



**HAL**  
open science

# Contribution à la modélisation des durées de séjour du CHU de Grenoble

Cécile Delhumeau

► **To cite this version:**

Cécile Delhumeau. Contribution à la modélisation des durées de séjour du CHU de Grenoble. Autre [q-bio.OT]. Université Joseph-Fourier - Grenoble I, 2002. Français. NNT: . tel-00004392

**HAL Id: tel-00004392**

**<https://theses.hal.science/tel-00004392>**

Submitted on 29 Jan 2004

**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 DE L'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I

**Spécialité :** Génie Biologique et Médical

Présentée et soutenue publiquement

par

Cécile DELHUMEAU

Le 6 décembre 2002

---

---

## CONTRIBUTION À LA MODÉLISATION DES DURÉES DE SÉJOUR DU CHU DE GRENOBLE

---

---

### Composition du jury :

M. le Professeur P. CINQUIN,	Président
M. le Professeur R. ÉCOCHARD,	Rapporteur
M. le Professeur B. FALISSARD,	Rapporteur
M. le Professeur M. HOMMEL,	Examineur
M. le Docteur J. FAUCONNIER,	Invité

**Directeur de thèse :** M. le Professeur J. DEMONGEOT

Thèse préparée au sein du laboratoire TIMC de l'Université Joseph Fourier.

À Genest,  
à mes parents,  
à mes frères,

## Remerciements

En premier lieu, je tiens à remercier vivement Monsieur Jacques Demongeot, Professeur de l'Université Joseph Fourier de Grenoble et Directeur du laboratoire TIMC, sans qui ce travail n'aurait pu aboutir. Merci à lui de m'avoir accueillie au sein de son laboratoire et de m'avoir permis de découvrir l'univers de la recherche. Merci aussi pour sa disponibilité (et je sais que ça n'était pas toujours facile ...), la qualité de son encadrement, ses remarques pertinentes et toujours amicales dans l'analyse et la rédaction de ce manuscrit.

J'exprime ma gratitude à Monsieur René Écochard, Professeur de l'Université Claude Bernard de Lyon et à Monsieur Bruno Falissard, Professeur de l'Université Paris XI, d'avoir rapporté mon manuscrit. Je leur suis extrêmement reconnaissante de l'intérêt et de l'enthousiasme qu'ils ont portés à ce travail.

Je remercie Philippe Cinquin, Professeur de l'Université Joseph Fourier de Grenoble, d'avoir accepté de présider mon jury de thèse.

Je tiens à remercier aussi les autres membres de mon jury de thèse. Monsieur Hommel, Professeur de l'Université Joseph Fourier de Grenoble, pour les discussions enrichissantes que nous avons eu concernant l'accident vasculaire cérébral. Le Docteur Fauconnier, responsable de la division PMSI au CHU de Grenoble, pour les précieux conseils et les idées concernant l'analyse des données issues du PMSI qu'il m'a fourni tout au long de ce travail.

À Monsieur Morabia, Professeur de l'Université de Genève et Directeur de la Division d'Épidémiologie Clinique (DEC) du CHU de Genève pour m'avoir embauché depuis 4 ans. Il m'a permis d'acquérir une expérience supplémentaire dans le domaine de l'épidémiologie. Je le remercie de sa confiance et de m'avoir accordé un emploi du temps flexible durant toute cette période pour réaliser et soutenir ce travail de recherche à Grenoble. J'espère que notre collaboration durera encore longtemps... Je remercie aussi tous mes collègues de la DEC pour leur soutien.

Je tiens également à remercier Bernard Espiau et Claude Lemaréchal, directeurs de recherche à l'INRIA Grenoble, pour la confiance qu'ils m'ont accordée sur la fin de ce travail.

Merci au projet SARDE de m'avoir hébergée, plus particulièrement à Daniel, Sara et Christophe.

Merci à Isabelle et à Chantal pour leur aide précieuse concernant la recherche bibliographique.

Merci enfin à tous ceux qui ont été complices des bons moments, heureusement les plus nombreux ..., mais aussi qui ont supporté mes crises d'angoisse, de découragement, et autres manifestations classiques et cycliques de l'état de stress (qu'il soit justifié ou non) dans lequel se trouvent, un jour ou l'autre, les thésards, toutes disciplines confondues. Je remercie donc tous mes amis, plus particulièrement Florence, Sepanta, Nathalie, Jean-Baptiste, Edmond, Stéphanie, Émilie, ..., mes correspondants "mailistiques" (trop nombreux pour les citer tous sans en oublier ...), les buveurs de café et/ou thé de l'INRIA, mes très patients voisins de bureau au cours de ces cinq années de dur labeur : Daniela, Isabelle, Myriam, Léonard, Mareck et Philippe. Aux anciens et nouveaux membres du projet IS2, j'ai rencontré des personnes formidables. J'ai peur d'en oublier. Je sais qu'elles se reconnaîtront. Elles ont contribué à ce que mon travail aboutisse, indépendamment de leur fonction, quelquefois juste grâce à leur sourire et leur jovialité.

Aux coureurs, randonneurs et skieurs "INRIAlpins" qui m'ont permis de me changer les idées au cours de cette période ...

Je tiens aussi à remercier tous ceux qui se sont déplacés spécialement pour l'occasion.

Que mes parents voient, à travers ce document, la concrétisation de tous leurs efforts.

À mes frères, pour leur soutien dans les passages difficiles ...

À Genest, pour le temps qu'il a consacré aux nombreuses relectures de ce document au lieu d'aller "promener le chien" en montagne sous le soleil... Il a toujours cru en l'aboutissement de ce travail, même dans les moments les plus difficiles, je le remercie de son amour, sa confiance et son soutien. Qu'il imagine, à travers la réalisation de ce manuscrit, la conquête d'un sommet qui lui tient à cœur, certes difficile à atteindre, mais réussie après une ascension éprouvante...

# Table des matières

<b>Introduction</b>	<b>19</b>
<b>1 Fonctionnement budgétaire d'un centre hospitalier et durées de séjour</b>	<b>25</b>
1.1 Mise en place du Programme de Médicalisation des Systèmes d'Information en France . . . . .	25
1.1.1 Contexte économique au début des années 80 . . . . .	25
1.1.2 Naissance du Programme de Médicalisation des Systèmes d'Information	26
1.2 Le Programme de Médicalisation des Systèmes d'Information (PMSI) . . . . .	28
1.2.1 Objectifs du PMSI . . . . .	28
1.2.2 Fonctionnement du PMSI . . . . .	29
1.3 Le CHU de Grenoble . . . . .	35
1.3.1 Le CHU de Grenoble en quelques chiffres . . . . .	35
1.3.2 Cas du CHU de Grenoble en 1996 . . . . .	36
1.4 Une variable pertinente du PMSI : la durée de séjour . . . . .	37
1.4.1 Notes bibliographiques liées aux intérêts méthodologiques de l'analyse des durées de séjour . . . . .	37
1.4.2 Notes bibliographiques liées aux raisons médico-économiques de l'analyse des durées de séjour . . . . .	41
1.4.3 Synthèse . . . . .	46
<b>2 Population étudiée et listes de référence</b>	<b>47</b>
2.1 Présentation des données . . . . .	47
2.1.1 Données issues de fichiers PMSI . . . . .	47
2.1.2 Variables d'intérêt pour construire le fichier des profils . . . . .	48
2.1.3 Critères d'exclusion . . . . .	48
2.2 Élaboration des profils grenoblois de distribution de DS . . . . .	48
2.2.1 Indicateurs utilisés pour l'élaboration des profils de distributions . . . . .	49
2.2.2 Démarche suivie . . . . .	49
2.2.3 Description des indicateurs de distribution de DS . . . . .	61
2.2.4 Intérêt des profils grenoblois de distribution de DS . . . . .	62
2.3 Listes de référence . . . . .	64
2.3.1 Construction des listes de référence . . . . .	64

2.3.2	Intérêt des listes de référence . . . . .	66
2.3.3	Description des listes de référence . . . . .	67
2.4	Synthèse . . . . .	70
2.4.1	Profils grenoblois de distribution de DS . . . . .	70
2.4.2	Listes de référence . . . . .	71
<b>3</b>	<b>Description et modélisation des profils grenoblois</b>	<b>73</b>
3.1	Description des profils grenoblois de distribution de DS . . . . .	73
3.1.1	Méthodes . . . . .	73
3.1.2	Résultats . . . . .	74
3.2	Modélisation des distributions de DS par des lois log-normales unimodales . . . . .	79
3.2.1	Démarche suivie . . . . .	79
3.2.2	Résultats . . . . .	82
3.3	Discussion . . . . .	86
3.3.1	Classification issue des profils empiriques . . . . .	86
3.3.2	Influence des paramètres sur la forme des profils . . . . .	86
3.3.3	Modélisation des 9 profils référents . . . . .	86
3.3.4	Synthèse . . . . .	87
<b>4</b>	<b>Classification des profils issue de l'approche factorielle</b>	<b>99</b>
4.1	Stratégie de modélisation . . . . .	99
4.2	Méthodes utilisées . . . . .	100
4.2.1	Principes de l'ACP . . . . .	100
4.2.2	Principe de la classification hiérarchique . . . . .	106
4.2.3	Principe de l'analyse discriminante . . . . .	108
4.3	Résultats . . . . .	110
4.3.1	ACP sur les 257 GHM grenoblois . . . . .	110
4.3.2	ACP sans les GHM grenoblois n°48 et n°230 . . . . .	112
4.3.3	Classification hiérarchique ascendante . . . . .	125
4.3.4	Analyse discriminante . . . . .	133
4.4	Discussion . . . . .	136
4.4.1	ACP . . . . .	136
4.4.2	Classification des profils des 257 GHM . . . . .	137
4.4.3	Les facteurs de l'ACP dans l'analyse discriminante . . . . .	141
4.4.4	Limites . . . . .	142
4.4.5	Synthèse . . . . .	143
<b>5</b>	<b>Classification des profils issue des modèles de mélange de lois</b>	<b>147</b>
5.1	Méthode . . . . .	147
5.1.1	Principe des modèles de mélange de lois normales . . . . .	147
5.1.2	Application à la classification des profils . . . . .	150
5.1.3	Stratégie de modélisation . . . . .	151
5.2	Résultats . . . . .	151

5.2.1	Classification résultant des modèles de mélange . . . . .	151
5.2.2	Description et interprétation du mélange de 3 lois normales . . . . .	152
5.2.3	Confrontation de la classification avec les listes de référence . . . . .	157
5.2.4	Confrontation de la classification avec celle issue l'approche factorielle	158
5.3	Discussion . . . . .	159
5.3.1	Apport de la classification en trois groupes . . . . .	159
5.3.2	Groupe des "mauvais M" . . . . .	159
5.3.3	Limites de l'analyse . . . . .	160
5.3.4	Synthèse . . . . .	160
<b>6</b>	<b>Modélisation des DS du GHM n°18 : accidents vasculaires cérébraux non transitoires</b>	<b>161</b>
6.1	Population et méthodes . . . . .	162
6.1.1	Population et variables d'intérêt . . . . .	162
6.1.2	Démarche suivie . . . . .	163
6.1.3	Interprétation du modèle de mélange . . . . .	165
6.2	Résultats . . . . .	168
6.2.1	Description des distributions des DS . . . . .	168
6.2.2	Description médico-sociale . . . . .	168
6.2.3	Modélisation des distributions de DS . . . . .	169
6.2.4	Description médico-sociale des composantes . . . . .	171
6.3	Discussion . . . . .	173
6.3.1	Principaux résultats . . . . .	173
6.3.2	Variables explicatives de la DS des composantes . . . . .	175
6.3.3	Validité interne de l'étude . . . . .	176
6.3.4	Validité externe de l'étude . . . . .	176
6.3.5	Limites de l'étude . . . . .	177
6.3.6	Synthèse . . . . .	177
	<b>Conclusion</b>	<b>187</b>
	<b>A Liste des CMD</b>	<b>193</b>
	<b>B Exemple : liste des GHM appartenant à la CMD 02</b>	<b>195</b>
	<b>C Informations contenues dans le RUM</b>	<b>197</b>
	<b>D Numéros, effectifs et libellés des GHM</b>	<b>199</b>
	<b>E Comparaison des quartiles de DS entre Grenoble et la base nationale</b>	<b>205</b>
	<b>F Liste de référence fondée sur les points ISA</b>	<b>209</b>
	<b>G Liste de référence fondée sur le "test des quantiles"</b>	<b>211</b>



H	Liste de référence fondée sur le “test du $\chi^2$ ”	213
I	Groupe des “mauvais FA”	215
J	Groupe des “mauvais FG”	217
K	Groupe des “autres F”	219
L	Groupe des “bons E”	223
M	Tableaux des coefficients de classement des GHM	227
N	Algorithme EM	229
O	Groupe des “mauvais M”	231
P	Groupe des “autres M”	233
Q	Groupe des “bons M”	237
	Glossaire	245

## Liste des tableaux

1.1	Répartition des différentes catégories d'établissements hospitaliers (au 1 <sup>er</sup> janvier 1996).	43
2.1	Quartiles de DS du GHM numéro 48 : interventions sur la rétine (Cas 1A).	53
2.2	Quartiles de DS du GHM numéro 51 : interventions sur le cristallin avec ou sans vitrectomie (Cas 2B).	53
2.3	Quartiles de DS du GHM numéro 874 : nouveau-nés de 2 500g et plus, sans problème significatif (Cas 3G).	53
2.4	Quartiles de DS du GHM numéro 52 : autres interventions extra-oculaires, âge supérieur à 17 ans (Cas 4E).	53
2.5	Quartiles de DS du GHM numéro 118 : embolies pulmonaires (Cas 1H).	53
2.6	Descriptif des proportions cumulées de patients grenoblois dans les 3 intervalles interquartiles nationaux de DS.	62
2.7	Intersection entre 2 listes A et B contenant des GHM.	67
2.8	Intersection entre les listes du CHU et des quartiles.	69
2.9	Intersection entre les listes du CHU et du $\chi^2$ .	69
3.1	Description des GHM grenoblois selon leur profil de distributions de DS.	77
3.2	Intersection des profils de distributions de DS avec la liste du CHU.	79
4.1	Exemple de fichier de données.	100
4.2	Moyennes et écart-types des proportions cumulées de patients grenoblois dans les 3 intervalles interquartiles nationaux de DS (n=257).	110
4.3	Matrice des corrélations entre les trois intervalles interquartiles.	110
4.4	Pourcentage de variance expliquée par chacune des composantes principales et inertie cumulée.	111
4.5	Matrice des vecteurs propres.	111
4.6	Corrélations des variables avec les 2 premières composantes principales retenues.	112
4.7	Corrélations des facteurs de l'ACP avec les variables indiquant les différences entre les aires grenobloise et nationale et le déficit en points ISA.	125
4.8	Moyennes et écarts-types de la différence d'aire et du déficit en points ISA des groupes de GHM issus de la classification hiérarchique.	129

4.9	Sensibilité et valeur prédictive positive entre la liste CHU (référence) et la classification issue de la classification hiérarchique. . . . .	132
4.10	Sensibilité et valeur prédictive positive entre la liste des quartiles (référence) et la classification issue de la classification hiérarchique. . . . .	132
4.11	Sensibilité et valeur prédictive positive entre la liste du $\chi^2$ (référence) et la classification issue de la classification hiérarchique. . . . .	133
4.12	Moyennes et écarts-types des facteurs issus de l'ACP et du déficit en points ISA pour les 154 GHM utilisés pour la discrimination. . . . .	133
4.13	Sélection des variables avec une analyse discriminante pas à pas avec la méthode du $V$ de Rao. . . . .	134
4.14	Test de signification (*) des fonctions discriminantes de la première analyse. . . . .	134
4.15	Sélection des variables avec une analyse discriminante pas à pas avec la méthode du $V$ de Rao. . . . .	134
4.16	Test de signification (*) des fonctions discriminantes de la deuxième analyse. . . . .	135
4.17	Comparaison des taux de bien classés pour deux analyses discriminantes, l'une réalisée avec les facteurs issus de l'ACP et le déficit en points ISA, l'autre avec les variables d'origine : pourcentages cumulés de patients dans les différents intervalles interquartiles. . . . .	135
4.18	Tableau de contingence entre les profils et les classes des 103 GHM utilisés pour construire la fonction de classement de l'analyse discriminante. Variables discriminantes utilisées : les 3 pourcentages cumulés de patients sur les intervalles interquartiles de DS. 82% de bien classés. . . . .	136
4.19	Analyse de la robustesse des taux de bien classés dans les différentes classes à partir de 10 nouvelles analyses des variables d'origine : proportions cumulées de patients dans les différents intervalles interquartiles. . . . .	136
4.20	Sensibilité et valeur prédictive positive de la discrimination (deuxième analyse). . . . .	137
4.21	Calcul des points ISA totaux du CHU de Grenoble avant et après avoir enlevé le groupe des 16 GHM. . . . .	139
5.1	Moyennes et écarts-types de la différence d'aire et du déficit en points ISA des groupes de GHM issus du modèle de mélange en 3 classes. . . . .	157
5.2	Sensibilités et valeurs prédictives positives entre les listes de référence : CHU, quartile et $\chi^2$ , et le groupe des "mauvais" GHM issu du modèle de mélange. . . . .	157
5.3	Intersection entre les classes du modèle de mélange et celles issues de l'approche factorielle. . . . .	158
5.4	Comparaison des caractéristiques médico-économiques entre les 3 groupes des classifications issues de l'approche factorielle et des modèles de mélange. . . . .	159
6.1	Description de la distribution des DS des bases grenobloise et nationale. . . . .	168
6.2	Caractéristiques du GHM n°18 : accidents vasculaires cérébraux non transitoires, pour les bases grenobloise et nationale. . . . .	179

6.3	Paramètres de la lois normale estimés sur les DS de la base nationale et résultats du test d'adéquation de KS. . . . .	180
6.4	Paramètres de la lois log-normale estimés sur les DS de la base nationale et résultats du test d'adéquation de KS. . . . .	180
6.5	Paramètre de la lois exponentielle estimé sur les DS de la base nationale et résultats du test d'adéquation de KS. . . . .	180
6.6	Paramètres de la lois gamma estimés sur les DS de la base nationale et résultats du test d'adéquation de KS. . . . .	180
6.7	Paramètres de la lois de Weibull estimé sur les DS de la base nationale et résultats du test d'adéquation de KS. . . . .	180
6.8	Valeurs des critères ICL et BIC pour les modèles de mélange de 2 ou 3 lois log-normales ajustés sur les DS de la base nationale et grenobloise. . . . .	181
6.9	Paramètres estimés des modèles de mélange de 2 lois normales sur le logarithme de DS des GHM de la base nationale et grenobloise. . . . .	181
6.10	GHM n°18, accidents vasculaires cérébraux non transitoires (base grenobloise) : probabilités conditionnelles d'appartenance à l'une ou à l'autre des classes du modèle selon les modalités des variables explicatives. . . . .	182
6.11	GHM n°18, accidents vasculaires cérébraux non transitoires (base nationale) : probabilités conditionnelles d'appartenance à l'une ou à l'autre des classes du modèle selon les modalités des variables explicatives. . . . .	183
6.12	GHM n°18 : caractéristiques médico-sociales des deux classes du modèle pour chacune des bases. . . . .	184
6.13	Résultat de la modélisation de la première composante du mélange pour la base grenobloise par une régression log-normale. . . . .	185
6.14	Résultat de la modélisation de la première composante du mélange pour la base nationale par une régression log-normale. . . . .	186
6.15	Résultat de la modélisation de la deuxième composante du mélange pour la base nationale par une régression log-normale. . . . .	186
E.1	Récapitulatif des 31 configurations possibles avec leurs effectifs. . . . .	207
M.1	Coefficients des fonctions discriminantes (première analyse). . . . .	227
M.2	Coefficients de classement : fonction linéaire discriminante de Fisher (première analyse). . . . .	227
M.3	Coefficients des fonctions discriminantes (deuxième analyse). . . . .	228
M.4	Coefficients de classement : fonction linéaire discriminante de Fisher, (deuxième analyse). . . . .	228



## Table des figures

1.1	Classement d'un séjour hospitalier dans un GHM. . . . .	33
1.2	Exemple de distribution de la variable DS. . . . .	38
2.1	Représentation graphique des quartiles d'une distribution. . . . .	50
2.2	Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n°48). . . . .	56
2.3	Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°48). . . . .	56
2.4	Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°48). . . . .	56
2.5	Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n°48). . . . .	56
2.6	Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n°51). . . . .	57
2.7	Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°51). . . . .	57
2.8	Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°51). . . . .	57
2.9	Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n°51). . . . .	57
2.10	Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n°874). . . . .	58
2.11	Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°874). . . . .	58
2.12	Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°874). . . . .	58
2.13	Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n°874). . . . .	58
2.14	Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n°52). . . . .	59
2.15	Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°52). . . . .	59

2.16	Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°52). . . . .	59
2.17	Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n°52). . . . .	59
2.18	Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n°118). . . . .	60
2.19	Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°118). . . . .	60
2.20	Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n°118). . . . .	60
2.21	Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n°118). . . . .	60
2.22	Représentation des proportions cumulées de patients sur les 3 intervalles interquartiles nationaux de DS. . . . .	63
2.23	Représentation des proportions cumulées de patients grenoblois sur le premier et le deuxième intervalles interquartiles nationaux de DS. . . . .	63
2.24	Représentation des proportions cumulées de patients grenoblois sur le premier et le troisième intervalles interquartiles nationaux de DS. . . . .	63
2.25	Représentation des proportions cumulées de patients grenoblois sur le deuxième et le troisième intervalles interquartiles nationaux de DS. . . . .	63
2.26	Représentation de la liste des 50 GHM grenoblois causant 80% du déficit du CHU sur les 2 premiers intervalles interquartiles. . . . .	68
2.27	Représentation de la liste des 50 GHM grenoblois causant 80% du déficit du CHU sur le premier et le troisième intervalles interquartiles. . . . .	68
2.28	Représentation de la liste des quartiles sur les 2 premiers intervalles interquartiles. . . . .	68
2.29	Représentation de la liste des quartiles sur le premier et le troisième intervalles interquartiles. . . . .	68
2.30	Représentation de la liste du $\chi^2$ sur les 2 premiers intervalles interquartiles. . . . .	68
2.31	Représentation de la liste du $\chi^2$ sur le premier et le troisième intervalles interquartiles. . . . .	68
3.1	Profil de type (a) . . . . .	75
3.2	Profil de type (b) . . . . .	75
3.3	Profil de type (c) . . . . .	75
3.4	Profil de type (d) . . . . .	75
3.5	Profil de type (e) . . . . .	75
3.6	Profil de type (f) . . . . .	75
3.7	Profil de type (g) . . . . .	75
3.8	Profil de type (h) . . . . .	75
3.9	Profil de type (neutre) . . . . .	75
3.10	Densités log-normales selon les variations de $\mu$ induites par $\beta$ . . . . .	83

3.11	Profils estimés selon les variations de $\beta$ . . . . .	83
3.12	Densités log-normales selon les variations de $\sigma_B$ induites par $\gamma$ . . . . .	84
3.13	Profils estimés selon les variations de $\gamma$ . . . . .	84
3.14	Densités log-normales selon les variations de $\mu$ et de $\sigma_B$ induites respectivement par $\beta$ et $\gamma$ . . . . .	88
3.15	Profils estimés selon les variations de $\beta$ et $\gamma$ (première expérience). . . . .	88
3.16	Densités log-normales selon les variations de $\mu$ et de $\sigma_B$ induites respectivement par $\beta$ et $\gamma$ . . . . .	88
3.17	Profils estimés selon les variations de $\beta$ et $\gamma$ (deuxième expérience). . . . .	88
3.18	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°118). . . . .	89
3.19	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°118). . . . .	89
3.20	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°118). . . . .	89
3.21	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°48). . . . .	90
3.22	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°48). . . . .	90
3.23	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°48). . . . .	90
3.24	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°372). . . . .	91
3.25	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°372). . . . .	91
3.26	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°372). . . . .	91
3.27	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°63). . . . .	92
3.28	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°63). . . . .	92
3.29	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°63). . . . .	92
3.30	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°274). . . . .	93
3.31	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°274). . . . .	93
3.32	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°274). . . . .	93
3.33	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°230). . . . .	94



3.34	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°230). . . . .	94
3.35	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°230). . . . .	94
3.36	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°197). . . . .	95
3.37	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°197). . . . .	95
3.38	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°197). . . . .	95
3.39	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°458). . . . .	96
3.40	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°458). . . . .	96
3.41	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°458). . . . .	96
3.42	Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n°623). . . . .	97
3.43	Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n°623). . . . .	97
3.44	Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n°623). . . . .	97
4.1	Projection des variables dans le plan factoriel $F1 \times F2$ (ACP réalisée sur les 257 GHM). . . . .	114
4.2	Projection des individus dans le plan $F1$ et $F2$ (ACP réalisée sur les 257 GHM). . . . .	115
4.3	Identification du nuage des individus avec les 9 profils caractéristiques du CHU de Grenoble. . . . .	116
4.4	Projection des variables dans le plan factoriel $F1 \times F2$ (ACP réalisée sans les GHM n°48 et n°230). . . . .	117
4.5	Représentation des individus (ACP réalisée sans les GHM n°48 et n°230). . . . .	118
4.6	Identification du nuage des individus avec la liste du CHU. . . . .	119
4.7	Identification du nuage des individus avec la liste fondée sur les quartiles. . . . .	120
4.8	Identification du nuage des individus avec la liste fondée sur le test du $\chi^2$ . . . . .	121
4.9	Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste CHU. . . . .	122
4.10	Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste des quartiles. . . . .	123
4.11	Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste du $\chi^2$ . . . . .	124
4.12	Box-plot des coordonnées sur le premier axe de l'ACP selon les profils empiriques. . . . .	126

4.13	Box-plot des coordonnées sur le deuxième axe de l'ACP selon les profils empiriques. . . . .	127
4.14	Dendrogramme de la classification hiérarchique de 137 GHM identifiés empiriquement comme "mauvais". La distance utilisée est la distance euclidienne et le critère d'agrégation est celui de la variation minimale d'inertie (méthode de Ward). . . . .	128
4.15	Proportions cumulées moyennes dans chacun des 3 intervalles interquartiles pour chacun des groupes de GHM: "bons E", "mauvais FA", "mauvais FG" et "autres F". . . . .	130
4.16	Représentation du déficit ou bénéfice en points ISA par rapport à la différence d'aire entre le profil grenoblois et la norme nationale pour la classification issue de l'approche factorielle. . . . .	131
4.17	Première analyse: projection des 154 GHM dans le plan des 2 facteurs discriminants significatifs (62% de bien classés). . . . .	144
4.18	Deuxième analyse: projection des 154 GHM dans le plan des 2 facteurs discriminants significatifs (82% de bien classés). . . . .	145
5.1	Exemple de densité d'un modèle de mélange de deux lois normales univarié.	148
5.2	Interprétation géométrique de la décomposition spectrale. La quantité $\lambda_k$ représente le volume de l'ellipse de concentration, $a_k = A_k(1,1)$ détermine sa forme et l'angle $\alpha_k$ de la matrice de rotation $D_k$ détermine son orientation.	149
5.3	Valeurs de la vraisemblance pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses. . . . .	153
5.4	Valeurs du critère BIC pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses. . . . .	153
5.5	Valeurs du critère ICL pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses. . . . .	153
5.6	Proportions cumulées moyennes dans chacun des 3 intervalles interquartiles pour chacun des groupes de GHM: "bons M", "mauvais M" et "autres M".	154
5.7	Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur les trois intervalles interquartiles nationaux. . . . .	155
5.8	Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le premier et le deuxième intervalle interquartile national. . . .	155
5.9	Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le premier et le troisième intervalle interquartile national. . . .	155
5.10	Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le deuxième et le troisième intervalle interquartile national. . .	155
5.11	Représentation des déficit ou bénéfice en points ISA par rapport à la différence d'aire entre le profil grenoblois et la norme nationale pour la classification issue des modèles de mélange. . . . .	156
6.1	Ajustement de la distribution des DS du GHM n°18 de la base grenobloise par différentes lois. . . . .	170

6.2	Ajustement de la distribution des DS du GHM n°18 de la base nationale par différentes lois. . . . .	170
-----	---	-----

# Introduction

## Historique

Le besoin de maîtrise des dépenses hospitalières est devenu pressant, dès le début des années 80, en raison de la crise financière qui a touché les systèmes de santé de la plupart des pays industrialisés. De toutes les solutions proposées, c'est l'introduction, par Fetter [49], d'une classification des séjours par pathologie qui a recueilli la plus large adhésion pour l'évaluation du produit hospitalier. Dans cette optique et sous l'impulsion du Directeur des Hôpitaux de l'époque, J. de Kervasdoué, une phase expérimentale, puis opérationnelle, Programme de Médicalisation des Systèmes d'Information (PMSI)<sup>1</sup> s'est progressivement mise en place en France : circulaire du 14 février 1985 relative à la comptabilité analytique [1], du 4 octobre 1985 relative aux Résumés Standard de Sortie (RSS) [6], du 5 août 1986 relative aux groupes homogènes de malades (GHM) [8] et circulaires du 6 janvier 1989 [10] et 24 juillet 1989 [9] relatives à l'informatisation des hôpitaux publics.

L'objectif de tarification des établissements hospitaliers par le PMSI est d'allouer un budget équivalent et comparable à tous les établissements (tout en tenant compte de certaines missions telles que la formation ou la recherche dans les Centres Hospitaliers Universitaires [CHU]), pour une activité équivalente et comparable mesurée par des points d'Indice Synthétique d'Activité (ISA, en quelque sorte l'unité d'œuvre des hôpitaux). À partir de la répartition des patients dans les GHM (dérivés des Diagnosis Related Groups américains issus de la classification de Fetter), les séjours sont ensuite transformés en nombres de points ISA. La variable de comparaison des établissements est le coût moyen de ce point ISA. Il est calculé en divisant le budget annuel de l'établissement par le nombre total de points ISA apportés par les patients au cours d'une année. Connaissant la valeur moyenne du coût du point ISA dans une région, le budget théorique des établissements est reconstitué et comparé avec le budget réel. Le Directeur de l'Agence Régionale Hospitalière a alors la possibilité de réaliser des arbitrages budgétaires à partir des chiffres obtenus. La traduction de l'activité en termes de points ISA est ainsi devenue une des caractéristiques des séjours dans les établissements de santé. Le nombre final de points est directement en relation avec la qualité du codage.

