de 2 jours	0,0545	4,08E-03	0,5453
07M02	Affections des voies biliaires	0,0563	4,22E-03	0,5495
14Z16	Faux travail et menaces d'accouchements prématurés	0,0569	4,27E-03	0,5538
04M20	Bronchopneumopathies chroniques surinfectées	0,0575	4,31E-03	0,5581
06M02	Autres gastroentérites et maladies diverses du tube digestif, âge inférieur à 18 ans	0,0589	4,42E-03	0,5625

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
01M20	Commotions cérébrales	0,0596	4,47E-03	0,5670
11C11	Interventions par voie transurétrale ou transcutanée pour lithiases urinaires	0,0603	4,52E-03	0,5715
07C14	Cholécystectomies sans exploration de la voie biliaire principale à l'exception des affections aiguës	0,0620	4,65E-03	0,5762
10M13	Explorations et surveillance pour affections endocriniennes et métaboliques	0,0628	4,71E-03	0,5809
21M10	Effets toxiques des médicaments et substances biologiques, âge supérieur à 17 ans	0,0646	4,84E-03	0,5857
23Z02	Soins Palliatifs, avec ou sans acte	0,0678	5,09E-03	0,5908
12C08	Circoncision	0,0690	5,17E-03	0,5960
20Z04	Ethylisme avec dépendance	0,0702	5,27E-03	0,6012
08C24	Prothèses de genou	0,0707	5,30E-03	0,6065
09C03	Greffes de peau et/ou parages de plaie à l'exception des ulcères cutanés et cellulites	0,0709	5,32E-03	0,6119
09C10	Autres interventions sur la peau, les tissus sous-cutanés ou les seins	0,0719	5,39E-03	0,6172
05K14	Mise en place de certains accès vasculaires pour des affections de la CMD 05, séjours de moins de 2 jours	0,0721	5,40E-03	0,6226
08C45	Ménisectomie sous arthroscopie	0,0727	5,45E-03	0,6281
11M04	Infections des reins et des voies urinaires, âge supérieur à 17 ans	0,0731	5,48E-03	0,6336
20Z05	Ethylisme aigu	0,0733	5,50E-03	0,6391
05M13	Douleurs thoraciques	0,0747	5,60E-03	0,6447
08C48	Prothèses de hanche pour des affections autres que des traumatismes récents	0,0757	5,68E-03	0,6504
11C13	Interventions par voie transurétrale ou transcutanée pour des affections non lithiasiques	0,0761	5,71E-03	0,6561
01M30	Accidents vasculaires intracérébraux non transitoires	0,0783	5,87E-03	0,6619
08C37	Interventions sur le pied, âge supérieur à 17 ans	0,0799	5,99E-03	0,6679
08C14	Résections osseuses localisées et/ou ablation de matériel de fixation interne au niveau d'une localisation autre que la hanche et le fémur	0,0808	6,06E-03	0,6740
05M08	Arythmies et troubles de la conduction cardiaque	0,0819	6,14E-03	0,6801
06M12	Douleurs abdominales	0,0820	6,15E-03	0,6863
05C17	Ligatures de veines et éveinages	0,0829	6,22E-03	0,6925
16M11	Autres troubles de la lignée érythrocytaire, âge supérieur à 17 ans	0,0860	6,45E-03	0,6989
14M03	Affections de l'ante partum sans intervention chirurgicale	0,0874	6,56E-03	0,7055
23M06	Autres facteurs influant sur l'état de santé	0,0888	6,66E-03	0,7122
06C12	Interventions réparatrices pour hernies inguinales et crurales, âge supérieur à 17 ans	0,0889	6,67E-03	0,7188

Racine de GHM	Libellé	Contribution absolue à l'entropie totale	Contribution relative à l'entropie totale	Contribution cumulée
01C15	Libérations du médian au canal carpien	0,0945	7,08E-03	0,7259
08C44	Autres interventions sur la main	0,0957	7,18E-03	0,7331
23K02	Explorations nocturnes et apparentées : séjours de moins de 2 jours	0,0983	7,37E-03	0,7404
05K06	Endoprothèses vasculaires sans infarctus du myocarde	0,0996	7,47E-03	0,7479
04M05	Pneumonies et pleurésies banales, âge supérieur à 17 ans	0,1044	7,83E-03	0,7557
14C08	Césariennes pour grossesse unique	0,1119	8,39E-03	0,7641
06M03	Autres gastroentérites et maladies diverses du tube digestif, âge supérieur à 17 ans	0,1123	8,42E-03	0,7725
17M06	Chimiothérapie pour autre tumeur	0,1160	8,70E-03	0,7812
05K10	Actes diagnostiques par voie vasculaire	0,1282	9,61E-03	0,7909
14Z08	Interruptions volontaires de grossesse : séjours de moins de 3 jours	0,1311	9,83E-03	0,8007
05M09	Insuffisances cardiaques et états de choc circulatoire	0,1458	1,09E-02	0,8116
23M20	Autres symptômes et motifs de recours aux soins de la CMD 23	0,1774	1,33E-02	0,8249
14Z13	Accouchements uniques par voie basse chez une primipare	0,1817	1,36E-02	0,8385
03K02	Affections de la bouche et des dents avec certaines extractions, réparations et prothèses dentaires	0,1970	1,48E-02	0,8533
14Z14	Accouchements uniques par voie basse chez une multipare	0,2438	1,83E-02	0,8716
06K02	Endoscopies digestives thérapeutiques et anesthésie : séjours de moins de 2 jours	0,2819	2,11E-02	0,8927
02C05	Interventions sur le cristallin avec ou sans vitrectomie	0,4382	3,29E-02	0,9256
15M05	Nouveau-nés de 3 300 g et âge gestationnel de 40 SA et assimilés (groupe nouveau-nés 1)	0,4429	3,32E-02	0,9588
06K04	Endoscopie digestive diagnostique et anesthésie, en ambulatoire	0,5497	4,12E-02	1,0000

VI. Publication – Medico-economic decision making: an international comparison

9. Medico-economic decision making: an international comparison

P. Landais, T. Boudemaghe, C. Suehs, G. Dedet, C. Lebihan-Benjamin

After reading this chapter, you should be able to:

- *identify and explain the key principles of health planning and hospital funding*
- *understand the principles of activity-based payment*
- *explain the origin of hospital stay rates*
- *understand the principle of DRGs and related pricings*
- *understand draft estimates of revenue and expenditure*
- *describe the basis for pricing activity and its assessment*
- *describe how hospital expenditures are regulated*
- *describe the similarities and differences in hospital funding among certain nations.*

9.1 Activity-based payment: general principles

9.1.1. Context

Hospital expenditures represent a large part of public health care costs, and for this reason, several related regulatory mechanisms exist. Activity-based payment has gradually become the most common system for hospital reimbursement in high income countries. Initially implemented as an indicator of relative hospital performance, today it mainly represents a hospital payment system. This was variously implemented over the last decade in order to achieve shared goals such as improving overall efficiency, quality, transparency, and to help in hospital management. Activity-based funding is also supposed to help target where and how money is being spent, and thus orient policy and decisions on the behalf of patients. The system also provides for payment adjustments and promotes high quality of care via reward payments. An efficient price can be used to maximize the volume of high-quality care through cost efficiency.

9.1.2. Motives

The motives underlying the introduction and the development of DRG systems vary greatly from country to country. The retained schemes have been adapted to each country according to their individual developmental contexts, or to their conception of a welfare state (Beveridge, Bismarck or mixed models). Variations relate to differences in the health system models used, the relationships between providers and funders, the degree of centralization, the separation between purchasing and provision, the structure of the hospital market, the type of facilities (e.g. profit versus non-profit), or the level of competition between public and private structures.

9.1.3. Implementation

The implementation of activity-based payment has also varied among countries. Hospital payment mechanisms include global budgets, service fees, daily rebates, as well as case payments; each of the latter differentially influences provider behaviour and efficiency through various incentives and disincentives. The models developed incentivize providers to improve coding quality. Activity-based payment was therefore beneficial in some aspects but also generated unexpected side-effects such as upcoding. Upcoding is the practice of maximizing activity-based reimbursement by exaggerating case severities when coding. Another unwanted effect is selecting patient populations in order to have the most profitable casemix. The casemix, which represents the distribution of DRGs for a facility, also illustrates the relative complexity and intensity of services required to treat patients in a hospital due to diagnosis, disease severity, and personal characteristics such as age.

Thus, contrasting hospital policies, such as responding to a public health need versus maximizing profits, can greatly affect their activity and casemix.

9.1.4. Assessment

Evaluation of the hospital activity-based payment program points towards unwanted shifts away from its intended purpose. The reasons for performing such evaluations include but are not limited to:

- Identifying those incentives which generate unwanted effects, such as a risk for decreased quality of care;
- Assessing incentives for productivity, effectiveness of the activity-based reform, and implementing additional funding to overcome the imperfections in the model;
- Ensuring facility sustainability without generating annuities;
- Maintaining the economic model (classification, rates, etc.);
- Ensuring system consistency via control procedures.

9.1.5. The Uniform Hospital Discharge Data Set

Activity-based payment is based on Diagnosis Related Groups (DRGs). Diagnosis Related Groups are assigned to stays for acute inpatient care. Information is collected for hospital stays in the fields of medicine, surgery and obstetrics, including ICU stays. Activity analysis can be carried out via the automated treatment of minimum and standard administrative information as declared in the Uniform Hospital Discharge Data Set (UHDDS; table 1).

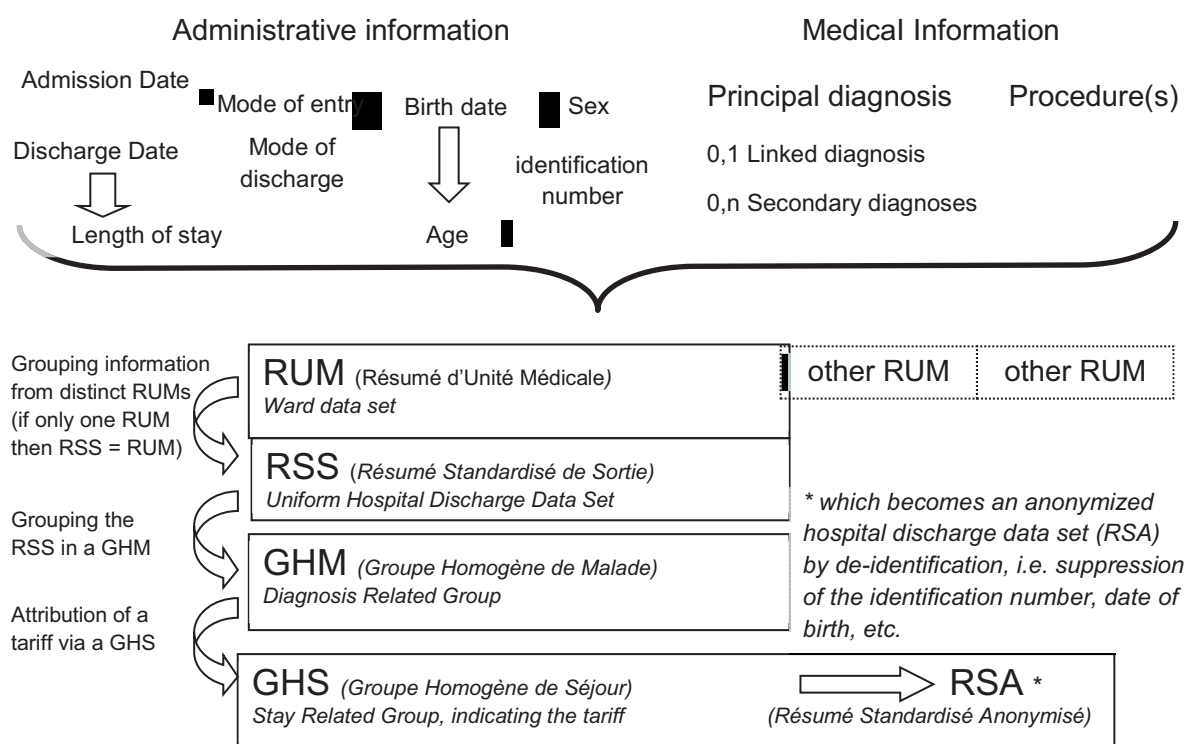


Table 1: The coupling administrative and medical information to generate DRGs': example of the French system.

The medical information contained within an UHDDS varies from country to country. The administrative data generally describe the main characteristics of stays, for example, facility identification, patient age, gender, postal code, stay identifier, the ward involved, date/mode of entry and discharge from the ward, or type of transfer if any. The medical data include the diagnoses and procedures carried out during the stay, and other country-specific data such as a severity index for patients in ICUs, intensive care procedures or continuous monitoring, or radiotherapy-specific information (dosimetry standards, type of device).

9.1.6. Diagnoses and procedures

Diagnoses are prioritized. As specifically related to the French system, the principal diagnosis is the diagnosis which motivated the hospital stay. With some exceptions, all principal diagnoses are dispatched into Major Diagnostic Categories (MDC) that generally correspond to a single organ system. For instance, Diseases and Disorders of the Nervous System, Diseases and Disorders of the Eye, Diseases and Disorders of the Ear, Nose, Mouth and Throat, and so forth. A major diagnostic category represents a category of diagnoses generally based on a single body system or disease aetiology that is associated with a particular medical specialty.

Some diagnoses have specifically been identified as associated with higher resource consumption, and are consequently termed “complications” in the coding lexicon. Depending on the intensity of this effect, several categories of complications have been defined, eventually splitting DRGs into different severity levels, priced accordingly.

Therefore the presence of a “complication” will result in a higher priced DRG (this is the primary mechanism via which upcoding can occur).

Diagnoses are coded according to the international classification of diseases (ICD), whereas procedures are coded according to country-specific classifications. This information is then submitted to a grouping algorithm which associates each standardized discharge summary with the corresponding DRG.

9.1.7. DRGs

Diagnosis related groups (DRGs) have been created in such a way as to result in homogeneity in terms of both (1) medical characteristics and (2) resource consumption. In consequence cases that share the same DRG are medically and economically similar.

The DRGs are assigned to inpatient stays via an algorithm (DRG grouper) using the principal diagnosis and additional diagnoses, the principal procedure and additional procedures, as well as gender and discharge status. One and only one DRG is assigned to each stay, and each DRG has a tariff.

Payment rates are expressed as a base rate (e.g. observed past cost) multiplied by a cost weight (adjustment for contextual evolutions).

9.1.8. Budget allocation

While differing from country to country, budget allocation has a general pattern that mixes one or several of the following elements.

- General principles of cost studies

The way tariffs are determined is based on 2 main principles. Top-down costing (or full absorption model) tries to ensure that the full cost of the stay is allowed for in the calculated cost. This requires simple data and fairly basic technical knowledge. Furthermore, it is cost-effective. Requirements include identifying the total, hospital-wide costs that are expected to be incurred in the year. This is calculated on the basis of resource inputs and utilization, expected levels of outputs, and any agreed surplus/deficits generated in a year. Then costs are classified in a standardized form such that they can be allocated to the service specialties that form the basis of the global budget. The whole process takes the total cost down to the level of the specialty or subspecialty of the budget. Conversely, bottom-up costing is a patient-based micro-costing model. It determines the unit cost of a medical service by summing the cost of all inputs used to provide said service in the most recent year, and then dividing the sum by the annual total number of provided services. This approach is precise but more delicate and demanding to carry out.

- DRG base rate

The hospital base rates are determined as a function of average stay costs per hospital across a group of hospitals. It has an attached monetary value even though cost-weights usually are a relative measure determined according to the relative resource use within a given DRG, where the average cost-weight is taken as '1'. Approximately, the hospital base rate identifies the reimbursement that a hospital would receive for treating the average patient.

- DRG weights

Each DRG may be assigned a weight. Weights are updated annually to reflect changes in medical practice patterns, use of hospital resources, diagnostic and procedural definitions and DRG assignment criteria. To estimate the reimbursement that a hospital will receive for a particular DRG, the hospital's base rate is multiplied by the DRG weight. Software which makes this calculation is called a "pricer". For convenience, providers can bundle the DRG grouper and the pricer within their software. However, one must not confuse grouping DRGs and pricing, the latter being based on tariffs.

- Fixed payment

Even though fixed payments may be accompanied by supplementary money that takes into account particular situations (for instance, a stay in an ICU), hospitals are paid a set fee for treating patients in a given DRG, regardless of the actual cost for that case. This is why accurate and complete coding by professionals is essential for appropriate DRG assignment and subsequent reimbursement.

9.2. International policies

Management of Medicare expenditures is multiple. A transposition from one country to another is not possible. The control of expenditure of health insurance holds to several factors, all of which are not reducible to a pattern of governance or to preformatted rules. Practices, involvement of stakeholders, public financial constraints, the facility considered or the period involved, differ enough to limit the comparability of national health systems. Moreover, detailed rules for the administration of the health budget are very dependent on the modality of budget allocation and periodicity.

Funding the health system is operated by compulsory deductions in several European countries. The prediction of the periodic national health expenditure is region driven in Denmark or Germany, or performed according to a particular type of activity or a specific medical act such as in Belgium. The procedure is most often annual or multi-year as in Denmark, or on a longer period, for instance triennial in the UK. There is most of the time a pretty good match between the authorized or voted health expenditures, and the money finally spent in Germany, Denmark, Belgium and the UK. The UK which has the most restrictive mechanism of regulation, also presented the largest increase of its expenditures.

Germany, Denmark and Belgium have frequent projections of future health expenditures. Thus, the Danish regions publish quarterly reports. German follow-up is based on frequent projections, but all of them are not made public. A group of experts in health expenditure predictions meets quarterly and publishes aggregated projections twice a year. The detailed projections are provided only to members of the group.

Compared to the framework of regulation defined over one year, the levers of interventions to correct possible overruns during the year are seldom clarified. However, local manager put a persistent pressure on professionals either health insurance funds in Germany, Primary Care Trustees in the United Kingdom, or regions and municipalities in Denmark.

In Belgium for instance, the management of health insurance expenditures is performed through the annual setting of an overall health insurance expenditure target. The target of the overall budget is determined by the Belgian "National Institute for Health Insurance and Invalidity" quarterly by sector, with details about classification codes subject to refund by the health insurance. Each of these steps can be a source of possible savings measures. Quarterly, in case of significant overrunning, a warning mechanism is provided for information purposes only. Finally, every six months audit reports for each sector are provided. The control of expenditures passed from an accounting approach to an analytical approach. These reports determine pre-emptively the potential causes of exceeding a target budget. This intra-year follow-up is only valid if decisions necessary to correct trajectories are controlled and observed. So far, this mechanism has never resulted in recommendations for new measures of money saving outside the annual process.

European countries have adopted rather different regulation and monitoring of health expenditure. Thus any attempt to evaluating the impact of activity-based payment in a given country or between countries was seldom performed or focused on specific aspects. Thus for instance, O'Reilly et al

showed that activity-based funding has been associated with an increased activity, a reduced length of stay and/or a diminished rate of growth in hospital expenditure.

Assessing and comparing activity-based funding is seldom available in the literature as well as the impact of DRG payment systems on cost-containment, efficiency and quality of care. To better understand the complexity of the framework it is however possible to describe the context of development, the diagnosis related groups they are based upon, the basis of setting up costs and tariffs, the fields of application and the regulatory mechanisms. As an example, five systems will be described, in the United States, Australia, the United Kingdom, France and Germany.