Le coût moyen du point ISA permet de hiérarchiser les hôpitaux publics. Ceux qui sortent du "tunnel" [-11%, +8%] autour de la valeur moyenne régionale du coût du point ISA

---

1. Un glossaire des abréviations est donné à la fin de la thèse

pourront, au cours de la négociation de leur budget, se voir pénalisés (s'ils sont au-dessus de la valeur moyenne régionale) ou bonifiés (s'ils sont au-dessous de la valeur moyenne régionale). Il est donc important pour un CHU de totaliser un maximum de points ISA au cours d'une année. Cependant, aucune notion qualitative n'accompagne pour le moment les calculs arithmétiques du PMSI, et aucune notion d'élasticité (permettant des prévisions médico-économiques) n'est actuellement dégagée des données PMSI, même au niveau de la base nationale des GHM.

D'autre part, l'utilisation des GHM en France pour le classement des patients suscite les réticences de certains acteurs du monde hospitalier, notamment en raison de leur hétérogénéité. L'hétérogénéité intra-GHM est inhérente à la variabilité des malades, aux pratiques des médecins et aux structures des établissements (équipements, filières de soins), mais aussi à la dispersion introduite par la description de l'information (codage des informations médicales, choix du type administratif du séjour, découpage du séjour retentissant sur sa durée, ...). Cette hétérogénéité a été montrée dans les distributions de durées de séjours (DS) et de coûts [41, 52, 72, 73, 74, 75]. Pourtant, le nombre de points ISA rapporté au CHU par un patient est toujours le même, qu'il reste 2 jours ou 15 jours, il dépend seulement du GHM d'affectation. C'est pourquoi il est important de repérer dans les GHM les patients qui restent plus longtemps que la normale, et de les décrire, car ils sont sans doute responsables du fort coût du point ISA pour l'établissement. Cependant, une démarche exhaustive de l'analyse des 582 GHM semble difficile à mettre en œuvre. Une alternative serait, d'abord, de repérer les GHM comportant une hétérogénéité au niveau de la distribution de leurs DS, et, seulement ensuite, d'en identifier les patients responsables. Cette démarche pourrait être rentable pour un CHU, en termes d'apport en points ISA.

## Origine du projet

Malgré une exhaustivité croissante du recueil des données pour le PMSI, le CHU de Grenoble continuait, en 1996, à figurer parmi les hôpitaux dont le coût du point ISA était fort, tant au niveau régional que national. La qualité de la collecte des données était certes perfectible, mais d'autres causes devaient être recherchées pour expliquer le coût élevé du point ISA au CHU de Grenoble par rapport à celui de la région Rhône-Alpes.

Une étude, réalisée en 1997 par le Dr. J. Fauconnier, a montré qu'en 1996, le nombre de journées réalisées par le CHU de Grenoble, à "case mix" équivalent par rapport à l'ensemble des CHU français, était manifestement trop important pour soigner ses patients. Le CHU utilisait 12 236 journées de plus que s'il avait eu, pour chaque GHM, la même durée de séjour moyenne que les CHU nationaux [44]. Si ces journées avaient été utilisées pour les mêmes GHM, le nombre de points ISA aurait augmenté de manière significative, et cela aurait entraîné la baisse de son coût.

## Objectifs de ce travail

La première étape de ce travail de thèse a consisté à mettre en œuvre une méthodologie statistique fiable, utilisant des méthodes classiques, pour éclairer les problèmes identifiés par le Dr. Fauconnier, et permettre ainsi l'identification des GHM grenoblois "atypiques" en termes de DS, par rapport à leurs homologues de la base nationale des CHU. En effet, les patients responsables de la valeur élevée du coût du point ISA appartiennent sans doute à ces GHM, sur lesquels il serait possible d'entreprendre une action visant à rapporter plus de points ISA au CHU, en examinant leur RSA ou leur RSS. Le but serait de ramener le coût moyen du point ISA du CHU de Grenoble au niveau de la référence nationale.

Ce travail se positionne plutôt dans une logique économique que médicale, c'est pourquoi nous avons adopté une approche pragmatique et non explicative des DS. Autrement dit, nous avons travaillé sur les distributions marginales des DS, c'est-à-dire que nous avons considéré toutes les DS, sans tenir compte du statut du patient à sa sortie de l'hôpital : vivant guéri, vivant non guéri ou décédé. Comme on peut le remarquer, la logique économique ne recouvre pas ici la logique médicale. En effet, d'un point de vue purement économique, un patient qui sort vivant ou décédé rapporte le même nombre de point ISA au CHU, cela ne dépend que de son GHM d'affectation. Cependant, il est évident que les DS concernant des patients décédés ne participent pas à la performance hospitalière d'un point de vue médical.

Nous avons donc cherché des méthodes statistiques nous permettant de dresser une typologie des GHM grenoblois, selon la forme de leurs distributions de DS par rapport à celle de leurs homologues de la base nationale. Notre objectif est d'étudier leur éventuelle structure en classes, en termes de profil de distribution de DS. Il s'agit en quelque sorte d'une analyse inter-GHM. Pour ce faire, nous allons considérer plusieurs approches : d'une part les approches factorielles (analyse en composantes principales, classification hiérarchique, analyse discriminante) et, d'autre part, les modèles de mélange. L'utilisation de ces méthodes est très classique dans les problèmes de classification, et présente l'intérêt de dégager, le cas échéant, la structure de classes des données.

Dans la deuxième étape, nous proposons une méthodologie statistique permettant de repérer et de caractériser les "outliers" (patients ayant des DS extrêmes, sans doute responsables du coût élevé du point ISA au CHU de Grenoble), à partir des distributions de DS des GHM grenoblois identifiés comme "atypiques" lors de la première étape. Pour ce faire, nous avons utilisé une approche par modèle de mélange de lois. Ces modèles permettent de prendre en compte l'hétérogénéité intra-GHM.

De manière plus générale, nous avons réalisé ce travail dans le but d'une meilleure estimation des besoins financiers du CHU, afin de faciliter la négociation des décideurs d'une part avec les tutelles, d'autre part avec les acteurs locaux du soin.

## Plan du manuscrit

Ce manuscrit comporte 6 chapitres dont nous détaillons le plan ci-dessous.

Le premier chapitre s'intéresse d'abord à la naissance, la mise en place et le fonctionnement du Programme de Médicalisation des Systèmes d'Information en France. Il détaille ensuite les raisons qui ont motivé notre étude centrée sur les DS du CHU de Grenoble en 1997. Pour finir, il évoque l'intérêt de notre recherche concernant les durées de séjour d'un point de vue méthodologique et médical.

Dans le deuxième chapitre, nous présentons les données avec lesquelles ce travail a été réalisé. Nous détaillons d'abord les méthodes qui nous ont permis d'élaborer les profils de distribution des DS des 257 GHM grenoblois par rapport à leurs homologues de la base nationale, et les listes de référence qui nous ont aidé à évaluer la pertinence médico-économique des futures classifications de profils. Enfin, nous les décrivons et nous les discutons d'un point de vue médico-économique.

Le troisième chapitre fournit d'abord une description détaillée des 257 profils de DS en termes statistiques puis médico-économiques, en quelque sorte une première classification empirique. Celle-ci est ensuite confrontée aux trois listes de référence pour en analyser la pertinence. Puis nous étudions l'influence des paramètres des lois log-normales unimodales sur la forme des profils. Finalement, l'application de ce modèle a été faite sur les distributions de DS correspondant aux profils types référents de notre échantillon.

Dans le quatrième chapitre, nous proposons d'abord la stratégie de modélisation utilisée pour obtenir une deuxième classification des profils des 257 GHM. Elle est issue cette fois-ci des approches factorielles que sont l'analyse en composantes principales, la classification hiérarchique ascendante et l'analyse discriminante. Nous décrivons ensuite cette nouvelle classification du point de vue statistique et médico-économique. Pour finir, nous discutons ses caractéristiques.

Le cinquième chapitre s'intéresse à une approche différente, utilisant les modèles de mélange de lois, pour établir une troisième classification des profils des 257 GHM. L'objectif de cette nouvelle classification est de valider la classification issue de l'approche factorielle. Elle est ensuite décrite d'un point de vue statistique et médico-économique, puis elle est confrontée à la deuxième classification. Enfin, nous analysons les résultats de cette confrontation, ainsi que ses caractéristiques.

Le sixième chapitre aborde une méthodologie utilisant les modèles de mélange de lois pour analyser plus spécifiquement les causes du mauvais classement des GHM identifiés comme "atypiques" par les classifications précédentes, en repérant et en donnant les caractéristiques médico-sociales des "outliers" dans les distributions de DS. Nous appliquons cette méthodologie aux données du GHM n°18 : accidents vasculaires cérébraux non transi-

toires. Ensuite, nous en présentons les résultats, et, finalement, nous les discutons.





# Chapitre 1

## Fonctionnement budgétaire d'un centre hospitalier et durées de séjour

L'introduction de ce travail de thèse étant faite, ce chapitre présente, dans un premier temps, la naissance, la mise en place et le fonctionnement du Programme de Médicalisation des Systèmes d'Information en France. Il détaille ensuite les raisons qui ont motivé notre étude centrée sur les durées de séjour du CHU de Grenoble en 1997 et il évoque l'intérêt de notre recherche concernant les durées de séjour d'un point de vue méthodologique et médical.

### 1.1 Mise en place du Programme de Médicalisation des Systèmes d'Information en France

Au début des années 80 est apparu le besoin d'une meilleure connaissance de la réalité de l'activité hospitalière. Le Programme de Médicalisation des Systèmes d'Information en France, né de ce besoin, est présenté ici dans ses différentes étapes.

#### 1.1.1 Contexte économique au début des années 80

La forte croissance des dépenses de santé est un phénomène auquel sont confrontés, depuis plusieurs décennies, tous les pays industrialisés. En 1990, la France se situe au deuxième rang mondial des dépenses de santé par rapport au PIB (derrière les États-Unis) avec 9% du PIB : elle est en première position en Europe avec une augmentation de 4.3% par an entre 1980 et 1990 [34]. Dans un contexte de ressources limités, le strict respect d'un quota entre l'évolution des dépenses de santé et la richesse nationale s'impose [53].

Depuis cette période, la mesure de l'activité médicale est donc considérée comme l'un des problèmes cruciaux par l'ensemble des spécialistes qui étudient les systèmes de santé [53] : médecins, épidémiologistes, économistes, administrateurs sanitaires, ...

### 1.1.2 Naissance du Programme de Médicalisation des Systèmes d'Information

Jusqu'à la publication du décret du 11 août 1983 [4], l'hôpital public était rémunéré sur la base de tarifs journaliers. Ces tarifs étaient définis annuellement, pour chaque hôpital, et pour chaque discipline médico-tarifaire. Ce type de rémunération avait pour conséquence un dérapage très élevé des dépenses de santé et des tarifs très disparates selon les établissements, plus proches de la "tête du client" [34] que de la juste rémunération d'une prestation. Cela a eu pour conséquence une inflation des séjours.

Il est donc décidé, à partir du décret de 1983, de ne plus attribuer de tarif par journée, mais une rémunération globale de la structure, pour un volume global prévu, afin de maîtriser les dépenses de santé. S'inspirant de l'expérience canadienne, il est décidé d'importer en France la rémunération sous forme de dotation globale. Chaque année, cette dotation est augmentée avec l'application d'un taux directeur [34]. Ainsi, les CHU qui étaient bien rémunérés, le restent et inversement. Il est décidé également de créer un indicateur d'activité permettant de repérer de façon plus pertinente le service réel rendu à la population. Cet indicateur se doit d'être médicalisé, afin de dépasser la limite des indicateurs classiques de mesure de l'activité qu'étaient les entrées ou les journées [34].

Sous l'impulsion du Directeur des Hôpitaux de l'époque, J. de Kervasdoué, l'idée d'un Projet de Médicalisation du Système d'Information hospitalier apparaît dès 1985 ; d'abord dans une phase expérimentale : circulaires du 14 février 1985 relative à la comptabilité analytique [1], du 4 octobre 1985 relative aux résumés standards de sorties (RSS, voir paragraphe 1.2.2)<sup>1</sup> [6] et du 5 août 1986 relative aux Groupes Homogènes de Malades (GHM) [8], puis, dans une phase opérationnelle, dès 1989, lorsque ce projet est promu programme national : circulaire du 6 janvier 1989 [10] relative à l'informatisation des hôpitaux publics.

Le Programme de Médicalisation du Système d'Information ou PMSI est né de la constatation qu'il était possible de grouper des patients similaires (en âge, en pathologie et en mode de diagnostic et de traitement) en GHM (voir paragraphe 1.2.2) : les études préliminaires ont été effectuées principalement par le professeur Robert Fetter [49], d'abord aux États-Unis à la fin des années 70, puis en France au début des années 80. Une de ces études a été réalisée au CHU de Grenoble, à partir des données du système DOSTAM (Dossier résumé de sortie des malades et statistiques médicales) [32, 86, 87, 88].

La circulaire du 24 juillet 1989 [9] et celle du 12 février 1990 [11] prévoient la généralisation du PMSI selon deux modalités : d'une part, l'installation d'un Département d'Information Médicale (DIM) dans tout hôpital ayant une activité significative de court séjour et, d'autre part, l'attribution de crédits spécifiques PMSI pour permettre notamment les créations de postes nécessaires pour faire fonctionner ces DIM. Le DIM devient une nouvelle structure hospitalière, destinée à la gestion de l'information médicale.

L'été 1991 voit les perspectives du PMSI se clarifier par le vote d'une nouvelle loi concernant la réforme hospitalière (Journal Officiel du 31 juillet 1991 [14]). Cette loi prévoit l'obli-

---

1. Un glossaire des abréviations est donné à la fin de la thèse.

gation, pour tous les établissements hospitaliers, d'une évaluation de leur activité médicale, en volume, en qualité et en coûts, à partir de la mise en place d'un système d'information médicalisé. Ces dispositions généralisent l'installation du PMSI dans tous les centres hospitaliers. La circulaire de juillet 1991 [12] et celle de septembre 1991 [13] rappellent clairement ces perspectives :

- tout hôpital doit produire, courant 1992, des Résumés Standards de Sortie (RSS) exhaustifs et fiables,
- tout hôpital doit s'engager sur un calendrier de montée en charge du PMSI et des coûts par Groupe Homogène de Malades (GHM),
- aucune demande de moyens supplémentaires (au delà du faible taux directeur de base appliqué chaque année sur le budget global, environ 1%) ne pourra être prise en compte hors de cette perspective de production de l'information PMSI.

Le PMSI est également étendu au secteur privé ; en 1994, le ministère de la Santé lance en Languedoc-Roussillon une expérience de tarification qui vise l'ensemble des établissements publics et privés de la région. L'objectif est de tester la faisabilité de nouveaux modes de tarification intégrant les données du PMSI, ce qui permet d'ajuster la dotation globale en fonction de ces données. Au vu des travaux accomplis, la circulaire du 10 mai 1995 [16], confirmée par la lettre-circulaire du 22 août 1995 [19], généralise aux établissements sous compétence tarifaire de l'État, ce dispositif technique, dès 1996.

D'autre part, la circulaire du 11 décembre 1995 [17] crée, dans chaque région, les Comités Techniques Régionaux de l'Information Médicale (COTRIM) dont les attributions sont de 2 ordres :

- ils veillent à la validité de l'information médicale PMSI produite par les établissements dispensant des soins de médecine, chirurgie ou obstétrique (MCO),
- ils discutent de la qualité des données médicales du PMSI,
- ils peuvent être consultés par les établissements, les caisses et les services déconcentrés sur les projets d'utilisation de cette information médicale.

Le COTRIM rassemble, sous l'égide du médecin-inspecteur régional, des médecins-inspecteurs de santé publique des Directions Départementales des Affaires Sanitaires et Sociales (DDASS) et de la Direction Régionale des Affaires Sanitaires et Sociales (DRASS), des médecins-conseils et des médecins de DIM.

Enfin, l'ordonnance du 24 avril 1996, portant réforme de l'hospitalisation publique et privée [20], remet l'accent sur le PMSI en :

- le généralisant à l'ensemble des établissements dispensant des soins de médecine, chirurgie ou obstétrique,
- situant aussi les enjeux de cette réforme dans deux domaines principaux :
  - \* la réforme du système de la tutelle des établissements hospitaliers,
  - \* l'institution d'une procédure d'accréditation des structures.

Ainsi, dans chacune des 24 régions françaises, est créée une Agence Régionale d'Hospitalisation (ARH). Elle a le statut de groupement d'intérêt public (GIP) entre l'État et les

organismes d'assurance maladie. Composée à parité de représentants de l'État et des Caisses régionales d'assurance-maladie, elle constitue une autorité de décision pour l'ensemble des problèmes de l'hospitalisation, publique comme privée. C'est désormais l'ARH qui établit le SROS (Schéma Régional d'Organisation Sanitaire), fixe les budgets des hôpitaux et les tarifs des établissements privés, attribue les équipements, signe les contrats d'objectifs et de moyens avec hôpitaux et cliniques, dont elle peut décider la répartition et le regroupement. Le PMSI va servir de base à ces négociations.

L'accréditation des structures, déjà en place dans plusieurs pays étrangers, constitue le second enjeu important de la réforme de 1996. Cette procédure sera conduite par l'Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES). Elle vise à octroyer un "label de qualité" aux établissements de soins. L'accréditation est présentée par certains auteurs comme un garde-fou des dérives induites par le PMSI (taux de réadmission trop élevé...) [34].

Même si, dès 1997, le PMSI a été généralisé à l'ensemble des établissements hospitaliers publics et privés, la mise en place de celui-ci dans le système hospitalier français a été lente. Les freins à cette mise en place ont pu être l'absence d'un système d'information hospitalier antérieur, l'insuffisance de ressources humaines qualifiées, la difficile implication du corps médical et les problèmes de coordination entre médecins et administratifs.

Depuis, le PMSI a été mis en place et généralisé aux établissements publics et privés dispensant des soins de suite et de réadaptation (SSR), grâce aux circulaires du 3 avril 1997 et du 9 mars 1998 [21, 22] sous l'appellation PMSI SSR. Quant au PMSI Psychiatrie, il est mis en place, à titre expérimental, depuis le 1er janvier 2002 et pour 1 an, dans les établissements psychiatriques de 4 régions pilotes (Aquitaine, Rhône-Alpes, Lorraine, Réunion, partiellement Île de France).

## 1.2 Le Programme de Médicalisation des Systèmes d'Information

Nous nous intéressons maintenant aux buts du PMSI, dont nous présentons ensuite le fonctionnement : du recueil des informations médicales des patients au calcul de la dotation globale de financement allouée aux hôpitaux.

### 1.2.1 Objectifs du PMSI

Dès la conception du PMSI, un lien étroit entre la maîtrise des dépenses, la répartition des moyens et la mesure médicalisée de l'activité est apparu [53]. Dès l'origine, le PMSI était, comme d'ailleurs aux États-Unis, un vecteur attendu de rémunération, sinon de rémunération par pathologie [34].

Le PMSI est une démarche visant à décrire qualitativement et quantitativement l'activité hospitalière et à évaluer l'utilisation des ressources collectives. Il vise à répondre à la question : que fait l'hôpital et à quel coût ? Le PMSI permet de disposer d'un outil de prévision et de planification pour allouer des ressources aux hôpitaux d'une façon judicieuse, donner un coup de pouce là où c'est nécessaire, et freiner les consommations lorsque l'activité tend à diminuer [53].

Ce programme permet aussi progressivement aux gestionnaires de mieux intégrer la dimension médicale et les soins infirmiers dans leurs décisions, et aux médecins et soignants de mieux intégrer la dimension économique dans leurs pratiques professionnelles. Les premiers ont besoin des seconds pour savoir à quel coût ils travaillent ; les seconds ont besoin des premiers pour leur expliquer les enjeux, les difficultés et les évolutions de leur pratique face aux patients. Est aujourd'hui en jeu la capacité de tous les acteurs de l'hôpital, notamment les médecins, à accepter que la santé a un coût et qu'une meilleure gestion par le biais du PMSI est garant d'une meilleure médecine [53].

Le PMSI a donc la triple mission suivante [34] :

- connaître l'activité et les coûts d'un hôpital : publication de références, comparaisons, analyse des écarts ;
- gommer les disparités de dotation financière entre les établissements ayant une activité qualitativement et quantitativement comparable ;
- contraindre à la productivité optimale, par la mise en évidence des écarts de productivité hospitalière globale, par les décisions effectives de réallocation de ressources, par une politique d'intéressement et par la sanction de l'action des décideurs (tutelle/gestionnaires/médecins).

La médicalisation du système d'information est donc un outil externe d'allocation des ressources aux différents établissements. Certains auteurs l'ont proposé comme un outil de pilotage interne, notamment au niveau des services [34], mais ceci est discutable, du fait des séjours multi-unités. À moyen terme, ce dispositif doit permettre de réduire les disparités budgétaires entre les différents hôpitaux, en comparant leur dotation à activité équivalente.

## 1.2.2 Fonctionnement du PMSI

### La classification en Groupes Homogènes de Malades

La classification en Groupes Homogènes de Malades (GHM) constitue l'outil de base du PMSI dans le champ de la seule activité hospitalière concernée par une médicalisation du système d'information : l'activité de court séjour, avec hébergement, en médecine, chirurgie et obstétrique et, depuis 1998 en SSR et depuis 2002, en psychiatrie.

Les GHM doivent répondre aux caractéristiques suivantes :

- être définis par des variables couramment disponibles ;
- être en nombre “maniable” et être mutuellement exclusifs ;
- recouvrir l'ensemble des maladies codifiées par l'Organisation Mondiale de la Santé (OMS) dans la classification internationale des maladies (CIM-10) [15] ;
- représenter un nombre significatif de malades ;
- avoir une cohérence interne en termes de classification anatomique, physiopathologique et thérapeutique.

La classification en GHM est issue d'une adaptation de la quatrième version de la classification des “Diagnosis Related Groups” (DRG) américains, résultant des travaux d'une équipe de chercheurs (économistes) de l'université de Yale, dirigée par le Professeur Robert Fetter en 1970 [45, 46, 47, 48, 49]. Elle permet de classer les séjours hospitaliers dans des groupes présentant une double homogénéité en termes économique et médical, suivant :

- le diagnostic principal,
- la présence de co-morbidités “classantes” : co-morbidités appartenant à une liste de 2000 diagnostics secondaires augmentant la durée de séjour (DS) d'au moins un jour dans au moins 75% des cas,
- la présence d'actes “classants” : actes appartenant à une liste de 4000 actes distinguant les situations médicales et chirurgicales,
- l'âge du patient.

L'homogénéité est d'abord économique, la logique médicale ne venant que secondairement. La classification est fortement corrélée à la durée de séjour, comme l'a montré la première analyse du Professeur Fetter portant sur des millions de dossiers [49]. La classification en GHM résulte du regroupement des malades dont l'hospitalisation a donné lieu à des explorations et des traitements qui ont mobilisé des ressources de même ampleur. Ces données médicales concernent l'état du malade (âge, diagnostic principal, diagnostics associés) et les processus de soins (actes diagnostics et thérapeutiques).

La première classification en GHM a été publiée en 1986 et elle comportait 471 GHM [7]. Elle correspondait aux premiers travaux français de validation et d'adaptation de la quatrième version des DRG américains. En 1992, une nouvelle classification est parue [23], plus détaillée, qui comportait 508 GHM. Elle compte aujourd'hui 582 groupes et en est à sa sixième version [38]. Cette nouvelle version est née du fait que la classification issue des versions antérieures décrivait mal les GHM, notamment les GHM médicaux. Les 582 GHM se répartissent en 25 Catégories Majeures de Diagnostics (CMD), plus deux Catégories Majeures (CM), les CM n°24 et CM n°90, correspondant l'une aux séances (chimiothérapie, dialyse, radiothérapie, ...) et séjours ambulatoires à l'hôpital et l'autre aux erreurs (voir Annexe A). Au sein des CMD, une subdivision place d'un côté les séjours sans mention d'acte opératoire, dits séjours médicaux, et de l'autre les séjours avec actes opératoires, dits séjours chirurgicaux, ordonnés selon une hiérarchie prédéterminée des actes. Qu'il s'agisse des GHM

médicaux ou des GHM chirurgicaux, quelques subdivisions supplémentaires apparaissent, en fonction de l'âge, de l'existence de Complications et Morbidités Associées (CMA) et de l'existence de Complications et Morbidités Associées Sévères (CMAS). Un exemple est donné en Annexe B, où nous pouvons voir l'intitulé des 13 GHM faisant partie de la CMD n°02 "Affections de l'œil". La sixième version de la classification en GHM semble entériner un modèle descriptif à "deux vitesses" : les groupes homogènes, au contenu médicalement et économiquement évocateur, et les groupes hétérogènes, pourtant toujours intitulés GHM, dont la vocation est de concentrer les "impuretés", qui perturbent les autres. Parmi ceux-là, notamment, les quelques 130 groupes avec CMA ou CMAS.

Depuis 1992, l'actualisation de la classification des GHM est confiée au Pôle d'Expertise et de Référence Nationale des Nomenclatures de Santé (PERNNS) créé par le Docteur Girardier [54] qui, avec l'aide du Centre de Traitement de l'Information du PMSI (CTIP) créé par Anne Strauss, pour la partie informatique, réalise une mise à jour régulière de la classification des GHM et du Catalogue des Actes Médicaux (CdAM).

### Classement d'un séjour hospitalier dans un GHM

Le classement d'un séjour hospitalier d'un patient dans un GHM se fait par l'intermédiaire de la fonction **groupage** du logiciel "GROUPEUR" (voir graphique 1.1). Ce logiciel est propriété de l'État, géré par le PERNNS et il se trouve dans les DIM. La fonction **groupage** est le noyau dur, elle contient une adaptation française de l'algorithme de la classification inventée par le Professeur Fetter [49], avec tous les arbres de décision (il existe un arbre de décision par CMD), et toutes les tables (listes des actes ou des diagnostics, utilisés à chaque nœud de décision). La fonction **groupage** est alimentée par les informations médico-administratives du séjour recueillies dans le Résumé Standardisé de Sortie (RSS). Elle permet d'affecter à un RSS, un numéro de GHM et un seul. Tout ce qu'il faut savoir sur le RSS a été défini dans une circulaire en octobre 1985, dite circulaire RSS [5].

Le RSS correspond, en fait, à un séjour hospitalier effectué dans la partie court séjour d'un établissement de soins. Il est constitué d'un (ou plusieurs) Résumé(s) d'Unité Médicale (RUM). Le RUM contient un nombre limité d'informations d'ordre administratif et médical, qui doivent être systématiquement renseignées (voir Annexe C). Les nomenclatures utilisées pour coder les RUM sont, pour le codage du diagnostic principal (et s'il y a lieu des diagnostics associés), la CIM-10 et, pour le codage des actes (médicaux et chirurgicaux), le CdAM. Le CdAM identifie les actes médicaux d'une part, et quantifie l'activité médicale ou médico-technique correspondant aux actes d'autre part.

La mention du diagnostic principal (DP) dans le RUM est essentielle, car elle influe sur l'affectation du séjour dans la CMD adéquate au moment du groupage et donc dans le GHM adéquat. Le DP est celui qui a mobilisé l'essentiel de l'effort médical et soignant au cours du séjour hospitalier. Dans le cas d'un séjour multi-unités, le DP est choisi en fonction d'un



algorithme qui prend en compte la présence ou non d'un acte classant sur les différents RUM produits pendant le séjour. Un acte est dit classant, s'il est marqueur d'une différenciation statistique de la consommation de ressources ; les actes classants sont typiquement des actes chirurgicaux et ils sont repérés par la lettre "Y" dans le CdAM. Toujours dans le cas de séjour multi-unités, les diagnostics associés (DA) correspondent en général aux DP inscrits dans les RUM et non retenus comme DP, mais aussi aux DA enregistrés dans ces RUM. En conséquence, le taux de RUM par RSS est un indicateur qui permet d'évaluer le pourcentage de dossiers pour lequel le choix du DP n'est plus le résultat d'une réflexion d'un médecin à la fin d'un séjour, mais celui d'un algorithme de traitement de données. Les DA sont synonymes de CMA et donc d'orientation du RSS vers un GHM avec complications.

Une des critiques que l'on peut porter à la classification des GHM concerne la simplification de la pathologie à laquelle elle procède [55]. En effet, les malades sont classés dans les GHM à partir d'un seul diagnostic, le DP. La classification est très pertinente pour les GHM chirurgicaux, puisque les patients classés dans ces GHM ont tous eu le même acte classant (technique très spécifique). Par contre, cette classification devient nettement moins performante pour les GHM médicaux qui ne comportent pas d'actes classants. Il est donc nécessaire de manipuler avec prudence des données médicales dont l'exhaustivité et la précision sont encore à améliorer, même si cette exhaustivité a finalement peu d'impact sur l'attribution d'un séjour à un GHM et sur le calcul des coûts.

### **Transmission des données du PMSI**

Les données médicales recueillies dans le cadre du PMSI sont protégées par le secret professionnel [2]. Le DIM, qui organise le recueil, la circulation et le traitement des données médicales, est placé sous l'autorité d'un médecin. Son rôle en matière de centralisation des informations médicales est prévu par l'article L.710.5 du code de la santé publique [2]. La création des fichiers et les traitements de données sont soumis à l'avis préalable de la Commission Nationale de l'Informatique et des Libertés (CNIL). La transmission d'informations médicales à la direction de l'établissement, ou à l'extérieur de celui-ci, s'opère sur la base de données agrégées ou de Résumés de Sortie Anonymes (RSA) obtenus par la transformation du RSS. La production de RSA est automatisée. À partir du fichier de RSS groupés, le médecin responsable du DIM utilise le logiciel "GENRSA" (GÉNÉrateur de RSA), propriété de l'État, pour produire le fichier de RSA. À la différence du RSS, le RSA constitue toujours un enregistrement rendu anonyme. Des informations sont supprimées, d'autres sont ajoutées. "L'établissement transmet à la DRASS, sur un support magnétique agréé par les services de l'État concernés, le fichier des RSA relatifs aux patients sortis pendant le semestre. Ce fichier est issu de la plus récente version de "GENRSA" ; il est transmis trois mois au plus tard après la fin du semestre considéré. Dans le mois suivant la réception, la DRASS transmet ces informations à la Caisse Régionale d'Assurance Maladie (CRAM) et à la caisse pivot dont dépend l'établissement" [2].