The management of healthcare expenditures is multifactorial, and direct comparisons between countries are not possible. Controlling health insurance expenditures depends on several factors that cannot be reduced to any one pattern of governance or preformatted rules. Differences in practices, stakeholder involvement, public financial constraints, the facility considered or the period involved limit the comparability of national health systems. Moreover, detailed rules concerning health budget administration depend on how and when budgets are allocated.

In several European countries, health system funding is generated via compulsory deductions. The determination of periodic national health expenditures is regionally driven in Denmark or Germany, or performed according to a particular type of activity or a specific medical act, such as in Belgium. The procedure is most often annual, but can be multi-year as in Denmark, or triennial as in the UK. Most of the time, there is a pretty good match between the authorized or voted health expenditures, and the money finally spent in Germany, Denmark, Belgium and the UK. The UK, which has the most restrictive mechanism of regulation, also had the largest increases in expenditures.

Germany, Denmark and Belgium frequently attempt to predict or project their future health expenditures. Thus, the Danish regions publish quarterly reports. The German system is based on frequent projections, but not all are made public. A group of experts in health expenditure predictions meets quarterly and publishes aggregated projections twice a year. The detailed projections are provided only to members of the group.

Regulatory frameworks are usually well defined for a given year. However, the same cannot be said for interventions aimed at correcting incorrect predictions, which are seldom clarified. However, local managers (e.g. health insurance funds in Germany, Primary Care Trustees in the UK, or regions and municipalities in Denmark) can persistently pressure the system. In Belgium, the management of health insurance expenditures is performed by setting the overall health insurance expenditure target for a given year. The target of the overall budget, determined by the Belgian "National Institute for Health Insurance and Invalidity", is divided into forty partial budget targets corresponding to separate sub-sectors of health care. Fixing the overall budget target implies fixing partial budget targets. A description of the total expenditure by sector is made monthly. A quarterly expenditure statement is made by sector, with details on classification codes subject to refund by the health insurance. Each of these steps can be a source of possible savings measures. Quarterly, when predicted targets are overrun, a warning mechanism is provided for information purposes only. Finally, every six months audit reports for each sector are provided.

Expenditure control is now considered from an analytical point of view, as compared to an approach based solely on accounting. These reports pre-emptively determine the potential causes of exceeding a target budget. This intra-year follow-up is only valid if decisions necessary to correct trajectories are controlled and observed. So far, this mechanism has never resulted in recommendations for new measures of money saving outside the annual process.

Thus, European countries have adopted rather different regulations and ways of monitoring health expenditures. Therefore, any attempt at evaluating the impact of activity-based payment in a given country or between countries was seldom performed or focused on specific aspects. For example, O'Reilly et al showed that activity-based funding has been associated with increased activity, a reduced length of stay and/or a diminished rate of growth in hospital expenditure.

Assessing and comparing activity-based funding among countries is seldom available in the literature. The same holds true for the impact of DRG payment systems on cost-containment, efficiency and quality of care. However, in order to better understand the complexity involved, it is possible to describe the developmental context, the diagnosis related groups they are based upon, the basis of setting up costs and tariffs, the fields of application and the related regulatory mechanisms. As an example, five systems will be described: the United States, Australia, the United Kingdom, France and Germany.

9.3. United States

9.3.1. Context

Activity-based payment was introduced in the United States in 1983 for the payment of elderly patients stays covered by Medicare, a health insurance programme set up in 1966. At that time, a large number of elderly patients could not afford their hospital care. The hospital funding mechanism was based on the principle of “1 US\$ refunded for 1 US\$ spent”. This system encouraged hospital expenditures since increased expenditures results in increased hospital income. In this context, Congress set up the “Prospective Payment System” (PSS) in 1983. The aim was to control hospital expenditures, which had almost doubled between 1970 and 1982. Presently, three disbursing organisms are involved in the funding of American hospitals, each one contributing for one third of the whole, namely (1) Medicare, (2) private insurance companies that negotiate their tariffs directly with hospitals, and finally (3) Medicaid, focussing on uninsured patients and those with poor resources. Hospitals can negotiate with the various funders and can be paid at different price levels for the same service.

9.3.2. Diagnostic Related Groups

The classification system presently used is called the “Medicare Severity - Diagnostic Related Groups” (MS-DRG). It takes into account the principal diagnosis, the secondary diagnoses and an index of gravity. The total number of DRGs reported was 470 in 1983 and 557 in 2006. In 2007 *Medicare* set up a new system in order to better target payments. DRGs were adjusted according to severity, as related to the presence of major or common complications. The total number of the MS-DRGs rose to 999. Patients who are not covered by Medicare are classified according to another system, the « all-patients-refined DRGs ». It includes 1100 categories that integrate disease severity, as well as the risk of mortality assessed on four levels (minor, moderate, major and extreme).

9.3.3. Costs and tariffs

Costs are based on record keeping performed by 3500 health facilities founded by Medicare, representing 60% of all American health facilities. It defines the average cost of health care delivery.

Costs are distinct from tariffs. A cost calculation is estimated on the average resources consumed by all Medicare patients belonging to the same DRG, and then it is standardized to eliminate the differences that might remain between structures and regions. This becomes the national baseline tariff, based on a single model used by both public and private health facilities. Several factors are taken into consideration:

- Exogenous factors, such as health worker salaries, hospital supplies, or a geographical adjustment factor;
- Indirect costs such as medical teaching. In this case an index is estimated which reflects the ratio between the number of medical residents and the number of beds;
- Additional costs in order to support specific populations, such as low-income or vulnerable groups;
- The type of health facility (rural hospital, reference hospital).

Resources are allocated on the basis of these references. However, a part of these resources is also negotiated directly by hospitals with the “Health Maintenance Organizations”. HMOs are health assistance organizations that manage health care by grouping the health care coverage provided by hospitals, doctors or by other suppliers under specific contracts. Finally, Medicare allows tariff

increases for severe and expensive cases whose stay costs are more than three times the standard deviation of the mean.

9.3.4. Fields of application

Activity-based payment is used for all medical, surgical and obstetric (MSO) in-patient facilities, and has been applied to full time hospitalization since 1985, and to ambulatory care since 2000. It does not apply to practitioner fees, to expensive drugs, rehabilitation, oncology, psychiatry, or to the direct costs related to education and research. It does not apply to small institutions, which are still funded according to their actual costs. The system provides additional funding not related to activity. Expensive drugs and prostheses are provided for by an additional lump sum payment. Considering health institutions catering for a high proportion of indigent patients, a bonus index is applied to their tariffs, as well as for university hospitals concerning their missions of education for medical students. For research, scholarships are offered.

9.3.5. Regulatory mechanisms

Health institutions are not submitted to a price/volume regulation of expenditures: unlike other countries, a given facility collects different tariffs for a given stay according to the insurance company implicated, which curtails regulatory measures at the local or national level. Facilities are responsible for their deficits; on the other hand, profits are retained.

In order to prevent facilities from orienting their treatment choices or selecting patient groups based on financial considerations rather than medical needs, DRG payments have been adjusted. However, DRG classification only explained 30% of the variation among stay costs. In 1995, the Commission of evaluation of activity-based payment, known as "ProPAC", recommended that specific secondary diagnoses that frequently affect health care costs, as well as differences in medical practices between establishments, be taken into consideration.

In 2005, MedPAC highlighted important variations in profitability between-DRGs and intra-DRGs. Despite similar clinical situations, DRGs related to surgery were more advantageous than those related to medical treatments. This cost variation was systematically related to the patient's secondary diagnosis. A recommendation was implemented in order to improve the adjustment of recruitment from one hospital to another, and to adjust prices in order to better reflect differences in severity between cases.

The DRG system generated a reduction in the length of stay in MSO facilities, an increase of ambulatory care and a shift to follow-up and rehabilitation care. These changes promoted the development of structures such as centres for ambulatory surgery. Case-based or day-based prospective payment increased efficiency payment systems in the fields of ambulatory, follow-up and rehabilitation care.

This funding mechanism made it possible to fund therapeutic innovations such as coronary angioplasty. Quality improvement programmes were also encouraged via the creation of quality of care indicators and penalties in case of non-quality, such as the occurrence of hospital acquired infections.

This tariffication system does not have a durable impact on cost growth or variation. Because DRGs tariffs are negotiated, a given DRG tariff can considerably vary from one health facility to another. Private institutions specialized in specific types of programmed and ambulatory surgeries offer rates well below those of conventional facilities. Thus, they become even more competitive.

There is also heterogeneity in costs of various stays despite the fact that DRG payment is adjusted for daily costs. Successive modifications of DRGs did not lower this heterogeneity, which explains only 30% of cost variability.

DRG refinements make classification more complex, requiring a very detailed information base. For facilities, the risk of not being fully paid increases if their data are not appropriately collected. A perverse effect can occur, resulting in better payment of large teaching institutions and poorer payment of smaller rural facilities that are already in trouble.

It should finally be emphasized that in a prospective payment system, the power of funders decreases when the number of payment categories increases. Indeed, bargaining power is lost when these categories are numerous and very specific.

9.4. Australia

9.4.1. Context

Initiated in 1984, the Australian Medicare plan is funded by taxes, under the aegis of the Australian federal government and by the six member states. However public hospitals are controlled solely by regional governments. Before the introduction of activity-based payment, hospitals received their global budgets according to their activity together with the number of patients in their waiting lists. Hospitals used these queues to negotiate an increase of their financings. This mechanism was criticized because it encouraged health facilities keeping patients in their waiting list, particularly for surgery. This led to an overhaul of the funding system. The main goal of the introduction of activity-based payment was to reduce these waiting lists. The State of Victoria, representing one quarter of the Australian population, initiated this process in 1993.

9.4.2. AN-DRG classification

Classification of hospital stays incorporates a hierarchy of benefits and medical situations. Stays are distributed in one of 23 classes and 667 AN-DRGs created in the 1990s. AN-DRGs classification has been adopted by most Australian states. The State of Victoria has gradually changed the AN-DRGs to finally create its own groups, the VIC-DRGs, for a total of 760 groups. They adjusted the payments of stays of less than 24 hours duration on the basis of the technical difficulty of treatment. An annual review is performed by a panel of clinicians.

9.4.3. Costs and tariffs

Costs of stays were calculated from patient data by a bottom-up method. This type of model aggregates all available information at the patient level by medical specialties in dedicated cost centres. This type of modelling is supposed to improve the accuracy of costs measurement for each patient compared to other methods the costs of which are allocated to patients solely on the basis of their length of stay.

Uniform tariffs for each AN-DRG are applied to hospitals that are similar regarding their patients' case-mix. They are used to negotiate price-volume contracts. Prices are also adjusted on several criteria, including the socio-demographic conditions of the population represented by the health facility. For each hospital a fixed tariff is generally applied to each AN-DRG for a given target level of activity determined by its historical activity.

To encourage health facilities to exceed their target level of activity, additional payments mechanisms were introduced. Regarding the management of patients requiring urgent care or waiting for a long time, a classification according to the rate of emergency was introduced: urgent, semi-urgent or non-urgent. Moreover, a ceiling on payments for additional activity was established. A pre-defined budget is shared between health facilities. It is correlated with each establishment's own contribution to the total additional activity. Finally, access to additional funding has been conditioned on the achievement of specific care goals related to the waiting list of patients. This pricing policy has proven effective as shown by the drastic reduction in the number of "urgent" patients waiting for more than one month.

9.4.4. Field of application

The Australian Health System is a mixture of both private and public sector health services providers, and a range of funding and regulatory systems. They involve:

- The Australian government. Its primary role is to develop broad national policies, regulation and funding. Funding includes three major national subsidy schemes; Medicare, the Pharmaceutical Benefits Scheme, as well as the thirty-percent Private Health Insurance Rebate.
- State and Territory, and Local governments. They are primarily responsible for the delivery and management of public health services. They are also in charge of maintaining direct relationships with most health care providers, including regulation of health professionals and private hospitals.

Under Medicare, the Australian and State Governments jointly fund public hospital services, providing free services to Australians who choose to receive treatment as public patients.

- Private practitioners including general practitioners, specialists and consultant physicians.
- Profit and non-profit organizations and voluntary agencies.

The Pharmaceutical Benefits Scheme and Medicare cover every Australian, subsidizing their payments for private medical services, as well as for a large portion of their prescribed medications. While these options are available to Australians, government funding of the thirty-percent rebate as well as additional key incentives support citizens choice to pursue and retain private health insurance.

Individuals make their contribution to the health care system through taxes and the Medicare levy based on their income, and through private financing such as private health insurance.

9.4.5. Regulatory mechanisms

In 2009, the final report of the National Health and Hospitals Commission made 123 recommendations, which included the following: hospital funding tied to performance; introduction of activity based funding for all hospital services, with the Commonwealth paying a fixed percentage of the efficient price; promoting a personal electronic health record and national eHealth system; public reporting of public and private hospital performance; a range of measures related to choice in care for the elderly; a permanent, independent and expanded Australian Commission on Safety and Quality in Health Care; National Access Targets for timeliness of care.

To ensure that the nationally efficient price is determined on a fair and equitable basis, an independent expert will set the nationally efficient price and advise the Government on appropriate timelines and transition processes for all hospital services. In setting the nationally efficient price, the expert will be required to strike an appropriate balance between reasonable access, clinical safety, efficiency and fiscal considerations.

A main objective is eliminating the unnecessary use of resources in the production and delivery of services. In hospital systems, operational efficiency can be achieved by reducing length of stay, increasing quality, and looking closely at the reasons for significant variations in clinical practice.

The increased operational efficiency from activity based funding will provide savings for taxpayers or help fund additional services. Although the precise efficiencies from activity based funding are difficult to estimate, the NHHRC estimates that the introduction of activity based funding will lead to savings of between \$0.5 billion and \$1.3 billion each year.

9.4.6. Evaluation

The use of activity-based funding is supposed to drive increased operational efficiency across the hospitals system as it explicitly links the funds allocated to the services provided. It also allows for easy identification of underperforming providers so that the cause of underperformance can be remedied, while lessons from high performance can be disseminated.

In 2012, a new report, the “national Health and hospitals network for Australia’s future” proposed several adjustments to the model. The Commonwealth Government will become the majority funder of the Australian public hospitals system. The Government will fund 60% of:

- the efficient price of every public hospital service provided to public patients;
- recurrent expenditure on research and training functions undertaken in public hospitals;
- capital expenditure, both operating capital and planned new capital investment, to maintain and improve public hospital infrastructure;
- and over time, up to 100 % of the efficient price of ‘primary health care equivalent’ outpatient services provided to public hospital patients.

In return for providing a secure funding base for public hospitals in the future, the Commonwealth will require the states to commit to system wide reform to improve public hospital governance, performance and accountability.

In establishing Local Hospital Networks, states will be asked to create hospitals groups that can ensure geographic linkages, management quality, economies of scale, an appropriate service mix, and referral pathways within the network.

9.5. United Kingdom

9.5.1. Context

In Great Britain, the National Health System (NHS) is managed by the Ministry of Health. It is a decentralized system. The NHS is organized around more than 304 local instances, called Primary Care Trusts (PCT), which provide primary and secondary care to citizens. PCTs are responsible for purchasing medical care for the whole population. Health care accessibility is "free" for everyone. The major specificity of the British system is that access to hospital care or specialists is controlled by general practitioners called "gatekeepers".

Before 2004, British hospital budgets were global contracts, called "blocks contracts". The predictable volume of care was used to determine the total amount a health facility might expect. At that time, British health expenditures were the lowest in Europe, but the reported quality of care was poor, including long waiting times for scheduled surgery.

To improve the health system, "Payment by Results" (PbR), similar to activity-based payment, has been introduced progressively since 2004. The main objective of this reform was to increase efficiency and productivity of care, while ensuring quality.

PbR was introduced at a time of relatively low levels of funding and high waiting lists. The explicit objective was to increase acute activity. In particular, surgical activity was required to reduce waiting lists and therefore the system was incentivized essentially by "piecework".

The English system is a mixed system. Health facility budgets are prospective, with 30% of the budget financed by an overall envelope and 70% funded by activity-based payment. UK hospitals are funded by a single-payer: PCTs. From taxes, the Ministry of Health allocates an annual budget to the NHS. Its amount is based on changes in care production's costs, previously allocated amounts and health indicators such as mortality rates.

The NHS allocates 80% of its budget to PCTs who manage primary and secondary care. PCTs allocate secondary care to public, private or nonprofit providers. These caregivers may be hospitals or groups of hospitals under joint management, "NHS Trusts", "Foundation Trusts" meeting performance criteria, independent treatment centres, private hospitals or ambulatory specialists.

9.5.2. HRG classification

Activity based payment is based on Healthcare Resource Groups (HRG), a classification which is close to the American model. This classification takes into consideration diagnosis and treatment complexity. It included 550 HRGs in its first version, and 1400 in the 2009 version.

9.5.3. Costs and tariffs

The accounting policy is similar to the French system. However, the national cost scale is based on all the hospitals involved in activity based payment. The price of each HRG is national, calculated from a baseline cost, using a top-down costing method. A reference cost results from a mean cost for a given HRG across all hospitals plus 350 Hospital Trustees, created in 1991 to introduce market rules. These Trustees are legal entities that often manage several health institutions.

Tariffs were designed to increase the level of activity beyond a pre-negotiated level; this is called the "Service level agreement". They reflect the average costs of care procedures, adjusted for several characteristics such as: scheduled stay or not, type of surgery, health procedure, ambulatory or not, income variations and consumer price index (using an index called the "Market Forces factor"), a geographic coefficient, costs generated by implementing the recommendations of the National Institute for Health and Clinical Excellence (NICE) on treatments and drugs considered as cost-effective. The payment unit is based on HRGs.

9.5.4. Field of application

Activity based payment applies to hospital departments, including day care hospitalization in medicine, surgery and obstetrics. It also applies to emergency departments. It is not used for practitioner's fees, expensive drugs, chemotherapy, radiotherapy, dialysis units, cardiology

intensive care, burn units or transplant units. Medical research and teaching is funded by the Ministry of Health and the Higher Education Funding Councils.

9.5.5. Regulatory mechanisms

The English system is not very flexible. Deficits are at the expense of the institutions. Any increase or decrease in activity will be charged according to the national scale, without any local negotiation. In addition, in order to regulate access to emergency departments, tariffs for emergency care services are reduced by 50%, as soon as the volume of activity exceeds that of the previous year by more than 3%. Conversely, there is no regulation of the volume of surgical activity.

Since the mid-1980s, the budgetary mechanism for health expenditure in the United Kingdom is multiannual and limiting. The health budget is voted by the Parliament as part of the State budget. Health expenditures included in the State budget are thus subject to the general evolution of the State revenue. These revenues do not depend, as in France, on the trends in total wages. There is therefore, by design, no possible deficit concerning social security insurance. The programming of expenditure is triennial. The preparation of multi-annual programming, or a "spending review", is held every two years. The third year is for the revision of the objectives. The amounts of expenditure are adopted in accordance with the objectives and priorities of the government, and available through multi-annual envelopes. A deficit or a surplus compared to the forecast of expenditures, gives the following year an amputation or an increase in the amount of the excess or surplus. As noted above, the consequences of this very strict and effective discipline were attenuated by the authorized increase of health spending - identified as health priorities - for a decade, despite a slowdown over the last two years.

9.5.6. Evaluation

Though the original goal of payment by results was to incentivize activity it has, with some modifications been subsequently used for additional policy objectives: enabling choice, transferring funding so that money follows the patient, enabling benchmarking, creating efficiency or incentivizing quality.

In hospitals where activity based payment has been introduced, the cost of care at a unit level fell faster than in other facilities, and length of stay decreased significantly, falling from 8.20 days in 2001-2002 to 5.98 days in 2005-2006. Nevertheless, the reduction in the length of stay cannot be formally attributed to the reform, since in parallel significant government pressures was exerted on hospitals to reduce waiting times. Result-based payment was associated with an increase in volume of care, but had no impact on the quality of care. The latter was measured on the basis of hospital's global mortality rate, mortality rate 30 days after surgery, and emergency readmission rates after hip fracture. The introduction of activity based payment caused heavy financial losses for certain hospitals.

9.6. France

9.6.1. Context

The French health system is decentralized and based on Bismarck's model. It has many particularities, such as: a free path for patients, an important for-profit private sector, and the coexistence of several hospitalization sectors. Activity-based payment was introduced in France in 2004. It aims at encouraging better response to health care supply and improving efficiency. This system is based on a competitive approach between health facilities and encourages them to optimize a cost-effective health care supply, as well as the quality of care and patient satisfaction. It is a powerful incentive towards productivity. The challenge lies in the ability of the system to combine cost containment with maintaining the quality of care and innovation.

In keeping with the peculiarities of the French health system, a mixed model has been set up. On the one hand it covers the financing of activities related to diagnosis, treatment and care, and on the other hand it compensates expenses related to missions of general interest.

In order to equitably allocate resources to institutions, taking into account their specific activities and respecting the notion of equal access to care, activity-based payment defined according to four additional compartments was implemented, in agreement with health professionals:

- Stay-based prospective payment is the basis of this system. It consists of prospective pricing with a refund based on real and all-in costs. Stays are grouped into diagnosis related groups (“Groupes Homogènes de Malades”, French GHM) according to medical criteria and applied similarly to all institutions of a given healthcare sector;

- The other three components are funded differently; they take into consideration the missions, activities and products that cannot be included either temporarily or permanently in the costing of the diagnosis related groups (DRGs’). These three components are:

-Annual all-in packages (emergency care units, organ procurement and transplantation activities);

-Extra funding for expensive drugs and medical devices included on a specific list and fully reimbursed;

- Finally, Missions of General Interest and Contracting Assistance (“Missions d’Intérêt Général et d’Aide à la Contractualisation”, French MIGAC) are funded, providing opportunity for financing specific services such as palliative care, consultation for drug users, halting tobacco use, care for prisoners,.. Some health care services are subject to specific invoices (intensive care supplement, dialysis package, or package for small equipment).

Since 2004, this reform is applied to both public and private sectors, but following different schemes, in particular with regard to costs and prices.

9.6.2. Classification system

The French classification system is based on GHM (“Groupes Homogènes de Malades”) which corresponds to the American’s Diagnosis Related Groups (DRGs). A GHM represent inpatient classifications on the basis of several items comprising diagnosis, procedure, age, gender, or discharge disposition. These groups were constructed to evaluate length-of-stay, which is related to illness severity and resource consumption.

Thus, DRGs provide a tool to analyze hospital stays. DRGs were designed to allow hospitals to operate on a more systematic basis, associated with resource allocation and cost analysis. DRGs are a tool used for the prediction of resource consumption for any given hospital stay, allowing comparisons between facilities.

This classification, which is the basis of the French payment system, was published for the first time in 1986. It is regularly updated. Its most important feature, compared to other classifications, is that it integrates day care hospital’s activities.

The original version included 784 groups, the current one (version 11) consists of 2296 GHMs. This last version is composed of 611 roots, each divided into up to 4 levels of severity to better take into account the complexity of care given patients.

9.6.3. Costs and tariffs

Based on GHMs, Stay Related Groups (“Groupes Homogènes de Séjours” or GHS) have been created. Almost all GHM correspond to only one GHS (only 3 exceptions). GHSs are tariffs associated with GHMs. Thus GHMs are a basis for measuring resource consumption in the form of a normalized weight. The construction of a tariff is critical and complex. It has to meet several objectives:

- A tariffs’ hierarchy conform to production costs in order to avoid under or over paid health care deliveries.

- An observance of the ONDAM (French national target for health insurance expenditures). The constructed hierarchy is thus subject to several constraints. The volume of all GHS can’t generate higher expenditure than the amount provided by the ONDAM. Therefore, the ONDAM determines the tariff level, which is the first rate scale. This raw rate scale allows funds distribution (according to what is planned by the ONDAM) to health facilities while respecting the hierarchy of the classification;

- A stimulation of the tariff policy in order to achieve public health objectives or encourage the evolution of medical practice. These incentives are called ‘marked tariffs’;

- Taking into account the adaptive capacity of health facilities and the need for stability while limiting the effects of tariffication scales on their revenues.

Tariffication is centralized and generated at a national level using a top-down approach. Tariff weighting is performed on the basis of average costs observed in a 122 voluntary sample of facilities. A statistical adjustment is realized in order to take into consideration the selection bias related to the voluntary participation of each health facility. Two scales are defined:

- One for the public sector, institutions previously under global budget allocation;
- One for the private sector (where medical fees are not included in prices).

Tariffs are set up at a national level, determined on the basis of standards in order to meet targeted spending levels. Pricing is adjusted over regions, using a geographical factor which takes into consideration only geographical characteristics; therefore characteristics such as size of the facility or status are not included in this adjustment. In 2009, activity based payment accounted for 90% of medicine, surgery and obstetrics expenditures and concerned 50% of all French health care facilities.

9.6.4. Field of application

Activity based payment is used for Medicine Surgery Obstetrics (MSO), and full time hospitalization. However, expensive drugs and medical devices are not included in GHS' tariffs, nor medical fees for private sector GHSs.

Another important element is that activity based payment is not used in Psychiatric hospitals follow-up and rehabilitation care.

Other remarks:

- Expensive drugs and prostheses have funding that can be added to those reported in the list of drugs and devices charged in addition to each GHS.
- MIG funding is provided by an allocation, as a compensation of public service and the fulfillment of general interest missions.
- A budget called 'Mission d'Enseignement de Recherche, de Référence et d'Innovation (MERRI) provides funding for research and teaching. This additional funding is adjusted according to the declared activity in this field.

MIG budgets can better fund certain additional costs incurred by health facilities. Activities grouped in this field as well as their corresponding resources, are subject to negotiation between the Regional Health Agency (ARS) and the facility, embodied by a multi-year agreement of objectives and funding ("Contrat Pluriannuel d'Objectifs et de Moyens").

9.6.5. Regulatory mechanisms

In France, the HPST law (2009-879) relative to hospital reform was dedicated to better articulating patients, health and territories. The creation of regional health agencies (ARS) aims at implementing, at the regional level, a coordinated set of programs and actions that put into practice the objectives of the national policy of health, the social and medico-social actions and the fundamental principles of the social security code. Thus, regional health agencies expand their regional expertise across the health field and create regional health projects. They were assigned the task of implementing health policy at the regional level, and in consultation with health professionals, regulate, direct and organize, the provision of health services. The creation, conversion, and the grouping of care activities, as well as heavy equipment are subject to authorization. ARSs establish a multi-year contract with each institution, which defines objectives and means responding to the guidelines of the regional health project, of the regional organization of care (SROS) scheme, or interregional schema.

The law anticipated that if in a given year the budget is exceeded, adjustments can be made. Therefore, when MSO activity expenses are exceeded, a regulatory mechanism lowers the levels of the national tariff for all facilities. This operation is carried out annually by adjusting the amount of rates of hospital stays to the projected business of the following year. Surplus gains and deficits are retained by the institutions.

In 1996, the public and private hospital system was reformed and established regulation through costs based on two new items: annual funding of social security laws and the National Insurance (ONDAM) spending target.

-the annual funding of social security laws: these funding laws determine the general conditions of financial equilibrium and, given the forecast of revenue (contributions), set its spending targets, under the conditions and subject to the reservations provided by an organic law. The Act of funding for a given year includes four parts relating to (1) the period ended, (2) the current fiscal year, (3) revenues, and (4) the projected equilibrium and spending for the coming year. Three types of data are presented by the law of funding: revenues, objectives of projected expenditure and cash advance limits.

-The National Insurance spending target (ONDAM). This is a budgetary control passed under the control of the Parliament which votes the Act of financing social security, which sets the ONDAM.

9.6.6. Evaluation

Regarding efficacy and productivity, the effect of activity based payment remains unclear. In 2005, public and private institutions increased their activity in different ways. Public hospitals have increased the number of inpatients by 1.5% and the number of outpatients by 5%. During the same period, private clinics have reduced the number of inpatients by 3% and increased their outpatient number by 9.5%. In addition, a study undertaken in 800 hospitals reported that once facilities had understood and integrated the logic of activity based payment, most of the time they decided to change their activity rather than seeking to improve their efficiency. Thus, private clinics tended to over-specialize their activity. For instance, some of them report that 80% of their global activity is invested into ambulatory surgery. Changes in the management of medical human resources were also introduced.

To date, it is not possible to measure the impact of activity based payment on quality of care. For this reason, indicators have been developed by the health authorities in order to measure the impact this reform has had on the quality of care in French hospitals.

Unlike the private sector, the public sector expenses have progressed faster than incomes between 2003 and 2007. Implementation of activity-based payment has been harmful to public hospitals, and their profitability has declined. Public facilities face difficulties in reducing their costs despite increases in activity. Conversely, the private sector, which has seen its overall turnover and profitability rising by 8% in 2005 and 4% in 2006, seems to have benefited from this reform.

9.7. Germany

9.7.1. Context

The German health system is a federal system. Federal states, called “Länder”, control public hospitals together with counties and municipalities. Hospital charging system borders were reformed in 2000. At this time, pricing was based on a day rate. Hospital expenditures accounted for 35% of health spending. The main objective of the tariff reform was to optimize acute care by increasing efficiency, quality and transparency, while reducing the growth of hospital expenditures. The process was achieved step by step, based on three main principles: solidarity, benefits in kind and self-administration.

The classification system was created in 2000-2002 with the implementation of costing, which became stable between 2003 and 2004. Activity based payment became mandatory in 2004 for the MSO sector. The charging system converged between 2005 and 2010, and was provided by two funders. On the one hand, Länders together with municipalities, manage capital expenditures and infrastructures, and develop hospital planning. On the other hand, public health insurance funds manage current spending on the basis of a budget prepared according to the nature and volume of care in a given hospital. Finally, part of the funding comes from insured people who have to pay a fixed daily hospital charge.

Health facility budgets are prospective, based on annual estimates of activity and expenditures. Such budgets are applied to public and private institutions according to similar modalities. The budget is mixed, with 20% determined by a fixed envelop, and 80% determined by activity based payment.

9.7.2. Classification system: G-DRGs

Activity based payment is based on a classification system called “G-DRG” for German Diagnosis Related Groups. It is modelled on the Australian AR-DRG. Principal and secondary diagnoses are coded using the International Classification of Diseases adapted to Germany (ICD-10-GM). The number of DRGs increased from 664 in 2002 to 1200 in 2010.

9.7.3. Costs and prices

The cost for each DRG is self-administered, established annually by the Federation of Health Facilities, the Deutsche Krankenhausgesellschaft and by the Institut für das Entgelt system Krankenhaus (InEK), which groups the federal union of health insurance funds and the federation of private health insurers. This Institute is the property of the hospitals and of the health insurance funds. Objectives include classification improvements, estimating the relative weights of the DRGs and the creation of additional payments. The national costs scale is established by a sample of 250 voluntary hospitals (12% of all facilities). The average costs for each DRG are published annually. Tariffs are negotiated between each hospital and the health insurance service for a one year period. Some geographic disparities are taken into consideration for certain tariffs, and may result in the creation of special funds. Regional tariffs are supposed to converge by 2014.

9.7.4. Fields of application

Activity based payment applies to acute care medical departments, including emergency units, surgery and obstetrics. It is supposed to extend to psychiatry departments by 2013. It does not apply to private practitioner’s fees or to expensive drugs or devices. For missions of general interest such as teaching, research or doctors’ training, DRG rates are increased and special funds are allocated. Innovative treatments prices are increased. Tariffs are supposed to converge by 2014.

9.7.5. Regulatory mechanisms

The federal government sets the regulatory framework for the health care system, including the rules of provision for social services and their financing. In case of a budget deficit, control is provided by the *Länders*, who are the key regulators of the hospital sector. When to hospital budget is exceeded, a *Land* can formulate a plan for restructuring activities over a group of hospitals. It can suspend certain activities, close facilities, or privatize a hospital. It can also offer financial assistance. Moreover, health insurance funds may minimize tariffs, increase tariffs of stays, and/or reduce the repayment of the funding the following year according to a principle known as “partial repayment”.

9.7.6. Evaluation

The G-DRG system has made hospitals more organized and operational, through greater interdisciplinary cooperation. Significant restructuring has reduced hospital numbers by 4% and the number of beds by 5%. The length of stay has decreased by 5%. Between 2004 and 2007, the bed occupancy rate increased by 2%. Some sectors, such as obstetrics and gynaecology, internal medicine or surgery, have decreased, and other specialties such as neurosurgery, plastic surgery or paediatrics surgery increased. One hundred public hospitals have been fully or partially privatized. Between 2000 and 2007, the proportion of public institutions has declined from 38% to 32%, and private non-profit from 41% to 38%. Conversely, the proportion of the fully private sector rose from 22% to 30%.

In 2007, 53.5% of public hospitals were mixed, public-private institutions. Nevertheless, activity did not increase as expected. Indeed, the number of hospitalized patients, adjusted on age, rose only by 0.4% in 2006. A transfer of activity has been observed, mainly towards ambulatory surgery.

In the majority of medical departments, quality remained stable or improved, which is not the case in some disciplines such as cardiac surgery. However, it is difficult to attribute this improvement to activity based payment, since simultaneous measures to improve quality were developed, such as accreditation procedures. Moreover, the working conditions of medical and paramedical workers

deteriorated, due to significantly increased workload. Usually, economic efficacy is better, despite additional costs such as expenses related to creation and maintenance of hardware. Between 2003 and 2006, the hospitals' total adjusted costs increased by 1.4% annually. This increase was less pronounced than during the 1991-2006 period (+3.7% every year). During the 2005 to 2008 period, health insurance spending increased by 2.6% every year. In 2007, 30% of hospitals were in deficit, 20% balanced, and 50% beneficiary. Large facilities and university hospitals were favoured to the detriment of small hospitals. The Act of March 17, 2009 thus proposed a 3-year aid program.

The G-DRG system is notable for its increased transparency and traceability of hospitals stays since it requires collection and publication of standardized and detailed data. The implementation of activity based payment has made the German system more effective and more organized, while maintaining quality of care. The strength of this system is due to the analysis and pricing of the G-DRG by the InEK. Nevertheless, it is difficult to attribute changes to just activity based payment implementation due to interactions with other economic parameters, such as rising wages, high inflation of energy costs, and the implementation of quality procedures.

9.8. Conclusion

Street et al have considered a range of studies evaluating the introduction of activity based funding. Like others, they point out the lack of rigorous evaluation and the difficulty of conducting any controlled studies. They note increases in volume of care, increasing day case activity and reductions in lengths of stay. However, they appear unable to find any evidence that the quality of patient care has improved; in a similar fashion, there is no evidence of deterioration. Interestingly, their overview of DRG system goals for 12 European countries does not list quality improvement.

Further adaptations of activity-based funding are necessary. For instance, emphasis on quality of care is essential, as well as value-based purchasing. Delivering integrated care through episode-based payment is also a field to explore. These new adaptations to the model will require extending the reimbursement mechanism beyond the hospital sector. Recent developments indicate that progress is underway towards adapting activity-based funding in order to meet future challenges.

9.9. Bibliography

- *Activity-based Payment, General Principles*
- Cylus J, Irwin R. The challenges of hospital payment systems. *Euro Observer*, Autumn 2010;12:1-3.
- Dexia and HOPE (2008), *Hospitals in the 27 Member States of the European Union*, Dexia Editions, France.
- Ettelt S. et al. Investing in hospitals of the future, eds. B. Rechel, S. Wright, N. Edwards, B. Dowdeswell, M. McKee, *Observatory Studies Series N°16*, European Observatory on health systems and policies 2009.
- Miraldo M, Goddard M, Smith P. The incentive effects of Payment by Results, *CHE Research Paper 19*, Centre for Health Economics, University of York, September 2006.
- O'Reilly J, Busse R, Häkkinen U, Or Z, Street A, Wiley M. Paying for hospital care: the experience with implementing activity-based funding in five European countries. *Health Econ Policy Law*. 2012 Jan;7(1):73-101.
- Or Z, Häkkinen U. *Qualité des soins et T2A : pour le meilleur ou pour le pire ?* Institut de recherche et documentation en économie de la santé. Irdes editor, Paris December 2012. DT n° 53.
- Paris V, Devaux M, Wei L. "Health Systems Institutional Characteristics: A Survey of 29 OECD Countries", *OECD Health Working Papers*, No. 50, OECD Publishing, 2010.
- Scheller-Kreinsen D, Geissler A, Busse R. 'The ABC of DRGs', *Euro Observer*, 2009;11(4): 1-5.
- Street A, O'Reilly J, Ward P, Mason A. 'DRG-Based Hospital Payment and Efficiency: Theory, Evidence, and Challenges', in R. Busse, A. Geissler, W. Quentin and M. Wiley (eds), *Diagnosis Related Groups in Europe. Moving Towards Transparency, Efficiency and Quality in Hospitals*, Maidenhead: Open University Press. 2011.
- Swiss DRG. *Les forfaits par cas dans les hôpitaux suisses*. Informations de base pour les professionnels de la santé. 2010.
- Wiley, M. From the origins of DRGs to their implementation in Europe. In Busse, R., Geissler, A., Quentin, W. & Wiley, M. (Eds.) *Diagnosis Related Groups in Europe - Moving towards transparency, efficiency and quality in hospitals*. Copenhagen, WHO Regional Office for Europe 2011.
- World Bank (2009) *Designing and Implementing Provider Payment Systems: How To Manuals*. John C. Langenbrunner, Cheryl Cashin, Sheila O'Dougherty Editors.
- *USA*

- Cash, R., M. Grignon, and D. Polton. L'expérience américaine et la réforme de la tarification hospitalière en France - commentaire de l'article de J. Newhouse. *Economie publique*, 2006. 13.
- Rosenthal M.B., Landon B.E., Normand S.L., Frank R.G and A. Epstein (2006), Pay for Performance in Commercial HMOs, *New England Journal of Medicine*, Vol. 355, N°18, pp. 1895-1902
 - *Australia*
 - Australian Government. A National Health And Hospitals Network For Australia's Future. Publications Number: P3-6430. Commonwealth of Australia 2010. ISBN: 978-1-74241-147-7.
 - *Great Britain*
 - Ellis, R. and M. Vidal. Activity-based payments and reforms of the English hospital payment system. 2007.
 - Mason, A. and P. Smith. Le système de santé anglais : régulation et rapports entre les différents acteurs. *RFAS*, 2006.2-3: p. 265-284.
 - Street A. and A. Maynard. Activity based financing in England: the need for continual refinement of payment by results, *Health Economics, Policy and Law*, 2007; No.2, pp. 419-427.
 - Farrar S., Sutton M., Chalkley M., Sussex J., Scott A. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *BMJ* 2009; 339:1-8.
 - Health Economics Research, Unit Office of Health Economics, University of Dundee, Payment by results. *Evaluation Newsletter* 3. March 2008.
 - *France*
 - Ministère de la santé et des sports, Rapport 2009 au Parlement sur la tarification à l'activité. 2009: p. 1-67.
 - IRDES, Principes et enjeux de la tarification à l'activité à l'hôpital (T2A). Enseignements de la théorie économique et des expériences étrangères. Document de travail. DT n°23, Mars 2009.
 - Durant, G., Financement des hôpitaux : des divergences mais surtout des convergences. *Gestion*, 2006. 1: p. 6-14.
 - DHOS MissionT2A, "T2A : Expériences Européennes". Conference on European Casemix-based Hospital Prospective Payment Systems 2005.
 - Commission des comptes de la sécurité sociale, Comparaison internationale des dispositifs de fixation des tarifs d'activité des établissements de santé. October 2009: p. 142-145.
 - Or, Z. Activity based payment in hospitals: Evaluation. *Health Policy Monitor*, 2010.
 - Ministère de la santé et des sports, Rapport 2010 au Parlement sur la tarification à l'activité (T2A). 2010: p. 1-56.
 - DREES, Second rapport d'activité du Comité d'évaluation de la T2A. Document de travail n°94, Septembre 2009.
 - IRDES, Les écarts des coûts hospitaliers sont-ils justifiables ? Réflexions sur une convergence tarifaire entre les secteurs public et privé en France. Document de travail. DT n°25, Mai 2009.
 - *Germany*
 - Busse, R. and A. Riesberg. Health Care Systems in Transition: Germany. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies, 2004; Vol. 6, No. 9.
 - Commission des comptes de la sécurité sociale, Evolution récente des dépenses hospitalières en Allemagne. Juin 2009: p. 134-137.
 - Hensen P, et al. Introduction of diagnosis-related groups in Germany: evaluation of impact on in-patient care in a dermatological setting. *European Journal of Public Health* 2008; 18, 85-91.
 - Neubauer G. and F. Pfister. "DRGs in Germany: Introduction of a comprehensive prospective DRG payment system by 2009", in J.R. Kimberly, G. de Pourville and T. D'Aunno (eds), *The Globalization of Managerial Innovation in Health Care*, Cambridge University Press, Cambridge. 2008.
 - German Institute for the Hospital Remuneration System (InEK), G-DRG-Impact evaluation according to sec. 17b para.8 Hospital Finance Act. Final Report of the first research cycle (2004 to 2006), March 2010.
 - Vogl M. Improving patient-level costing in the English and the German 'DRG' system. *Health Policy*. 2012 Oct 12. pii: S0168-8510(12)00260-6.
 - Herwartz H, Strumann C. On the effect of prospective payment on local hospital competition in Germany. *Health Care Manag Sci*. 2012 Mar; 15(1):48-62.