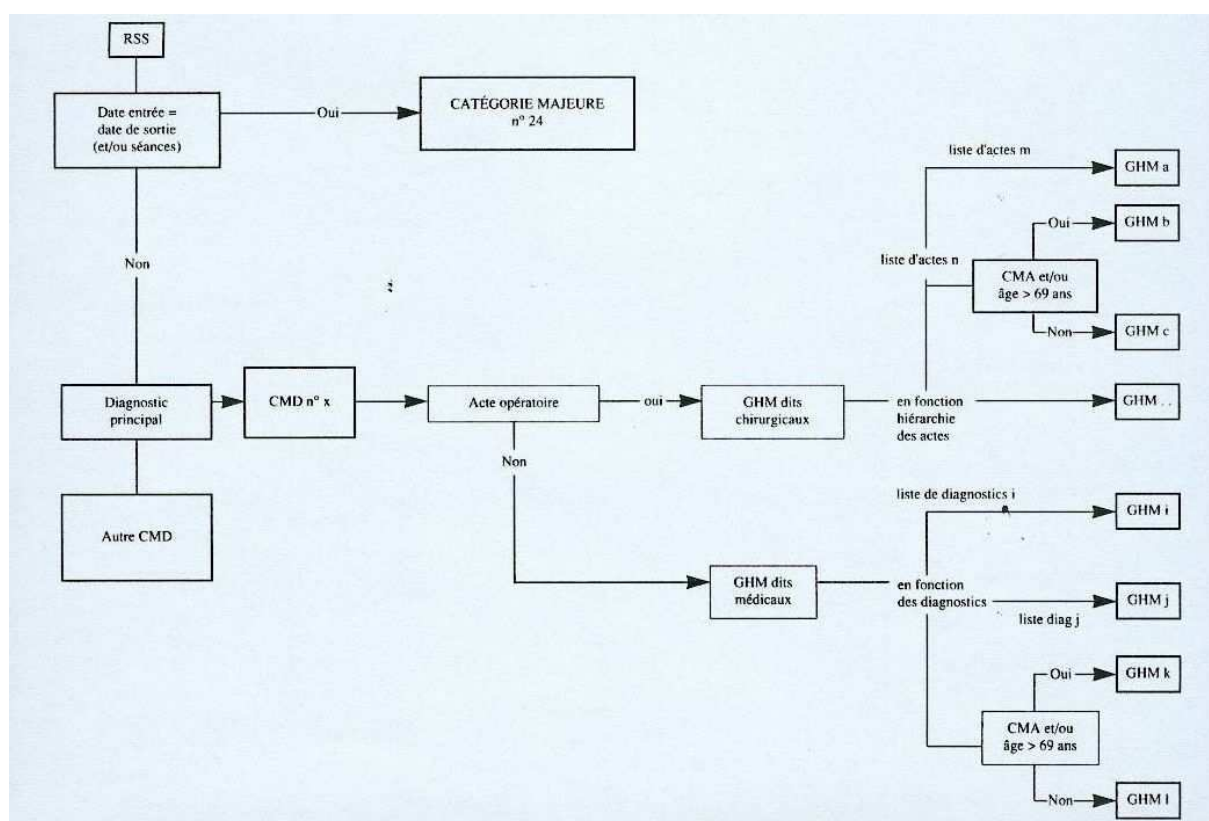


FIG. 1.1 – Classement d'un séjour hospitalier dans un GHM.

(Source : Dubois-Lefrère J. et Coca É. *Maîtriser l'Évolution des dépenses hospitalières : le PMSI*, p 95 [34]).

Les GHM permettent donc de décrire, de manière synthétique et qualitative, l'activité médicale, chirurgicale et obstétricale d'un établissement de court séjour.

### Point ISA et analyse de l'activité médico-économique

L'analyse de l'activité médico-économique des établissements hospitaliers s'appuie sur la pondération économique des 582 GHM de la classification. L'échelle nationale de coûts, construite à partir des données médico-économiques fournies par un échantillon d'établissements publics dotés d'une cellule d'analyse de gestion très performante, attribue à chaque GHM (l'unité de production hospitalière) un certain nombre de points Indice Synthétique d'Activité (ISA) qui permettent d'évaluer, dans une unité commune, l'activité produite dans le cadre des courts séjours hospitaliers. Le point ISA est comparable à une unité d'œuvre. La première version de l'échelle nationale des coûts est sortie en mai 1995 [18] ; depuis, elle est actualisée chaque année, afin de tenir compte de l'évolution de la classification en GHM, des procédures de soins et des coûts. Elle permet de calculer des coûts de référence par GHM, à l'usage des services de l'État et des établissements hospitaliers. Ainsi, une pathologie grave et compliquée, avec une co-morbidité associée sévère, oriente la classification vers un GHM lourd, induisant une cotation en points ISA élevée. En fonction du nombre de points ISA attribués à un GHM, celui-ci se positionne sur l'échelle nationale de coûts par rapport au GHM "calant" numéro 540 : "Accouchement par voie basse sans complication", doté arbitrairement de 1000 points ISA.

Les données recueillies dans le PMSI conditionnent donc le nombre de points ISA des établissements et, par conséquent, les moyens financiers qui leur sont attribués.

### Calcul de la valeur du point ISA et dotation budgétaire

Chaque année, les DRASS calculent la valeur financière du point ISA des hôpitaux de leur région et la dotation théorique de chaque établissement, en traitant les informations transmises par les établissements (fichier des RSA), par le logiciel "MAHOS" : Mesure de l'Activité HOSPitalière. La valeur régionale du point ISA est obtenue en divisant l'enveloppe régionale de court-séjour par la somme des points réalisés par l'ensemble des établissements concernés dans la région. On observe alors l'écart entre le budget théorique calculé *ex post* et la dotation globale de financement, pour déterminer si l'établissement est correctement, sur- ou sous-doté. Voici en détail les étapes du calcul réalisé par le logiciel "MAHOS" :

- constitution de la base régionale de RSA valorisée en points ISA ;
- redressement, s'il y a lieu, de la valorisation en points des RSA transmis,
- calcul du nombre moyen de points ISA par séjour pour l'ensemble des établissements de la région,
- contrôle et redressement, s'il y a lieu, de l'exhaustivité des RSA transmis,
- valorisation de l'activité externe,

- calcul de la valeur régionale du point ISA,
- calcul de la dotation théorique de chaque établissement au regard de son activité en court séjour,
- estimation, s’il y a lieu, du niveau de sur- ou sous-dotation de l’établissement,
- et simulation des ajustements budgétaires pour les établissements de la région.

Les enseignements du calcul du coût moyen du point ISA par établissement sont les suivants :

- il permet de hiérarchiser les hôpitaux publics,
- les hôpitaux qui sortent du “tunnel” [  $-11%$  ;  $+8%$  ] autour de la valeur moyenne de ce coût du point ISA par établissement, calculée au niveau régional, pourront, au cours de la négociation de leur budget, se voir pénalisés (s’ils sont au-dessus de la valeur moyenne régionale) ou bonifiés (s’ils sont au-dessous de la valeur régionale),
- aucune notion qualitative n’accompagne pour le moment les calculs arithmétiques du PMSI,
- aucune notion d’élasticité (permettant des prévisions médico-économiques) n’est actuellement dégagée des données PMSI, même au niveau de la base nationale des GHM.

Ainsi, le but du point ISA est de persuader les hôpitaux de rapprocher leurs coûts d’un “coût cible” équivalent au coût moyen national, en restant dans les limites des prévisions budgétaires précédentes.

## 1.3 Le CHU de Grenoble

Après avoir présenté le PMSI d’une manière générale, nous nous intéressons plus spécifiquement au CHU de Grenoble : nous le décrivons en quelques chiffres, révélateurs de son activité actuelle. Ensuite, nous étudions plus précisément les données du PMSI de Grenoble en 1996, qui sont à l’origine de ce travail de thèse sur l’analyse des durées de séjour. Ces données laissent apparaître trois raisons possibles pouvant expliquer la forte valeur du point ISA à cette époque et, par conséquent, le “mauvais” classement, d’un point de vue économique, du CHU de Grenoble parmi les autres CHU français.

### 1.3.1 Le CHU de Grenoble en quelques chiffres

Le CHU de Grenoble joue le rôle d’hôpital de proximité pour la population de l’agglomération grenobloise, d’hôpital de recours pour les activités de pointe au service des structures hospitalières de la région Rhône-Alpes, et sert enfin de référence universitaire. Les données analysées sont celles de 1997, dont les caractéristiques sont: [35] :

- 2 065 lits d’hospitalisation de court séjour, répartis sur 22 services de médecine, 19 services de chirurgie et 4 services de réanimation,
- 97 126 entrées, correspondant à 569 597 journées d’hospitalisation et à 78 572 RSS ;

- durée moyenne de séjour de 5.9 jours et coefficient global d'occupation des lits de 78.1% ;
- budget d'environ 27 millions d'euros d'investissements et de 360 millions d'euros de dépenses d'exploitation.
- Le personnel médical est composé de 1 600 personnes et le personnel non médical est composé de plus de 5 800 personnes.

D'autre part et contrairement à la majorité des hôpitaux, le CHU de Grenoble fonctionne en codage centralisé pour les données PMSI.

### 1.3.2 Cas du CHU de Grenoble en 1996

Malgré une exhaustivité croissante du recueil des données pour le PMSI, l'établissement continuait à figurer parmi les hôpitaux dont le coût du point ISA était fort, tant au niveau régional que national. La qualité de la collecte des données était certes perfectible (environ 20% de taux d'erreur lors des contrôles qualités), mais d'autres causes devaient être recherchées pour expliquer le coût élevé du point ISA au CHU de Grenoble par rapport à celui de la région Rhône-Alpes.

Une première hypothèse pouvait être l'éventail des cas traités au CHU ou "case mix". Un recrutement particulier de patients pouvait expliquer un coût du point ISA particulièrement élevé, en raison de la prédominance de GHM mal valorisés par l'échelle des coûts. Ce type d'effet, s'il est souvent en cause dans les hôpitaux à forte spécialisation (maternités, centres de lutte contre le cancer), ne peut être important pour un hôpital comme celui de Grenoble, dont le recrutement est assez varié. De plus, on devrait le retrouver dans d'autres CHU, ce qui n'apparaît pas dans les comparaisons utilisant la base de données (BDD) PMSI de tous les CHU français. L'éventail des cas n'était donc pas la première hypothèse à explorer.

Une deuxième hypothèse possible impliquait la structure de l'établissement et, notamment, les coûts induits par la prise en charge des différentes pathologies. Ainsi, à "case mix" équivalent, un établissement peut avoir un coût du point ISA supérieur à un autre en raison des charges plus importantes nécessitées par certains GHM. Cette analyse demandait, là aussi, d'avoir une référence pour pouvoir être menée, ainsi que des clés de répartition des charges indirectes (structure et personnel) qui soient identiques, pour pouvoir être comparables entre les différents établissements. En 1996, seuls les établissements faisant partie de l'Étude Nationale des Coûts avaient une standardisation de la comptabilité analytique et du recueil des données médicales suffisante pour permettre de telles comparaisons ; le CHU de Grenoble n'en faisait pas partie, il était donc difficile de vérifier cette deuxième hypothèse.

Une troisième hypothèse concernait l'activité du CHU de Grenoble. Il se peut que, pour un éventail de cas identique à celui d'autres établissements universitaires, le CHU garde, pour des raisons médicales ou de performances, les patients plus longtemps, et donc en soigne un nombre moins important. Cela aboutit, pour un coût équivalent, à une moindre

production de points ISA. C'est cette troisième hypothèse qui a motivé ce travail de thèse (voir paragraphe "Origine du projet" dans l'introduction).

## 1.4 Une variable pertinente du PMSI : la durée de séjour

La durée de séjour (DS) hospitalière joue un rôle prédominant dans le PMSI : c'est un indicateur important pour la planification et la gestion des lits hospitaliers, elle permet également des études sur les coûts [58, 70, 72, 73, 74, 75]. Elle est en effet fortement corrélée avec les coûts d'hébergement. Les DS sont ainsi un bon reflet de la charge de travail et des besoins financiers d'un service. Elles permettent aussi de mesurer le taux d'occupation des lits à l'intérieur d'un service et, par conséquent, de calculer son "turn-over" ainsi que ses besoins.

Le choix de la DS s'explique aussi par le fait qu'elle est facilement mesurable et qu'elle a l'avantage d'être comparable avec d'autres établissements. Cependant, nous savons que le classement des séjours en GHM n'explique qu'une partie seulement des variations des DS [29, 69, 89].

Nous allons voir de manière plus approfondie les intérêts méthodologique et médico-économique de l'analyse des DS.

### 1.4.1 Notes bibliographiques liées aux intérêts méthodologiques de l'analyse des durées de séjour

Après avoir présenté la pertinence de la variable durée de séjour, nous étudions dans ce paragraphe les différents intérêts méthodologiques liés à l'analyse de cette variable.

#### Forme de la distribution des durées de séjour

Les distributions de DS sont le plus souvent de **forme asymétrique** et de grande variance, due à la présence d'**outliers** (patients ayant des DS extrêmes). Sa forme générale est caractérisée par une longue queue de distribution située à droite (Fig 1.2); par conséquent, il semble difficile d'approximer les distributions des DS par de simples lois normales. Cela empêche aussi l'utilisation des techniques d'inférence statistique fondées sur l'hypothèse de normalité (régression linéaire, comparaison de moyennes et analyse de variance) [40]. Cependant, l'indicateur le plus utilisé actuellement est la DS moyenne : étant donné l'asymétrie de la distribution des DS, la médiane serait un indicateur de position plus approprié [40]. On constate aussi que, plus la DS moyenne est grande, plus l'écart-type est grand, et donc plus l'incertitude de la mesure de la DS moyenne est grande du fait du poids important des DS extrêmes dans le calcul de la moyenne [70]. Les comparaisons de DS moyenne doivent donc tenir compte de la dispersion des distributions. D'autre part, plus l'effectif d'un GHM est petit et plus sa DS moyenne sera grande [70].

La majorité des études publiées s'intéressent aux facteurs explicatifs de la DS, mais peu discutent des types de lois ajustant les distributions de DS.

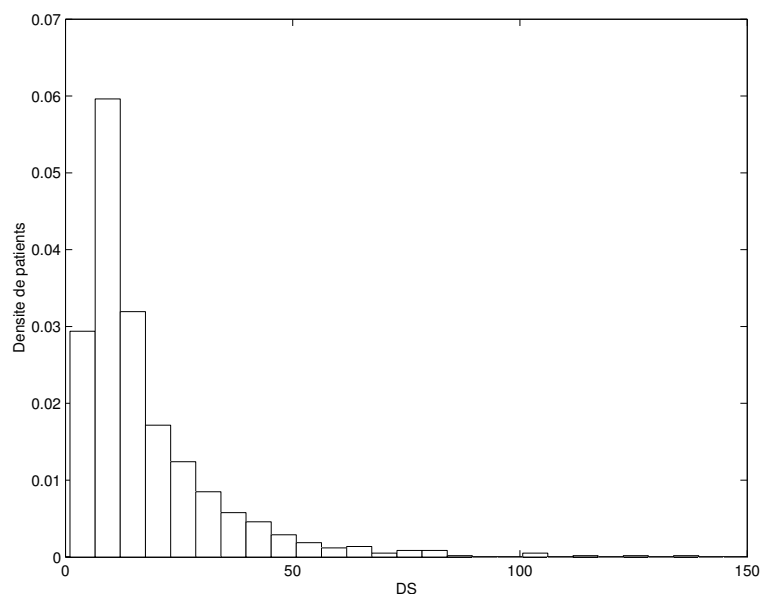


FIG. 1.2 – Exemple de distribution de la variable DS.

### Modèles étudiant les facteurs explicatifs de la durée de séjour

Différents modèles ont été étudiés pour rechercher les variables influant significativement sur les DS: de nombreuses études se ramènent à une situation classique de régression linéaire. En général, cela passe par une transformation logarithmique [24, 81]. L'hypothèse faite est que le logarithme des DS suit une loi normale unimodale (synonyme d'homogénéité de la distribution des DS). L'avantage de la loi log-normale est de pouvoir ajuster des distributions où cohabitent des événements courts et des événements longs, ce qui est le cas des DS. L'étude menée par Vergnenègre et coll. [89] concerne les indicateurs explicatifs de la DS dans un service de pneumologie. Ils ont utilisé un modèle de régression linéaire, mais aussi une analyse factorielle simple sur le logarithme des DS, pour voir quelles étaient les variables qui expliquaient sa variance. Ils ont aussi utilisé un modèle de régression logistique, en créant une nouvelle variable à deux classes calculée à partir de la DS moyenne: les courts séjours correspondaient à toutes les DS strictement inférieure à la DS moyenne, et les longs séjours correspondaient à toutes les DS supérieures ou égales à la DS moyenne. Cette dernière analyse a permis de voir quels étaient les facteurs qui avaient une influence sur les longs séjours. Il s'avère que l'étude des DS entre souvent mal dans le cadre standard de la régression linéaire, même après une transformation logarithmique ou une régression logistique [24, 68, 81, 85, 89]. Dans toutes ces études, aucun test d'adéquation n'a été fait entre les distributions de DS et l'hypothèse faite pour les modèles utilisés.

L'étude de Nighoghossian et coll. utilise un modèle de Cox pour connaître les facteurs influençant la DS des patients victimes d'un accident vasculaire cérébral [69]. L'inconvénient de ce modèle est qu'il suppose que l'effet de chaque variable est le même, quelle que soit la

DS (hypothèse des risques proportionnels non vérifiée).

L'étude de Weil et coll., réalisée sur les DS moyennes des services du CHU de Grenoble entre 1980 et 1990, avait montré qu'à partir de 1984, on voyait apparaître une hétérogénéité des services en terme de distribution de DS moyenne : ceux-ci se répartissaient en 2 groupes (DS moyenne  $\leq 7$  jours et DS moyenne  $> 7$  jours), qu'il est difficile de confirmer 10 ans plus tard, du fait de l'absence de mention du service dans le RSA. Pour ce faire, ils avaient utilisé une analyse en composante principale.

### Lois ajustant les distributions de durées de séjour

Nous allons à présent nous intéresser aux lois statistiques qui peuvent régir les distributions de DS.

L'étude de Pérez et coll. [71] a analysé quelles étaient les distributions statistiques pouvant ajuster les DS d'un service d'urgence, suite à une admission pour une crise d'asthme ou une obstruction pulmonaire chronique liée à la pollution atmosphérique. Ils ont confronté trois modèles de survie paramétriques : exponentiel, Weibull et log-normal. Le modèle log-normal s'est révélé le meilleur.

D'après d'autres études, les lois qui régissent les DS sont plutôt de type exponentiel ou gamma unimodale [25, 50, 58]. Les facteurs influençant les DS sont alors analysés par des modèles linéaires généralisés avec une hypothèse de distribution exponentielle ou gamma.

Le Professeur Marazzi et ses collaborateurs ont cherché quelles étaient les lois qui ajustaient les distributions de DS des DRG (Diagnosis Related Groups : équivalent des GHM français) de 5 pays européens, soit 4 758 347 séjours [62]. Ils ont montré qu'en utilisant la méthode basée sur les "M-estimateurs" [61], la majorité d'entre elles pouvait être modélisée par une ou plusieurs de ces 3 lois unimodales : log-normale, gamma et Weibull. La loi log-normale était en fait celle qui approximait le plus d'échantillons de DS, d'après le test de Cox et le critère ATAR. Cela confirmait les résultats d'une étude concernant les distributions de DS de 230 DRG (87 673 séjours) déjà réalisée en 1993 [78]. Par contre, l'approximation par la loi log-normale des distributions de DS avait tendance à surestimer la DS moyenne par rapport à celles obtenues après une modélisation par la loi gamma ou Weibull. De plus, comme les approximations fournies par les lois gamma et Weibull étaient similaires, il était possible de s'affranchir de la loi gamma, qui a le désavantage d'être difficile à calculer. Certains pays "préféraient" la loi log-normale, d'autres la loi Weibull. Ce phénomène semblait représenter une différence systématique des pratiques médicales dans les hôpitaux entre pays.

Un point très intéressant soulevé par cet article concerne les outliers : souvent, les données concernant ces outliers sont exclues de l'échantillon d'analyse, alors qu'ils sont très intéressants d'un point de vue économique. En effet, ce sont ces patients qui coûtent cher à l'hôpital (patients ayant souvent des pathologies multiples et des DS extrêmes). Ce sont souvent eux qui empêchent une bonne gestion des lits dans les services hospitaliers. Il est donc important de les identifier et de les décrire par des méthodes statistiques adaptées. L'outil statistique fondé sur les "M-estimateurs" utilisé pour cette analyse a l'avantage de prendre en compte ces outliers, mais il n'arrive pas à les isoler ni à les décrire.



Ainsi, si quelques distributions de DS s'avèrent unimodales, une majorité d'entre elles sont plurimodales, indiquant la présence de sous-groupes de patients, en particulier au sein des GHM à larges spectres [53]. En Suisse, dès 1984, la VESKA (association suisse des hôpitaux) a analysé de manière systématique les formes des distributions de DS de tous les DRG de plus de 200 séjours et a permis de mettre en évidence que 20 DRG avaient une distribution bimodale [70]. En France, des études réalisées par le Professeur Quantin et son équipe ont montré que les distributions des DS et des coûts de la plupart des GHM étaient plurimodales et se décomposaient en trois sous-groupes [52, 72, 73, 74, 75], faisant ainsi la preuve d'une hétérogénéité des distributions de DS.

L'inconvénient des modèles présentés jusqu'ici est que, même s'ils prennent en compte l'asymétrie des distributions de DS, ils ne considèrent pas leur éventuelle hétérogénéité.

### **Modèles décrivant l'hétérogénéité des distributions de durées de séjour**

Nous voyons donc qu'il est important d'utiliser des méthodes statistiques robustes, qui prennent en compte l'hétérogénéité intra-GHM de la distribution des DS et son asymétrie, afin de détecter une majorité d'outliers et de pouvoir les décrire. Cette hétérogénéité est inhérente à la variabilité des malades, aux pratiques des médecins et aux structures des établissements, mais aussi à la dispersion introduite par la description de l'information médico-sociale des patients dans le PMSI [72].

La méthode statistique répondant à ce problème est le modèle de mélange de distributions. Il permet d'isoler des sous-groupes de façon non arbitraire, d'en estimer les proportions et d'en rechercher les critères cliniques explicatifs.

Au Royaume-Uni, le Professeur Millard (responsable du service de gériatrie de l'hôpital Saint Georges à Londres) et ses collaborateurs se sont beaucoup intéressés aux modèles de mélange, pour comprendre comment se structuraient les DS dans les services de gériatrie, afin de caractériser les flux de patients, permettre une meilleure prise en charge des personnes âgées, mais aussi une meilleure gestion des lits [63]. Ils ont ensuite mené le même travail dans d'autres secteurs hospitaliers. L'hypothèse faite est que les distributions de DS suivent un mélange de lois exponentielles. L'estimation des paramètres des modèles a été faite par la méthode des moindres carrés [82]. Ils ont montré que le flux des patients dans les services de gériatrie, de psychiatrie et les maisons de retraite se répartissait en trois sous-groupes, et que le flux des patients dans les services de court séjour se structurait en deux sous-groupes [67]. À cette occasion, S. Wyatt a développé le logiciel BOMPS (Bed Occupancy Management and Planning System) qui automatise cette méthode [91].

En France, le Professeur Quantin et ses collaborateurs se sont intéressés aux mélanges pour modéliser des distributions de DS par GHM. Ils ont baptisé leur méthode : "Justification clinique des séjours coûteux d'un GHM" [41]. Elle consiste, dans un premier temps, à déterminer les principaux sous-groupes qui composent un GHM, chaque sous-groupe étant

caractérisé par sa moyenne, en terme de coût ou de DS, et par la proportionnalité des séjours du GHM qu'il représente. Dans un deuxième temps, la proportion de ces sous-groupes est déterminée pour chaque établissement. Il est alors possible de distinguer des situations divergeant de la référence nationale. Ainsi, les hôpitaux qui présentent une proportion du sous-groupe coûteux significativement plus élevée que la proportion nationale, peuvent être considérés comme "hors normes" pour le GHM étudié.

Ils ont appliqué cette méthode à trois études :

- la première s'intéressait à la modélisation des DS et des coûts dans deux GHM de malades : hématologie (GHM n°590) et pneumologie (GHM n°128), afin de caractériser cliniquement les patients de longue durée et de coût élevé [72] ;
- la deuxième traitait de l'analyse des séjours de longue durée pour des GHM de malades de néonatalogie (GHM n°873 à n°876) [52] ;
- la troisième évoquait la modélisation des facteurs de coûts dans le cas de patients atteints de leucémies ou de lymphomes (GHM n°589 et n°590) [73].

La première étude a montré que les distributions de DS sont mieux ajustées par des mélanges de lois de Weibull que par des mélanges de lois de Poisson. En effet, les vraisemblances des mélanges de lois de Weibull étaient légèrement supérieures à celles des modèles construits à partir des mélanges de lois de Poisson. L'hypothèse faite par les deux études suivantes est que les DS suivent un mélange de lois de Weibull. L'estimation des paramètres des mélanges a été faite par la méthode du maximum de vraisemblance. Le choix du nombre de sous-groupes dans les modèles a été analysé par un test du rapport de vraisemblance au seuil de 5%. Ces trois études ont montré que chacune des distributions de DS ou de coûts se structurait en trois sous-groupes, quels que soient les GHM d'intérêt.

## Synthèse

Nous voyons que les lois (log-normale, Weibull, gamma, exponentielle) régissant les distributions de DS et le nombre de leurs modes varient selon les pays et les GHM. Il faudra donc déterminer la loi qui ajuste le mieux ces distributions, mais aussi utiliser des méthodes statistiques robustes qui prennent en compte et décrivent l'hétérogénéité intra-GHM des DS, comme les modèles de mélange.

Concernant des approches statistiques mises au point pour la classification des GHM en terme de distribution de DS, mais aussi pour l'identification des "mauvais" GHM parmi un groupe de GHM donné, aucune référence bibliographique n'a été trouvée.

### 1.4.2 Notes bibliographiques liées aux raisons médico-économiques de l'analyse des durées de séjour

Après avoir présenté les intérêts méthodologiques liés à l'étude des DS, nous nous intéressons aux raisons médico-économiques de cette analyse. En effet, les DS sont la résultante d'un

phénomène multivarié. Ainsi les quelques variables dont on dispose dans le PMSI ne suffisent pas à les expliquer complètement car elles sont le résultat d'autres variables que nous ignorons.

### Catégories d'établissements hospitaliers

Le secteur public hospitalier (27.7% des établissements) représente 69.8% de la capacité d'accueil de ce secteur et gère la moitié des lits de courts séjours [30]. Il est intéressant de regarder, en détail, les différents établissements hospitaliers et leurs activités.

Les Centres Hospitaliers Régionaux (CHR), pour la plupart CHU, représentent 2.7% (29) des établissements publics, mais regroupent à eux seul 21.4% des lits du secteur public (voir tableau 1.1). Ils correspondent à des hôpitaux de recours régional, et, pour les CHU, de référence universitaire. Ils se consacrent de plus en plus aux traitements lourds et innovants, qui réclament d'importants investissements techniques et humains. Les CHR concentrent 70% des lits de chirurgie digestive et de neurologie, la plupart des lits de réanimation, de surveillance continue et de chirurgie thoracique cardio-vasculaire.

Les centres hospitaliers (CH) représentent la majorité du secteur public (52.9%) et regroupent 53.5% des lits du secteur public (voir tableau 1.1). Ils correspondent à des hôpitaux de recours départemental. Ils ont développé des secteurs complémentaires à ceux des CHR : médecine générale, chirurgie générale, pédiatrie, rhumatologie et pneumologie. De plus, en liaison avec leurs services d'urgence, les services de traumatologie des CH prennent en charge les traumatismes les plus graves.

Les hôpitaux locaux représentent 32.9% du secteur public mais ne gèrent que 12.3% des lits de ce secteur (voir tableau 1.1). Ce sont des établissements à vocation rurale ou dits de proximité, qui n'assurent que des soins aigus de médecine.

Le CHU de Grenoble est un des hôpitaux de référence universitaire pour la région Rhône-Alpes. Pour cette raison, il a tendance à recruter une proportion plus élevée de patients compliqués qui engendrent des séjours coûteux pour l'hôpital. Il est donc important de pouvoir quantifier le taux de ces patients.

### Hétérogénéité intra-GHM

Comme nous l'avons vu précédemment, le PMSI conduit à un système de régulation financière dit "prospectif" fondé sur les GHM (type de pathologie). Il est vrai que la construction en GHM pour décrire l'activité hospitalière constitue un progrès important, mais il reste toutefois perfectible : l'interprétation médicale des groupes constitués apparaît parfois difficile et, à l'intérieur d'un même GHM, la dispersion du volume des ressources consommées par séjour semble importante. Beaucoup reconnaissent aujourd'hui que la variabilité du coût par séjour ne reflète pas uniquement l'efficacité de la prise en charge des cas traités, appréhendée par la classification par GHM [37]. Même si la classification par GHM a été établie dans un souci d'homogénéité intra-groupe, celle-ci reste imparfaite. En effet, des études ont montré

TAB. 1.1 – Répartition des différentes catégories d'établissements hospitaliers (au 1<sup>er</sup> janvier 1996).

Catégorie d'établissement	Établissements		Lits et places				Ensemble des lits et places (%)
	Effectif (%)	Lits en hospitalisation complète	Lits d'hébergement	Places en hospitalisation de jour ou de nuit	Places en anesthésie ou chirurgie ambulatoire	Places en hospitalisation à domicile	
<b>Public</b>							
Centre hospitalier régional	29 (2.7)	88 866	3 640	4 885	464	317	98 172 (21.4)
Centre hospitalier	562 (52.9)	171 929	63 306	7 275	399	324	246 233 (53.5)
Hôpital local	350 (32.9)	23 292	33 403	51	-	-	56 746 (12.3)
Centre hospitalier spécialisé en psychiatrie	97 (9.1)	39 743	1 371	14 575	-	554	56 243 (12.2)
Autres établissements	25 (2.4)	1 681	688	13	6	10	2 398 (0)
<b>Total public</b>	<b>1 063 (100)</b>	<b>328 511</b>	<b>102 408</b>	<b>26 799</b>	<b>869</b>	<b>1 205</b>	<b>459 791 (100)</b>
<b>Privé</b>							
Établissement de soins de courte durée	1 297 (44.2)	111 150	465	1 227	-	-	112 842 (56.5)
Établissement d'enfants à caractère sanitaire	82 (3.0)	4 707	66	132	-	-	4 905
Établissement de soins de suite et réadaptation, de soins de longue durée	727 (26.2)	48 284	1 498	2 684	16	-	52 482 (26.3)
Hôpital psychiatrique privé faisant fonction de public	21 (0.8)	8 714	119	2 093	-	-	10 926 (5.5)
Autre établissement de lutte contre les maladies mentales et toxicomanie	230 (8.3)	15 724	-	3 299	13	5	19 041 (9.5)
Traitements et soins à domicile, dialyse ambulatoire	496 (17.9)	48	-	-	-	1 213	1 261 (0)
<b>Total privé</b>	<b>2 773 (100)</b>	<b>179 564</b>	<b>2 824</b>	<b>10 422</b>	<b>5 666</b>	<b>1 271</b>	<b>199 747 (100)</b>
<b>Ensemble</b>	<b>3 836</b>	<b>508 075</b>	<b>105 232</b>	<b>37 221</b>	<b>6 535</b>	<b>2 476</b>	<b>659 189</b>

Source : Ministère de l'Emploi et de la Solidarité - SESI - Données issues de l'Annuaire des statistiques sanitaires et sociales 1998, page 219

qu'il existait une hétérogénéité intra-GHM en termes de DS ou de coûts, inhérente à la variabilité des malades, aux pratiques médicales, et à la structure des établissements (équipements, environnement socio-économique de l'hôpital, taille de celui-ci et filières de soins, existence de plusieurs services traitants dans la même spécialité médicale ou chirurgicale, ...) [41].