9.10. Exercise

A 18 year old patient presenting with an Alport syndrome (hereditary nephropathy associated with a collagen abnormality), responsible for end-stage renal disease, was hospitalized for a kidney transplantation. Transplantation was performed on March 1, 2011. The lack of diuresis in the post-operative recovery led to a transfer to a resuscitation unit where acute renal failure necessitated 2 hemodialysis sessions. At day 3, the resumption of diuresis permitted the transfer to a continuous monitoring care unit.

The postoperative was marked by:

- high blood pressure (HBP) requiring an anti-hypertensive treatment.
- discovery by ultrasonography of a perirenal urinoma infected by enterobacteriaceae requiring surgical drainage. A retrograde pyelogram showed a calyceal injury due to the JJ catheter. The catheter was then removed.

- digestive hemorrhage of the lower digestive tract. A complete colonoscopy was then performed which proved to be normal and an upper gastrointestinal endoscopy which showed a hemorrhagic duodenal ulcer.

- a 15 kg loss of weight which required an enteral nutrition for 4 days, then a parenteral nutrition for 30 days.

On day 50, the patient left the continuous monitoring care unit for the transplantation ward. The discharge to home occurred on day 60.

Questions

Q1: Propose a coding for diagnoses, while specifying the principal diagnosis and procedures which occurred in the 3 wards: transplantation ward, intensive care unit and continuous monitoring care unit, respectively.

Q2: Find the ICD-10 diagnostic codes for the stay. You can consult the ICD10 site on internet, for example: http://www.icd10.ch/ebook/FR_OMS_FR/FRS_Index.asp

Q3: Describe all the procedures used during the stay.

Q4: Below is the activity of the transplantation ward (French DRG root 27C06) of a given facility and the corresponding national data. Comment the results of the facility compared to national data.

DRGs	N	Facility		National data	
		Mean length of stay	Mean age	Mean length of stay	Mean age
27C061	6	11,8	50,6	13,0	47,7
27C062	42	17,3	52,4	16,7	50,1
27C063	15	31,3	44,2	22,9	53,3
27C064	7	15,3	39,4	29,0	51,1

N: number of stays, DRGs according to the French classification which gives four levels of severity (increasing order from 1 to 4) for these DRGs dedicated to renal transplantation

Responses

R1. Identify diagnoses and procedures related to the stays in the different clinical units.

- Nephrology/Transplant ward:

Principal diagnosis (PD): end-stage renal disease; *Associated diagnosis (AD):* Alport syndrome; *Procedure:* renal transplantation.

- intensive care unit:

Principal diagnosis: Acute renal failure due with acute tubular necrosis; *Procedure:* two sessions of hemodialysis for acute renal failure.

- For the continuous monitoring care unit:

PD: postoperative surveillance; *AD:* HBP due to a nephropathy, infected perirenal urinoma, infection due to enterobacteriaceae, calyceal injury due to a JJ catheter, hemorrhagic duodenal ulcer, malnutrition.

Procedures: renal ultrasonography, surgical drainage, retrograde pyelogram, upper gastrointestinal endoscopy, total colonoscopy, enteral nutrition then parenteral nutrition.

- Nephrology/transplantation ward:

PD: postoperative surveillance; *AD:* HBP due to a nephropathy, duodenal ulcer.

R2

Diagnosis	CIM 10 Code
End-stage renal disease	N18.5
Alport Syndrome	Q87.8
Acute renal failure with acute tubular necrosis	N17.0
Postoperative surveillance	Z48.9
HBP linked to a nephropathy	I12.0
Infected perirenal urinoma	N15.1

Enterobacteriaceae infection	B96.8
Calyceal injury caused by a JJ catheter	S37.00 + T83.8
Hemorrhagic duodenal ulcer	K26.0
Duodenal ulcer	K26.3
Severe denutrition	E43

R3

Procedure

Kidney transplantation
 Hemodialysis session for acute renal failure
 Renal ultrasonography
 Surgical drainage of a perirenal collection
 Retrograde pyelogram
 Upper gastrointestinal endoscopy
 Complete coloscopy (crossing the ileo-colic orifice)
 Enteral nutrition
 Parenteral nutrition

R4

For the facility data, mean age is lower for stays of severity levels 3 and 4 (27C063 and 27C064 DRGs). Mean length of stay is higher for severity level 3, and lower for severity level 4.

When comparing facility data to national data for each French DRG, we noted that:

- For severity levels 1 and 2, the facility has results similar to the national data for both duration and age.
- For severity level 3, stays are longer and mean age is lower for the facility than for national data. One may wonder whether there is a relationship between age and length of stay?
- Severity level 4: for the facility, the length of stay is lower than for severity level 3 (but pay attention to the restricted number of stays), while for the national data the length of stay lengthens with the level of severity. It would be interesting to have additional information such as the percentage of deaths or early discharge by transfer to another facility.

This picture raises the question of comparability to a reference activity, when the numbers are low and in the absence of additional information.

Authors:

Landais Paul MD, PhD. Université Montpellier 1, Faculté de Médecine

paul.landais@univ-montp1.fr

Boudemaghe Thierry MD. CHU Nîmes

thierry.boudemaghe@chu-nimes.fr

Suehs Carey PhD. CHU Nîmes

carey.suehs@chu-nimes.fr

Dedet Guillaume MD, BSc. London School of Hygiene & Tropical Medicine

guillaume.dedey@hotmail.com

Le Bihan-Benjamin Christine MD, MSc. Institut National du Cancer

clebihan@institutcancer.fr

VII. Publication – Outcomes After Hip Fracture Surgery Compared With Elective Total Hip Replacement

Outcomes After Hip Fracture Surgery Compared With Elective Total Hip Replacement

Yannick Le Manach, MD, PhD^{1,2,3}; Gary Collins, PhD⁴; Mohit Bhandari, MD, PhD^{2,3,5}; Amal Bessissow, MD^{3,6}; Jacques Boddaert, MD, PhD^{7,8}; Frédéric Khiami, MD, PhD⁹; Harman Chaudhry, MD^{3,5}; Justin De Beer, MD⁵; Bruno Riou, MD, PhD^{10,11,12,13}; Paul Landais, MD, PhD^{14,15}; Mitchell Winemaker, MD⁵; Thierry Boudemaghe, MD, MSc^{14,15}; P. J. Devereaux, MD, PhD^{2,3,16}

¹Department of Anesthesia, Michael DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada

²Department of Clinical Epidemiology and Biostatistics, Michael DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada

³Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Ontario, Canada

⁴Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, Botnar Research Center, University of Oxford, Oxford, United Kingdom

⁵Department of Surgery, Michael DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada

⁶Division of General Internal Medicine, McGill University, Montreal, Quebec, Canada

⁷Department of Geriatrics, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique Hôpitaux de Paris, Paris, France

⁸Département Hospitalo-Universitaire "Fight Against Stress and Aging", Université Pierre et Marie Curie-Paris 06, Sorbonne Universités, Paris, France

⁹Department of Orthopedic Surgery, Groupe Hospitalier Pitié-Salpêtrière Assistance Publique Hôpitaux de Paris, Paris, France

¹⁰Unité Propre de Recherche de l'Enseignement Supérieur Equipe d'Accueil 2415, Clinical Research Institute, Faculty of Medicine, Montpellier University, Montpellier, France

¹¹Department of Emergency Medicine and Surgery, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique Hôpitaux de Paris, Paris, France

¹²Institute of Cardiovascular metabolism and Nutrition, Université Pierre et Marie Curie-Paris 06, Sorbonne Universités, Paris, France

¹³Unité Mixte de Recherche 1166, Institut National de la Santé et de la Recherche Médicale, Paris, France

¹⁴Unité Propre de Recherche de l'Enseignement Supérieur Equipe d'Accueil 2415, Clinical Research Institute, Faculty of Medicine, Montpellier University, Montpellier, France

¹⁵Department of Biostatistics, Clinical Research and Medical Information, Nîmes University Hospital, Nîmes, France

¹⁶Department of Medicine, Michael DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada

JAMA. 2015;314(11):1159-1166. doi:10.1001/jama.2015.10842.

ABSTRACT

Importance Patients undergoing surgery for a hip fracture have a higher risk of mortality and major complications compared with patients undergoing an elective total hip replacement (THR) operation. The effect of older age and comorbidities associated with hip fracture on this increased perioperative risk is unknown.

Objective To determine if there was a difference in hospital mortality among patients who underwent hip fracture surgery relative to an elective THR, after adjustment for age, sex, and preoperative comorbidities.

Design, Setting, and Participants Using the French National Hospital Discharge Database from January 2010 to December 2013, patients older than 45 years undergoing hip surgery at French hospitals were included. The *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)*, codes were used to determine patients' comorbidities and complications after surgery. A population matched for age, sex, and preoperative comorbidities of patients who underwent elective THR or hip fracture surgery was created using a multivariable logistic model and a greedy matching algorithm with a 1:1 ratio.

Exposure Hip fracture.

Main Outcomes and Measures Postoperative in-hospital mortality.

Results A total of 690 995 eligible patients were included from 864 centers in France. Patients undergoing elective THR surgery (n = 371 191) were younger, more commonly men, and had less comorbidity compared with patients undergoing hip fracture surgery. Following hip fracture surgery (n = 319 804), 10 931 patients (3.42%) died before hospital discharge and 669 patients (0.18%) died after elective THR. Multivariable analysis of the matched populations (n = 234 314) demonstrated a higher risk of mortality (1.82% for hip fracture surgery vs 0.31% for elective THR; absolute risk increase, 1.51% [95% CI, 1.46%-1.55%]; relative risk [RR], 5.88 [95% CI, 5.26-6.58]; $P < .001$) and of major postoperative complications (5.88% for hip fracture surgery vs 2.34% for elective THR; absolute risk increase, 3.54% [95% CI, 3.50%-3.59%]; RR, 2.50 [95% CI, 2.40-2.62]; $P < .001$) among patients undergoing hip fracture surgery.

Conclusions and Relevance In a large cohort of French patients, hip fracture surgery compared with elective THR was associated with a higher risk of in-hospital mortality after adjustment for age, sex, and measured comorbidities. Further studies are needed to define the causes for these differences.

INTRODUCTION

Although hip surgery can improve mobility and pain, it can be associated with major postoperative medical complications and mortality.¹ Patients undergoing surgery for a hip fracture are at substantially higher risk of mortality and medical complications compared with patients undergoing an elective total hip replacement (THR).^{2,3} The increased risk might be due to the advanced age and comorbidities of hip fracture patients relative to elective THR patients.²⁻⁴ Although patient characteristics may explain the higher risk for a poor outcome among patients undergoing hip fracture surgery, it is possible that physiological processes associated with hip fracture (eg, the acute inflammatory, stress, hypercoagulable, and catabolic states)⁵⁻⁹ may account for some of the increased risk.^{10,11} These processes may account for some of the perioperative morbidity and mortality and, therefore, may represent modifiable risk factors. For example, surgery may be performed earlier in a patient's disease course to minimize the time patients are exposed to these intrinsic factors to improve outcomes.¹² To better understand the potential contribution of these processes to adverse outcomes, we determined the difference in in-hospital mortality between patients undergoing hip fracture surgery and elective THR after adjusting for the known risk factors age, sex, and patient comorbidities.

METHODS

The Agence Technique de l'Information Sur l'Hospitalisation (ATIH) waived the need for consent according to the Enforcement (decree No. 94-666). Since 1996, all French hospitals caring for medical and surgical patients have submitted anonymized patient data to the French Hospital Discharge Database (FHDD). Each discharge summary submitted to the FHDD is linked to a national grouping algorithm leading to a French diagnosis related group¹³; thereby allowing patient comorbidities to be recorded and linked.¹⁴ The study was conducted according to the approval given by the ATIH. Authorization was also obtained from the Commission Nationale de l'Informatique et des Libertés (agreement No. 1375062). The data provided were anonymized.

We identified and included all patients older than 45 years who underwent elective THR or hip fracture surgery requiring anesthesia, performed in France from January 1, 2010, to December 31, 2013. Patients with multiple trauma were excluded. For each patient the following information was extracted: age, sex, primary diagnosis (ie, reason for hospital admission), patient comorbidities (coded according to the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision [ICD-10]*),¹⁵ procedures performed during the hospital stay (coded according to the French classification for medical procedures in the *Classification Commune des Actes Médicaux [CCAM]*),¹⁶ length of stay (LOS) in the hospital (days), and in-hospital complications and mortality after surgery. Using a previously established process, 3 physicians, with expertise in *ICD-10* coding, independently aggregated *ICD-10* codes into broader disease groups to resemble the clinical observations commonly recorded during preoperative assessments (eTable 1 in the Supplement). All disagreements were resolved through a consensus process that involved discussions among the 3 physicians. The Simplified Acute Physiology Score (SAPS II)¹⁷ was used to measure the severity of disease for patients admitted to an intensive care unit (ICU).

The primary outcome was in-hospital mortality, defined as death after surgery and prior to hospital discharge, regardless of LOS. Secondary outcomes included in-hospital major postoperative complications (ie, a composite of myocardial infarction, heart failure, stroke, renal failure, sepsis, and mortality) and each of the individual components of the composite except for mortality. We also evaluated readmission for any cause within 72 hours after initial hospital discharge and in-hospital mortality during readmission.

Continuous and categorical variables were compared using parametric or nonparametric methods as appropriate. A matched population of patients who underwent elective THR and hip fracture surgery was created using a multivariable logistic model and a greedy matching algorithm with a 1:1 ratio and a caliper width of 10^{-5} with no replacement.¹⁸ The matching procedure was carried out by categorizing age (in 5-year age intervals) in which elective THR and hip fracture patients were matched within each age strata. All preoperative variables were included in the matching model and are presented in [Table 1](#). Patients receiving hip fracture surgery and elective THR for whom no match was found were discarded from the matched analyses. After matching, balance of covariates between the groups was assessed using the standardized differences expressed as a percentage for each covariate,¹⁹ and globally using the L_1 measure (in which a value of 0 indicates identical distributions between groups and 1 indicates complete imbalance) and post-matching Cstatistic (in which a value of 0.5 indicates perfect balance).²⁰ Any baseline covariate with an absolute standardized difference greater than 5% was considered to be imbalanced, because such a difference may have a statistical effect on the results depending on the prognostic importance of the covariate.

Table 1. Preoperative Patient Characteristics in the Unmatched and Matched Study Populations for Hip Fracture Surgery vs Total Hip Replacement

Table 1. Preoperative Patient Characteristics in the Unmatched and Matched Study Populations for Hip Fracture Surgery vs Total Hip Replacement

	Population					
	Unmatched (n = 690 995)			Matched (n = 234 314)		
	Group, No. (%)		Standard Difference, %	Group, No. (%)		Standard Difference, %
Hip Fracture Surgery (n = 319 804)	Elective THR (n = 371 191)	Hip Fracture Surgery (n = 117 157)		Elective THR (n = 117 157)		
Demographics						
Age, mean (SD), y	81.7 (10.6)	70.2 (10.2)	110.9	75.0 (10.6)	75.0 (10.6)	0.02
Men	80 640 (25.2)	164 732 (44.4)	-41.1	36 876 (31.5)	36 709 (31.3)	0.31
Public center ^a	248 103 (77.6)	132 750 (35.8)	93.1	77 914 (66.5)	77 904 (66.5)	0.02
Residing in own home prior to surgery	299 337 (93.6)	368 955 (99.4)	-31.9	115 511 (98.6)	115 448 (98.5)	0.45
Medical history						
Hypertension	116 239 (36.3)	137 871 (37.1)	-1.6	37 967 (32.4)	37 844 (32.3)	0.22
Ischemic heart disease	21 585 (6.7)	15 960 (4.3)	10.7	4774 (4.1)	4607 (3.9)	0.73
Cardiac arrhythmia	19 021 (5.9)	12 076 (3.2)	12.9	4096 (3.5)	4046 (3.4)	0.23
Chronic heart failure	3730 (1.2)	1638 (0.4)	8.1	371 (0.3)	376 (0.3)	-0.08
Heart valve disease	8421 (2.6)	7811 (2.1)	3.5	2034 (1.7)	2034 (1.7)	0.00
Peripheral vascular disease	6827 (2.1)	5391 (1.4)	5.1	1542 (1.3)	1419 (1.2)	0.94
Dementia	52 399 (16.4)	2214 (0.6)	59.1	1868 (1.6)	1888 (1.6)	-0.14
Cerebrovascular disease	7040 (2.2)	1513 (0.4)	15.9	709 (0.6)	660 (0.6)	0.55
Hemiplegia or paraplegia	4930 (1.5)	791 (0.2)	14.3	456 (0.4)	416 (0.4)	0.56
Chronic obstructive pulmonary disease	10 113 (3.2)	8270 (2.2)	5.8	2385 (2.0)	2325 (2.0)	0.36
Pulmonary circulation disorder	1109 (0.3)	290 (0.9)	5.8	74 (0.1)	75 (0.1)	-0.03
Chronic respiratory failure	4716 (1.5)	2134 (0.6)	8.9	662 (0.6)	651 (0.6)	0.13
Chronic alcohol abuse	9189 (2.9)	3430 (0.9)	14.3	1613 (1.4)	1597 (1.4)	0.12
Cancer	14 784 (4.6)	9055 (2.4)	11.8	4065 (3.5)	4013 (3.4)	0.24
Cancer with metastasis	2825 (0.9)	1227 (0.3)	7.1	588 (0.5)	597 (0.5)	-0.11
Diabetes	35 446 (11.1)	33 203 (8.9)	7.1	10 508 (9.0)	10 389 (8.9)	0.36
Obesity	6829 (2.1)	38 099 (10.3)	-34.2	3401 (2.9)	3358 (2.9)	0.22
Chronic renal failure	12 075 (3.8)	5186 (1.4)	15.3	1718 (1.5)	1736 (1.5)	-0.13
Preoperative chronic dialysis	1513 (0.5)	318 (0.1)	7.3	115 (0.1)	114 (0.1)	0.03

Abbreviation: THR, total hip replacement.