Au vu de cette hétérogénéité des patients dans certains GHM, il semble nécessaire de tenir compte des séjours particulièrement coûteux. Dans le cadre d'une dotation globale de financement, indexée sur l'activité médicale (valeur du point ISA) exprimée par le PMSI, certains établissements, comme les CHU, ont des coûts, pour des GHM donnés, plus élevés que les coûts de référence nationale, du fait de proportions anormalement élevées de patients coûteux. Cependant, ces écarts par rapport à la référence ne sont pas pris en compte par la méthode d'allocation budgétaire utilisée par le PMSI. Ce système pourrait être sujet à caution si les variations de sévérité des pathologies des patients, de pratiques médicales ou de structures d'établissements conduisent à la présence de sous-groupes de différentes DS et de différentes dépenses à l'intérieur des GHM [52, 72, 73, 74, 75]. En effet, seules les variations dues à la sévérité des cas sont acceptables, les variations liées aux pratiques médicales ou aux structures ne sont pas tolérables. Ainsi, dès lors que les surcoûts sont imputables à des sous-groupes de patients coûteux, l'allocation budgétaire basée sur ces informations serait insuffisante pour couvrir l'ensemble des dépenses justifiées par ces patients [41]. Cela pose alors le problème de l'équité de la répartition des allocations entre les hôpitaux. La méthode de "justification clinique des patients coûteux" a l'avantage de repérer les établissements considérés à tort comme sur-dotés du fait de la prise en charge d'une proportion très élevée de patients coûteux. Elle permet aussi d'affiner la description de l'activité hospitalière sans remettre en cause la classification en GHM [75].

D'autre part, ces patients coûteux repérés par leur DS peuvent aussi correspondre à une nouvelle entité typologique, donnant alors une dimension épidémiologique au PMSI. On trouvera dans [84] une étude épidémiologique utilisant les données du poids de naissance issues du PMSI.

En conséquence, il semble intéressant de répondre au problème suivant : le mauvais classement des hôpitaux, du point de vue de la valeur moyenne de leur point ISA, cas du CHU de Grenoble en 1996, reflète-il des variations de pratiques médicales ou un recrutement de patients plus complexes ?

Pour ce faire, il est nécessaire de mettre en œuvre des méthodes permettant de repérer les GHM comportant une proportion anormalement élevée de patients coûteux, en termes de DS, (présence d'outliers). Il est également important d'en analyser l'origine. Ainsi, différents malades appartenant à un même GHM auront des DS très diverses et consommeront des valeurs très variées de ressources hospitalières, suivant le degré de gravité de leur maladie et le motif d'hospitalisation [55]. Or, les variables descriptives des GHM ne rendent compte ni de l'un, ni de l'autre. Cela pourrait fournir aux établissements des outils de négociation d'allocation complémentaire par la justification clinique des surcoûts qu'ils supportent. Comme nous l'avons vu dans les travaux des professeurs Millard et Quantin [52, 63, 67, 72, 73, 74, 75, 91], la

DS peut être un indicateur simple et pertinent pour repérer les patients coûteux, à condition de choisir des modèles statistiques adaptés (modèle de mélanges de lois) pour les analyses (voir paragraphe 1.4.1).

De plus, lorsqu'un hôpital apprend qu'il est identifié comme étant sur-doté ou sous-doté au niveau de son budget global, il a besoin de comprendre les éléments qui font que, in fine, il bénéficie d'un tel jugement global de productivité [33]. Un des premiers éléments qui vient à l'esprit est l'interrogation sur la DS moyenne. En effet, puisque le nombre de points ISA est accordé à un séjour, indépendamment de sa durée concrète dans tel hôpital, le fait de garder ce type de malades plus longtemps qu'ailleurs en hospitalisation peut être une forte raison d'un surcoût qui peut être rédhibitoire, quelles que soient les bonnes données sur les autres indications. Le coût à la journée, surtout lorsqu'elle est inutile, crée une part probablement importante des surcoûts. Avec le PMSI, on dispose d'un outil performant de ce point de vue. Mais dans certains cas, le séjour est légitimement long et donc mal pris en compte par le GHM dans lequel il est classé.

### Capacité en lits et niveau d'activité

Depuis 1980, un nombre important de lits de courts séjours ont fermé, principalement dans le secteur public. L'instauration du budget global a accentué cette évolution. L'augmentation de l'activité a alors été rendue possible par une diminution de la DS [64]. Ainsi, une étude réalisée au CHU de Grenoble par le professeur Demongeot et son équipe montrait que dès 1984, 2 catégories de services étaient identifiables : ceux qui essayaient de faire sortir leurs patients avant le week-end, mais qui souffraient d'un faible taux d'occupation, et les autres. Ces services avaient la particularité d'être dirigés, dans la plupart des cas, par des professeurs de la nouvelle génération, préférant faire sortir les patients le plus rapidement possible [90]. On peut donc penser que, dès le début des années 80, il préexistait une tendance lourde à la diminution des DS, dans les hôpitaux, due à l'arrivée d'une génération de chefs de service plus sensibilisé à l'utilisation du PMSI. Cette diminution a aussi plusieurs autres explications : les progrès technologiques, l'intérêt évident pour les hôpitaux à faire sortir les patients plus vite et la réalisation d'un maximum d'examen en ambulatoire. En particulier, les DS préopératoires ont été considérablement raccourcies.

Comme nous l'avons vu précédemment, la production d'un établissement correspond aux nombres de séjours répartis par GHM. Le producteur de soins adopte un comportement de minimisation des coûts, qui est entraîné par le budget global [64]. Pour porter un jugement complet sur l'existence d'un comportement de minimisation des coûts, il faudrait connaître l'évolution du coût des séjours ajusté sur la lourdeur des cas. Or cette information n'est pas encore disponible dans le PMSI.

D'autre part, les données du PMSI étant anonymisées, il n'est pas possible de rattacher les séjours aux unités de production que sont les services, ce qui ne nous permet pas d'ajuster les données sur la relative adéquation entre la capacité en lits et le niveau de l'activité (le taux d'occupation) [64]. Malheureusement, il n'a pas été possible de calculer les coûts par

séjour car l'hôpital de Grenoble ne faisait pas partie de l'étude nationale des coûts.

### 1.4.3 Synthèse

Il est donc nécessaire, dans un premier temps, de mettre en œuvre des méthodes statistiques fiables permettant de faire un premier screening et un tri des GHM grenoblois afin d'identifier ceux qui posent un problème en termes de DS, comparativement à leurs homologues de la base de référence nationale. C'est en quelque sorte une analyse inter-GHM. C'est ce que nous avons étudié dans les chapitres 3, 4 et 5. Dans un deuxième temps, il est important d'analyser les causes de l'allongement des DS pour les GHM posant des problèmes, grâce aux informations contenues dans les RSA des patients : est-elle due à des variations de pratiques médicales ou à un recrutement de patients plus complexes au CHU de Grenoble ? Pour répondre à cette question, il faut appliquer des méthodes statistiques permettant de prendre en compte l'hétérogénéité des DS afin de repérer les RSS à aller examiner de plus près, en quelque sorte un approche intra-GHM. C'est l'objet du chapitre 6. À terme, cela pourrait fournir au CHU, un outil de négociation d'allocation complémentaire par la justification clinique des surcoûts que ces patients entraînent.

# Chapitre 2

## Population étudiée et listes de référence

Après avoir présenté les buts et le fonctionnement du PMSI, la problématique liée au CHU de Grenoble et les raisons qui nous ont amenés à privilégier la variable DS et ses caractéristiques, nous allons maintenant nous intéresser aux données avec lesquelles ce travail a été réalisé. Nous allons d'abord présenter les méthodes qui nous ont permis d'élaborer les profils de distributions de DS des 257 GHM grenoblois et les listes de référence qui vont nous aider à évaluer la pertinence médico-économique des futures classifications de profils. Ensuite, nous les décrirons et nous les discuterons d'un point de vue médico-économique.

### 2.1 Présentation des données

Ce paragraphe présente les données que nous avons étudiées.

#### 2.1.1 Données issues de fichiers PMSI

Nous avons travaillé sur les données de deux bases de données (BDD) du PMSI concaténées :

- l'une est celle du CHU de Grenoble,
- l'autre correspond à un tirage aléatoire de 5% de la BDD des 29 CHU français (BDD nationale).

Ce fichier contient toutes les données médico-administratives anonymisées de 140 825 résumés de séjours de patients admis entre le 1er janvier 1997 et le 31 décembre 1997. Ils se répartissent comme suit :

- 54 650 RSA pour le CHU de Grenoble,
- 86 175 RSA pour la BDD nationale.



### 2.1.2 Variables d'intérêt pour construire le fichier des profils

Pour établir le fichier regroupant les profils de distribution de DS des GHM grenoblois, nous avons gardé les variables du PMSI suivantes :

- la DS du patient au CHU (en jours), variable “phare” de l'étude ;
- le numéro du GHM dans lequel les patients ont été enregistrés à leur sortie du CHU ;
- le type de GHM : médical ou chirurgical,
- le numéro de la BDD qui vaut :
  - \* “0”, si le patient appartient à la BDD du CHU de Grenoble ;
  - \* “1”, si le patient appartient à la BDD nationale.

Ce travail se positionnant plutôt dans une logique économique que médicale, nous avons donc considéré une approche pragmatique et non explicative des DS. Autrement dit, nous avons travaillé sur les distributions marginales des DS, c'est à dire que nous avons considéré toutes les DS sans tenir compte du statut du patient à sa sortie de l'hôpital : vivant guéri, vivant non guéri ou décédé. Comme on peut le remarquer, la logique économique ne recouvre pas ici la logique médicale. En effet, d'un point de vue purement économique, un patient qui sort vivant ou décédé rapporte le même nombre de point ISA au CHU, cela dépend de son GHM d'affectation. Cependant, on est conscient que les DS concernant des patients décédés ne parlent pas beaucoup en terme de performance hospitalière d'un point de vue médical.

### 2.1.3 Critères d'exclusion

N'ont pas été inclus dans le fichier d'analyse, les patients qui :

- appartiennent à des GHM ambulatoires correspondant à la Catégorie Majeure (CM) numéro 24 (DS = 0) ;
- appartiennent à un des GHM de la CM numéro 90 (erreurs et autres séjours inclassables) ;
- sont restés hospitalisés plus de 200 jours (DS > 200) ;
- ne sont venus que pour des séances ponctuelles sous le même numéro de séjour (typiquement les séances de chimiothérapie). La variable “Nombre de Séances” du fichier PMSI est alors  $\neq 0$ .

D'autre part, ce fichier ne contient que les GHM communs aux 2 BDD et dont l'effectif de patients est supérieur ou égal à 50. Au total, 257 GHM ont été retenus pour l'étude (voir annexe D).

## 2.2 Élaboration des profils grenoblois de distribution de DS

Après avoir présenté les données sur lesquelles nous avons travaillé, ce paragraphe présente la démarche qui nous a permis d'élaborer un fichier de données **unique**, permettant de

comparer, pour chaque GHM grenoblois, son profil de distribution de DS à celui de son homologue de la BDD nationale (la référence). Nous décrivons, ci-après, la méthode utilisée.

### 2.2.1 Indicateurs utilisés pour l'élaboration des profils de distributions

Pour établir les profils de distribution de DS, nous avons choisi de résumer les distributions de DS à partir de 3 indicateurs : les trois premiers intervalles interquartiles de DS en proportions cumulées.

### 2.2.2 Démarche suivie

À partir des fichiers PMSI de 1997, du CHU de Grenoble et de la base nationale, nous avons calculé, pour chaque GHM :

- les quartiles de DS ( $q_{(0.25)}$ ,  $q_{(0.50)}$ , et  $q_{(0.75)}$ ), à partir de leur fonction de répartition empirique, pour le CHU Grenoble et la base nationale,
- les proportions de patients grenoblois dans les 4 intervalles interquartiles de DS nationaux formés autour des quartiles nationaux  $q_{(0.25)}$ ,  $q_{(0.50)}$  et  $q_{(0.75)}$ ,
- les proportions cumulées de patients grenoblois ayant des DS inférieures ou égales aux DS correspondant à 25%, 50% et 75% des patients de la base nationale.

#### Définition d'un quartile :

Les quartiles  $q_{(u)}$  ( $u = 0.25, 0.50$  et  $0.75$ ) d'une variable  $\mathbf{X}$  correspondent aux 3 valeurs  $x_u$  qui partagent en quatre parties d'effectifs égaux (25%), la série des valeurs observées de  $\mathbf{X}$  et classées par ordre croissant (voir graphique 2.1). Ils vérifient les 2 égalités suivantes :

$$\begin{aligned} P(X < x_u) &\leq u \text{ et} \\ P(X > x_u) &\geq (1 - u). \end{aligned}$$

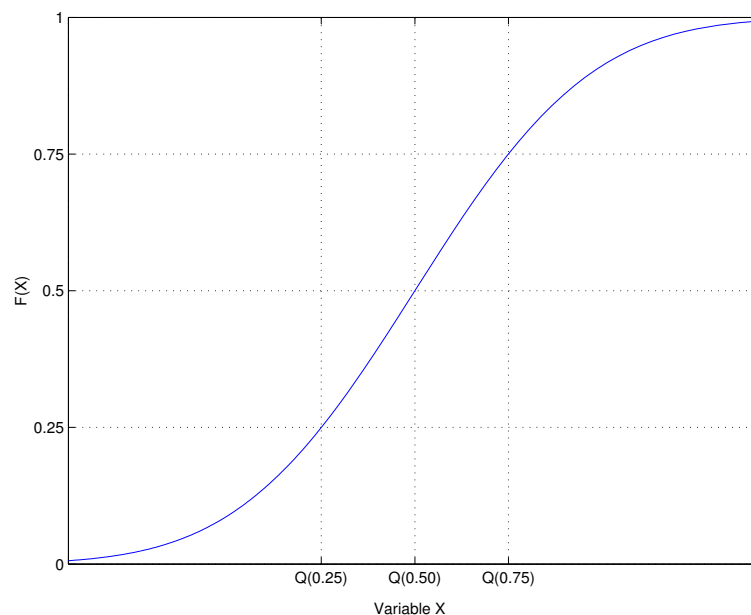


FIG. 2.1 – Représentation graphique des quartiles d'une distribution.

### Comparaison des quartiles de DS entre les bases nationale et grenobloise

Pour établir les méthodes permettant de calculer les proportions de patients grenoblois dans les intervalles interquartiles de DS nationaux, nous avons commencé par comparer, pour chaque GHM, les 3 quartiles de DS entre le CHU de Grenoble et la base nationale. Il y a 31 configurations possibles réparties en 4 grandes familles et, pour chacune d'entre elles, 7 ou 8 cas (détail de ces 31 configurations en annexe E). Nous en avons rencontré 19 dans notre échantillon. Cela a abouti à 5 méthodes de calcul qui ont regroupé les configurations de la manière suivante :

- Méthode 1 (212 GHM) : cas **1A**, 1B, 1C, 1D, 1E, 1F et 1G,
- Méthode 2 (14 GHM) : cas 2A, **2B**, 2E, 2F et 2G,
- Méthode 3 (2 GHM) : cas 3C et **3G**,
- Méthode 4 (2 GHM) : cas **4E**,
- Méthode 5 (27 GHM) : cas **1H**, 2H, 3H et 4H.

**En gras**, nous avons noté les 5 configurations utilisant les différentes méthodes de calcul, que nous allons présenter.

Notons :

- $q_{(0.25)}^G$  : la DS correspondant à 25% des patients grenoblois,
- $q_{(0.50)}^G$  : la DS correspondant à 50% des patients grenoblois,
- $q_{(0.75)}^G$  : la DS correspondant à 75% des patients grenoblois,

- $q_{(0.25)}^N$  : la DS correspondant à 25% des patients de la base nationale,
- $q_{(0.50)}^N$  : la DS correspondant à 50% des patients de la base nationale,
- $q_{(0.75)}^N$  : la DS correspondant à 75% des patients de la base nationale,
  
- $I_{(1)}$  : le premier intervalle interquartile<sup>1</sup> qui réunit tous les patients grenoblois avec une  $DS \leq q_{(0.25)}^N$ ,
- $I_{(2)}$  : le deuxième intervalle interquartile qui rassemble tous les patients grenoblois tels que  $q_{(0.25)}^N < DS \leq q_{(0.50)}^N$ ,
- $I_{(3)}$  : le troisième intervalle interquartile qui regroupe tous les patients grenoblois tels que  $q_{(0.50)}^N < DS \leq q_{(0.75)}^N$ ,
- $I_{(4)}$  : le quatrième intervalle interquartile qui réunit tous les patients qui ont une  $DS > q_{(0.75)}^N$ ,
  
- $Ic_{(1)}$  : le premier intervalle interquartile en proportion cumulée qui réunit tous les patients grenoblois dont la  $DS \leq q_{(0.25)}^N$ ,
- $Ic_{(2)}$  : le deuxième intervalle interquartile en proportion cumulée qui rassemble tous les patients grenoblois avec une  $DS \leq q_{(0.50)}^N$ ,
- $Ic_{(3)}$  : le troisième intervalle interquartile en proportion cumulée qui regroupe tous les patients grenoblois dont la  $DS \leq q_{(0.75)}^N$ ,
- $Ic_{(4)}$  : le quatrième intervalle interquartile en proportion cumulée qui réunit l'ensemble des patients grenoblois. Il est toujours égal à 100% et n'apporte donc pas d'information supplémentaire.

\* Cas 1A :

- $q_{(0.25)}^N \neq q_{(0.50)}^N \neq q_{(0.75)}^N$  et
- $q_{(0.25)}^N \neq q_{(0.25)}^G$ ,
- $q_{(0.50)}^N \neq q_{(0.25)}^G$ ,
- $q_{(0.75)}^N \neq q_{(0.25)}^G$ .

Nous illustrerons ce cas avec le GHM n°48 : interventions sur la rétine (voir quartiles de DS, tableau 2.1, et fonctions de répartition empiriques, figure 2.2).

\* Cas 2B :

- $q_{(0.25)}^N = q_{(0.50)}^N \neq q_{(0.75)}^N$  et
- $q_{(0.25)}^G = q_{(0.25)}^N$ ,
- $q_{(0.50)}^G \neq q_{(0.50)}^N$ ,

---

1. Les 4 intervalles interquartiles sont en fait relatifs à la base nationale, puisqu'ils sont calculés en fonction des quartiles de DS nationaux.

$$- q_{(0.75)}^G \neq q_{(0.75)}^N.$$

Nous illustrerons ce cas avec le GHM n°51 : interventions sur le cristallin avec ou sans vitrectomie (voir quartiles de DS, tableau 2.2, et fonctions de répartition empiriques, figure 2.6).

\* Cas 3G :

$$- q_{(0.25)}^N \neq q_{(0.50)}^N = q_{(0.75)}^N \text{ et}$$

$$- q_{(0.25)}^G = q_{(0.25)}^N,$$

$$- q_{(0.50)}^G \neq q_{(0.50)}^N,$$

$$- q_{(0.75)}^G = q_{(0.75)}^N.$$

Nous illustrerons ce cas avec le GHM n°874 : nouveau-nés de 2 500g et plus, sans problème significatif (voir quartiles de DS, tableau 2.3 et fonctions de répartition empiriques, figure 2.10).

\* Cas 4E :

$$- q_{(0.25)}^N = q_{(0.50)}^N = q_{(0.75)}^N \text{ et}$$

$$- q_{(0.25)}^G = q_{(0.25)}^N,$$

$$- q_{(0.50)}^G = q_{(0.50)}^N,$$

$$- q_{(0.75)}^G \neq q_{(0.75)}^N.$$

Nous illustrerons ce cas avec le GHM n°52 : autres interventions extra-oculaires, âge supérieur à 17 ans (voir quartiles de DS, tableau 2.4 et fonctions de répartition empirique, figure 2.14).

\* Cas 1H :

$$- q_{(0.25)}^N \neq q_{(0.50)}^N \neq q_{(0.75)}^N \text{ et}$$

$$- q_{(0.25)}^G = q_{(0.25)}^N,$$

$$- q_{(0.50)}^G = q_{(0.50)}^N,$$

$$- q_{(0.75)}^G = q_{(0.75)}^N.$$

Nous illustrerons ce cas avec le GHM n°118 : embolies pulmonaires (voir quartiles de DS, tableau 2.5 et fonctions de répartition empiriques, figure 2.18).

### Calcul des proportions de patients grenoblois dans les intervalles interquartiles de DS nationaux

Ces calculs nous ont permis d'obtenir un fichier unique de données avec comme référence la base nationale.

TAB. 2.1 – *Quartiles de DS du GHM numéro 48 : interventions sur la rétine (Cas 1A).*

BDD	$q(0.25)$	$q(0.50)$	$q(0.75)$
Nationale	2	4	6
Grenoble	7	9	12

TAB. 2.2 – *Quartiles de DS du GHM numéro 51 : interventions sur le cristallin avec ou sans vitrectomie (Cas 2B).*

BDD	$q(0.25)$	$q(0.50)$	$q(0.75)$
Nationale	2	2	3
Grenoble	2	3	5

TAB. 2.3 – *Quartiles de DS du GHM numéro 874 : nouveau-nés de 2 500g et plus, sans problème significatif (Cas 3G).*

BDD	$q(0.25)$	$q(0.50)$	$q(0.75)$
Nationale	4	5	5
Grenoble	4	4	5

TAB. 2.4 – *Quartiles de DS du GHM numéro 52 : autres interventions extra-oculaires, âge supérieur à 17 ans (Cas 4E).*

BDD	$q(0.25)$	$q(0.50)$	$q(0.75)$
Nationale	2	2	2
Grenoble	2	2	3

TAB. 2.5 – *Quartiles de DS du GHM numéro 118 : embolies pulmonaires (Cas 1H).*

BDD	$q(0.25)$	$q(0.50)$	$q(0.75)$
Nationale	7	10	15
Grenoble	7	10	15

Le cas 1A, illustré par les patients ayant eu une intervention sur la rétine [GHM n°48], ne pose pas de problème pour le calcul des proportions de patients grenoblois dans les 4 intervalles interquartiles nationaux de DS. En effet, les quartiles de DS de la base nationale sont tous différents les uns des autres (voir figure 2.2 et tableau 2.1). Nous avons utilisé la méthode 1, la plus générale et :

- le premier intervalle interquartile,  $I_{(1)}$ , rassemble les patients grenoblois ayant une DS inférieure ou égale à 2 jours ( $q_{(0.25)^N}$ ), soit 1.7%,
- le deuxième intervalle interquartile,  $I_{(2)}$ , regroupe les patients ayant une DS strictement supérieure à 2 jours et inférieure ou égale à 4 jours ( $q_{(0.50)^N}$ ), soit 3.9%,
- dans le troisième intervalle interquartile,  $I_{(3)}$ , se trouvent les patients ayant une DS strictement supérieure à 4 jours et inférieure ou égale à 6 jours ( $q_{(0.75)^N}$ ), soit 12.7%,
- le quatrième intervalle interquartile,  $I_{(4)}$ , réunit les patients avec une DS strictement supérieure à 6 jours ( $q_{(0.75)^N}$ ), soit 81.7%.

Par contre, les cas 2B, 3G et 4E sont un peu plus complexes. En effet, deux ou trois des quartiles de DS de la base nationale sont égaux (voir figures 2.6, 2.10, 2.14 et tableaux 2.2, 2.3, 2.4); nous avons dû mettre en œuvre 3 méthodes de calculs, adaptées à chacun de ces cas. Elles font l’hypothèse suivante: les proportions de patients grenoblois sont **égales** dans les intervalles interquartiles formés entre les quartiles nationaux de DS. Pour le calcul des proportions sur les autres intervalles interquartiles, rien ne change. Ainsi, pour ces 3 configurations :

- Cas 2B (méthode 2), illustré par les patients ayant eu une intervention sur le cristallin avec ou sans vitrectomie [GHM n°51] (voir figure 2.7) : les proportions de patients grenoblois dans  $I_{(1)}$  et  $I_{(2)}$  sont équivalentes (13.0%); elles correspondent à la moitié de la proportion de patients ayant une DS inférieure ou égale à 2 jours ( $q_{(0.50)^N}$ ).
- Cas 3G (méthode 3), illustré par les mères ayant eu un nouveau-né de 2 500g et plus, sans problème significatif [GHM n°874] (voir figure 2.11) : les proportions de patientes sont égales dans  $I_{(2)}$  et  $I_{(3)}$  (8.0%); elles correspondent à la moitié des patientes grenobloises ayant une DS entre 4 et 5 jours ( $q_{(0.50)^N}$  et  $q_{(0.75)^N}$ ).
- Cas 4E (méthode 4), illustré par les patients âgés de plus de 17 ans et ayant eu une intervention extra-oculaire [GHM n°52] (voir figure 2.15) : les proportions de patients grenoblois dans  $I_{(1)}$ ,  $I_{(2)}$  et  $I_{(3)}$  sont identiques (21.8%), elles sont égales au tiers de la proportion de patients grenoblois ayant une DS inférieure ou égale à 2 jours ( $q_{(0.25)^N}$ ,  $q_{(0.50)^N}$  et  $q_{(0.75)^N}$ ).

Certes, les hypothèses faites pour les 3 méthodes de calcul des proportions de patients grenoblois dans les intervalles interquartiles ont tendance à avantager les GHM grenoblois. En effet, avec ces 3 méthodes, la proportion de patients est répartie de manière égale dans les intervalles formés entre les quartiles de DS égaux, alors qu’en appliquant la méthode 1, on obtiendrait des effectifs nuls dans un des deux, voire des trois intervalles égaux. Cela aurait été trop pénalisant pour le CHU de Grenoble, n’aurait pas représenté la réalité et aurait posé

certaines problèmes lors des analyses statistiques (convergence d’algorithme, etc.)

Dans le cas 1H, illustré par les patients ayant fait une embolie pulmonaire [GHM n°118], nous avons fait une hypothèse très simple : les trois quartiles de DS étant égaux entre les bases nationale et grenobloise (voir figure 2.18 et tableau 2.5), les distributions de DS sont identiques. Ainsi, les proportions de patients dans les 4 intervalles interquartiles,  $I_{(1)}$ ,  $I_{(2)}$ ,  $I_{(3)}$  et  $I_{(4)}$ , sont toutes **égales** à 25.0% (voir tableau 2.19). Elles correspondent au quart de la proportion totale de patients grenoblois.

Certes, l’hypothèse faite est très forte, mais cela n’aura pas beaucoup d’effet sur la classification. En effet, ces GHM, classés comme “neutres”, ne nous intéressent pas vraiment d’un point de vue médico-économique car ils ne sont pas “atypiques” en terme de profil de distribution de DS, tout au moins sur les indicateurs que nous avons choisis. Cette hypothèse a même tendance à pénaliser les GHM grenoblois, étant donné qu’ils pourraient compter plus de 25% de patients sur certains des intervalles si on appliquait la méthode 1.



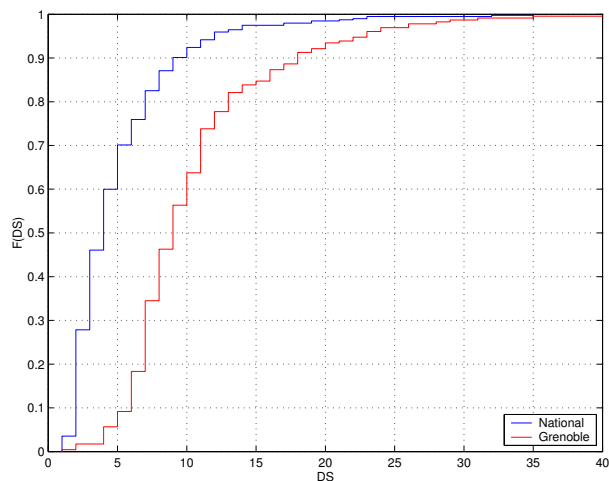


FIG. 2.2 – Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n° 48).

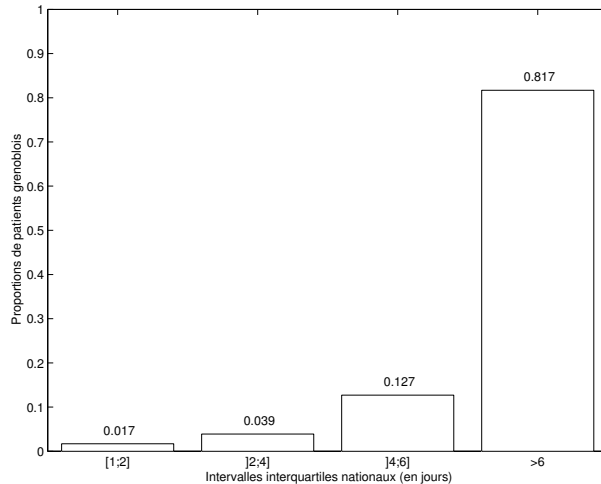


FIG. 2.3 – Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 48).

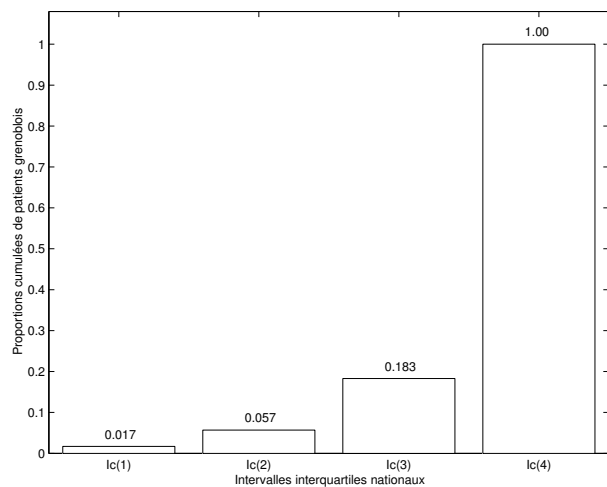


FIG. 2.4 – Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 48).