^a Hospitals owned and managed by the French government.

Relative risk (RR) and the 95% CIs for primary and secondary outcomes were computed using conditional Poisson regression. *P* values were 2-tailed, and values less than .05 were considered significant.

To explore the potential statistical effect of unmeasured confounders related to patient frailty on the primary outcome, we conducted a sensitivity analysis, replicating the main analysis for the primary and secondary outcome on patients younger than 60 years and residing at home prior to surgery (ie, the group of patients at lowest risk of frailty). We also conducted sensitivity analyses in which RR estimations were adjusted for the hospital LOS.

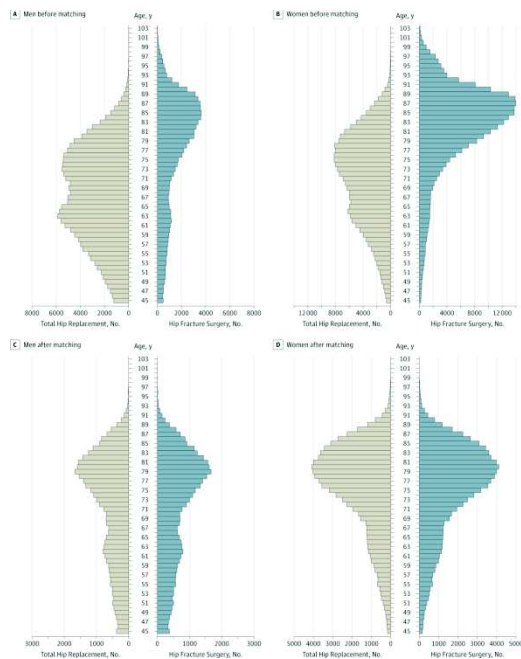
Statistical analyses were carried out in R software (R Foundation), version 3.1, and matching procedures were performed using MatchIt package (R Foundation), version 2.4-21.

RESULTS

During the study period, 690 995 eligible patients from 864 centers in France were identified and included in this study. Elective THR was carried out in 371 191 patients and hip fracture surgery in 319 804 patients. [Table 1](#) reports the baseline patient characteristics of these 2 groups, and the [Figure](#) shows the age and sex distributions.

Figure. 1

Comparison of Age Distribution in Patients Who Underwent Total Hip Replacement or Hip Fracture Surgery Before and After Matching



In the unmatched study population, patients undergoing an elective THR were younger and had fewer preoperative comorbidities. The mean LOS for patients who underwent hip fracture surgery was longer than for those who underwent an elective THR (12.09 days [95% CI, 12.07-12.13] for hip fracture surgery vs 7.80 days [95% CI, 7.79-7.81] for elective THR; $P < .001$). Over the study period, the mean length of hospital stay decreased in the hip fracture group (from 12.77 days [95% CI, 12.70-12.81] in 2010 to 11.72 days [95% CI, 11.67-11.77] in 2013; $P < .001$) and the elective THR group (from 9.19 days [95% CI, 9.13-9.23] in 2010 to 8.11 days [95% CI, 8.08-8.12] in 2013; $P < .001$). Postoperatively, patients who underwent surgery for a hip fracture were less frequently discharged to their home (35.8% for hip fracture surgery vs 60.0% for elective THR; $P < .001$). Readmissions within 72 hours of discharge after the index surgery were more common among patients who underwent hip fracture surgery (3897 patients [1.3%]) compared with patients who underwent an elective THR (1647 patients [0.4%]; $P < .001$).

[Table 2](#) reports the in-hospital outcomes for the unmatched study groups. After hip fracture surgery, 10 931 patients (3.42% [95% CI, 3.35%-3.47%]) died and, after elective THR surgery, 669 patients (0.18% [95% CI, 0.17%-0.19%]) died (absolute risk increase, 3.23% [95% CI, 2.98-3.51]; RR, 18.96 [95% CI, 17.54-20.50]; $P < .001$). All the other individual complications were more common in the patients receiving hip fracture surgery compared with those receiving the elective THR, with RRs ranging from 3.41 to 7.23. At least 1 major postoperative complication occurred in 32 913 patients receiving hip fracture surgery (10.29% [95% CI, 10.18%-10.40%]) and in 5955 patients receiving elective THR (1.60% [95% CI, 1.56%-1.64%]; absolute risk increase, 8.69% [95% CI, 8.38%-8.94%]; RR, 6.41 [95% CI, 6.24-6.59]; $P < .001$).

Table 2. Postoperative Outcomes in the Unmatched Study Population (n = 690 995)

	Group, No. (%)		RR (95% CI)	P Value
	Elective THR (n = 371 191)	Hip Fracture Surgery (n = 319 804)		
In-hospital mortality	669 (0.18)	10 931 (3.42)	18.96 (17.54-20.50)	<.001
Postoperative				
Myocardial infarction	612 (0.16)	1797 (0.56)	3.41 (3.11-3.74)	<.001
Heart failure	2855 (0.77)	16 678 (5.22)	6.78 (6.52-7.05)	<.001
Stroke	332 (0.09)	2067 (0.65)	7.23 (6.44-8.11)	<.001
Renal failure	823 (0.22)	3355 (1.05)	4.73 (4.38-5.11)	<.001
Sepsis	291 (0.08)	1122 (0.35)	4.48 (3.93-5.09)	<.001
Any postoperative complication	5955 (1.60)	32 913 (10.29)	6.41 (6.24-6.59)	<.001
ICU admission	1015 (0.27)	4856 (1.52)	5.55 (5.19-5.94)	<.001
Readmission within 72 h of discharge	1647 (0.44)	3897 (1.22)	2.75 (2.59-2.91)	<.001
In-hospital mortality during readmission	25 (1.52)	488 (12.52)	9.24 (6.16-13.88)	<.001

Abbreviations: ICU, intensive care unit; RR, relative risk; THR, total hip replacement.

In the unmatched study population, postoperative ICU admission was also more frequent in patients who underwent a hip fracture surgery compared with an elective THR (1.52% for hip fracture surgery vs 0.27% for elective THR, $P < .001$), and the SAPS II at ICU admission was higher among patients receiving hip fracture surgery compared with those receiving elective THR (45.37 [95% CI, 44.65-46.08] for hip fracture surgery vs 34.97 [95% CI, 33.69-36.25] for elective THR; $P < .001$). Among patients readmitted within 72 hours after the initial hospital discharge, in-hospital mortality during the second stay was higher in those receiving hip fracture surgery (12.5% for hip fracture surgery vs 1.5% for elective THR; $P < .001$).

The matching procedure retained 117 157 patients who underwent hip fracture surgery and 117 157 patients who underwent an elective THR. When compared with the unmatched populations, patients who underwent hip fracture surgery included in the matched population were younger (mean age: 75.0 years [95% CI, 74.94-75.06] for matched population vs 81.7 years [95% CI, 81.66-81.74] for unmatched population; $P < .001$) and patients who underwent elective THR included in the matched population were older (mean age: 75.0 years [95% CI, 74.94-75.06] for matched population vs 70.2 [95% CI, 70.17-70.23] for unmatched population; $P < .001$). After matching, there were no clinically important differences in baseline characteristics across the patient groups (Table 1 and Figure). Standardized differences in baseline characteristics were all less than 1%. Global balance metrics, the L_1 measure and Cstatistic, demonstrated minimal imbalances between groups after matching (L_1 measure: 0.592 before matching vs 0.427 after matching; Cstatistic: 0.882 before matching vs 0.503 after matching).

Table 3 reports the in-hospital outcomes for the matched study groups. In the matched population, 2130 patients died after hip fracture surgery (1.82% [95% CI, 1.75%-1.90%]) and 362 patients died after elective THR (0.31% [95% CI, 0.28%-0.34%]; absolute risk increase: 1.51% [95% CI, 1.46%-1.55%]; RR, 5.88 [95% CI, 5.26-6.58]; $P < .001$). All other individual complications were more common in patients who underwent hip fracture surgery compared with those who underwent elective THR, with RRs ranging from 1.62 to 3.44. A major postoperative complication occurred in 6890 patients receiving hip fracture surgery (5.88% [95% CI, 5.75%-6.01%]), and 2741 patients receiving elective THR (2.34% [95% CI, 2.25%-2.43%]; absolute risk increase, 3.54% [95% CI, 3.50%-3.59%]; RR, 2.50 [95% CI, 2.40-2.62]; $P < .001$).

Table 3. Postoperative Outcomes in Matched Study Population (n = 234 314)

	Group, No. (%)		RR (95% CI)	P Value
	Elective THR (n = 117 157)	Hip Fracture Surgery (n = 117 157)		
In-hospital mortality	362 (0.31)	2130 (1.82)	5.88 (5.26-6.58)	<.001
Postoperative				
Myocardial infarction	259 (0.22)	419 (0.36)	1.62 (1.39-1.89)	<.001
Heart failure	1368 (1.17)	3114 (2.66)	2.28 (2.14-2.43)	<.001
Stroke	171 (0.15)	461 (0.39)	2.70 (2.26-3.21)	<.001
Renal failure	346 (0.30)	763 (0.65)	2.21 (1.94-2.50)	<.001
Sepsis	104 (0.09)	322 (0.27)	3.10 (2.48-3.86)	<.001
Any postoperative complication	2741 (2.34)	6890 (5.88)	2.50 (2.40-2.62)	<.001
ICU admission	435 (0.37)	1496 (1.28)	3.44 (3.09-3.83)	<.001
Readmission within 72 h of discharge	636 (0.54)	1135 (0.97)	1.78 (1.62-1.97)	<.001
In-hospital mortality during readmission	14 (2.20)	95 (8.37)	4.05 (2.30-7.17)	<.001

Abbreviations: ICU, intensive care unit; RR, relative risk; THR, total hip replacement.

The sensitivity analysis, which focused on patients younger than 60 years who were residing at home before surgery, included 12 683 patients who underwent hip fracture surgery (4.0% of all patients who underwent hip fracture surgery) and 12 683 patients who underwent elective THR (3.4% of all patients who underwent elective THR). In this matched population, 58 patients died after hip fracture surgery (0.46% [95% CI, 0.34%-0.57%]), and 10 patients died after elective THR (0.08% [95% CI, 0.03%-0.13%]; absolute risk increase, 0.38% [95% CI, 0.16%-0.83%]; RR, 5.80 [95% CI, 2.96-11.35]; $P < .001$). eTable 2 in the Supplement reports the primary and secondary outcomes for this younger group who resided at home before surgery. As per the results of the primary outcome, patients with a hip fracture also demonstrated higher risk-adjusted associations

with the secondary outcomes (except for postoperative myocardial infarction and postoperative heart failure).

The mean hospital LOS for patients was longer among patients who underwent hip fracture surgery compared with patients who underwent an elective THR (11.61 days [95% CI, 11.56-11.64] for hip fracture surgery vs 9.3 days [95% CI, 9.28-9.32] for elective THR; $P < .001$). eTable 3 in the Supplement reports the results of the analyses that did and did not include hospital LOS in the adjusted analyses. The results were similar for these analyses. For example, after adjustment for LOS, the RR for in-hospital mortality associated with hip fracture surgery relative to elective THR was similar to our primary analysis not adjusted for LOS (in-hospital mortality RR: 5.67 [95% CI, 5.07-6.34] for LOS-adjusted analysis vs 5.88 [95% CI, 5.26-6.58] for unadjusted primary analysis; $P < .001$ for both analyses).

Patients in the hip fracture surgery group were more frequently admitted to the ICU postoperatively (1.3% for hip fracture surgery vs 0.4% for elective THR; $P < .001$) and stayed longer in the ICU than patients who underwent an elective THR (6.3 days [95% CI, 5.8-6.8] for hip fracture surgery vs 4.5 days [95% CI, 3.8-5.2] for elective THR; $P < .001$). Patients in the hip fracture surgery group and the elective THR group demonstrated no significant difference in SAPS II values at ICU admission (40.1 [95% CI, 39.0-41.1] for hip fracture surgery vs 39.0 [95% CI, 37.1-40.9] for elective THR; $P = .261$) and in-hospital mortality in patients admitted to the ICU (18.1% for hip fracture surgery vs 15.9% for elective THR; $P = .264$). Hospital readmission within 72 hours was more frequent in patients who underwent hip fracture surgery (1.0% for hip fracture surgery vs 0.5% for elective THR; $P < .001$) and in-hospital mortality during the second admission was higher in patients receiving hip fracture surgery compared with those receiving elective THR (8.4% vs 2.2%; $P < .001$).

DISCUSSION

Patients undergoing surgery for a hip fracture were older and had more comorbidities than patients who underwent an elective THR, and these differences accounted for some of the difference in outcomes between these groups. The primary finding of this study was that among patients who received hip fracture surgery or elective THR and were matched for age, sex, and preoperative medical conditions, the risks were higher after hip fracture surgery for in-hospital mortality (absolute risk increase, 1.51% [95% CI, 1.46%-1.55%]; RR, 5.88 [95% CI, 5.26-6.58]) and major postoperative complications (absolute risk increase, 3.54% [95% CI, 3.50%-3.59%]; RR, 2.50 [95% CI, 2.40-2.62]). If the absolute risk increases of 1.51% for in-hospital mortality and 3.54% for major postoperative complications were modifiable, they would be consistent with the number needed to treat of 59 patients for in-hospital mortality and 28 patients for major postoperative complications. Hip fracture may be associated with physiologic processes that are not present in circumstances leading to elective THR and increase the risk of morbidity and mortality following surgery.

Strengths of the study include our evaluation of a large contemporary sample of patients who underwent hip surgery in France. We took into account an extensive list of comorbidities to obtain matched patient groups who underwent an elective THR or a hip fracture surgery. The balance of covariates between the groups suggests that the RR estimates were not biased by imbalances in the recorded variables.

This study had several limitations including potential misclassification of outcomes. In-hospital mortality was defined as the primary outcome to minimize the statistical effect of these limitations. We were unable to evaluate the cause of death because these data were not available. Misclassification of secondary outcomes and potential underestimation of their frequencies remains possible; however, this would be expected to be similar in both groups. Nevertheless, we cannot

completely exclude differential misclassification between the surgical groups, which could have biased the RR estimates.

Another limitation of this study is the nonstandardized follow-up period. The follow-up period was limited to the in-hospital admission period, and it is different between the 2 patient groups. Because the absolute difference in the duration of follow-up was short (ie, a difference of 2.3 days), it is unlikely that this difference in follow-up explains the magnitude of risk associated with hip fracture surgery relative to an elective THR. Although there is no national mortality database that we could link to the FHDD to capture 30-day mortality, we evaluated hospital readmissions within 72 hours of the index hospital discharge and assessed mortality during readmissions. Readmissions and in-hospital mortality during readmission were higher in the patients who underwent a hip fracture surgery relative to an elective THR. Moreover, the sensitivity analyses, adjusting for LOS, produced similar results as the main analyses and supported our main findings (eTable 3 in the Supplement).

An additional limitation of our study is the possibility of residual confounding (ie, that there were unmeasured factors affecting one group or another that were not measured by the data available for analysis). There is also the possibility of a selection bias that adjustment cannot resolve because patients having an elective THR are selected for their ability to withstand the stresses of surgery before the operation. Moreover, we did not have information on the mechanism of injury (eg, a fall due to instability, orthostatic hypotension, or poor nutritional status) or concomitant injuries. These unmeasured comorbidities in patients with hip fracture may have influenced outcomes. Although residual confounding is probable, physiologic mechanisms, such as acute stress and inflammatory states resulting from the fractures, likely explain our findings because the diverse array of preoperative comorbidities that we adjusted for in our analyses, the magnitude of effect we found, and our sensitivity analyses restricted to younger patients living at home prior to surgery (ie, the group of patients at lowest risk of frailty before surgery) demonstrated results similar to the overall analyses.

We were not able to adjust for whether surgery was performed by a consultant or surgical resident because this information was not available. In France, surgical residents are not allowed to operate on patients alone without consultant supervision. Even if that were the case, resident-performed procedures like the ones assessed in this study have clinical outcomes equivalent to those when a consultant (attending surgeon) performs the procedures alone.²¹⁻²³ Furthermore, hip fracture surgery is generally considered less invasive than total hip arthroplasty. These points suggest that if the surgeons performing the hip fracture surgeries were relatively less experienced than the surgeons performing the elective THRs, the effect on our results was likely minimal.

A meta-analysis of prospective cohort studies (22 studies among women and 17 studies among men) of age- and sex-matched patients with and without hip fractures demonstrated, after adjustment for comorbidities, that patients experiencing a hip fracture had an increased risk of mortality at 3 months (relative hazard, 5.75 [95% CI, 4.94-6.67] among women and 7.95 [95% CI, 6.13-10.30] among men).²⁴ Although these data establish a patient experiencing a hip fracture is at an increased risk of mortality compared with a patient who has not sustained a hip fracture, these data do not determine if the risk relates to surgery or something intrinsic to the hip fracture. A large administrative database study evaluated outcomes between patients undergoing surgery for a hip fracture and an elective THR between the years 1990 and 2007.²⁵ Similar to our findings, this study demonstrated that hip fracture surgery was associated with worse outcomes than THR (eg, in-hospital mortality, 1.8% for hip fracture surgery vs 0.2% for THR). In contrast to our study, the study by Sassoon et al²⁵ only reported unadjusted results.

In a recent study, patients with hip fracture admitted to a dedicated geriatric unit with a multidisciplinary approach after surgery had a lower risk-adjusted 6-month mortality compared

with patients admitted to an orthopedic floor.²⁶ In our matched analysis, patients admitted to the ICU in both groups had similar disease severity (ie, SAPS II) and experienced similar in-hospital mortality. Our results also suggest the possibility that the medical specialty (ie, intensivists) and location of postoperative care after a hip fracture may be associated with outcomes.

Despite the benefits of hip surgery, it increases a patient's risk of major morbidity and mortality. The more comorbidity a patient has prior to surgery, the higher their risk of a major complication after surgery.¹ Patients undergoing hip fracture surgery are older and have higher burdens of comorbidities compared with patients undergoing elective THR. Some authors have assumed these differences account for the higher risk of complications in patients undergoing hip fracture surgery relative to elective THR. Our data suggest this is not the entire explanation and that factors intrinsic to a hip fracture may also influence the outcomes.

A hip fracture results in trauma, pain, bleeding, and immobility. These factors initiate inflammatory, hypercoagulable, stress, and catabolic states^{5-7,9,27-30} that have the potential to cause complications (eg, myocardial infarction, pulmonary embolism, pneumonia, sepsis, stroke, major bleeding, disability, or mortality).^{10,31} It is possible to minimize a patient's exposure to these harmful factors through rapid surgery.^{12,32,33} Studies suggest that reducing the delay between a hip fracture and surgery may limit the consequences of the fracture, and the risk-adjusted mortality and major complications after hip fracture surgery may then more closely approximate those of patients undergoing elective THR.³³

CONCLUSIONS

In a large cohort of French patients, hip fracture surgery compared with elective THR was associated with a higher risk of in-hospital mortality after adjustment for age, sex, and measured comorbidities. Further studies are needed to define the causes for these differences.