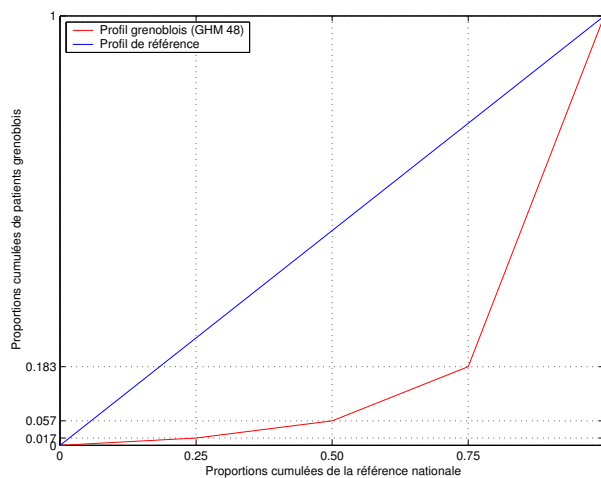


FIG. 2.5 – Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n° 48).

**Cas 1A : Démarche suivie pour établir le profil grenoblois de distribution de DS par rapport à la référence nationale, avec l'exemple du GHM n° 48 : interventions sur la rétine.**

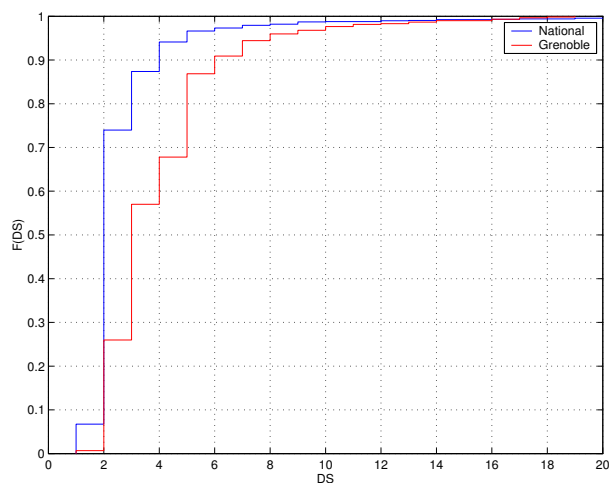


FIG. 2.6 – Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n° 51).

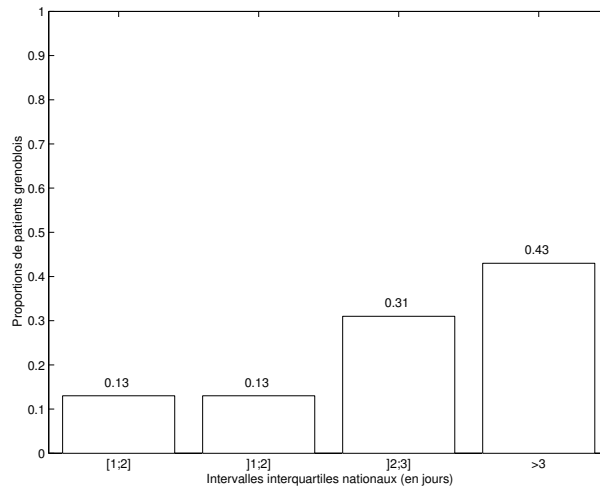


FIG. 2.7 – Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 51).

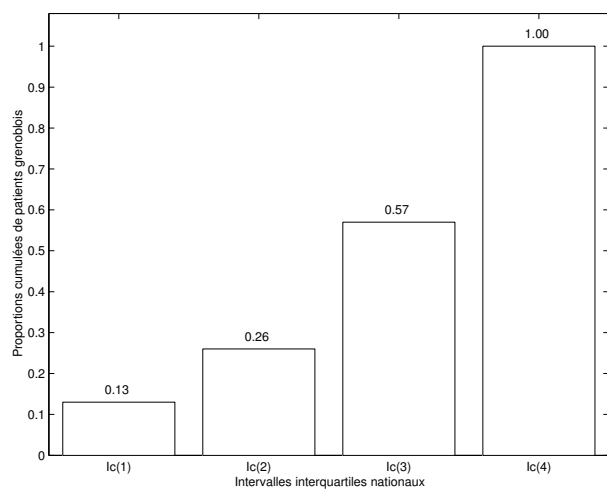


FIG. 2.8 – Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 51).

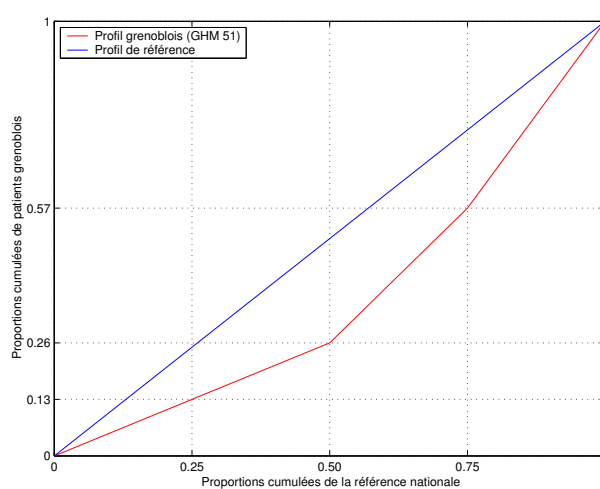


FIG. 2.9 – Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n° 51).

**Cas 2B: Démarche suivie pour établir les profils grenoblois de distributions de DS par rapport à la référence nationale, avec l'exemple du GHM n° 51: interventions sur le cristallin avec ou sans vitrectomie.**

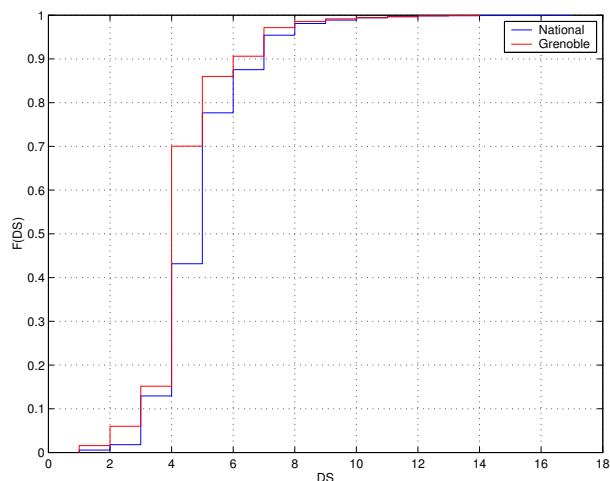


FIG. 2.10 – Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n° 874).

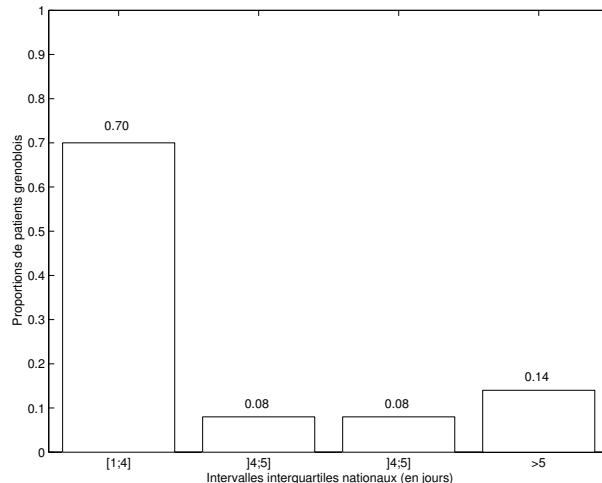


FIG. 2.11 – Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 874).

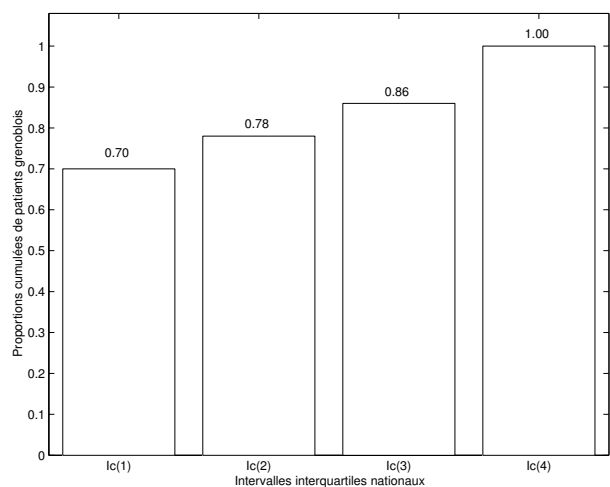


FIG. 2.12 – Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 874).

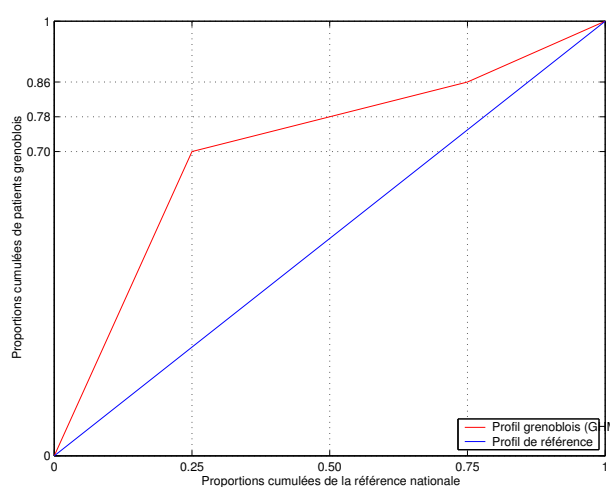


FIG. 2.13 – Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n° 874).

**Cas 3G : Démarche suivie pour établir les profils grenoblois de distributions de DS par rapport à la référence nationale, avec l'exemple du GHM n° 874 : nouveau-nés de 2500g et plus, sans problème significatif.**

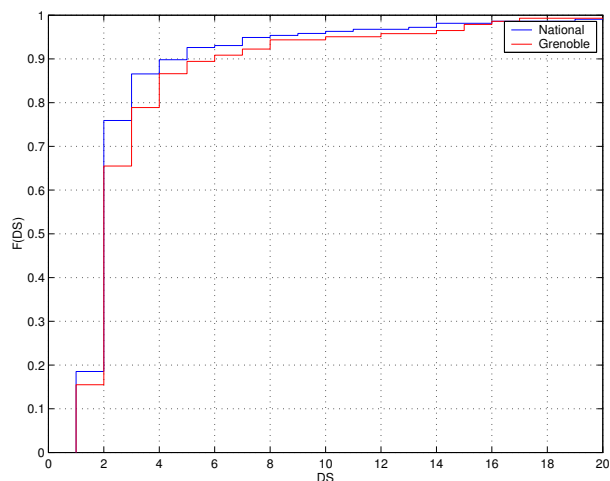


FIG. 2.14 – Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n° 52).

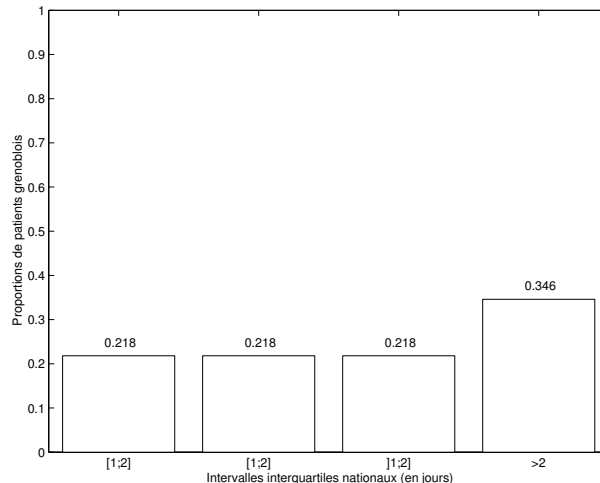


FIG. 2.15 – Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 52).

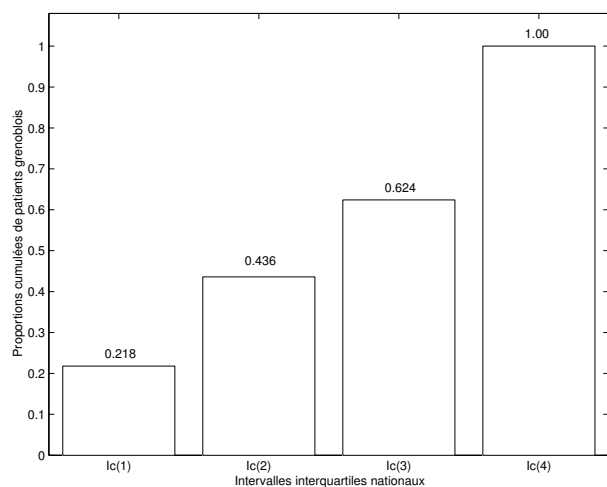


FIG. 2.16 – Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 52).

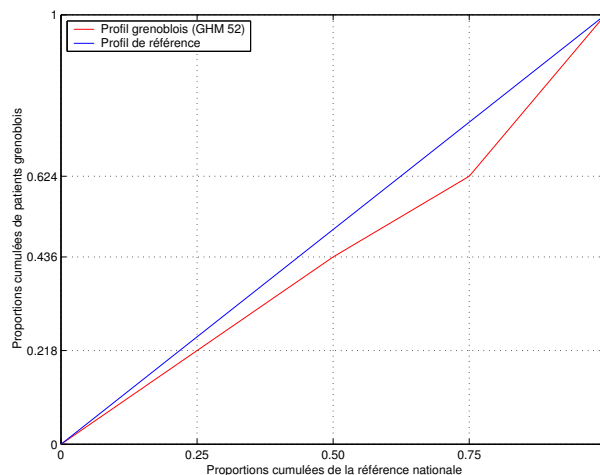


FIG. 2.17 – Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n° 52).

**Cas 4E : Démarche suivie pour établir les profils grenoblois de distribution de DS par rapport à la référence nationale, avec l'exemple du GHM n° 52 : autres interventions extra-oculaires, âge supérieur à 17 ans.**

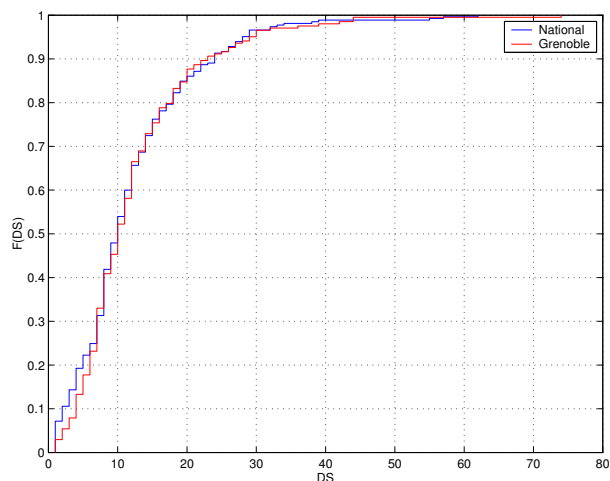


FIG. 2.18 – Fonctions de répartition empiriques des DS des BDD nationale et grenobloise (GHM n° 118).

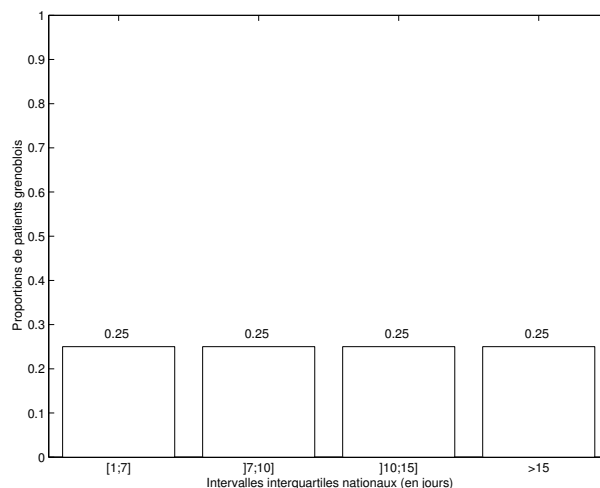


FIG. 2.19 – Proportions de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 118).

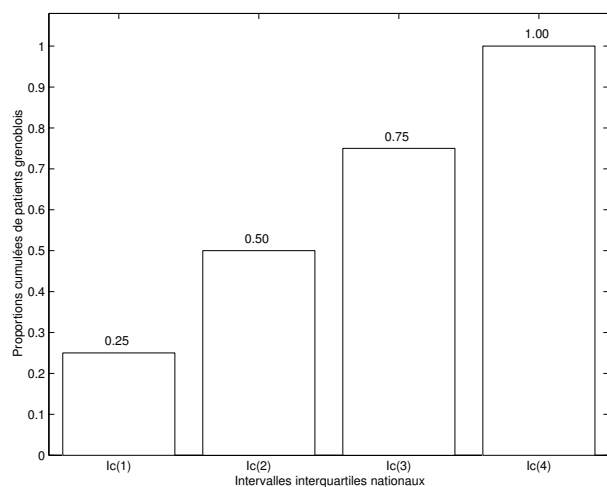


FIG. 2.20 – Proportions cumulées de patients grenoblois dans chaque intervalle interquartile de DS national (GHM n° 118).

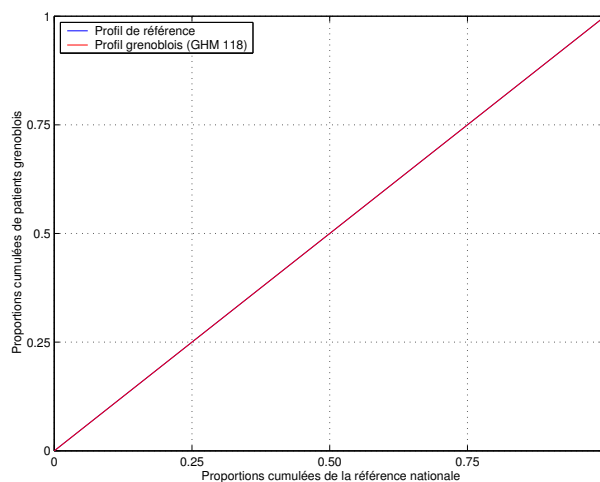


FIG. 2.21 – Représentation du profil grenoblois de distribution de DS par rapport à la référence nationale (GHM n° 118).

**Cas 1H: Démarche suivie pour établir les profils grenoblois de distribution de DS par rapport à la référence nationale, avec l'exemple du GHM n° 118: embolies pulmonaires.**

### Calcul des proportions cumulées de patients grenoblois dans les intervalles interquartiles de DS nationaux

Le calcul des proportions cumulées de patients grenoblois dans les intervalles interquartiles de DS nationaux s'est fait de la manière suivante, dont nous ne présentons que le cas 1A (patients ayant eu une intervention sur la rétine [GHM N°48], voir figure 2.4) :

- le premier intervalle interquartile de DS national en proportion cumulée,  $Ic_{(1)}$ , rassemble 1.7% des patients grenoblois qui ont une DS inférieure ou égale à 2 jours (DS correspondant à 25% des patients de la base nationale)
- le deuxième intervalle interquartile en proportion cumulée,  $Ic_{(2)}$ , réunit 5.6% (1.7% + 3.9%), des patients grenoblois ayant une DS inférieure ou égale à 4 jours (DS correspondant à 50% des patients de la base nationale)
- le troisième intervalle interquartile en proportion cumulée,  $Ic_{(3)}$ , regroupe 18.3% (1.7% + 3.9% + 12.7%) des patients grenoblois qui ont leur DS inférieure ou égale à 6 jours (DS correspondant à 75% des patients de la base nationale),
- le quatrième intervalle interquartile en proportion cumulée,  $Ic_{(4)}$ , rassemble 100% (1.7% + 3.9% + 12.7% + 81.7%) des patients.

Les 4 autres cas sont illustrés dans les figures 2.8, 2.12, 2.16 et 2.20.

C'est à partir des proportions cumulées de patients grenoblois dans les 4 premiers intervalles interquartiles de DS nationaux, que nous avons élaboré le fichier de données rassemblant les profils de distribution de DS des 257 GHM grenoblois.

### 2.2.3 Description des indicateurs de distribution de DS

#### Méthode utilisée

Les trois indicateurs ayant permis d'élaborer les profils de distributions de DS, à savoir les proportions de patients grenoblois dans les 3 premiers intervalles interquartiles nationaux,  $Ic_{(1)}$ ,  $Ic_{(2)}$  et  $Ic_{(3)}$ , ont été décrits avec les paramètres usuels : moyenne, écart-type, minima et maxima, et représentés par des nuages de points.

#### Résultats de l'analyse des indicateurs de distribution de DS

La description univariée de l'ensemble des 3 intervalles interquartiles nationaux en proportions cumulées montre que le profil "moyen" des distributions des DS des 257 GHM grenoblois est comparable à celui de la base nationale et même meilleur sur le premier intervalle interquartile. En effet, les GHM grenoblois comptent en moyenne 30.1% + / - 14.4% de patients sur le premier intervalle interquartile, 52.6% + / - 15.3% de patients sur le deuxième et 75.8% + / - 10.2% sur le troisième (voir tableau 2.6). Cela se vérifie sur les figures 2.22, 2.23, 2.25 et 2.24; près de la moitié des GHM grenoblois (46.7%) ont des proportions supérieures ou égales à la référence : 0.25%, 0.50% et 0.75% (voir tableau 3.1). Les proportions de patients nationaux sont très hétérogènes sur les intervalles interquartiles. Sur les graphiques,

nous pouvons aussi remarquer que les données se répartissent dans un cône, cela est dû au fait que nous travaillons sur les proportions cumulées de patients.

TAB. 2.6 – *Descriptif des proportions cumulées de patients grenoblois dans les 3 intervalles interquartiles nationaux de DS.*

Variables	Effectif	Moyenne (en %)	Écart-type (en %)	Minimum (en %)	Maximum (en %)
Premier intervalle de DS	257	30.1	14.4	0	73.6
Deuxième intervalle de DS	257	52.6	15.3	5.7	91.7
Troisième intervalle de DS	257	75.8	10.2	18.3	94.7

## 2.2.4 Intérêt des profils grenoblois de distribution de DS

Les variables du fichier d’analyse ont l’avantage de :

- résumer la forme des distributions de DS des GHM grenoblois en 3 indicateurs : les proportions cumulées de patients grenoblois sur les 3 premiers intervalles interquartiles nationaux de DS,  $Ic_{(1)}$ ,  $Ic_{(2)}$  et  $Ic_{(3)}$ ,
- de mettre en exergue les profils grenoblois “atypiques” par rapport aux proportions de référence nationale, 25%, 50% et 75%, significatives de distributions de DS équivalentes entre Grenoble et la base nationale.

Pour visualiser les profils de distributions de DS obtenus pour les GHM grenoblois, nous avons réalisé des “PP-plots”. Ces figures ont l’avantage de montrer simplement la forme des profils grenoblois en proportions cumulées par rapport à la référence nationale. Sur les graphiques 2.5, 2.9, 2.17, 2.13 et 2.21 :

- la courbe (en trait continu rouge) représente la forme du profil grenoblois pour un GHM donné,
- la bissectrice (en trait continu bleu),  $Ic_{(1)} = 25\%$ ,  $Ic_{(2)} = 50\%$  et  $Ic_{(3)} = 75\%$ , indique le profil de référence.

Ainsi, si la courbe est située en dessous de la bissectrice, c’est que ce GHM n’est pas “bon”, en terme de distribution de DS, par rapport à la référence nationale (voir figures 2.5, 2.9 et 2.17) et inversement le GHM est “mauvais” si la courbe est en dessous de la bissectrice (voir figure 2.13). Si, par contre, cette courbe est confondue avec la bissectrice, cela signifie que les distributions de DS sont équivalentes entre les bases nationale et grenobloise (voir figure 2.21). Nous verrons ce point, de manière plus approfondie, dans le paragraphe 3.1 du chapitre 3.

Le fichier d’analyse définitif compte 257 individus (les 257 GHM retenus) et 5 variables :

- le numéro des GHM,
- le type de GHM (médical ou chirurgical),

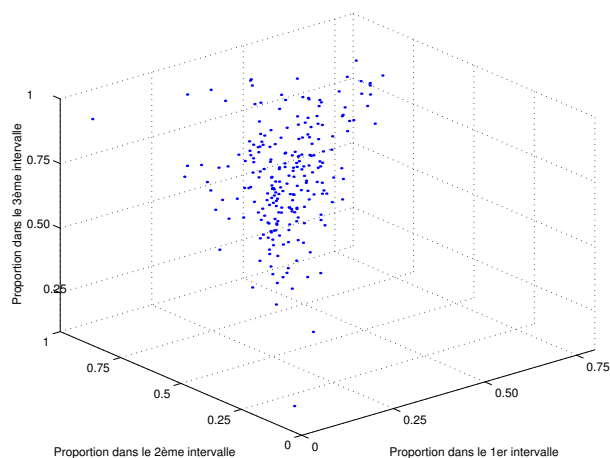


FIG. 2.22 – Représentation des proportions cumulées de patients sur les 3 intervalles interquartiles nationaux de DS.

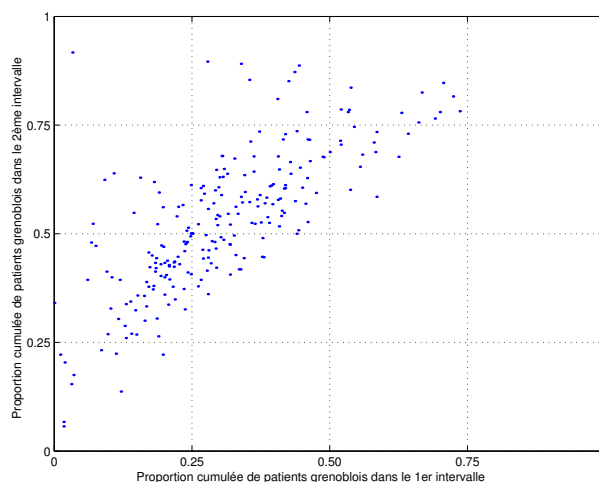


FIG. 2.23 – Représentation des proportions cumulées de patients grenoblois sur le premier et le deuxième intervalles interquartiles nationaux de DS.

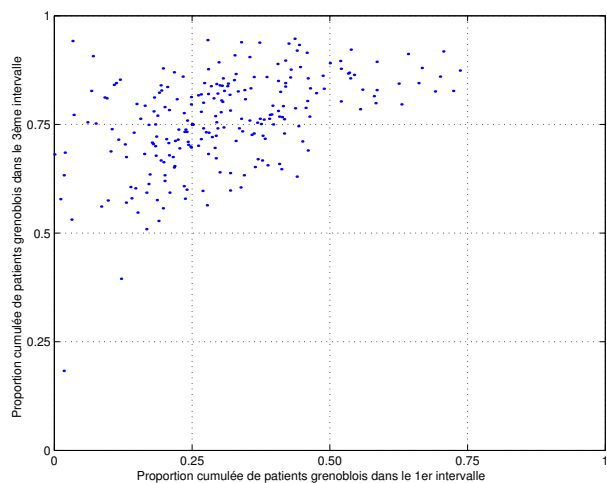


FIG. 2.24 – Représentation des proportions cumulées de patients grenoblois sur le premier et le troisième intervalles interquartiles nationaux de DS.

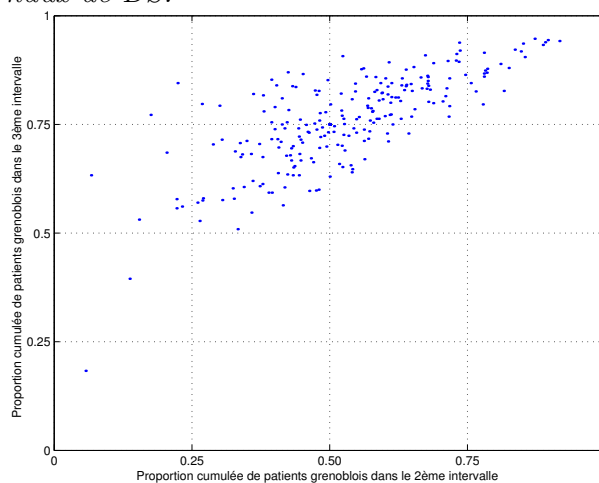


FIG. 2.25 – Représentation des proportions cumulées de patients grenoblois sur le deuxième et le troisième intervalles interquartiles nationaux de DS.

- les proportions cumulées de patients grenoblois observées dans les 3 premiers intervalles interquartiles nationaux de DS.

C'est à partir de ce fichier que nous avons réalisé les analyses des chapitres 3, 4 et 5.



## 2.3 Listes de référence

Après avoir présenté, dans le paragraphe précédent, les méthodes utilisées pour élaborer notre fichier d’analyse et l’avoir décrit brièvement, nous nous intéressons, dans ce paragraphe, aux différentes listes de référence que nous avons utilisées pour valider les classifications.

### 2.3.1 Construction des listes de référence

Nous avons utilisé 3 listes de référence. Ce paragraphe présente d’abord les méthodes mises en œuvre pour obtenir ces 3 listes, puis les décrit.

#### Liste 1 : liste du CHU de Grenoble fondée sur les déficits en points ISA

Nous avons établi une première liste de référence, baptisée “liste du CHU” (voir annexe F). Elle regroupe les 50 GHM grenoblois qui sont responsables de 80% du déficit du CHU en points ISA par rapport à leur DS moyenne (indicateur couramment utilisé dans le secteur hospitalier). Ils correspondent aux 50 GHM qui coûtent le plus cher au CHU, ou encore à ceux qui ont le plus grand déficit de points ISA. Pour ce faire, nous avons, pour chacun des 257 GHM de notre échantillon :

- calculé le nombre de journées réalisées au CHU de Grenoble ( $N_i^G \overline{DS}_i^G$ ),
- calculé le nombre théorique de journées; il correspond au nombre de journées qui auraient été effectuées si le CHU de Grenoble avait eu la même DS moyenne que l’ensemble des CHU français ( $N_i^G \overline{DS}_i^N$ ),

Le nombre théorique de journées était soustrait du nombre de journées réalisées; nous avons ainsi obtenu une différence de journées qui pouvait être positive (bénéfice de journées) ou négative (déficit de journées). Enfin, nous avons fait l’hypothèse que cette différence de journées aurait pu être utilisée pour soigner d’autres patients. Pour calculer ce nombre théorique de patients, nous avons divisé cette différence de journées par la DS moyenne de la base nationale. Il suffisait alors de multiplier ce nombre de patients théorique par le nombre de points ISA correspondant au GHM étudié ( $ISA_i$ ), pour obtenir le bénéfice ou le déficit en points ISA lié à une DS moyenne différente entre le CHU de Grenoble et la base nationale.

$$DifISA(GHM_i) = N_i^G \frac{\overline{DS}_i^G - \overline{DS}_i^N}{\overline{DS}_i^N} ISA_i$$

Cette méthode de calcul nous a été fournie par la division du PMSI du CHU de Grenoble.