ARTICLE INFORMATION

Corresponding Author: Yannick Le Manach, MD, PhD, Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Perioperative Medicine and Surgical Research Unit, 237 Barton St E, Hamilton, ON L8L 2X2, Canada (yannick.lemanach@phri.ca).

Author Contributions: Drs Le Manach and Boudemaghe had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Le Manach, Bhandari, De Beer, Riou, Devereaux.

Acquisition, analysis, or interpretation of data: Le Manach, Collins, Bessissow, Boddaert, Khiami, Chaudhry, Riou, Landais, Winemaker, Boudemaghe, Devereaux.

Drafting of the manuscript: Le Manach, Collins, Devereaux.

Critical revision of the manuscript for important intellectual content: Le Manach, Collins, Bhandari, Boddaert, Khiami, Chaudhry, De Beer, Riou, Landais, Winemaker, Boudemaghe, Devereaux.

Statistical analysis: Le Manach, Collins.

Obtained funding: Le Manach.

Administrative, technical, or material support: Le Manach, Khiami, Landais, Boudemaghe.

Study supervision: Le Manach, Bhandari, Boddaert, Khiami, De Beer, Riou, Landais.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Bhandari reports grant funding from Smith and Nephew, DePuy, Eli Lilly, Bioventus, Stryker, Zimmer Biomet, and Amgen and consulting fees

from Smith and Nephew, Stryker, Amgen, Zimmer Biomet, Moximed, Bioventus, Merck, Eli Lilly, and sanofi-aventis. Dr Boddaert reports receiving travel accommodations from Novartis. Dr Riou reports receiving personal fees from Crossject, Thermo Fisher Scientific/BRAHMS, LFB, sanofi-aventis, and Sangart. Dr Devereaux reports receiving grant funding from Abbott Diagnostics, Boehringer Ingelheim, Covidien, Octopharma Plasma, Roche Diagnostics, and Stryker. No other disclosures were reported.

Funding/Support: This work is supported by the Hamilton Anesthesia Associates (Dr Le Manach) and the Canadian Network and Centre for Trials Internationally (Dr Le Manach), by grant G1100513 from the United Kingdom Medical Research Council (Dr Collins), and by a Heart and Stroke Foundation of Ontario Career Investigator Award (Dr Devereaux).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1

Devereaux PJ, Chan MT, Alonso-Coello P, et al; Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*. 2012;307(21):2295-2304.

2

Hunt LP, Ben-Shlomo Y, Clark EM, et al; National Joint Registry for England, Wales and Northern Ireland. 90-day mortality after 409 096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. *Lancet*. 2013;382(9898):1097-1104.

3

White SM, Moppett IK, Griffiths R. Outcome by mode of anaesthesia for hip fracture surgery: an observational audit of 65 535 patients in a national dataset. *Anaesthesia*. 2014;69(3):224-230.

4

Ford S. Outcome differences after elective or emergency hip operations. *Anaesthesia*. 2014;69(6):643.

5

Beloosesky Y, Hendel D, Weiss A, et al. Cytokines and C-reactive protein production in hip fracture—operated elderly patients. *J Gerontol A Biol Sci Med Sci*. 2007;62(4):420-426.

6

Desborough JP. The stress response to trauma and surgery. *Br J Anaesth*. 2000;85(1):109-117.

7

Miller RR, Shardell MD, Hicks GE, et al. Association between interleukin-6 and lower extremity function after hip fracture—the role of muscle mass and strength. *J Am Geriatr Soc*. 2008;56(6):1050-1056.

8

Onuoha GN, Alpar EK. Elevation of plasma CGRP and SP levels in orthopedic patients with fracture neck of femur. *Neuropeptides*. 2000;34(2):116-120.

9

Svensén CH. Vascular endothelial growth factor (VEGF) in plasma increases after hip surgery. *J Clin Anesth*. 2004;16(6):435-439.

10

LeBlanc ES, Hillier TA, Pedula KL, et al. Hip fracture and increased short-term but not long-term mortality in healthy older women. *Arch Intern Med*. 2011;171(20):1831-1837.

11

Boddaert J, Raux M, Khiami F, Riou B. Perioperative management of elderly patients with hip fracture. *Anesthesiology*. 2014;121(6):1336-1341.

12

Hip Fracture Accelerated Surgical Treatment and Care Track (HIP ATTACK)
Investigators. Accelerated care vs standard care among patients with hip fracture: the HIP ATTACK pilot trial. *CMAJ*. 2014;186(1):E52-E60.

13

Agence Technique de l'Information Sur l'Hospitalisation. *Classification en groupes homogènes de malades (GHM), ministère de la Santé et des Sports, Paris, France*. <http://www.atih.sante.fr/manuel-des-ghm-version-definitive-11g>. Accessed August 24, 2015.

14

Landais P, Boudemaghe T, Suehs C, Dedet G, Lebihan-Benjamin C. Computerized Medico-Economic Decision Making: An International Comparison. In: Venot A, Burgun A, Quantin C, eds. *Medical Informatics, e-Health: Fundamentals and Applications*. Paris, France: Springer;2014:191-220.

15

World Health Organization. *International Classification of Diseases (ICD)*. <http://www.who.int/classifications/icd/en/>. Accessed July 14, 2013.

16

Classification Commune des Actes Médicaux. *Sécurité Sociale, l'Assurance Maladie [in French]*. <http://www.ameli.fr/accueil-de-la-ccam/index.php>. Accessed May 29, 2015.

17

Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270(24):2957-2963.

18

Lunt M. Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *Am J Epidemiol*. 2014;179(2):226-235.

19

Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083-3107.

20

Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med*. 2014;33(10):1685-1699.

21

Robinson RP. The impact of resident teaching on total hip arthroplasty. *Clin Orthop Relat Res*. 2007;465(465):196-201.

22

van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. *BMC Med.* 2012;10:65.

23

Woolson ST, Kang MN. A comparison of the results of total hip and knee arthroplasty performed on a teaching service or a private practice service. *J Bone Joint Surg Am.* 2007;89(3):601-607.

24

Haentjens P, Magaziner J, Colón-Emeric CS, et al. Meta-analysis: excess mortality after hip fracture among older women and men. *Ann Intern Med.* 2010;152(6):380-390.

25

Sassoon A, D'Apuzzo M, Sems S, Cass J, Mabry T. Total hip arthroplasty for femoral neck fracture: comparing in-hospital mortality, complications, and disposition to an elective patient population. *J Arthroplasty.* 2013;28(9):1659-1662.

26

Boddaert J, Cohen-Bittan J, Khiami F, et al. Postoperative admission to a dedicated geriatric unit decreases mortality in elderly patients with hip fracture. *PLoS One.* 2014;9(1):e83795.

27

Chuang D, Power SE, Dunbar PR, Hill AG. Central nervous system interleukin-8 production following neck of femur fracture. *ANZ J Surg.* 2005;75(9):813-816.

28

Esmon CT. The interactions between inflammation and coagulation. *Br J Haematol.* 2005;131(4):417-430.

29

Hess JR, Brohi K, Dutton RP, et al. The coagulopathy of trauma: a review of mechanisms. *J Trauma.* 2008;65(4):748-754.

30

Tsangari H, Findlay DM, Kuliwaba JS, Atkins GJ, Fazzalari NL. Increased expression of IL-6 and RANK mRNA in human trabecular bone from fragility fracture of the femoral neck. *Bone.* 2004;35(1):334-342.

31

Prevention of pulmonary embolism and deep vein thrombosis with low dose aspirin: Pulmonary Embolism Prevention (PEP) trial. *Lancet.* 2000;355(9212):1295-1302.

32

Bessissow A, Chaudhry H, Bhandari M, Devereaux PJ. Accelerated vs standard care in hip fracture patients: does speed save lives? *J Comp Eff Res.* 2014;3(2):115-118.

33

Simunovic N, Devereaux PJ, Sprague S, et al. Effect of early surgery after hip fracture on mortality and complications: systematic review and meta-analysis. *CMAJ.* 2010;182(15):1609-1616.

VIII. Publication – Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI)

Article soumis le 17/08/2016

Demande de 2 corrections mineures le 29/09/2016

Comments from reviewer:

1. Pg 3 Ln 29 - authors need to explain what "collision probability" means
2. I don't think Fig 1 offers much more than what is in the text and so can be cut. Most IJE readers will be familiar with de-identifying processes.

Corrections effectuées et adressées le 29/09/2016



Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI)

Journal:	<i>International Journal of Epidemiology</i>
Manuscript ID	IJE-2016-08-0972.R1
Manuscript Type:	Data Resource Profile
Date Submitted by the Author:	n/a
Complete List of Authors:	BOUDEMAGHE, Thierry; Centre Hospitalier Universitaire de Nimes, BESPIM BELHADJ, Ihssen; Centre Hospitalier Universitaire de Nimes, BESPIM
Key Words:	PMSI, French claims database, Record linkage



Data Resource Profile: The French National Uniform Hospital Discharge

Data resource basics

All major industrialized countries have set up a prospective payment system for piloting healthcare costs.¹ This system involves the production and transmission of structured and standardized billing information, including medical information concerning patients, their health problems and treatments, commonly referred to as the Uniform Hospital Discharge Data Set (UHDDS). This information is eventually computerized and archived, and has led to the constitution of massive databases.

The collection of UHDDS records was initiated as a part of the implementation of the French Prospective Payment System (PPS), called the *Programme National de Médicalisation des Systèmes d'Information* (PMSI). Following a period of testing, experimentation, and progressive implementation during the 80's²⁻⁷ and the 90's⁸⁻¹⁰, UHDDS production by all health care facilities became mandatory in 1996¹¹, representing 20-25 million records par year.

In 2001, a deterministic anonymous linkage identifier (LinkId) was implemented¹², thus making it possible to link UHDDS records and subsequently study patient re-hospitalization pathways.

Since 2004, tariffs have been assigned to DRGs. The collection of UHDDSs has taken an increasing part in the determination of the income of French hospitals as concerns acute inpatient activity, until finally in 2008 all such hospital reimbursements became entirely based upon the UHDDSs.

Furthermore, the UHDDSs are subject to quality and accuracy controls led by the French Social Security administration, potentially resulting in fines for those hospitals not respecting certain coding criteria. This process triggered significant improvement in data quality¹³.

Following major changes in March 2009, the methodology and terminologies for the production of UHDDSs have now been stable for 7 years. Currently, the UHDDSs describe all inpatient hospital stays in a standardized data set. The latter identifies the hospital and summarizes information about the patient and the stay, notably by describing diagnoses and procedures. It also provides additional information about specific aspects of the stay (for instance, a stay in an intensive care unit). Finally, the data set also contains the Diagnosis Related Group (DRG) the stay has been assigned to.

Including the year 2015, the PMSI database amounts to approximately 180 million stays and 40 million LinkIds.

Data collected

Dataset production and linkage

A UHDDS is produced by every hospital for every acute inpatient stay, for ambulatory care requiring a hospitalization (e.g. carpal tunnel procedure) and for session treatments (e.g. dialysis session).

A part of the data is directly generated from the administrative information system (e.g. date of entry, date of discharge). Medical information is coded from different sources (e.g. lab reports, discharge reports) according defined terminologies and coding rules. The Medical Information Department (MID) is in charge of the collection, control, transmission and analysis of the UHDDSs.

The UHDDSs are then anonymized, essentially by removing the patient's date of birth and the dates of entry and discharge, and uploaded to a government agency named the *Agence Technique de l'Information sur l'Hospitalisation* (ATIH)¹⁴; they are then further processed in order to issue payments for hospitals according to their activity.

Data linkage

The LinkId is created by hashing a combination of several identifying traits via a non-reversible algorithm. This process is technically as safe as possible with a collision probability (the occurrence of an identical hash for different patients) of 10^{-48} , and since the identifying information is also combined with several secret keys, dictionary attack is virtually impossible^{15,16}.

LinkIds are created in a two stage process. At the hospital level, when transmitting the data to the ATIH, a linkage program is used to hash source data supposedly unique to each patient (birth date, sex, and social security number) with a secret key. At the ATIH level, the resulting character string is again combined with a secret key and re-hashed.

Additionally, a chronological identifier is created at the hospital level: a) for a given LinkId, a hidden, random reference date is generated; b) the difference in days between this reference date and the date of admission is computed, and c) the result, labelled as the "stay number" is associated with the record pertaining to this LinkId. Each time the hashing leads to this particular LinkId, the same reference date will be used, resulting in successive stays having increasing stay numbers. The stay number may then be used to chronologically ordinate the records associated with a given LinkId.

Measures

Each record contains information mainly about:

- the hospital: national entity identifier;
- the patient: gender, age, location
- the patient's condition and management: pathologies, procedures;
- the stay: DRG, length of stay, type of admission and discharge (home, referral), month and year of discharge. Additional information about specific wards (e.g. intensive care unit) attended is given.

Pathologies are described by a principal diagnosis, and optionally secondary diagnoses. They are coded using the French version of the ICD 10¹⁷. In specific cases, i.e. essentially session-treatments, a “related diagnosis” item is used

Procedures are coded using the *Classification Commune des Actes Médicaux* (CCAM).¹⁸ Laboratory tests and drug prescriptions are not recorded in the UHDDS.

Data resource use

Three types of data resource use have successively developed over the years.

Simple use. Most of these studies are focused on a clinical theme, using the PMSI database as a source for identifying cases or analysing the main characteristics of a given population. Over time, some of these studies have incorporated other data sources (e.g. actual medical files, epidemiological registries) to consolidate them.

Assessment of the PMSI database as a reliable source of data for studies. The first domain that addressed the reliability of the PMSI database for disease monitoring as early as 2002 was cancer¹⁹. This topic will remain the most explored over the years, with more and more optimistic conclusions concerning the reliability of the system as time passes and data quality improves²⁰.

Use of PMSI data with methodological adjustments. In this case, the PMSI database is used with methodological adjustments. Example studies again mainly involve cancer^{21,22}. However, a noteworthy case of adjusted use of the data occurred in association with the “Mediator® scandal”; the latter concerned the increase of valvular heart disease caused by benfluorex. Following several case-control and case report articles^{23–27}, an increase in the risk of valvular heart disease for diabetic patients treated with benfluorex was assessed for the French population in a study based on the PMSI database²⁸. This study significantly contributed to the re-assessment of the benefit-risk ratio for this product by the French Medicine drug agency, and ultimately to the suspension of its market authorization in France, shortly followed by a withdrawal recommendation from the European Medicines Agency²⁹.

Table 1 provides example articles for each category.

Theme	Category	Use of linkage	Use of other data source
Drug-induced hypersensitivity syndrome ³⁰	Simple use	No	No
Relationship between in-hospital mortality rate and patient volume for total hip arthroplasty ³¹	Simple use	No	No
Evaluation of prescription practices ³²	Simple use	No	Yes
Frequency of hospitalizations related to drug abuse ³³	Simple use	Yes	Yes
Short-term risk of bleeding and arterial thromboembolic events ³⁴	Simple use	Yes	Yes
Cost-effectiveness model comparing dual-mobility to fixed-bearing designs for total hip replacement in France ³⁵	Simple use	Yes	No
Reliability of diagnostic coding ³⁶	Assessment as a reliable source	No	No
Estimation of the incidence of breast cancer from medico-administrative databases ³⁷	Assessment as a reliable source	Yes	Yes
Identification of incident central nervous system tumors ³⁸	Assessment as a reliable source	Yes	Yes
Role in epidemiological studies ³⁹	Assessment as a reliable source	Yes	Yes
Identification of incident cases of colorectal cancer ⁴⁰	Assessment as a reliable source	Yes	Yes
Breast cancer incidence ²¹	Adjusted use	Yes	Yes
Benfluorex and valvular heart disease ²⁸	Adjusted use	Yes	Yes
Estimation of cancer incidence ²²	Adjusted use	No	Yes
Outcomes after hip fracture surgery ⁴¹	Adjusted use	Yes	No

Table 1. Example articles of PMSI database use.

Strengths and weaknesses

Strengths

The UHDDSs are used to determine financial resources, and are frequently and thoroughly verified by both its producer and the paying party, with possible financial and legal consequences. It is thus reasonable to assume that this database is of an acceptable quality, taking into account the natural limits to precision imposed by the methodology and terminologies used to code conditions and procedures.

Because it is comprehensive, the PMSI database gives the opportunity to conduct studies on a national scale. The volume of data ensures statistical power and precision. Combined with the study of medical information conveyed by the UHDDS, record linkage gives access to dynamic factors for the study of pathologies and patient behaviour (e.g. re-hospitalization rates and delays, attended hospital type).

Weaknesses

The PMSI database is not built according to the thorough criteria required by clinical research. The quality of its data meets the requirement of the billing process, which relies more on completeness of core information and basic consistency (e.g. age or sex consistency with certain procedures). These data should simply be considered as raw data from the point of view of scientific research. Immediate use of the data is hampered by technical and methodological issues that can be classified into four categories:

- content data quality: the data may include unusable records containing errors, surrogate data (e.g. generic location code when patient location is unknown) or missing data;
- hospital identification: in an ever changing environment, every year hospitals close, open, merge or change identifiers;
- coding rules: each year the official coding guide is updated, sometimes with important impacts on the way patient disease and patient management are represented through coding;
- terminology updates: new codes are regularly added to the French version of the ICD10, along with new coding rules.

These issues need to be addressed before using the database, according to the required study design.

Data resource access

A public website presents predetermined queries on coarse parameters to the public⁴². For example, it is possible to obtain the case-mix of a given hospital for a given year. However, it is not possible to combine patient parameters (age, sex, location) with disease. This resource is not useful for scientific research.

The complete database is available online via a habilitated provider⁴³, provided that regulatory requirements are met concerning data security and dissemination⁴⁴. The cost of access currently varies between a few hundred to a few thousand euros, depending on the complexity of the considered project. A more efficient solution is to establish collaboration with the Health Information Management department of a French university hospital. Since HIM departments access these data on a daily basis to analyse hospital activity, most of them have developed considerable expertise as concerns the data, the accessibility platform, regulatory requirements and associated limitations. This ensures methodology expertise and safety.

Profile in a nutshell

- The PMSI database was initiated during the establishment of the French Prospective Payment System (PPS). Nation-wide record production is mandatory since 1996. The database contains a record for each acute in-patient stay and represents about 25 million records per year.
- The records describe the stays in a standardized data set. This data set summarizes information about the patient and the stay, describing diagnoses and procedures and also providing additional information about specific aspects of the stay (for instance, a stay in an intensive care unit). Finally, the data set contains the Diagnosis Related Group (DRG) for the stay. Since 2001, an anonymous patient identifier enables record linkage.
- The usage of these data has risen and evolved over the years, ranging from simple auxiliary usage to studies based on a specific adjusted exploitation. The most elaborate studies are focused on the epidemiology of cancer.
- This database is available upon registration with and payment to a habilitated provider, or through collaboration with a French university hospital Health Information Management department.