**Liste 2 : liste fondée sur le “test des quartiles”**

Une deuxième liste de référence, baptisée “liste des quartiles” a été constituée (voir annexe G). Elle rassemble les GHM grenoblois dont les proportions cumulées de patients vérifient les inégalités :

$$\begin{aligned} Ic_{(1)} &\leq 0.25, \\ Ic_{(2)} &\leq 0.50, \\ Ic_{(3)} &\leq 0.75, \end{aligned}$$

et dont les trois quartiles de DS sont supérieurs de manière significative à ceux des GHM nationaux au seuil  $\alpha = 5\%$ .

Pour ce faire, nous avons testé l’hypothèse suivante :

$$H0 : q_{(0.25)}^G = q_{(0.25)}^N \quad \text{contre} \quad H1 : q_{(0.25)}^G \neq q_{(0.25)}^N.$$

Nous avons réalisé, pour chacun de ces GHM, un tableau de contingence de la forme suivante :

	$\leq q_{(0.25)}^N$	$> q_{(0.25)}^N$	Total
Grenoble	$n_{11}$	$n_{12}$	$n_{1.}$
National	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n$

Ainsi, pour chacun de ces GHM, nous avons calculé la statistique de test suivante :

$$V = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \frac{n_i \cdot n_{.j}}{n})^2}{\frac{n_i \cdot n_{.j}}{n}},$$

où :

$V$  suit asymptotiquement une loi du  $\chi^2$  à 1 degré de liberté;

$n_{ij}$  correspond à l’effectif observé de patients appartenant à la base  $i$ ,  $i = \{G; N\}$ , et ayant une DS de type  $j$ ,  $j = \{DS \leq q_{(0.25)}^N; DS > q_{(0.25)}^N\}$ ;

$n_{i.}$  est l’effectif de patients appartenant à la base  $i$ ;

$n_{.j}$  est l’effectif de patients ayant une DS de type  $j$ , pour un GHM donné;

$n$  est l’effectif total de patients.

Si l’un au moins des  $(n_{i.} n_{.j})/n < 5$ , nous avons appliqué le test exact  $\mathcal{F}$  de Fisher.

Nous avons appliqué le même test pour  $q_{(0.50)}$  et  $q_{(0.75)}$ . Nous avons retenu les GHM grenoblois dont les 3 tests concluaient  $H1$ , c’est-à-dire les GHM dont les 3 quartiles de DS étaient significativement supérieurs à ceux des GHM de la base nationale.

**Liste 3 : liste des proportions fondée sur le “test du  $\chi^2$ ”**

Une troisième liste de référence, baptisée “liste des proportions”, a été créée (voir annexe H). Parmi les GHM grenoblois dont les proportions cumulées de patients sont :

$$Ic_{(1)} \leq 0.25,$$

$$Ic_{(2)} \leq 0.50,$$

$$Ic_{(3)} \leq 0.75,$$

elle rassemble tous ceux dont les proportions de patients dans les 4 intervalles interquartiles de DS,  $I_{(1)}$ ,  $I_{(2)}$ ,  $I_{(3)}$  et  $I_{(4)}$  sont significativement différentes de  $n/4$ ,  $n$  correspondant à l'effectif des patients grenoblois pour un GHM donné.

Pour ce faire, nous avons testé les hypothèses suivantes :

$$H0 : I_{(1)} = I_{(2)} = I_{(3)} = I_{(4)} = n/4$$

contre

$$H1 : I_{(1)} \neq I_{(2)} \neq I_{(3)} \neq I_{(4)} \neq n/4$$

L'hypothèse  $H0$  représente le fait qu'un profil grenoblois soit semblable à la référence nationale.

Pour chacun de ces GHM, nous avons calculé la statistique de test suivante :

$$\mathbf{Z} = \sum_{j=1}^4 \frac{(I_{(j)} - \frac{n}{4})^2}{\frac{n}{4}},$$

où :

$\mathbf{Z}$  suit asymptotiquement une loi du  $\chi^2$  à 3 degrés de liberté;

$I_{(j)}$  correspond à l'effectif de patients observé à Grenoble dans le  $j^{\circ}$  intervalle interquartile;  $n$  correspond à l'effectif total de patients grenoblois.

Ce test nous a permis de repérer s'il existait des écarts de distribution de DS significatifs entre les GHM grenoblois et les GHM nationaux. En effet, si les DS grenobloises sont distribuées comme celles de la base nationale, on devrait trouver un quart de patients dans chaque intervalle interquartile.

**2.3.2 Intérêt des listes de référence**

Dans le cadre de notre étude, ces listes de référence nous ont servi à évaluer la pertinence, d'un point de vue médico-économique, des classifications obtenues avec les différentes approches, que sont les analyses factorielles (classification hiérarchique et analyse discriminante) et les lois de mélange. Pour ce faire, nous avons confronté ces trois listes de référence :

“CHU”, “quartiles” et “proportions” aux groupes des “mauvais” GHM issus des différentes classifications.

### 2.3.3 Description des listes de référence

#### Méthode utilisée

Pour décrire les listes de référence, nous les avons représentées dans les différents intervalles interquartiles. Ensuite, nous avons confronté les listes basées sur les quartiles et sur le  $\chi^2$  à la liste du CHU de Grenoble. Pour ce faire, nous avons utilisé deux indices : la sensibilité et la spécificité (voir tableau 2.7).

TAB. 2.7 – *Intersection entre 2 listes A et B contenant des GHM.*

		Liste A		
		Mauvais GHM	Bons GHM	Total
Liste B	Mauvais GHM	a	b	a + b
	Bons GHM	c	d	c + d
	Total	a + c	b + d	a + b + c + d

La sensibilité,  $Se$ , est la probabilité qu’un GHM identifié comme “mauvais” dans la liste de référence A soit aussi classé comme “mauvais” dans la liste B :

$$Se = a/(a + c).$$

La spécificité,  $Sp$ , est la probabilité qu’un GHM soit considéré comme “bon” dans la liste B quand il est aussi identifié comme “bon” dans la liste de référence A :

$$Sp = d/(b + d).$$

La valeur prédictive positive,  $VPP$ , est la probabilité qu’un GHM soit identifié comme “mauvais” dans la liste A quand il est repéré comme “mauvais” par la liste B.

$$VPP = a/(a + b).$$

La sensibilité et la spécificité dépendent uniquement de la qualité de la liste B, liste que l’on veut comparer à la liste de référence A : dans notre étude, elle correspond à la liste “CHU”, mais aussi les listes de quartiles et du  $\chi^2$  dans les chapitre suivant.

D’autre part, nous avons décrit ces listes selon 2 variables explicatives qui sont :

- le type de GHM,
- le déficit ou le bénéfice en points ISA que les GHM occasionnent au CHU.

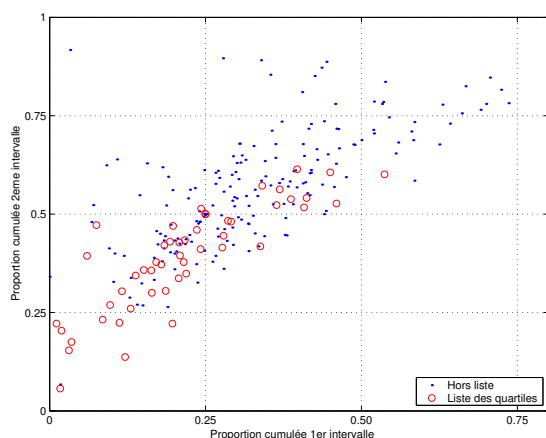


FIG. 2.26 – Représentation de la liste des 50 GHM grenoblois causant 80% du déficit du CHU sur les 2 premiers intervalles interquartiles.

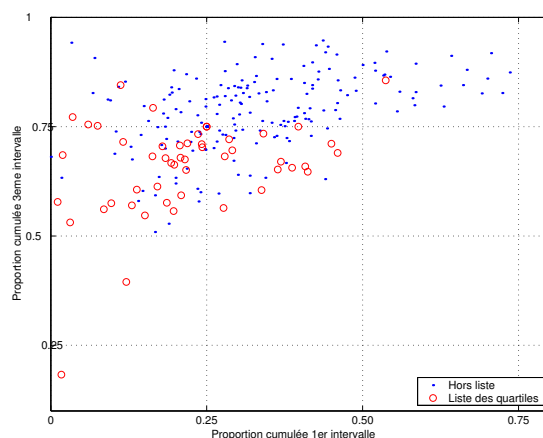


FIG. 2.27 – Représentation de la liste des 50 GHM grenoblois causant 80% du déficit du CHU sur le premier et le troisième intervalles interquartiles.

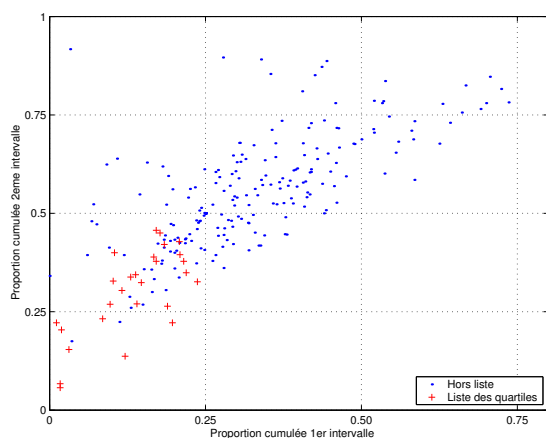


FIG. 2.28 – Représentation de la liste des quartiles sur les 2 premiers intervalles interquartiles.

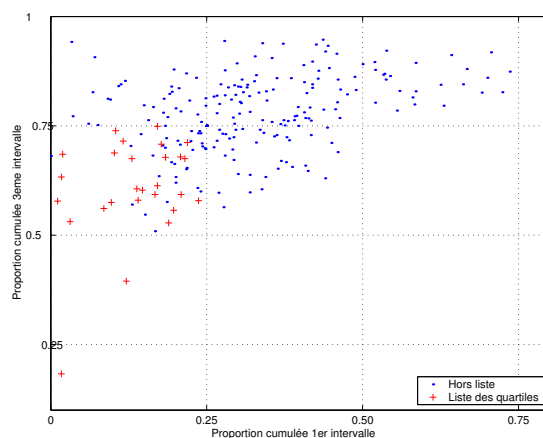


FIG. 2.29 – Représentation de la liste des quartiles sur le premier et le troisième intervalles interquartiles.

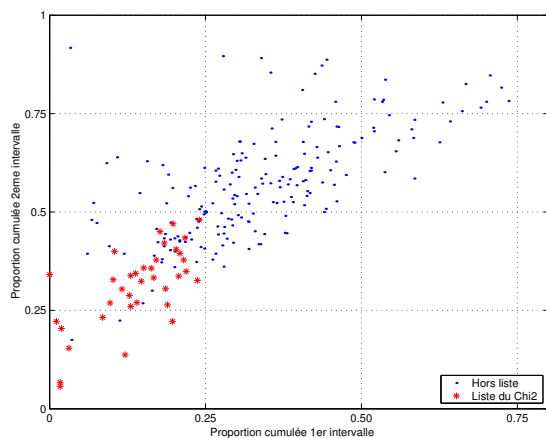


FIG. 2.30 – Représentation de la liste du  $\chi^2$  sur les 2 premiers intervalles interquartiles.

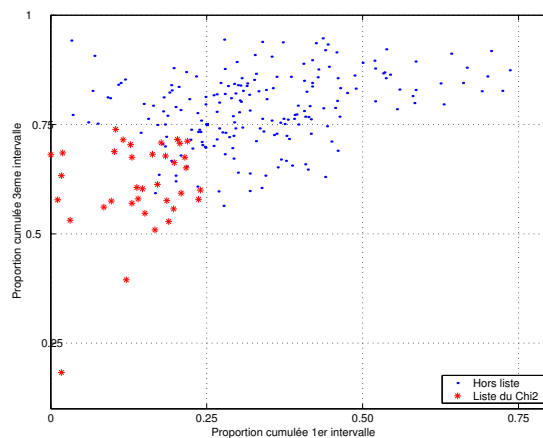


FIG. 2.31 – Représentation de la liste du  $\chi^2$  sur le premier et le troisième intervalles interquartiles.

## Résultats

### Liste 1 : liste du CHU de Grenoble

D'après les figures 2.26 et 2.27, nous voyons que les 50 GHM grenoblois causant 80% du déficit du CHU en points ISA (ronds de couleur rouge sur le graphique) ont, pour la majorité, des proportions cumulées de patients sur les 3 intervalles (voir annexe F) inférieures aux proportions de références égales à 25%, 50% et 75%. Ces GHM sont d'ailleurs tous situés en bas à gauche des graphiques. De plus chaque GHM occasionne en moyenne un déficit de points ISA égal à 111 034 +/- 93 877 au CHU. Cependant, ce déficit semble être très hétérogène selon les GHM car l'écart-type est grand. Ce sont en majorité des GHM médicaux (62%).

### Liste 2 : liste des quartiles

TAB. 2.8 – Intersection entre les listes du CHU et des quartiles.

		Liste CHU	
		Mauvais GHM	Bons GHM
Liste quartiles	Mauvais GHM	16	11
	Bons GHM	34	196
		$Se = 32\%$	$Sp = 95\%$

Cette liste (voir annexe G) regroupe les GHM grenoblois dont les 3 quartiles de DS sont supérieurs de manière significative aux 3 quartiles de DS de leurs homologues de la base nationale. Elle compte 27 GHM, dont 16 font partie de la liste des 50 GHM du CHU de Grenoble (voir annexe F et tableau 2.8). Tous les GHM appartenant à cette liste se situent en bas et à gauche des figures 2.28 et 2.29 (croix de couleur rouge). Chaque GHM entraîne en moyenne un déficit en points ISA de 91 541 +/- 112 610 au CHU. Néanmoins, ce déficit s'avère très hétérogène selon les GHM, l'écart-type étant grand. Cette liste comprend un peu plus de la moitié de GHM médicaux (56%).

### Liste 3 : liste du $\chi^2$

TAB. 2.9 – Intersection entre les listes du CHU et du  $\chi^2$ .

		Liste CHU	
		Mauvais GHM	Bons GHM
Liste $\chi^2$	Mauvais GHM	22	14
	Bons GHM	28	193
		$Se = 44\%$	$Sp = 93\%$

Cette liste (voir annexe H) réunit les GHM grenoblois dont les 3 proportions cumulées de patients grenoblois sont inférieures de manière significative aux 3 proportions de référence de la base nationale. Elle compte 36 GHM dont 22 appartiennent à la liste des 50 GHM fournie par le CHU de Grenoble (voir annexe F et tableau 2.9). Comme pour la liste des quartile, les GHM de cette liste se situent tous en bas à gauche des figures 2.30 et 2.31. Ils engendrent un déficit en points ISA de 101 346 +/- 112 712 au CHU. Notons là encore que ce déficit paraît très hétérogène selon les GHM car l'écart-type est grand. Un peu plus de la moitié des GHM de cette liste sont de type chirurgical (58%).

Les 2 listes de référence basées sur les quartiles et le  $\chi_2$  ont un coefficient de corrélation très élevé (0.74). Cela s'explique par le fait que les méthodes pour les élaborer sont très proches et dépendent des 3 nouvelles variables.

Leurs taux de sensibilité,  $Se$ , égaux à 32% pour la liste des quartiles et à 44% pour celle du  $\chi^2$ , indiquent qu'elles ont du mal à repérer les GHM grenoblois responsables de 80% du déficit du CHU en point ISA (voir tableaux 2.8 et 2.9). On pouvait s'y attendre, étant donné que ces listes sont peu corrélées avec celle du CHU, respectivement 0.34 et 0.42. La liste qui repère le mieux les "mauvais" GHM au sens du déficit en points ISA est celle des quartiles, mais cela reste très faible, elle n'en repère même pas la moitié.

En revanche, ces 2 listes repèrent bien les GHM qui ne posent pas de problème en termes de déficit de points ISA ( $Sp = 95%$  pour la liste des quartiles et  $Sp = 93%$  pour celle du  $\chi_2$ ).

## 2.4 Synthèse

### 2.4.1 Profils grenoblois de distribution de DS

Les nouvelles variables que nous avons construites ont deux avantages. Le premier est qu'elles permettent de résumer la forme des distributions de DS des GHM grenoblois en 3 indicateurs, qui sont les proportions cumulées de patients grenoblois sur les 3 premiers intervalles interquartiles nationaux de DS,  $Ic_{(1)}$ ,  $Ic_{(2)}$  et  $Ic_{(3)}$ . Le deuxième est qu'elle permettent de mieux visualiser les profils grenoblois "atypiques" par rapport aux proportions de référence nationale, 25%, 50% et 75%, significatives de distributions de DS équivalentes ou non entre Grenoble et la base nationale.

Cependant, nous sommes conscients qu'en appliquant les méthodes de calcul mises en œuvre pour élaborer ces profils, nous avons tendance à avantager la forme des profils grenoblois. En effet, alors qu'un GHM grenoblois peut apparaître comparable à la base nationale en termes de quartiles de DS (égalité de certains de ces quartiles), il se peut qu'il compte plus de 25% de patients dans le ou les intervalles interquartiles concernés. Il faudra donc y prêter attention lors de l'interprétation de nos résultats.

### 2.4.2 Listes de référence

Les GHM appartenant aux trois listes de référence, ont comme caractéristiques communes :

- d’être souvent de types médicaux,
- d’occasionner un déficit en points ISA important pour le CHU de Grenoble.

La liste CHU est à l’origine de la première partie de ce travail. En effet, en 1996, des médecins de la division du PMSI du CHU de Grenoble avaient identifié, par l’intermédiaire de cette liste, un certain nombre de “mauvais” GHM, avec des distributions de DS “atypiques”. Ils avaient tous tendance à regrouper des patients ayant des DS plus longues que ceux de la base nationale, mais à différents endroits de la distribution de leur DS : au début, au milieu et/ou à la fin. C’est pourquoi ils nous ont demandé de mettre en place des méthodes fiables de repérage des “mauvais” GHM, mais aussi de regarder si ces derniers ne se structuraient pas en classes (ce que nous verrons dans les chapitres 4 et 5).

La liste CHU a l’avantage de prendre en compte le point ISA (l’unité économique des hôpitaux). Cependant, cette liste se fonde sur la DS moyenne de chaque GHM grenoblois pour calculer le bénéfice ou le déficit en points ISA qu’ils causent au CHU de Grenoble par rapport à la DS moyenne base nationale. Cette méthode est critiquable d’un point de vue statistique, mais fondée d’un point de vue médical. En effet, s’il n’est pas juste de se fonder sur la DS moyenne car les distributions de DS ne suivent pas des lois normales (comme nous l’avons vu dans le chapitre 1), les gestionnaires des hôpitaux utilisent exclusivement cet indicateur. C’est la raison pour laquelle cette liste nous a plutôt servi pour la validation “économique” de nos classifications.

Les listes fondées sur les quartiles de DS et les effectifs ont été d’abord construites pour regarder, si avec des méthodes simples, il n’était pas possible de repérer les GHM grenoblois “atypiques”, sans pour autant se servir des points ISA, qui sont difficiles à utiliser car peu corrélés avec la DS. Dans un deuxième temps, ces listes nous ont servi de référence pour valider les classifications.

Pour la suite de l’étude, ces 3 listes de référence nous ont aidé à valider, d’un point de vue médico-économique, les différentes classifications des profils grenoblois de distribution de DS obtenus.





# Chapitre 3

## Description et modélisation des profils grenoblois

Maintenant que la méthode nous ayant permis d’obtenir les profils de distribution de DS des 257 GHM grenoblois par rapport à la base nationale est exposée, il est nécessaire de les étudier plus finement. Une première étape va être de décrire ces profils : combien en existe-t-il ? Quelle forme ont-ils ? Combien de profils peuvent être considérés comme “mauvais” ? Quelles sont leurs caractéristiques médico-économiques ? Nous cherchons en quelque sorte à trouver une première classification d’origine empirique. Celle-ci va être confrontée aux trois listes de référence pour en analyser sa pertinence. Une deuxième étape va être d’étudier, avec des lois log-normales unimodales, l’influence de leurs paramètres sur la forme des profils. On appliquera ensuite cette approximation aux distributions de DS correspondant aux profils types référents de notre échantillon..

### 3.1 Description des profils grenoblois de distribution de DS

À travers ce paragraphe, nous présentons les méthodes qui nous ont permis de visualiser et de décrire la forme des différents profils des GHM grenoblois.

#### 3.1.1 Méthodes

Les profils caractéristiques des distributions de DS des GHM grenoblois ont été décrits par des “PP-plots”. Nous avons choisi, pour chaque profil, un GHM grenoblois représentatif. Ils correspondent aux profils types observés dans notre échantillon. Dans chaque figure sont représentés :

- en abscisse, les proportions cumulées de patients appartenant à la base nationale correspondant à chacun des quartiles de DS,

- en ordonnée, les proportions cumulées de patients grenoblois ayant une DS inférieure à chacun des quartiles de DS nationaux.

La bissectrice, en trait continu bleu sur les graphiques, représente le profil que devraient avoir les profils des GHM grenoblois, s'ils étaient équivalents à la norme nationale. Les neuf courbes tracées en trait continu rouge sur les graphiques, correspondent au profil observé de la distribution des DS des GHM grenoblois référents.

Pour savoir si, pour un GHM donné, le profil grenoblois observé est significativement différent du national, nous avons effectué un test de Kolmogorov-Smirnov (K-S) [3] au seuil de 5% entre les distributions de DS grenobloise et nationale.

Pour quantifier l'éloignement global d'un profil grenoblois par rapport à la norme nationale, nous avons calculé l'aire correspondant à la différence entre les surfaces des profils grenoblois (aire sous la courbe représentant le profil entre 0 et 1) et national (aire sous la bissectrice entre 0 et 1 égale à 0.5). Cette nouvelle variable, baptisée "différence d'aire", indique si le profil grenoblois est très éloigné (négativement ou positivement) de la référence nationale.

Nous avons regardé pour chaque profil, le déficit ou le bénéfice moyen en points ISA qu'ils entraînent au CHU de Grenoble, mais aussi le type de GHM qu'ils regroupent.

### 3.1.2 Résultats

#### Description des profils grenoblois

L'approche empirique dégage 9 profils types de distribution de DS, tous observés dans notre échantillon, dont 7 sont identifiables à de "mauvais" profils (voir figures 3.1.2). Voilà la description détaillée de ces profils :

- profil de type *(a)* (voir figure 3.1 [21.4% des GHM], illustré par le GHM n°48 : interventions sur la rétine, regroupe les GHM grenoblois identifiés comme "mauvais" sur toute la distribution des DS ( $Ic_{(1)} < 0.25$ ,  $Ic_{(2)} < 0.50$  et  $Ic_{(3)} < 0.75$ );
- profil de type *(b)* (voir figure 3.2) [7.4% des GHM], illustré par le GHM n°372 : interventions sur le sein pour des affections non malignes autres que les actes de biopsie et d'excision locale, réunit les GHM grenoblois reconnus comme peu performants pour les séjours de courte et moyenne durées ( $Ic_{(1)} < 0.25$  et  $Ic_{(2)} < 0.50$ );
- profil de type *(c)* (voir figure 3.3) [7.0% des GHM], illustré par le GHM n°63 : autres diagnostics portant sur le système respiratoire, âge supérieur à 69 ans et/ou CMA, rassemble les GHM grenoblois considérés comme "mauvais" pour les séjours de moyenne et longue durées ( $Ic_{(2)} < 0.50$  et  $Ic_{(3)} < 0.75$ );
- profil de type *(d)* (voir figure 3.4) [1.6% des GHM], illustré par le GHM n°274 : cholécystectomies sans exploration de la voie biliaire principale, âge inférieur à 70 ans

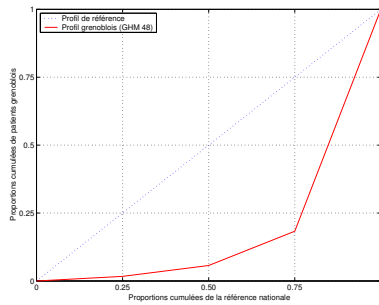


FIG. 3.1 – Profil de type (a)

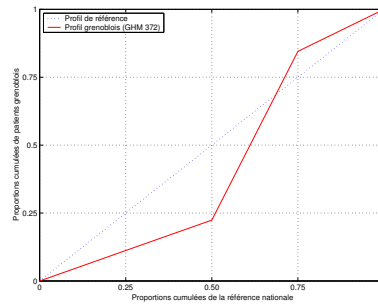


FIG. 3.2 – Profil de type (b)

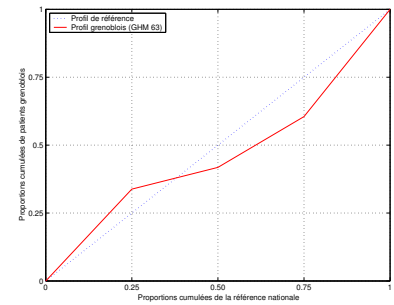


FIG. 3.3 – Profil de type (c)

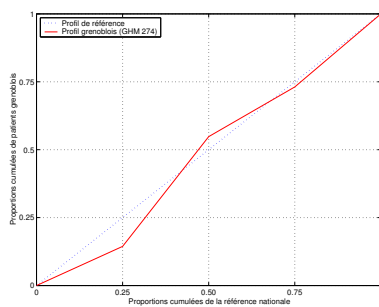


FIG. 3.4 – Profil de type (d)

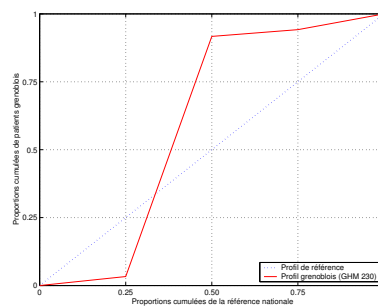


FIG. 3.5 – Profil de type (e)

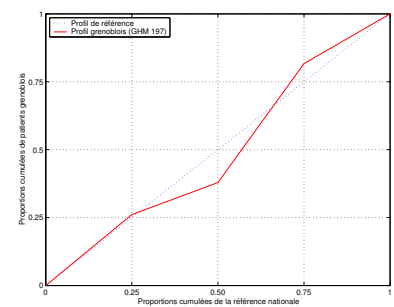


FIG. 3.6 – Profil de type (f)

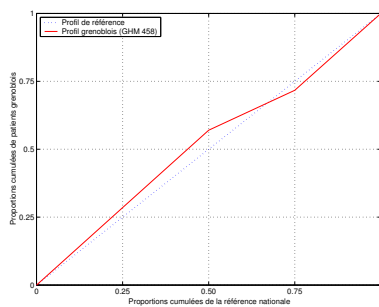


FIG. 3.7 – Profil de type (g)

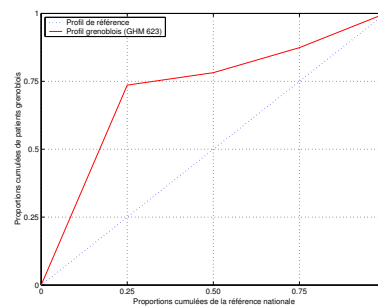


FIG. 3.8 – Profil de type (h)

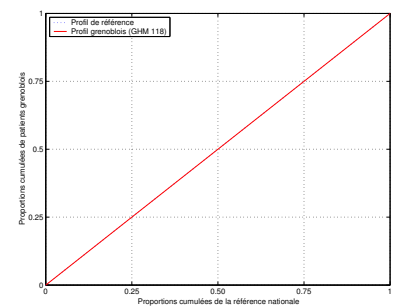


FIG. 3.9 – Profil de type (neutre)

*Représentation des neuf profils types de distributions des DS en proportions cumulées.*

- sans CMA, regroupe les GHM grenoblois reconnus comme peu performants pour les séjours de courte et longue durées ( $Ic_{(1)} < 0.25$  et  $Ic_{(3)} < 0.75$ );
- profil de type  $(e)$  (voir figure 3.5) [5.0% des GHM], illustré par le GHM n°230 : interventions réparatrices pour hernies et éventrations, âge inférieur à 18 ans, réunit les GHM grenoblois identifiés comme “mauvais” pour les séjours de courte durée uniquement ( $Ic_{(1)} < 0.25$ );
  - profil de type  $(f)$  (voir figure 3.6) [3.5% des GHM], illustré par le GHM n°197 : angine de poitrine, rassemble les GHM grenoblois reconnus comme peu performants pour les séjours de moyenne durée uniquement ( $Ic_{(2)} < 0.50$ );
  - profil de type  $(g)$  (voir figure 3.7) [7.4% des GHM], illustré par le GHM n°458 : lithiases urinaires, âge inférieur à 70 ans sans CMA regroupe les GHM grenoblois classés comme “mauvais” pour les séjours de longue durée uniquement ( $Ic_{(3)} < 0.75$ );
  - profil de type  $(h)$  (voir figure 3.8) [36.2% des GHM], illustré par le GHM n°623 : psychoses, réunit les GHM grenoblois considérés comme “bons” sur toute la distribution des DS ( $Ic_{(1)} > 0.25$ ,  $Ic_{(2)} > 0.50$  et  $Ic_{(3)} > 0.75$ );
  - profil de type  $(neutre)$  (voir figure 3.9) [10.5% des GHM], illustré par le GHM n°118 : embolies pulmonaires, réunit les GHM grenoblois ayant une distribution de DS équivalente à celle de la base nationale ( $Ic_{(1)} = 0.25$ ,  $Ic_{(2)} = 0.50$  et  $Ic_{(3)} = 0.75$ ).