Reference

1. Landais P, Boudemaghe T, Suehs C, Dedet G, Lebihan-Benjamin C. Computerized Medico-Economic Decision Making: An International Comparison. In: Venot A, Burgun A, Quantin C, editors. *Med Inform E-Health* [Internet]. Springer Paris; 2014 [cited 2015 Aug 17]. p. 191–220. Available from: http://link.springer.com/chapter/10.1007/978-2-8178-0478-1_9
2. Kervasdoué J de. Appel des candidatures des établissements hospitaliers à l'expérience de résumés de sorties du projet de médicalisation des systèmes d'information (PMSI). 1982.
3. Arrêté du 3 octobre 1985 autorisant l'informatisation des résumés de sortie standardisés dans les établissements d'hospitalisation publics et privés participant au service public hospitalier [Internet]. Mar 10, 1985. Available from: http://www.atih.sante.fr/sites/default/files/public/content/990/Arrete_1985.pdf
4. Circulaire n° 160 du 5 août 1986 relative à la diffusion de la classification des groupes homogènes de malades (G.H.M.) [Internet]. Aug 5, 1986. Available from: http://www.atih.sante.fr/sites/default/files/public/content/990/1986_Cir_V0_GHM.pdf
5. Circulaire n° 119 du 4 octobre 1985 relative à la mise en place dans les établissements hospitaliers des résumés de sortie standardisés (R.S.S.) [Internet]. Oct 4, 1985. Available from: http://www.atih.sante.fr/sites/default/files/public/content/990/Cir_4-10-85.pdf.pdf
6. Circulaire 178 du 11 mars 1987 relative à la collecte et à l'exploitation des R.S.S. Nov 3, 1987.
7. Circulaire DH/P.M.S.I. n° 303 du 24 juillet 1989 relative à la généralisation du programme de médicalisation des systèmes d'information (P.M.S.I.) et à l'organisation de l'information médicale dans les hôpitaux publics. Jul 24, 1989.
8. Circulaire 325 du 12 février 1990 relative aux modalités de mise en place des structures de gestion de l'information médicale dans les établissements hospitaliers publics et privés participant au service public. Dec 2, 1990.
9. Loi n° 91-748 du 31 juillet 1991 portant réforme hospitalière [Internet]. 91-748, SPSX9000155L Jul 31, 1991. Available from: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000720668&dateTexte=20160505>
10. Circulaire DH/DSS/95/23 du 10 mai 1995 relative à la généralisation du dispositif technique, expérimenté en Languedoc-Roussillon, de mesure de l'activité et des coûts des établissements hospitaliers sous compétence tarifaire de l'état, à l'aide des informations provenant du système d'information médicalisé. Oct 5, 1995.
11. Arrêté du 22 juillet 1996 relatif au recueil et au traitement des données d'activité médicale, visées à l'article L. 710-6 du code de la santé publique, par les établissements de santé visés à l'article L. 710-16-2 du même code et à la transmission, visée à l'article L. 710-7 du code de la santé publique, aux agences régionales de l'hospitalisation, aux organismes d'assurance maladie et à l'Etat d'informations issues de ce traitement [Internet]. Available from: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000549743&categorieLien=id>
12. Circulaire n° 106/DHOS-PMSI du 22 février 2001 relative au chaînage des séjours en établissements hospitalier. Feb 22, 2001.
13. C. Le Bihan-Benjamin PL, G.Chatelier. Linking hospital stays in the national PMSI MCO database improved between 2006 and 2009: Analysis and consequences. *J Econ Medicale*. 2012;**30**(1):17–30.
14. ATIH : Agence technique de l'information sur l'hospitalisation [Internet]. [cited 2016 May 6]. Available from: <http://www.atih.sante.fr/>

15. Quantin C, Fassa M, Coatrieux G, Riandey B, Trouessin G, Allaert FA. Chaînage de bases de données anonymisées pour les études épidémiologiques multicentriques nationales et internationales : proposition d'un algorithme cryptographique. *Rev Dépidémiologie Santé Publique*. 2009 Feb;**57**(1):33–39.
16. Trombert-Paviot B, Couris CM, Couray-Targe S, Rodrigues JM, Colin C, Schott AM. [Quality and usefulness of an anonymous unique personal identifier to link hospital stays recorded in French claims databases.]. *Rev Epidemiol Sante Publique*. 2007;**55**(3):203–211.
17. WHO | International Classification of Diseases (ICD) [Internet]. *WHO* [cited 2015 Aug 24]. Available from: <http://www.who.int/classifications/icd/en/>
18. Classification Commune des Actes Médicaux. Assurance Maladie [Internet]. [cited 2015 Aug 24]. Available from: <http://www.ameli.fr/accueil-de-la-ccam/index.php>
19. Schott A-M, Hajri T, Colin C, et al. [Usefulness of the French DRG based information system (PMSI) in the measurement of cancer activity in a multidisciplinary hospital: the Hospices Civils de Lyon]. *Bull Cancer (Paris)*. 2002 Nov;**89**(11):969–973.
20. Olive F, Gomez F, Schott A-M, et al. [Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: a longitudinal approach becomes possible]. *Rev Dépidémiologie Santé Publique*. 2011 Feb;**59**(1):53–58.
21. Couris CM, Polazzi S, Olive F, et al. Breast cancer incidence using administrative data: correction with sensitivity and specificity. *J Clin Epidemiol*. 2009 Jun;**62**(6):660–666.
22. Mitton N, Colonna M, Trombert B, et al. A Suitable Approach to Estimate Cancer Incidence in Area without Cancer Registry. *J Cancer Epidemiol*. 2011;**2011**:1–11.
23. Frachon I, Etienne Y, Jobic Y, Le Gal G, Humbert M, Leroyer C. Benfluorex and unexplained valvular heart disease: a case-control study. *PloS One*. 2010;**5**(4):e10128.
24. Connolly HM, Crary JL, McGoon MD, et al. Valvular Heart Disease Associated with Fenfluramine–Phentermine. *N Engl J Med*. 1997 Aug 28;**337**(9):581–588.
25. Fowles RE, Cloward TV, Yowell RL. Endocardial Fibrosis Associated with Fenfluramine–Phentermine. *N Engl J Med*. 1998 Apr 30;**338**(18):1316–1317.
26. Cannistra LB, Cannistra AJ. Regression of Multivalvular Regurgitation after the Cessation of Fenfluramine and Phentermine Treatment. *N Engl J Med*. 1998 Sep 10;**339**(11):771–771.
27. Boutet K, Frachon I, Jobic Y, et al. Fenfluramine-like cardiovascular side-effects of benfluorex. *Eur Respir J*. 2009 Mar;**33**(3):684–688.
28. Weill A, Païta M, Tuppin P, et al. Benfluorex and valvular heart disease: a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf*. 2010 Dec 1;**19**(12):1256–1262.
29. European Medicines Agency - News and Events - European Medicines Agency recommends withdrawal of benfluorex from the market in European Union [Internet]. [cited 2016 Aug 6]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2010/01/news_detail_000977.jsp&mid=WC0b01ac058004d5c1
30. Sparsa A, Loustaud-Ratti V, Mousset-Hovaere M, et al. [Drug-induced hypersensitivity syndrome in internal medicine: diagnostic and therapeutic traps. Eight observations]. *Rev Médecine Interne Fondée Par Société Natl Française Médecine Interne*. 2000 Dec;**21**(12):1052–1059.
31. Blatier JF, Fauconnier J. [Relationship between in-hospital mortality rate and patient volume for total hip arthroplasty in the Rhône-Alpes public hospitals]. *Rev Dépidémiologie Santé Publique*. 2001 Apr;**49**(2):173–182.

32. Kuss G, Drogue N, Michel B. [Evaluation of prescription practices for drugs charged in addition to DRG-based fees in Alsace]. *Thérapie*. 2012 Jun;**67**(3):205–211.
33. Jouanjus E, Pourcel L, Saivin S, Molinier L, Lapeyre-Mestre M. Use of multiple sources and capture-recapture method to estimate the frequency of hospitalizations related to drug abuse. *Pharmacoepidemiol Drug Saf*. 2012 Jul;**21**(7):733–741.
34. Maura G, Blotière P-O, Bouillon K, et al. Comparison of the short-term risk of bleeding and arterial thromboembolic events in nonvalvular atrial fibrillation patients newly treated with dabigatran or rivaroxaban versus vitamin K antagonists: a French nationwide propensity-matched cohort study. *Circulation*. 2015 Sep 29;**132**(13):1252–1260.
35. Epinette J-A, Lafuma A, Robert J, Doz M. Cost-effectiveness model comparing dual-mobility to fixed-bearing designs for total hip replacement in France. *Orthop Traumatol Surg Res OTSR*. 2016 Apr;**102**(2):143–148.
36. Misset B, Nakache D, Vesin A, et al. Reliability of diagnostic coding in intensive care patients. *Crit Care Lond Engl*. 2008;**12**(4):R95.
37. Remontet L, Mitton N, Couris CM, et al. Is it possible to estimate the incidence of breast cancer from medico-administrative databases? *Eur J Epidemiol*. 2008;**23**(10):681–688.
38. Coureau G, Baldi I, Savès M, et al. [Performance evaluation of hospital claims database for the identification of incident central nervous system tumors compared with a cancer registry in Gironde, France, 2004]. *Rev Dépidémiologie Santé Publique*. 2012 Aug;**60**(4):295–304.
39. Bernier M-O, Mezzarobba M, Maupu E, et al. [Role of French hospital claims databases from care units in epidemiological studies: the example of the ‘Cohorte Enfant Scanner’ study]. *Rev Dépidémiologie Santé Publique*. 2012 Oct;**60**(5):363–370.
40. Quantin C, Benzenine E, Hägi M, et al. [Evaluation of medical information systems as a mean of identification of incident cases of colorectal cancer]. *Santé Publique Vandoeuivre-Lès-Nancy Fr*. 2014 Feb;**26**(1):55–63.
41. Le Manach Y, Collins G, Bhandari M, et al. Outcomes after hip fracture surgery compared with elective total hip replacement. *JAMA*. 2015 Sep 15;**314**(11):1159–1166.
42. Scan Santé - Stats ATIH [Internet]. [cited 2016 Jul 25]. Available from: <http://www.scansante.fr/>
43. CASD - Centre d'accès sécurisé aux données [Internet]. [cited 2016 Jul 27]. Available from: <https://casd.eu/>
44. Commande de bases | Publication ATIH [Internet]. [cited 2016 Feb 2]. Available from: <http://atih.sante.fr/bases-de-donnees/commande-de-bases>

IX. Publication – Identifying Life-Threatening Admissions for Drug Dependence or Abuse (ILIADDA): Derivation and Validation of a Model

Article en cours de soumission

Identifying Life-Threatening Admissions for Drug Dependence or Abuse (ILIADDA): Derivation and Validation of a Model.

Tri-Long Nguyen, Pharm.D. Ph.D.^{1*}, Thierry Boudemaghe, M.D. M.Sc.², Géraldine Leguelinel-Blache, Pharm.D. Ph.D.³, Céline Eiden, Pharm.D. Ph.D.⁴, Jean-Marie Kinowski, Pharm.D. Ph.D.⁵, Yannick Le Manach, M.D. Ph.D.⁶, Hélène Peyrière, Pharm.D. Ph.D.⁷, Paul Landais, M.D. Ph.D.⁸

^{1*} Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University of Montpellier, France; Laboratory of Clinical Pharmacy, Faculty of Pharmacy, University of Montpellier, France; Department of Pharmacy, Nîmes University Hospital, France.

² Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University of Montpellier, France; Department of Biostatistics, Epidemiology, Public Health and Medical Informatics, Nîmes University Hospital, France.

³ Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University of Montpellier, France; Department of Pharmacy, Nîmes University Hospital, France.

⁴ Department of Medical Pharmacology and Toxicology, Addictovigilance Centre, Montpellier University Hospital, Montpellier University, France.

⁵ Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University of Montpellier, France; Department of Pharmacy, Nîmes University Hospital, France.

⁶ Assistant Professor. Departments of Anesthesia & Clinical Epidemiology and Biostatistics, Michael DeGroote School of Medicine, Faculty of Health Sciences, McMaster University and the Perioperative Research Group, Population Health Research Institute, Hamilton, Ontario, Canada.

⁷ Assistant Professor. Laboratory of Clinical Pharmacy, Faculty of Pharmacy, University of Montpellier, France; Department of Medical Pharmacology and Toxicology, Addictovigilance Centre, Montpellier University Hospital, Montpellier University, France.

⁸ Professor and Chairman. Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University of Montpellier, France; Department of Biostatistics, Epidemiology, Public Health and Medical Informatics, Nîmes University Hospital, France.

Short title: Identifying Life-Threatening Admissions for Drug Dependence or Abuse.

Corresponding author:

Tri-Long NGUYEN, Pharm.D. Ph.D.

Department of of Pharmacy, Nîmes University Hospital, Nîmes, France; Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, University Institute of Clinical Research, Montpellier University, Montpellier, France; Laboratory of Clinical Pharmacy, Faculty of Pharmacy, Montpellier University, Montpellier, France.

Address: 641 avenue du Doyen Giraud, Institut Universitaire de Recherche Clinique, EA2415, Faculté de Médecine, Université de Montpellier, 34093 Montpellier Cedex 5, France.

Email: longbacon.nguyen@gmail.com; phone: (+33)6.82.69.55.69

Data sharing: No additional data available.

Word count: Abstract: 200 words; text ('Methods' excluded): 1,964 words.

Keywords: Drug abuse; drug dependence; intensive care unit; predictive modeling.

ABSTRACT

Given that drug abuse and dependence are common reasons for hospitalization, we aimed to derive and validate a model allowing an early identification of life-threatening hospital admissions for drug dependence or abuse. Using the French National Hospital Discharge Data Base, we extracted 66,101 acute inpatient stays for substance abuse, dependence, mental disorders or poisoning caused by medicines or illicit drugs intake, recorded between January 1st, 2009 and December 31st, 2014. We split our study cohort at the center level to create a derivation cohort and a validation cohort. We developed a multivariate logistic model including patient's age, sex, entrance mode and diagnosis as predictors of a composite primary outcome of in-hospital death or ICU admission. A total of 2,747 (4.2%) patients died or were admitted to ICU. The risk of death or ICU admission was mainly associated with the consumption of opioids, followed by cocaine and other narcotics. Particularly, methadone poisoning was associated with a substantial risk (OR: 35.70, 95% CI [26.94–47.32], $P < 0.001$). In the validation cohort, our model achieved good predictive properties in terms of calibration and discrimination (c-statistic: 0.847). This allows an accurate identification of life-threatening admissions in drug users to support an early and appropriate management.

INTRODUCTION

Worldwide, illicit drug use and dependence are important contributors to the global burden of disease, accounting for 20 million disability-adjusted life years¹. This burden is mainly associated with opioid dependence and increases in highest incomes countries¹. In the United States, drug overdose deaths increased significantly within the last years². A wide variety of healthcare, social and educational settings are involved in the identification and the treatment of substance abuse- and dependence-related disorders³. Admissions to acute care hospitals for drug abuse or dependence are common and likely associated with complications.

Recently, it has been suggested that drug users were at higher risk for intensive care unit (ICU) admission than the general population⁴. Life-threatening hospital admissions need an early and appropriate management, and sometimes require supportive care⁵⁻⁷. In drug-overdosed patients, the immediate need for ICU monitoring can be assessed with tools, such as the Glasgow coma score⁸. However, the latency of certain toxic syndromes can bias the initial evaluation and thus lead to overestimating the prognosis. Particularly, a secondary deterioration could be observed in patients initially admitted for behavioral disorders due to an acute intoxication, if no appropriate and early management is initiated. Accurately identifying at an early stage the potential life-threatening cases at admission is therefore of utmost importance. In particular, it could result in the respiratory and hemodynamic stabilization of the patient before ICU admission⁶. We propose to provide early prognostic information based on the patients' characteristics at hospital entrance, including the suspected responsible agent. To the best of our knowledge, such an approach has not been previously described in the medical literature for identifying life-threatening cases in drug users admitted to acute care hospitals.

In the present study, we report both the derivation and the validation of a model allowing an early identification of life-threatening admissions for drug dependence or abuse (ILIADDA).

METHODS

Data source and study population

In France, the National Hospital Discharge Data Base (NHDDDB) records complete anonymized information on all patients, in all French private and public hospitals. It provides discharge summaries of acute hospitalizations, which are grouped into Diagnosis-Related Groups (DRGs) ⁹. For each DRG, a tariff is annually defined at the national level based on average costs. The study was conducted according to the approval given by the *Agence pour le Traitement de l'Information Médicale* (ATIH). Authorization was also obtained from the *Commission Nationale de l'Informatique et des Libertés* (agreement No. 1375062). The data provided were de-identified.

We extracted all of the discharge summaries, recorded within the NHDDDB from January 1st, 2009 to December 31st, 2014, reporting a DRG for substance abuse, dependence, mental disorders or poisoning caused by medicines or illicit drugs intake (see *Supplemental Digital Content* for the DRG codes). Admissions related to tobacco or alcohol disorders were not included.

Outcomes and predictors

We defined the primary endpoint as a composite outcome of ICU admission or in-hospital death.

In order to allow the clinician to predict at hospital admission the patient's risk for the occurrence of the primary outcome, we considered the potential predictors as those available at entrance (time origin): age, sex, entrance by transfer and the diagnosis that motivated the hospitalization. This diagnosis at admission was recorded according to the International Classification of Diseases, 10th Revision (ICD-10) ¹⁰.

Statistical analysis

Means (standard deviations, SD) and frequencies (percentages) were reported for continuous and categorical variables, respectively. For descriptive purposes, we mapped the geographical distribution of the acute admissions related to drug abuse or dependence. The number of admissions by location was standardized on the number of inhabitants reported by the *Institut National de la Statistique et des Études Économiques* ¹¹. The map was carried out with © Géoclip 2015 – IGN GéoFla.

For predictive purposes, it is recognized that implementing a prognostic model in clinical practice routine must be preceded by three steps: the model derivation, the model validation and the clinical impact assessment ¹²⁻²¹. In this study, we reported both the derivation and the validation of a model, according to the TRIPOD statement ^{12, 13}. Given that the predictive performance of a model is likely to be optimistic relatively to the data used for the model derivation, it must be evaluated on an independent set of patients. To this end, we split our study cohort into a derivation cohort and a validation cohort. There is a clear consensus as to the inefficiency of randomly splitting a cohort at the patient level, because it creates two very similar cohorts ^{12, 13, 22-24}. Splitting by time (temporal or “narrow” validation) or by location (geographic or “broad” validation) is a more reliable approach ^{12, 13}. To conduct a broad validation, we split the entire cohort at the center level (ratio 2:1 for derivation and validation, respectively). This allowed us to assess the performance of our model in centers, within which practices and measurements were likely to differ from those of the derivation cohort.

In the derivation cohort, we performed a multivariate logistic regression adjusting the risk of death or ICU admission for several predictors. Due the large sample size, selecting predictors according to the P-value would have led to a non-parsimonious model. Therefore, we limited the set of binary predictors to those for which 10 events (or more) occurred. We used a restricted cubic spline to handle the non-linearity of age effect. To provide a parsimonious and clinically usable final model, we removed the predictors associated with less than 0.25% of the model deviance, without damaging the predictive performance. In all analyses, multiple admissions for the same patient were considered as independent stays. Indeed, including the number of previous admissions per patient did not improve the model performance, while taking into account the database left-truncation reduced dramatically the sample size. In both the derivation cohort and the validation cohort, we reported the model discrimination (c-statistic, that is, the area under the curve plotting *sensitivity* against $1 - \textit{specificity}$), accuracy (Brier score) and calibration (local weighted regression curve, slope and intercept). All statistical analyses were carried out in R software (version 3.3).

Patient and funding organization involvement

No patients were involved in setting the research question; nor were they involved in the design of the study or the outcome measures. No funding organizations were involved in the study.