À première vue, les GHM grenoblois peuvent être identifiés comme “mauvais”, en termes de distribution de DS :

- à des degrés divers : sur l’ensemble de la distribution (profil de type  $(a)$ ), sur les deux tiers (profils de types  $(b)$ ,  $(c)$  et  $(d)$ ) ou sur un tiers seulement (profils de types  $(e)$ ,  $(f)$  et  $(g)$ )
- pour des séjours de différentes durées : courte ( $Ic_{(1)} < 0.25$ ), moyenne ( $Ic_{(2)} < 0.50$ ) et/ou longue ( $Ic_{(3)} < 0.75$ ).

Les résultats du test de K-S, au seuil de 5% , montrent que seulement 46.3% des GHM grenoblois ont une distribution de DS significativement différente de leur référence nationale (voir tableau 3.1). Ce pourcentage varie de 0% (profil de type  $(d)$ ) à 69.1% (profil de type  $(a)$ ). Ainsi, le profil de type  $(d)$  ne semble pas être significativement différent de la référence nationale. Les GHM identifiés comme “mauvais” semblent donc se structurer en 6 groupes. Plus la forme d’un profil est proche de la bissectrice, moins celle-ci est significativement différente de la référence nationale

D’autre part, plus le profil d’un GHM grenoblois est identifié comme “mauvais” ou “bon”, plus celui-ci se trouve éloigné de la bissectrice (la norme nationale) et plus sa différence d’aire avec la bissectrice est importante (et respectivement, négative ou positive). C’est l’inverse pour les profils identifiés comme “bons”, voir tableau 3.1. Les GHM les plus éloignés négativement, en terme de différence d’aire moyenne, sont principalement les GHM identifiés comme “mauvais” :

- sur toute la distribution de leurs DS (-8.7%),

TAB. 3.1 – Description des GHM grenoblois selon leur profil de distributions de DS.

Type de profil	Description du profil grenoblois par rapport aux proportions de référence nationale (25%, 50%, 75%)	Distributions de DS				
		GHM <i>N</i> (%)	significativement ≠ entre les 2 bases <i>N</i> (%)	Différence d'aire (en %) $\bar{X}(ET(X))$	Déficit en points ISA $\bar{X}(ET(X))$	GHM de type chirurgical <i>N</i> (%)
Profil a	---	55 (21.4)	38 (69.1)	-8.7 (5.7)	-76 163 (97 940)	23 (41.8)
Profil b	--+	19 (7.4)	11 (57.9)	-3.8 (3.3)	-23 929 (60 704)	10 (52.6)
Profil c	+--	18 (7.0)	9 (50.0)	-1.6 (2.0)	-42 118 (47 236)	3 (16.7)
Profil d	-+-	4 (1.6)	0 (0)	-0.7 (1.1)	-35 790 (59 269)	4 (100.0)
Profil e	-++	13 (5.0)	6 (46.2)	2.6 (2.6)	25 793 (45 681)	8 (61.5)
Profil f	+ - +	9 (3.5)	2 (22.2)	1.5 (2.0)	2 084 (25 711)	5 (55.6)
Profil g	++-	19 (7.4)	4 (21.1)	2.9 (2.1)	-34 323 (51 186)	2 (10.5)
Profil h	+++	93 (36.2)	46 (49.5)	11.1 (5.4)	48 671 (66 560)	23 (24.7)
Profil neutre	identique	27 (10.5)	3 (11.1)	0 (0)	6 969 (35 636)	8 (29.6)
Total		257 (100)	119 (46.3)	2.1 (8.8)	86 (33.5)	

- sur les séjours de courte et moyenne durées (-3.8%),
- sur les séjours de moyenne et longue durées (-1.6%),
- sur les séjours de courte et longue durées (-0.7%).

Néanmoins, la différence entre les aires des profils grenoblois et la norme nationale n'est pas un indicateur très pertinent pour les GHM grenoblois identifiés comme "mauvais" sur un seul des 3 intervalles interquartiles de DS car celle-ci est alors globalement positive.

Les GHM qui causent en moyenne le plus gros déficit en points ISA au CHU de Grenoble ont plutôt un profil de type *(a)* (-76163 +/- 97 940), voir tableau 3.1. Nous remarquons aussi que les GHM identifiés comme "mauvais" sur les séjours de longue durée (profils de types *(c)*, *(d)* et *(g)*) engendrent des déficits en points ISA important, respectivement -42 118 +/- 51 186, -35 790 +/- 59 269, -34 323 +/- 51 186. Néanmoins, ce déficit semble être très hétérogène selon les GHM car les écarts-types sont grands.

Près de la moitié des GHM ayant un profil de type *(a)* sont de type chirurgical (voir tableau 3.1). Par contre, les profils de types *(g)* et *(c)* comptent peu de GHM chirurgicaux, respectivement 10.5% et 16.5%. Cependant, cette variable n'a pas l'air très liée avec les types de profil. Ce n'est pas parce les profils comptent peu de GHM chirurgicaux qu'ils sont "mauvais" en termes de profils de DS.

Après avoir décrit les 9 profils types que nous avons trouvés dans notre échantillon, il est intéressant de les confronter aux listes de référence.

### Confrontation de différents profils grenoblois avec les listes de référence

Cent pour cent des GHM appartenant aux listes fondées sur les quartiles et le  $\chi^2$  ont des profils de type *(a)*. Cela s'explique par la méthode utilisée pour construire ces listes. Un peu plus de la moitié (54%) des GHM de la liste CHU ont un profil de type *(a)* et 16% un profil de type *(g)* (voir tableau 3.2). On remarque que 4 GHM de cette liste sont identifiés comme "bon" GHM : profils de type *(h)* ou *neutre*. La raison pour laquelle les GHM appartenant à la liste CHU ont plusieurs profils différents est que la liste CHU ne tient pas compte de la forme des distributions de DS, mais simplement de la DS moyenne nationale et grenobloise.

Pour aller plus loin, il est intéressant de regarder une approche théorique, sous l'angle de la modélisation des distributions de DS par des lois log-normales unimodales afin d'étudier l'influence des paramètres sur la forme des profils et d'appliquer cette méthode aux 9 GHM grenoblois référents et à leurs homologues de la base nationale. C'est l'objet du paragraphe suivant.

TAB. 3.2 – *Intersection des profils de distributions de DS avec la liste du CHU.*

Type de profil	Liste CHU N (%)
Profil a	27 (54.0)
Profil b	5 (10.0)
Profil c	5 (10.0)
Profil d	1 (2.0)
Profil e	0 (0)
Profil f	0 (0)
Profil g	8 (16.0)
Profil h	2 (4.0)
Profil neutre	2 (4.0)
Total	50 (100)

## 3.2 Modélisation des distributions de DS par des lois log-normales unimodales

Nous avons tenté de modéliser les distributions de DS par des lois log-normales unimodales, ce sont souvent ces hypothèses qui sont faites dans la littérature [24, 68, 81, 85, 89]. D'autre part, la loi log-normale a l'avantage de pouvoir ajuster des distributions où cohabitent des événements courts et des événements longs, ce qui est le cas des DS. Cependant, nous sommes conscients qu'en utilisant des lois unimodales, nous faisons une hypothèse forte d'homogénéité des patients des GHM concernés, ce qui n'est pas forcément le cas.

L'objectif de cette modélisation n'est pas de trouver la meilleure loi ajustant les distributions de DS, mais :

- d'étudier avec l'influence des paramètres de lois log-normales sur la forme des profils,
- d'appliquer ce modèle aux distributions de DS des 9 GHM référents pour vérifier s'il est possible de retrouver la forme de leurs profils empiriques.

### 3.2.1 Démarche suivie

#### Influence des paramètres sur la forme des profils

Pour analyser la manière dont les paramètres influent sur la forme des profils grenoblois, nous avons procédé comme suit :



- $DS_i^N \sim \mathcal{LN}(\mu_B, \sigma_B^2)$ ,  $\mu_B = 1.608$  et  $\sigma_B^2 = 0.597^2$ ,
- $\ln(DS_i^N) \sim \mathcal{N}(\mu_B, \sigma_B^2)$ ,
- $\beta = \{-0.8, -0.5, -0.2, 0.2, 0.5, 0.8\}$ ,
- $\gamma = \{-0.5, -0.3, -0.2, 0.2, 0.3, 0.5\}$ ,
- $DS_i^G \sim \mathcal{LN}(\mu_B \pm \beta, \sigma_B^2 \pm \gamma^2)$
- $\ln(DS_i^G) \sim \mathcal{N}(\mu_B + \beta, \sigma_B^2 + \gamma^2)$ ,

où :

- $DS_i^N$  correspond aux DS de la base nationale  $N$ ,
- $\mu_B$  : le premier paramètre de la loi log-normale régissant  $DS_i^N$  et correspondant à la moyenne de leur logarithme,
- $\sigma_B^2$  : le deuxième paramètre de la loi log-normale régissant  $DS_i^N$  et correspondant à la variance de leur logarithme,
- $\beta$  : effet du CHU de Grenoble sur  $\mu_B$  (le paramètre de position),
- $\gamma$  : effet du CHU de Grenoble sur  $\sigma_B^2$  (le paramètre de dispersion),
- $DS_i^G$  correspond aux DS de la base grenobloise  $G$ .

La densité de la loi normale suivie par le logarithme des  $DS_i^N$  la base nationale est égale à :

$$f(\ln(DS_i^N)) = \frac{1}{\sigma_B \sqrt{2\pi}} \exp(-(\ln(DS_i^N) - \mu_B)^2 / 2\sigma_B^2).$$

La densité de la loi normale suivie par le logarithme  $DS_i^G$  de la base grenobloise est égale à :

$$f(\ln(DS_i^G)) = \frac{1}{\sqrt{(\sigma_B^2 + \gamma^2)} \sqrt{2\pi}} \exp(-(\ln(DS_i^G) - (\mu_B + \beta))^2 / 2(\sigma_B^2 + \gamma^2)).$$

Nous avons calculé la densité de la loi normale pour les bases nationale et grenobloise.

Concernant Grenoble, nous avons fait varier  $\beta$  ou  $\gamma$  de différentes manières, pour étudier l'effet de  $\mu_B$  et/ou  $\sigma_B^2$  sur la forme des profils grenoblois :

- d'abord simplement  $\beta$ ,
- puis simplement  $\gamma$ ,
- et enfin  $\beta$  et  $\gamma$  simultanément.

Pour obtenir les profils de distribution de DS grenoblois, nous avons d'abord calculé, à partir de la densité nationale  $f(\ln(DS_i^N))$ , les 3 DS correspondant aux quartiles,  $q_{(0.25)^N}$ ,  $q_{(0.50)^N}$  et  $q_{(0.75)^N}$ . Ensuite, nous avons calculé, à partir de la densité grenobloise  $f(\ln(DS_i^G))$ , les proportions cumulées de patients grenoblois,  $p$ , ayant une DS inférieure ou égale aux DS correspondant respectivement à  $q_{(0.25)^N}$ ,  $q_{(0.50)^N}$  et  $q_{(0.75)^N}$  où :

$$q_{(u)} = \exp(\Phi^{-1}(u) * \sigma_B + \mu_B),$$

et

$$p_{(\leq u)} = \Phi \left( \frac{\ln q(u) - \mu_B - \beta}{\sqrt{\sigma_B^2 + \gamma^2}} \right).$$

avec :

$u = 0.25, 0.50$  et  $0.75$ ,

$\Phi$  correspondant à la fonction de répartition de la loi normale.

Pour la référence nationale, ces proportions de patients sont respectivement égales à 0.25, 0.50 et 0.75.

Nous avons représenté ces profils en utilisant des “PP-plot”.

Trois figures regroupent les densités de DS des bases nationale et grenobloise obtenues selon les variations de  $\beta$  et/ou  $\gamma$ , trois autres réunissent les profils grenoblois correspondants.

### Application du modèle aux 9 GHM référents

Nous avons appliqué cette méthodologie aux distributions des DS des GHM correspondant aux 9 profils grenoblois observés dans notre échantillon afin de vérifier s’il est possible de retrouver les profils empiriques. Nous avons procédé comme suit.

Nous sommes parties des hypothèses suivantes :

–  $DS_k^j \sim \mathcal{LN}(\mu_{kj}, \sigma_{kj}^2)$ .

–  $\ln(DS_k^j) \sim \mathcal{N}(\mu_{kj}, \sigma_{kj}^2)$ .

où :

- $DS_k^j$  : les DS du GHM représentant le profil  $k$ ,  $k = \{1, \dots, 9\}$ , de la base  $j$ ,  $j = \{N, G\}$  avec  $N$  : nationale,  $G$  : grenobloise,
- $\mu_{kj}$  : le premier paramètre de la loi log-normale régissant  $DS_k^j$ ,
- $\sigma_{kj}^2$  : le deuxième paramètre de la loi log-normale régissant  $DS_k^j$ .
- La densité de la loi normale suivie par le logarithme des DS du GHM $_k$  de la base  $j$  est égale à :

$$f((\ln DS_k^j)) = \frac{1}{\sigma_{kj} \sqrt{2\pi}} \exp(-(\ln(DS_k^j) - \mu_{kj})^2 / 2\sigma_{kj}^2).$$

Pour ce faire, pour chacun des 9 GHM étudiés :

- nous avons estimé les paramètres des lois normales ajustant le logarithme des DS des bases nationales et grenobloise.

- Nous avons représenté, sur un même graphique, les histogrammes de DS des bases nationale et grenobloise avec leur courbe respective de densité log-normale.
- À partir de la densité nationale  $f(\ln(DS_k^N))$ , nous avons calculé les 3 quartiles de DS  $q_{(0.25)}^N$ ,  $q_{(0.50)}^N$  et  $q_{(0.75)}^N$ .
- À partir de la densité grenobloise,  $f(\ln(DS_k^G))$ , nous avons estimé les proportions cumulées de patients,  $p$ , ayant une DS inférieure ou égale aux DS correspondant respectivement à  $q_{(0.25)}^N$ ,  $q_{(0.50)}^N$  et  $q_{(0.75)}^N$ , où :

$$q(u) = \exp(\Phi^{-1}(u) * \sigma_{kj} + \mu_{kj})$$

et

$$p_{(\leq u)} = \Phi\left(\frac{\ln q(u) - \mu_{kj}}{\sigma_{kj}}\right)$$

avec :

$$u = 0.25, 0.50 \text{ et } 0.75,$$

$\Phi$  correspondant à la fonction de répartition de la loi normale

Rappelons que pour la base nationale, ces proportions sont respectivement égales à 0.25, 0.50 et 0.75.

- Nous avons réalisé un “PP-plot” des proportions estimées et observées de patients grenoblois et celles de la référence nationale.

Cette méthode nous a permis d’obtenir 9 nouveaux profils de distribution de DS estimés à partir d’une loi log-normale que nous avons comparés à ceux que nous avons construits empiriquement. Nous avons alors confronté les profils 2 à 2 afin de vérifier s’ils étaient proches des 9 profils grenoblois types.

### 3.2.2 Résultats

#### Influence des paramètres $\beta$ et $\theta$ sur la forme des profils

Les figures 3.10 et 3.11 montrent que les profils de types *(a)* et *(h)* sont obtenus en faisant varier positivement ou négativement le paramètre  $\beta$  correspondant à l’effet du CHU de Grenoble sur le paramètre  $\mu_B$ . La distribution grenobloise des DS se décale sur la droite si la variation du paramètre  $\beta$  sur  $\mu_B$  est positive (profil de type *(a)*): GHM “mauvais” quelle que soit la durée de leurs séjours, voir les densités et les profils de couleur verte sur les figures 3.10 et 3.11) et inversement si la variation du paramètre  $\beta$  sur  $\mu_B$  est négative (profil de type *(h)*): GHM bons quelle que soit la durée de leurs séjours, voir les densités et les profils de couleur bleue sur les figures 3.10 et 3.11). Plus l’intensité de l’effet du CHU de Grenoble,  $\beta$ , sur  $\mu_B$  est forte, plus le profil se trouve éloigné de la norme nationale. Ainsi, les profils de types *(a)* et *(h)* s’expliquent en grande partie par la seule variation de  $\mu_B$  si la distribution

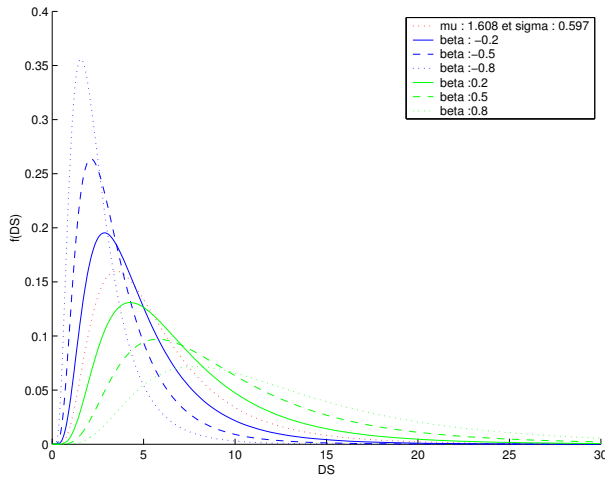


FIG. 3.10 – Densités log-normales selon les variations de  $\mu$  induites par  $\beta$ .

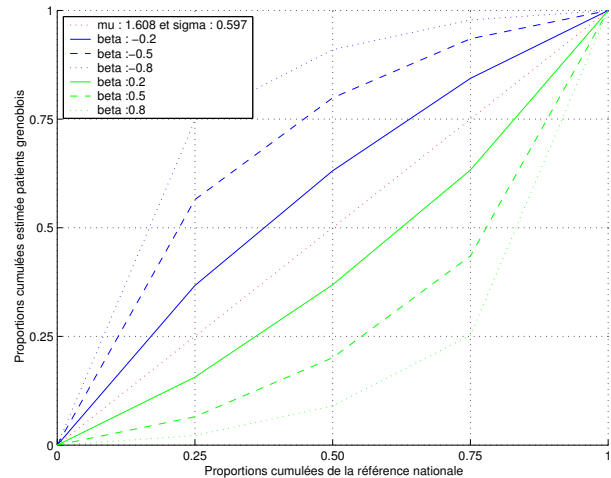


FIG. 3.11 – Profils estimés selon les variations de  $\beta$ .

des DS est modélisée par une loi log-normale.

Les figures 3.12 et 3.13 montrent qu’une partie seulement de la forme des profils de types (b), (c), (e) et (g) sont modélisés. En effet, si on considère simplement l’effet correspondant au CHU de Grenoble, le paramètre  $\gamma$ , sur le paramètre  $\sigma_B$ , tous les profils croisent la bissectrice à 0.5 (2<sup>e</sup> intervalle interquartile égal à 50%), alors que, ce n’est pas ce qui est observé dans notre échantillon. Cette valeur est toujours supérieure ou inférieure à 50%. On remarque néanmoins que, si la variation de  $\gamma$  sur  $\sigma_B$  est négative, on obtient le début et la fin des profils de types (b) et (e) (voir les densités et les profils de couleur verte sur les figures 3.12 et 3.13) : GHM identifiés comme médiocres sur les séjours de courte durée et performants sur les séjours de longue durée (distribution de DS décalée à droite et peu dispersée) et inversement si la variation de  $\gamma$  est positive, on observe les profils de types (c) et (g) (distribution de DS décalée à gauche et très dispersée, voir les densités et les profils de couleur bleue sur les figures 3.12 et 3.13).

Pour réussir à modéliser ces 4 profils, il faut considérer simultanément une variation modérée (positive ou négative) du paramètre  $\beta$  correspondant au CHU de Grenoble, sur le paramètre national  $\mu_B$ , et une variation importante (positive ou négative) du paramètre  $\gamma$  correspondant à l’effet du CHU de Grenoble, sur le paramètre de dispersion  $\sigma_B$  (voir figures 3.14 et 3.15). Pour obtenir un profil de type (b) (GHM identifiés comme médiocres sur les séjours de courte et de moyenne durées, voir la densité et le profil en trait pointillé de couleur bleue sur les figures 3.14 et 3.15), il faut simultanément considérer un effet positif  $\beta$  sur  $\mu_B$  et un effet négatif  $\gamma$  sur  $\sigma_B$ , et inversement pour le profil de type (g) (GHM “mauvais” sur les séjours de longue durée, voir densité et profil en trait pointillé de couleur verte sur les figures 3.14 et 3.15). Pour parvenir au profil de type (c) (GHM repérés comme peu

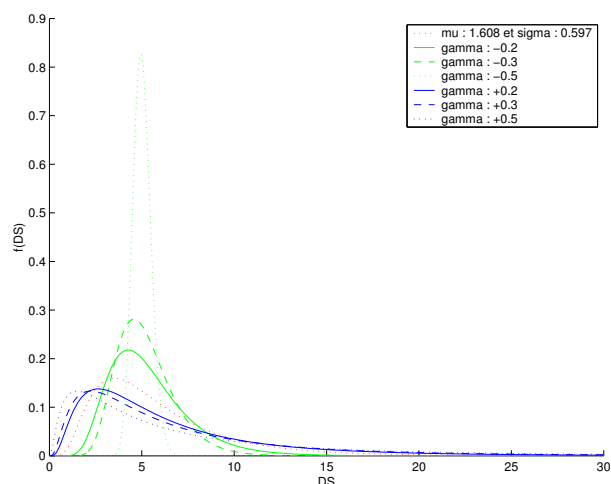


FIG. 3.12 – Densités log-normales selon les variations de  $\sigma_B$  induites par  $\gamma$ .

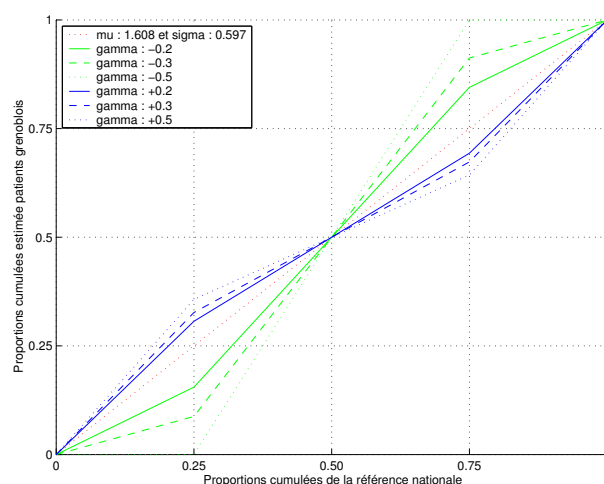


FIG. 3.13 – Profils estimés selon les variations de  $\gamma$ .

performants sur les séjours de longue durée, voir la densité et le profil en trait continu de couleur bleue sur les figures 3.14 et 3.15), il faut simultanément ajouter un effet positif  $\beta$  sur  $\mu_B$  et  $\gamma$  sur  $\sigma_B$ , et inversement pour le profil de type (e) (GHM médiocres sur les séjours de moyenne et longue durée, voir la densité et le profil en trait pointillé long de couleur verte sur les figures 3.14 et 3.15).

Nous remarquons qu'en inversant les conditions de modélisation, on obtient un profil différent ("mauvais" sur un tiers de la distribution contre "mauvais" sur les deux tiers) et non un profil inversé. Ces 4 profils sont plus dépendants de la variation observée sur l'écart-type que sur celle de la moyenne.

En considérant une variation (positive ou négative) importante du paramètre  $\beta$  sur  $\mu_B$  et une variation (positive ou négative) modérée du paramètre  $\gamma$  sur  $\sigma_B$ , on retrouve les profils de type (a) et de type (h) (voir figures 3.16 et 3.17). Plus les variations des paramètres  $\beta$  sur  $\mu_B$  et  $\gamma$  sur  $\sigma_B$  sont importantes, plus le profil se retrouve éloigné de la norme nationale.

Cependant, ces profils semblent dépendre plus souvent de la variation du paramètre  $\beta$  que du paramètre  $\gamma$ .

On observe le profil de type (neutre) quand il n'y a aucun effet du CHU de Grenoble sur  $\mu_B$  et  $\sigma_B$  (voir figures de 3.10 à 3.17, densité et profil de couleur rouge). Par contre, il n'est pas possible de modéliser les profils de type (d) (GHM médiocres sur les séjours de courte et longue durées) et (f) (GHM peu performants sur les séjours de moyenne durée) en utilisant une simple loi log-normale. Cela s'explique par le fait que les distributions correspondant à ces profils sont plutôt de forme bi-modale.

Après avoir vu quels étaient les paramètres dont dépendaient la forme des profils en les modélisant pas une simple loi log-normale, nous allons maintenant nous intéresser aux résultats donnés par la modélisation des 9 profils référents par une loi log-normale.

### Application du modèle aux 9 GHM référents

D'après les histogrammes de DS 3.18, 3.19, 3.21, 3.22, 3.24, 3.25, 3.27, 3.28, 3.30, 3.31, 3.33, 3.34, 3.36, 3.37, 3.39, 3.40, 3.42 et 3.43 nous voyons qu'un ajustement des distributions de DS de ces 9 GHM de référence par une simple loi log-normale n'est pas exact d'un point de vue statistique. Cependant, notre démarche n'était pas de trouver la loi qui ajustait le mieux les DS, mais d'appliquer le modèle réalisé avec des simulations aux données des 9 GHM référents et regarder si, avec une modélisation simple, il était possible de retrouver les profils que nous avons construits de manière empirique.

Les profils estimés pour les GHM n°118 (type (*neutre*)), n°48 (type (*a*)), n°230 (type (*e*)) et n°623 (type (*h*)) sont proches ou restent de même type que ceux que nous avons construits (voir figures 3.20, 3.23, 3.35 et 3.44). Cependant, les profils estimés des GHM n°230 et n°623 ont tendance à être d'amplitude moindre que ceux que nous avons obtenus.

Les profils estimés pour les GHM n°372, n°63 et n°458 sont tous de type (*a*) (voir figures 3.26, 3.29 et 3.41) : GHM "mauvais" quelle que soit la longueur de leurs séjours, alors que nous avons obtenu des profils de type :

- (*b*) pour le GHM n°372 : GHM "mauvais" sur les séjours de courte et de moyenne durées,
- (*c*) pour le GHM n°63 : GHM médiocres sur les séjours de moyenne et longue durées,
- (*g*) pour le GHM n°458 : GHM peu performants sur les séjours de longue durée.

Comme nous l'avons vu dans le paragraphe précédent, il n'est pas exact de modéliser les profils de type (*d*) (GHM médiocres sur les séjours de courte et longue durées) et (*f*) (GHM peu performants sur les séjours de moyenne durée) obtenus pour les GHM n°274 et n°197 par de simples lois log-normales, étant donné que leurs distributions sont plutôt bi-modales, comme le montrent les figures 3.30 et 3.36. Ainsi, les profils estimés sont de type (*c*) (GHM "mauvais" sur les séjours de courte et moyenne durées) pour le GHM n°274 et de type (*a*) pour le GHM n°197 (voir figures 3.32 et 3.38).

Néanmoins, le type de profil importe peu quand ce dernier est très proche de la norme nationale, cela étant synonyme d'une similitude des distributions de DS entre les bases nationale et grenobloise (cas des GHM n°118, n°197 et n°458).

### 3.3 Discussion

#### 3.3.1 Classification issue des profils empiriques

L'approche empirique a dégagé 9 profils de distributions de DS, tous observés dans notre échantillon. Ainsi, les 257 GHM grenoblois semblent se structurer en 9 groupes. Ils peuvent être identifiés comme "mauvais", en termes de distribution de DS, à des degrés divers : sur l'ensemble de la distribution (profil  $(a)$ ), sur les deux tiers (profils de types  $(b)$ ,  $(c)$  et  $(d)$ ) ou sur un tiers seulement (profils de types  $(e)$ ,  $(f)$  et  $(g)$ ). Un peu moins de la moitié des 257 GHM grenoblois étudiés (46.3%) ont une distribution de DS significativement différente de leur référence nationale. Cependant, le profil de type  $(d)$  n'est pas significativement différent de la référence nationale. Ainsi, les profils identifiés comme "mauvais" se structureraient en 6 groupes.

D'autre part, plus le profil est éloigné de la bissectrice (norme nationale), plus la différence d'aire entre le profil d'un GHM et la référence nationale est grande et négative, plus le GHM est identifié à juste titre comme "mauvais".

Les GHM appartenant aux 3 listes de référence se trouvent pour la majorité dans le groupe réunissant les profils de type  $(a)$ .

#### 3.3.2 Influence des paramètres sur la forme des profils

En modélisant les DS avec une simple loi log-normale, les simulations montrent que :

- les profils de types  $(a)$  et  $(h)$  dépendent essentiellement de l'intensité des effets du CHU de Grenoble,  $\beta$ , sur  $\mu_B$  correspondant au paramètre de la référence nationale,
- les profils de types  $(b)$ ,  $(c)$ ,  $(e)$  et  $(g)$  dépendent de l'intensité des effets simultanés du CHU de Grenoble,  $\beta$  et  $\gamma$ , sur les paramètres nationaux respectifs,  $\mu_B$  et  $\sigma_B$ ,
- les profils de types  $(d)$  et  $(f)$  ne sont pas modélisables avec une loi log-normale unimodale, la forme de leurs profils évoquant des distributions de DS bi-modales.

Ce résultat est très dépendant de l'hypothèse de log-normalité que nous avons faite. Mais il est intéressant de regarder l'influence des paramètres sur la forme des profils. Le modèle log-normal semble ainsi trop simple pour obtenir les profils de type  $(d)$  et  $(f)$ , qu'il faudra donc approcher avec des modèles de mélange.

#### 3.3.3 Modélisation des 9 profils référents

Cette analyse a montré que la modélisation des 9 profils référents par une simple loi log-normale n'était pas très adaptée. Un premier résultat intéressant est que les distributions de DS des 9 GHM référents, ne suivent pas tous une loi log-normale, alors que la majorité des études publiées dans la littérature font cette hypothèse [24, 68, 81, 85, 89].

Concernant le GHM n°118 (profil de type  $(neutre)$ ), le profil estimé par la loi log-normale ne diffère pas significativement de la première bissectrice. Dans le cas du GHM n°48, mauvais

sur toute la distribution des DS (profil de type (a)), le profil estimé est identique au profil observé. Dans le cas du GHM n°372, mauvais sur le premier tiers de la distribution (profil de type (b)), les faibles effectifs et l'asymétrie inverse de la distribution des DS grenobloise qu'une loi log-normale ne peut pas capter, peuvent expliquer la différence entre les profils observé et estimé. Dans le cas du GHM n°63, mauvais sur les 2 derniers tiers de la distribution (profil de type (c)), la différence peut s'expliquer par la bi-modalité de la distribution des DS grenoblois qu'une loi log-normale unimodale ne modélise pas. Dans le cas du GHM n°274, la différence entre le profil observé et le profil estimé peut s'expliquer par le fait que la loi log-normale n'arrive pas à prendre en compte le décentrage de + 3 jours de la distribution grenobloise. Pour le GHM n°197 (profil de type (f)), la bi-modalité de la distribution des DS peut expliquer la différence entre les profils estimé et observés. Les GHM n°230 (profil de type (e)), 458 (profil de type (g)) et 623 (profil de type (h)) ont des profils observés et estimés qualitativement comparables : les différences peuvent s'expliquer par le faible nombre de patients.