RESULTS

Acute hospitalization in drug users

From January 2009 to December 2014, we reported a total of 66,101 acute admissions for drug abuse or dependence in 956 centers (*Figure 1*). During this period, there was a substantial increase in the number of hospital admissions (from 8,834 admissions in 2009 to 14,118 admissions in 2014, +59.8%), which was associated with a widespread geographical distribution (*Figure 2*). The mean age was 37.4 years (SD: 16.8) and 39,325 patients (59.5%) were male. Nine hundred ninety-five (1.5%) entered by transfer from another structure. The mean length of stay was 3.6 days (SD: 5.7). A large majority of these admissions (61,875, 93.4%) occurred in public hospitals with a mean tariff of € 1,704 (SD: 1,895).

As summarized in *Table 1*, the main reason for admission was drug rehabilitation (19,319 admissions, 29.2%), including rehabilitation after drug abuse, drug detoxification, drug replacement therapy, drug withdrawal and treatment of withdrawal symptoms. Others were mostly related to the consumption of opioids (15,951 admissions, 24.1%), sedative or hypnotics (8,973 admissions, 13.6%), cannabinoids (7,445 admissions, 11.3%) and cocaine (2,709 admissions, 4.1%).

Risk of ICU admission or in-hospital death

In the entire cohort, the primary outcome occurred in 2,747 stays (4.2%), with 2,602 admissions to ICU and 211 in-hospital deaths. Of these deaths, 66 occurred in ICU (mortality rate in ICU: 2.5%). The mean SAPS II score at admission to ICU was equal to 35.0 (SD: 17.4).

For risk modeling, we split the study cohort at the hospital center level: 41,987 admissions in 637 centers were used for the model derivation, and 24,114 admissions to 319 centers for the model validation. A total of 1,633 (3.9%) and 1,114 (4.6%) deaths or ICU admissions occurred in the derivation cohort and in the validation cohort, respectively.

We reported a multivariate logistic model including 17 predictors (*Table 2*). Poisoning and acute intoxication by opioids were associated with a substantial increased risk of death or ICU admission: methadone poisoning OR: 35.70 (95% CI [26.94 – 47.32], $P < 0.001$), heroin poisoning OR: 22.72 (95% CI [15.84 – 32.61], $P < 0.001$), opium poisoning OR: 18.26 (95% CI [10.45 – 31.89], $P < 0.001$), poisoning by other opioids OR: 20.17 (95% CI [15.69 – 25.92], $P < 0.001$) and behavioral disorders due to an acute opioids intoxication OR: 17.61 (95% CI [13.49 – 22.97], $P < 0.001$). Poisoning and acute intoxication by other narcotics than opioids also increased the outcome risk: cocaine poisoning OR: 12.70 (95% CI [8.09 – 19.93], $P < 0.001$), behavioral disorders due to an acute cocaine intoxication OR: 3.12 (95% CI [1.69 – 5.79], $P < 0.001$), poisoning by other synthetic narcotics OR: 8.99 (95% CI [6.26 – 12.91], $P < 0.001$) and poisoning by other and unspecified narcotics OR: 18.97 (95% CI [13.26 – 27.14], $P < 0.001$). Other drugs had a lower, but still substantial effect on the risk of death or ICU admission (*Table 2*). In contrast, rehabilitation of drug users was associated with a decreased risk of occurrence of the primary outcome, OR: 0.18 (95% CI [0.11 – 0.29], $P < 0.001$), as well as the female sex, OR: 0.86 (95% CI [0.77 – 0.96], $P < 0.001$). The age effect followed an “n-shaped” curve with a maximal risk at 44 years old, OR: 7.98 (*Supplemental Digital Content*).

In the derivation cohort, our model achieved good predictive performances: c-statistic 0.851 (95% CI [0.844 – 0.859]) and Brier score 0.034 (95% CI [0.033 – 0.036]). This was confirmed in the validation cohort: c-statistic 0.847 (95% CI [0.838 – 0.855]) and Brier score 0.040 (95% CI [0.038 – 0.043]). The overall calibration slope in the validation cohort (0.950) depicted a slight over-fitting issue of the model development. Using a local weighted regression curve, we compared the predicted probabilities with the observed outcome frequencies (*Figure 3*). In the validation cohort, we reported a small overestimation of the predicted probabilities beyond 0.2, which concerned only a few patients.

DISCUSSION

We report the derivation and the validation of a model, which demonstrates an accurate prediction of death or ICU admission in hospitalized drug users.

Drug abuse or dependence is a common reason for admission into acute care hospitals. In a large French nationwide cohort, we found that 4.2% of hospitalizations led to death or ICU admission. This is concordant with the 3% of empoisoned patients requiring critical care in the United States⁷. Although these life-threatening cases are limited, they require an early and appropriate management^{5, 6, 25}. In response to this issue, we provide a model allowing their identification. Predictors include the early diagnosis, which is commonly revealed by specific clinical syndromes (“toxidromes”) suggesting particular substances. Life-threatening cases typically correspond to male adults within the range of 30-50 years old and involving an acute intoxication or poisoning by opioids. A rapid identification is essential, as patient’s respiratory and hemodynamic stabilization can be achieved before ICU admission⁶. As previously reported in other Western countries^{26, 27}, methadone poisoning corresponds to the greatest risk, followed by the other opioids. In France, the nationwide DRAMES study (*Décès en Relation avec l’Abus de Médicaments et de Substances – Death Related to Medication and Substance Abuse*) identifies, for several years, methadone as the first drug responsible for death before heroin and cocaine^{28, 29}. This high risk associated to methadone poisoning might be related to its long duration of action. Hall and colleagues showed that drug overdose deaths were mostly associated with diversion of opioid analgesics, and particularly methadone²⁷. To address the problem of prescription drug abuse, the American College of Physicians has recently published a position paper that provides guidance to prescribers and policymakers³⁰. According to recent American studies, there has been a trendy downturn in opioid analgesics abuse within the decade, which was associated with an increased shift to heroin abuse in a more widespread geographical distribution³¹⁻³⁴. In term of prognosis, our data suggest that the risk of death or ICU admission associated with heroin poisoning is lower than with methadone poisoning, but higher than other opioids.

If opioids appear as primarily responsible for life-threatening admissions, cocaine and other narcotics come next. The risk of death or ICU admission related to cocaine poisoning is about twice lower than the risk associated with methadone, but is still substantial³⁵. Other drugs involved are hallucinogens, sedatives or hypnotics, and cannabis or derivatives. In contrast to acute poisoning or intoxication, drug rehabilitation is associated with a reduced risk of outcome. Further studies are needed to demonstrate whether chronic care management is efficient³⁶.

To the best of our knowledge, we provide the first model allowing an early identification of life-threatening cases in drug users admitted to acute care hospitals. In response to clinical issues, this model includes only predictors that are readily available at hospital admission. Though we recognize comorbidities and biological markers are potential predictors, we did not take them into account in our model since they require specific investigations that are not systematically conducted in every patient at entrance. Had we considered those potential predictors would have led to limiting the usability of our model to patients whose comorbidities and biological markers are recorded at admission. Without including those variables, our model still demonstrates good predictive properties in terms of discrimination and calibration. In contrast with diagnosis scores (e.g., Glasgow coma score), which assess the immediate need for intensive care and are likely to vary according to the latency of the toxidrome, our model provides accurate prognostic information at hospital entrance. In routine practice, this might be a useful tool supporting the clinical decision-making to improve the early management of critically ill patients. One of the strengths of our study is using a complete nationwide database. The large sample size enabled a broad model validation. It assessed the performances of our model on centers that were not used for the model derivation. We reported all of our analyses with transparency.

Our study has to be considered in the light of its limitations. First, the lack of sensitivity of the French medico-administrative database has been discussed for identifying drug-dependent patients³⁷. However, this has been advocated in the context of a routine detection of drug users. We agree that the French NHDDDB should not be used for that purpose, as the population admitted to acute care hospitals does not reflect the overall drug-addicted population, who is likely to benefit from a wide variety of non-healthcare, social and educational settings. However, it did not concern our study, since we focused on the most severe acute admissions, which were specifically motivated by substance abuse, dependence, mental disorders or poisoning caused by medicines or illicit drug intake. About the specificity of our database, the hypothesis of false-positive cases should be dismissed, as the French NHDDDB is primarily used for an activity-based payment and is unlikely inclined to adulteration of the diagnoses³⁷. As the French NHDDDB includes diagnoses that are finally recorded at the end of hospitalization, one could question whether these match with

the diagnoses that are firstly established at admission. Though medico-administrative databases can be regarded as relevant tools for describing admissions specifically related to illicit drugs consumption³⁸, their potential use for clinical toxicology is still limited^{37, 39, 40}. In certain situations, the information on drugs responsible for admission were limited to the drug classes stated by the ICD-10, which did not allowed us to explore the risk related to some specific drugs, in particular, prescription opioids (e.g. oxycodone). Additionally, predictors such as individual socio-demographic characteristics were not available in our database. In spite of these limits, our model demonstrated good predictive performances. Nevertheless, measuring the effect of its implementation on the decision-making process is still required. A further study is therefore needed to assess its clinical impact, before it be implemented in routine clinical practice.

CONCLUSIONS

We report a simple model, which predicts the risk of death or ICU admission in drug users with good discriminative and calibration properties. This model, which includes patients' characteristics at hospital admission might be useful in clinical practice for early identification of life-threatening cases.

REFERENCES

1. Degenhardt L, Whiteford HA, Ferrari AJ, et al. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382:1564-1574.
2. Rudd RA, Aleshire N, Zibbell JE, Gladden RM. Increases in Drug and Opioid Overdose Deaths — United States, 2000–2014. *Centers for Disease Control and Prevention — Morbidity and Mortality Weekly Report*. 2016;64:1378-1382.
3. McLellan AT, Meyers K. Contemporary addiction treatment: a review of systems problems for adults and adolescents. *Biological psychiatry*. 2004;56:764-770.
4. Suchyta MR, Beck CJ, Key CW, Jephson A, Hopkins RO. Substance dependence and psychiatric disorders are related to outcomes in a mixed ICU population. *Intensive care medicine*. 2008;34:2264-2267.
5. Alapat PM, Zimmerman JL. Toxicology in the critical care unit. *Chest*. 2008;133:1006-1013.
6. Zimmerman JL. Poisonings and overdoses in the intensive care unit: general and specific management issues. *Critical care medicine*. 2003;31:2794-2801.
7. Mokhlesi B, Leiken JB, Murray P, Corbridge TC. Adult toxicology in critical care: part I: general approach to the intoxicated patient. *Chest*. 2003;123:577-592.
8. Hamad AE, Al-Ghadban A, Carvounis CP, Soliman E, Coritsidis GN. Predicting the need for medical intensive care monitoring in drug-overdosed patients. *J Intensive Care Med*. 2000;15:321–328.
9. Manuel des Groupes Homogènes de Malades. Ministère des Affaires Sociales et de la Santé. http://social-sante.gouv.fr/fichiers/bos/2014/sts_20140005_0003_p000.pdf.
10. Organisation WH. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). 1992.
11. Institut National de la Statistique et des Études Économiques. Recensement de la population, 2012. <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/2012/donnees-detaillees-recensement-2012.htm>.
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj*. 2015;350:g7594.
13. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162:W1-73.
14. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *Bmj*. 2013;346:e5595.
15. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj*. 2013;346:e5793.
16. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS medicine*. 2013;10:e1001380.
17. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS medicine*. 2013;10:e1001381.
18. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *Bmj*. 2009;338:b605.

19. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Bmj*. 2009;338:b606.
20. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *Bmj*. 2009;338:b375.
21. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *Bmj*. 2009;338:b604.
22. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Statistical methods in medical research*. 2014.
23. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*. 2003;56:441-447.
24. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
25. Mokhlesi B, Garimella PS, Joffe A, Velho V. Street drug abuse leading to critical illness. *Intensive care medicine*. 2004;30:1526-1536.
26. O'Dowd A. Deaths from drug poisoning in English and Welsh men reach five year peak. *Bmj*. 2008;337:a1521.
27. Hall AJ, Logan JE, Toblin RL, et al. Patterns of abuse among unintentional pharmaceutical overdose fatalities. *Jama*. 2008;300:2613-2620.
28. Jouanjus E, Gibaja V, Kahn JP, Haramburu F, Daveluy A. Signal identification in addictovigilance: the functioning of the French system. *Therapie*. 2015;70:123-131.
29. Richard N, Arditti J, Pépin G, Mallaret M, Castot A. Drames (Décès en relation avec l'abus de médicaments et de substances) : outil d'identification des décès liés à la drogue en France. *BEH*.40-41.
30. Kirschner N, Ginsburg J, Sulmasy LS, Health, Public Policy Committee of the American College of P. Prescription drug abuse: executive summary of a policy position paper from the American College of Physicians. *Annals of internal medicine*. 2014;160:198.
31. Dart RC, Surratt HL, Cicero TJ, et al. Trends in opioid analgesic abuse and mortality in the United States. *The New England journal of medicine*. 2015;372:241-248.
32. Cicero TJ, Ellis MS, Harney J. Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States. *The New England journal of medicine*. 2015;373:1789-1790.
33. Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. The changing face of heroin use in the United States: a retrospective analysis of the past 50 years. *JAMA psychiatry*. 2014;71:821-826.
34. McCarthy M. Illicit drug use in the US holds steady, but heroin use is on rise. *Bmj*. 2013;347:f5544.
35. Single E, Rehm J, Robson L, Truong MV. The relative risks and etiologic fractions of different causes of death and disease attributable to alcohol, tobacco and illicit drug use in Canada. *CMAJ*. 2000;162:1669-1675.
36. Saitz R, Cheng DM, Winter M, et al. Chronic care management for dependence on alcohol and other drugs: the AHEAD randomized trial. *Jama*. 2013;310:1156-1167.
37. Jouanjus E, Pourcel L, Saivin S, Molinier L, Lapeyre-Mestre M. Use of multiple sources and capture-recapture method to estimate the frequency of hospitalizations related to drug abuse. *Pharmacoepidemiology and drug safety*. 2012;21:733-741.

38. Chisolm DJ, Kelleher KJ. Admission to acute care hospitals for adolescent substance abuse: a national descriptive analysis. *Substance abuse treatment, prevention, and policy*. 2006;1:17.
39. Daveluy A, Miremont-Salame G, Kostrzewa A, et al. Identification of abuse and dependence cases through a hospital database. *Pharmacoepidemiology and drug safety*. 2012;21:1344-1349.
40. Jouanjus E, Leymarie F, Tubery M, Lapeyre-Mestre M. Cannabis-related hospitalizations: unexpected serious events identified through hospital databases. *British journal of clinical pharmacology*. 2011;71:758-765.

AUTHOR CONTRIBUTIONS

- **Study concept and design:** T.L.N., T.B., H.P., P.L.
- **Analysis and interpretation:** T.L.N., T.B., H.P., P.L.
- **Drafting of the manuscript:** T.L.N., T.B., H.P., P.L.
- **Critical revision of the manuscript for important intellectual content:** T.L.N., T.B., G.L.B., C.E., J.M.K., Y.L.M., H.P., P.L.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

FUNDING

None.

FIGURE LEGENDS

Figure 1. Admissions flowchart.

Figure 2. Admission to acute care hospitals for drug abuse or dependence in France, according to the National Hospital Discharge Data Base (2009 – 2014).

Figure 3. Calibration plot comparing predicted *versus* observed deaths or ICU admissions in the validation sample.

TABLES

Table 1. Primary diagnoses at admission.

Diagnosis (ICD-10 code)	No. (%)
Drug rehabilitation (Z503)	19,319 (29.2%)
<i><u>Opioids</u></i>	<i><u>15,951 (24.1%)</u></i>
Poisoning by opium (T400)	214 (0.3%)
Poisoning by heroin (T401)	548 (0.8%)
Poisoning by other opioids (T402)	6,953 (10.5%)
Poisoning by methadone (T403)	1,358 (2.1%)
Mental or behavioral disorders due to use of opioids (F11)	6,871 (10.4%)
Finding of opiate drug in blood (R781)	7 (0.0%)
Mental or behavioral disorders due to the use of sedatives or hypnotics (F13)	8,973 (13.6%)
<i><u>Cannabinoids</u></i>	<i><u>7,445 (11.3%)</u></i>
Poisoning by cannabis (derivatives) (T407)	1,694 (2.6%)
Mental or behavioral disorders due to the use of cannabinoids (F12)	5,751 (8.7%)
<i><u>Cocaine</u></i>	<i><u>2,709 (4.1%)</u></i>
Poisoning by cocaine (T405)	553 (0.8%)
Mental or behavioral disorders due to use of cocaine (F14)	2,147 (3.2%)
Finding of cocaine in blood (R782)	9 (0.0%)
Poisoning by other synthetic narcotics (T404)	1,519 (2.3%)
Poisoning by other and unspecified narcotics (T406)	1,072 (1.6%)
Poisoning by lysergide [LSD] (T408)	110 (0.2%)
Poisoning by other and unspecified psychodysleptics (T409)	434 (0.7%)
Mental or behavioral disorders due to the use of hallucinogens (F15)	393 (0.6%)
Mental or behavioral disorders due to the use of other stimulants (F16)	415 (0.6%)
Mental or behavioral disorders due to the use of volatile solvents (F18)	176 (0.3%)
Mental or behavioral disorders due to multiple drug use and use of psychoactive substances (F19)	5,785 (8.8%)
Others	1,800 (2.7%)

Table 2. Model for identifying life-threatening admissions for drug dependence or abuse (ILIADDA). A restricted cubic spline with three knots was used to describe the effects of age (knots at 19, 36 and 59 years).

	Regression coefficient	Odds ratio [95% CI]	P-value
<i>Constant</i>	-6.615		
Age (y)	0.060	1.06 [1.05 - 1.07]	<0.001
(Age) ³	-0.061	0.94 [0.93 - 0.95]	<0.001
Female sex	-0.155	0.86 [0.77 - 0.96]	0.006
Transfer at entrance	0.957	2.60 [1.99 - 3.41]	<0.001
<i>Reason for hospital admission (ICD-10 code)</i>			
Drug rehabilitation (Z503)	-1.712	0.18 [0.11 - 0.29]	<0.001
<i>Opioids</i>			
Poisoning by methadone (T403)	3.575	35.70 [26.94 - 47.32]	<0.001
Poisoning by heroin (T401)	3.123	22.72 [15.84 - 32.61]	<0.001
Poisoning by opium (T400)	2.905	18.26 [10.45 - 31.89]	<0.001
Poisoning by other opioids (T402)	3.004	20.17 [15.69 - 25.92]	<0.001
Mental and behavioral disorders due to use of opioids, acute intoxication (F110)	2.868	17.61 [13.49 - 22.97]	<0.001
<i>Cocaine and other narcotics</i>			
Poisoning by cocaine (T405)	2.542	12.70 [8.09 - 19.93]	<0.001
Mental and behavioral disorders due to use of cocaine, acute intoxication (F140)	1.139	3.12 [1.69 - 5.79]	<0.001
Poisoning by other synthetic narcotics (T404)	2.196	8.99 [6.26 - 12.91]	<0.001
Poisoning by other and unspecified narcotics (T406)	2.943	18.97 [13.26 - 27.14]	<0.001
<i>Others</i>			
Poisoning by cannabis (derivatives) (T407)	1.207	3.34 [1.93 - 5.78]	<0.001
Poisoning by other and unspecified psychodysleptics [hallucinogens] (T409)	2.105	8.21 [4.71 - 14.30]	<0.001
Mental and behavioral disorders due to use of sedatives or hypnotics, acute intoxication (F130)	1.817	6.16 [4.73 - 8.02]	<0.001
Mental and behavioral disorders due to multiple drug use and use of psychoactive substances, acute intoxication (F190)	1.908	6.74 [4.89 - 9.28]	<0.001