### 3.3.4 Synthèse

Les profils des GHM grenoblois semblent ainsi se structurer empiriquement en 9 groupes, dont 6 peuvent être considérés comme “mauvais” à des degrés divers : sur l'ensemble de la distribution, sur les deux tiers ou sur un tiers seulement, le profil de type (d) n'étant pas significativement différent de la référence nationale. Ces GHM empiriquement mauvais seront par la suite dénommés les “**mauvais E**”.

Par la suite, les profils (*neutre*) n'apportant que peu d'information intéressante dans notre étude des mauvais GHM, nous les confondrons avec les profils de type (h). Tous ces profils empiriquement bons seront appelés “**bons E**”. Nous ne considérerons donc dorénavant plus que 8 profils types.

Une modélisation par une loi log-normale unimodale ne suffit pas pour retrouver la majorité des profils des 8 GHM référents. Toutefois, si on fait cette hypothèse, les simulations montrent que :

- les profils de types (a) et (h) dépendent essentiellement de l'intensité des effets du CHU de Grenoble,  $\beta$ , sur  $\mu_B$  correspondant au paramètre de la référence nationale,
- les profils de types (b), (c), (e) et (g) dépendent de l'intensité des effets simultanés du CHU de Grenoble,  $\beta$  et  $\gamma$ , sur les paramètres nationaux respectifs,  $\mu_B$  et  $\sigma_B$ ,
- les profils de type (d) et (f) ne sont pas modélisables par une loi log-normale unimodale.

Il est intéressant d'étudier si, avec des méthodes de classification plus sophistiquées, les GHM se structurent selon leur profil empirique.



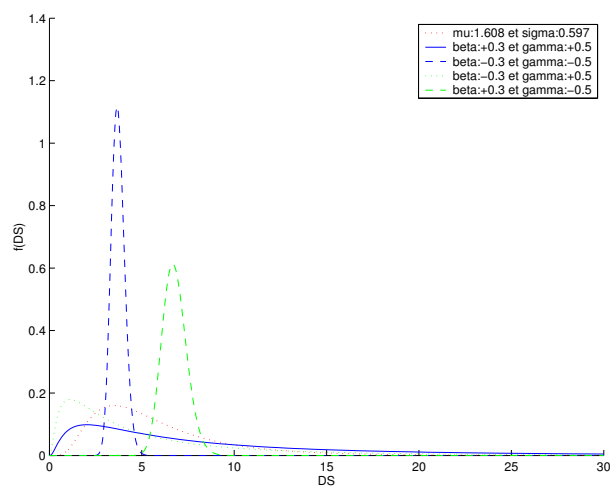


FIG. 3.14 – Densités log-normales selon les variations de  $\mu$  et de  $\sigma_B$  induites respectivement par  $\beta$  et  $\gamma$ .

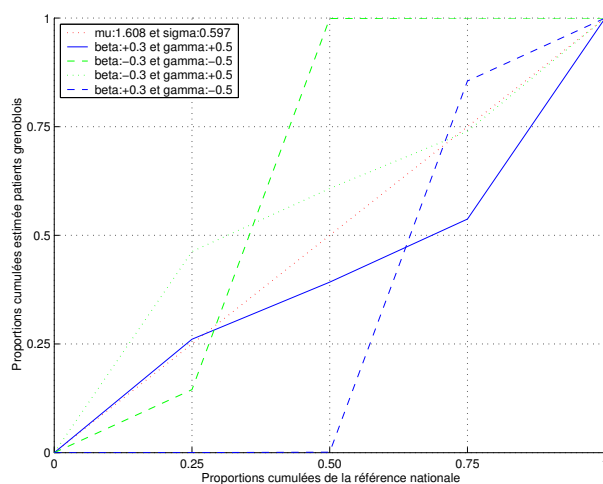


FIG. 3.15 – Profils estimés selon les variations de  $\beta$  et  $\gamma$  (première expérience).

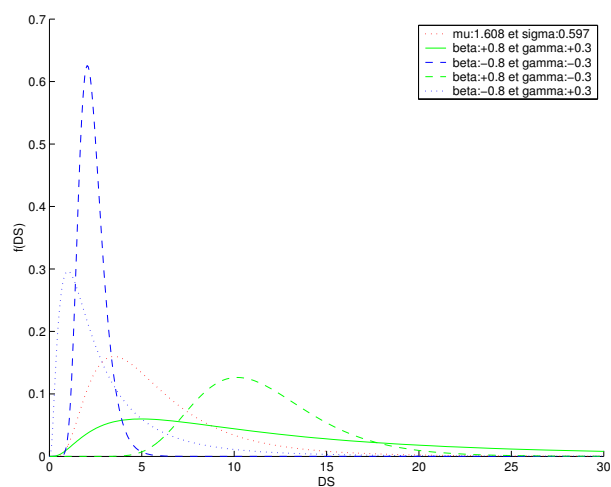


FIG. 3.16 – Densités log-normales selon les variations de  $\mu$  et de  $\sigma_B$  induites respectivement par  $\beta$  et  $\gamma$ .

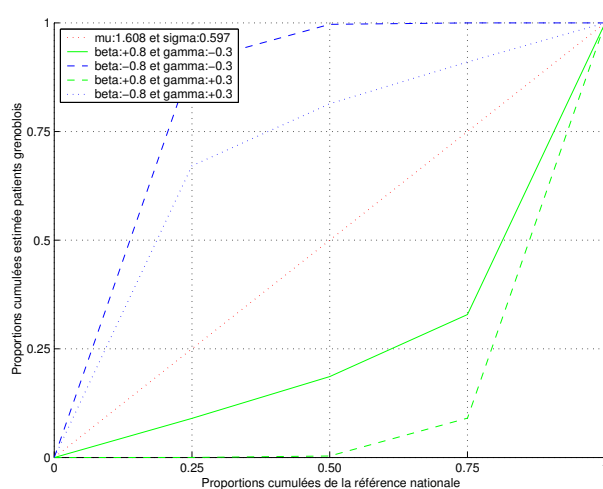


FIG. 3.17 – Profils estimés selon les variations de  $\beta$  et  $\gamma$  (deuxième expérience).

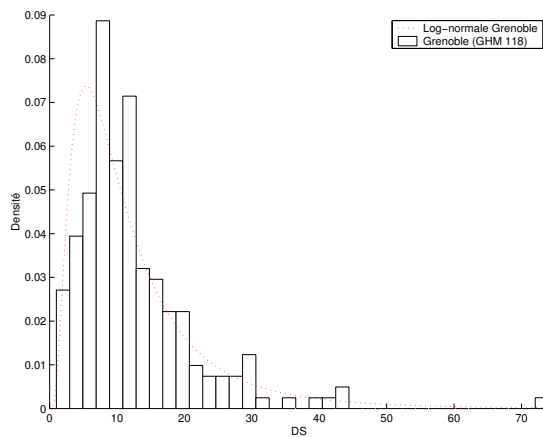


FIG. 3.18 – *Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 118).*

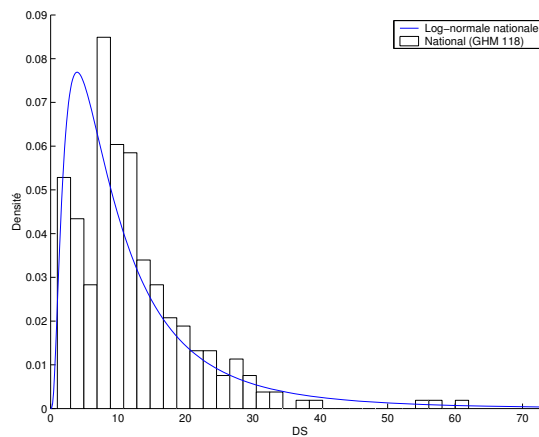


FIG. 3.19 – *Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 118).*

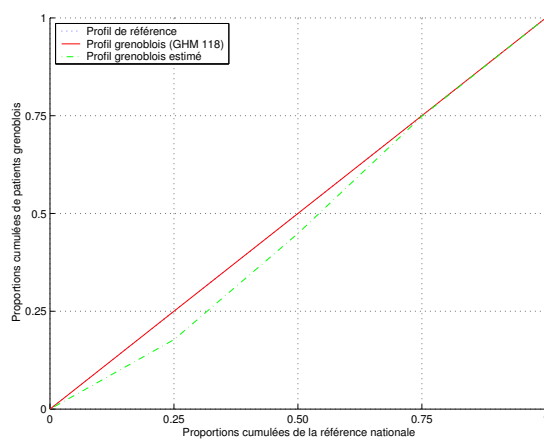


FIG. 3.20 – *Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 118).*

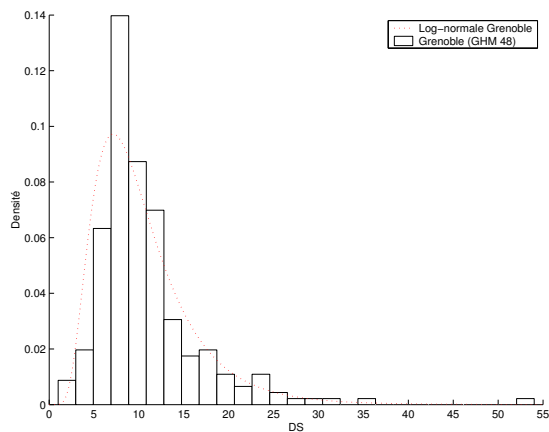


FIG. 3.21 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 48).

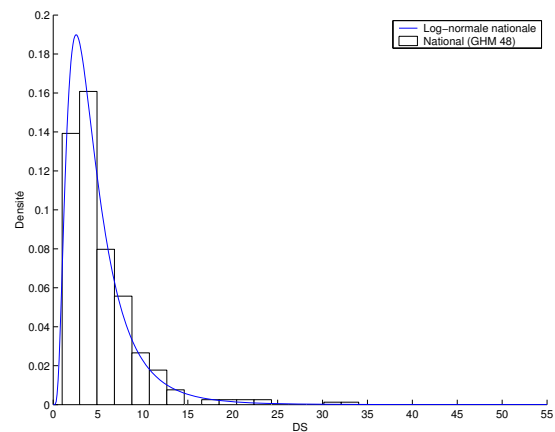


FIG. 3.22 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 48).

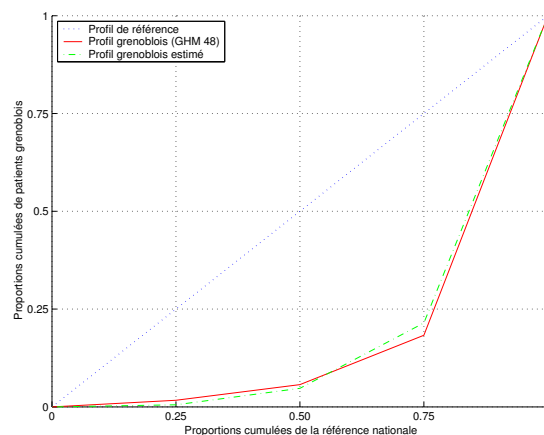


FIG. 3.23 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 48).

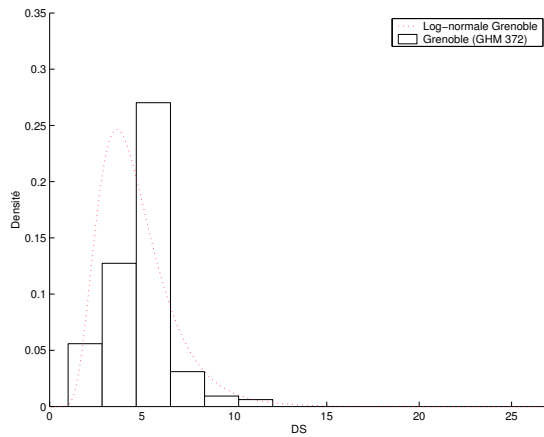


FIG. 3.24 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 372).

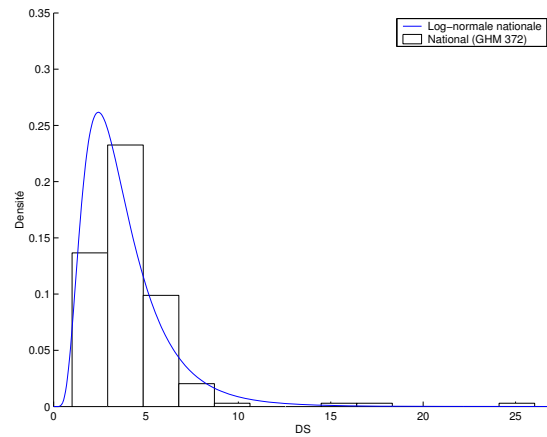


FIG. 3.25 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 372).

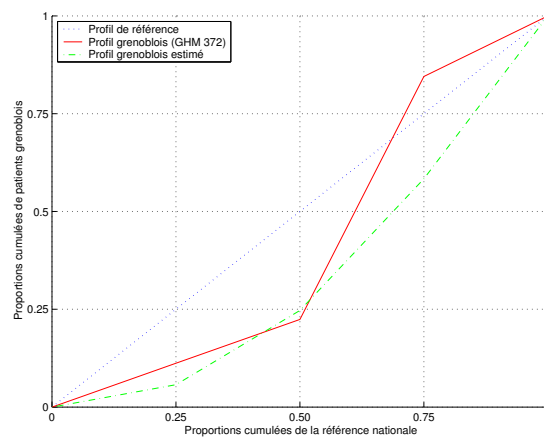


FIG. 3.26 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 372).

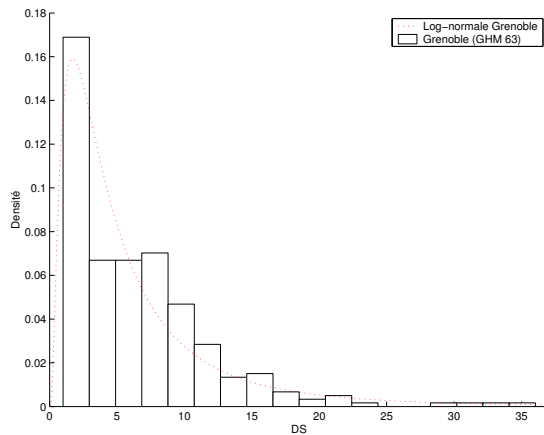


FIG. 3.27 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 63).

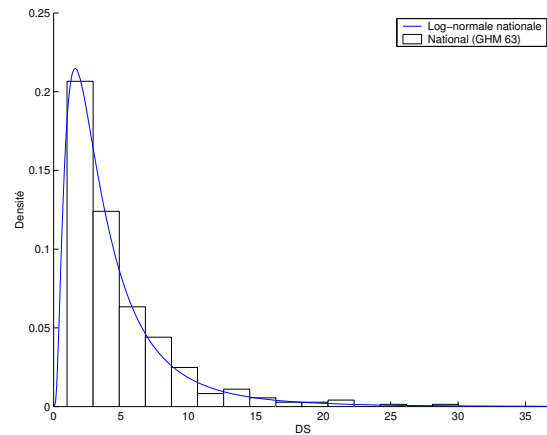


FIG. 3.28 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 63).

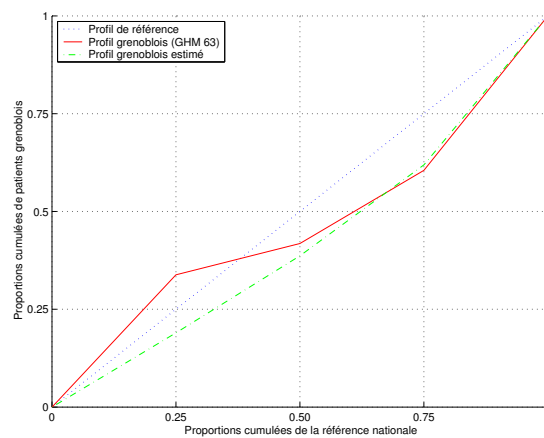


FIG. 3.29 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 63).

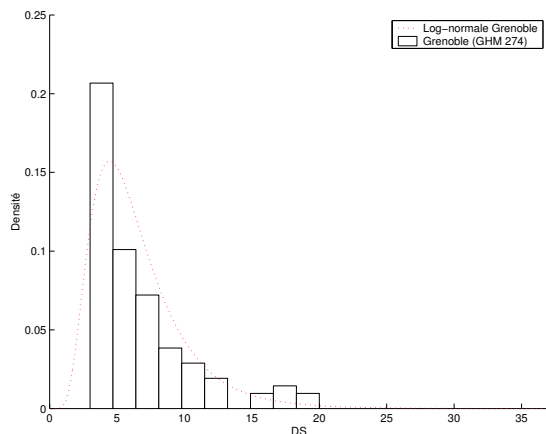


FIG. 3.30 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 274).

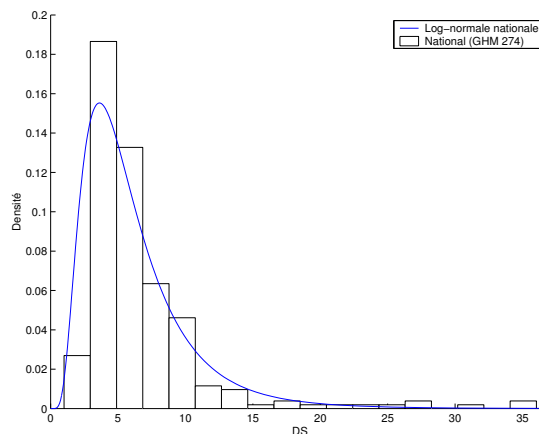


FIG. 3.31 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 274).

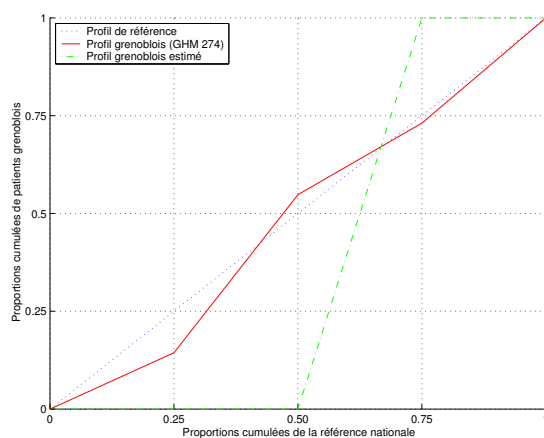


FIG. 3.32 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 274).

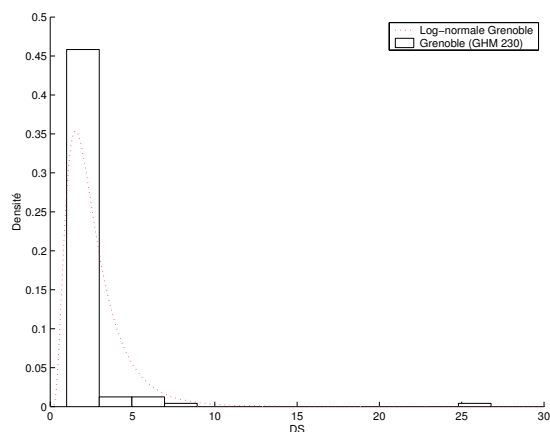


FIG. 3.33 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 230).

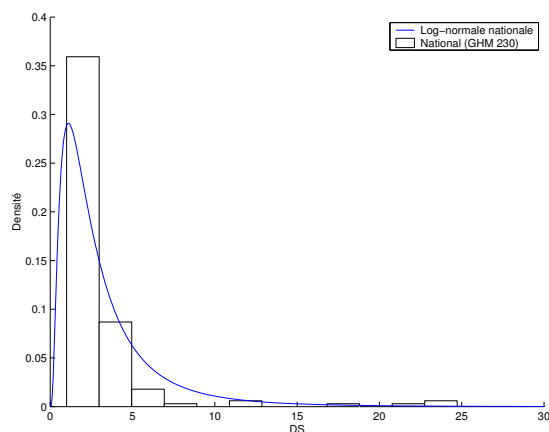


FIG. 3.34 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 230).

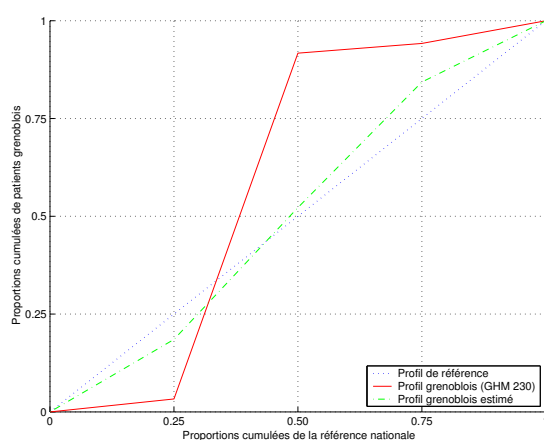


FIG. 3.35 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 230).

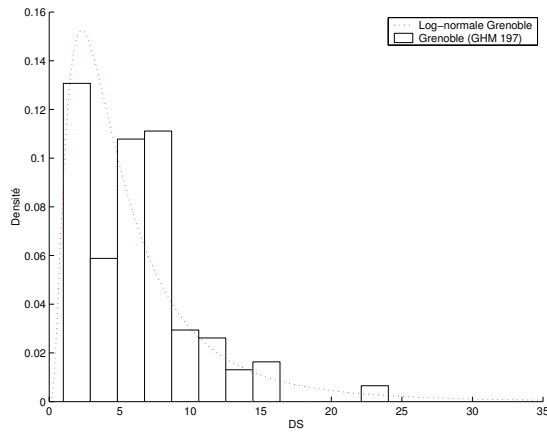


FIG. 3.36 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 197).

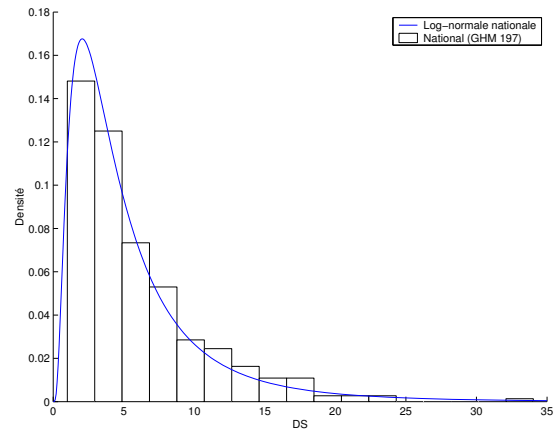


FIG. 3.37 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 197).

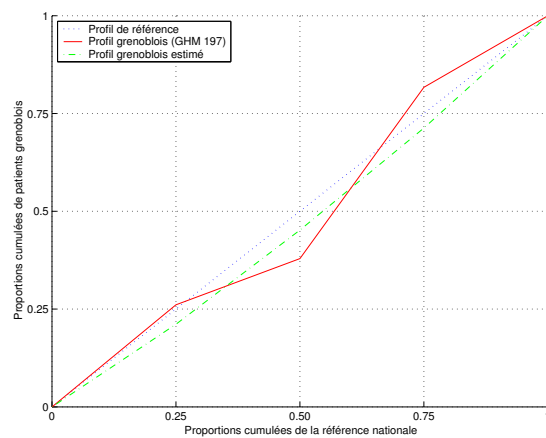


FIG. 3.38 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 197).



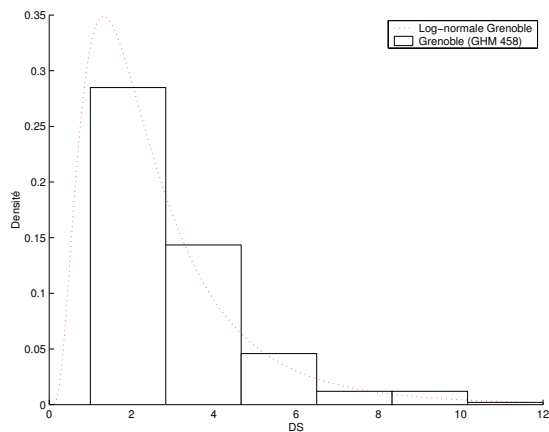


FIG. 3.39 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 458).

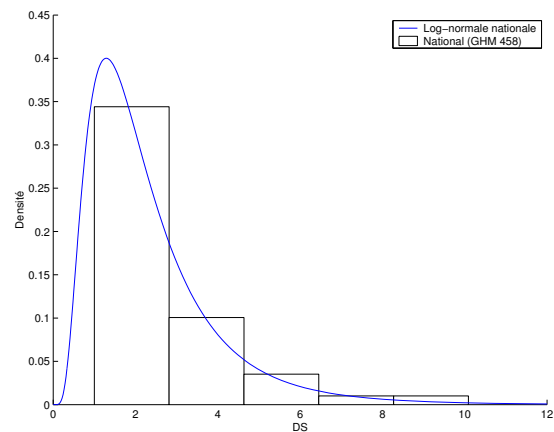


FIG. 3.40 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 458).

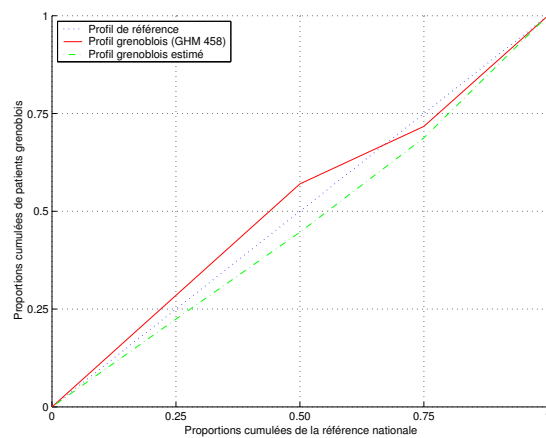


FIG. 3.41 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 458).

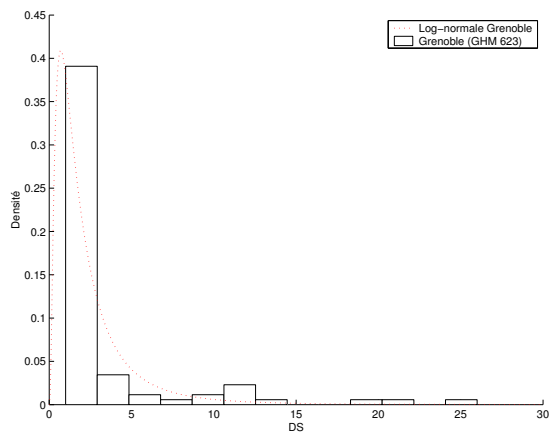


FIG. 3.42 – Histogramme des DS de la base grenobloise et approximation par une loi log-normale (GHM n° 623).

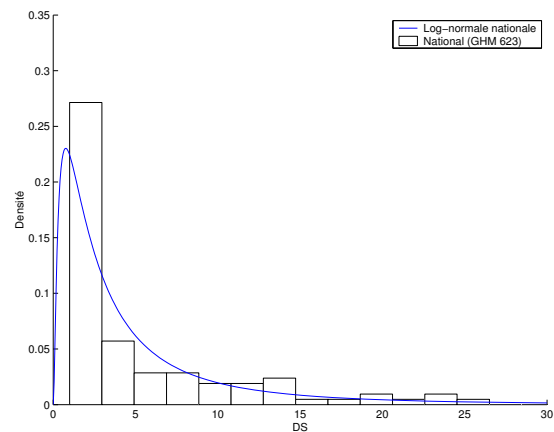


FIG. 3.43 – Histogramme des DS de la base nationale et approximation par une loi log-normale (GHM n° 623).

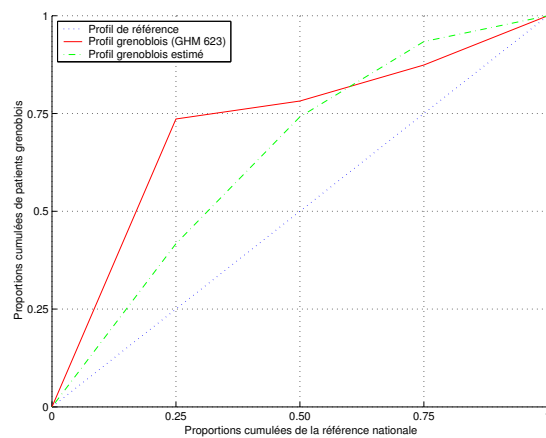


FIG. 3.44 – Profils de distribution des DS grenobloises, observé et estimé par une loi log-normale, par rapport à la référence nationale (GHM n° 623).



# Chapitre 4

## Classification des profils issue de l'approche factorielle

L'approche empirique nous a permis de montrer qu'il existait huit profils types de distributions de DS pour les 257 GHM grenoblois (six d'entre-eux pouvaient être considérés comme mauvais), mais également de préciser leurs caractéristiques médico-économiques. De manière à disposer d'une vue d'ensemble de ces différents profils et afin d'étudier s'ils se structurent effectivement en classes, nous avons utilisé différentes méthodes factorielles, comme l'analyse en composantes principales, la classification hiérarchique ascendante et l'analyse discriminante. Ces méthodes ont l'intérêt de montrer si des données se structurent en classes, mais aussi de les identifier, de les décrire et de rechercher une fonction de classement automatique des données. Nous verrons cela de manière plus détaillée à travers ce chapitre.

Nous présenterons d'abord la stratégie de modélisation que nous avons adoptée. Nous définirons, ensuite, les principes mathématiques sur lesquels reposent les différentes méthodes que nous avons utilisées. Pour finir, nous présenterons les résultats de ces analyses et nous les discuterons.

### 4.1 Stratégie de modélisation

Pour avoir une vue d'ensemble des profils et déterminer s'ils se structuraient en classes, nous avons utilisé l'analyse en composantes principales (ACP). Ensuite, comme les facteurs de l'ACP laissaient présager l'existence de groupes de GHM, nous avons utilisé la classification hiérarchique ascendante pour les identifier. Nous avons étudié la signification de ces groupes de GHM en termes physique (type de profil) et économique (déficit en points ISA). Pour finir, nous avons tenté d'affecter un GHM donné à une classe correspondant aux différents groupes ainsi constitués, avec une faible probabilité d'erreur, en utilisant l'analyse discriminante.

## 4.2 Méthodes utilisées

Le but de ce paragraphe n'est pas de "redémontrer" les concepts mathématiques des trois méthodes utilisées : l'ACP, la classification hiérarchique ascendante et l'analyse discriminante, mais de présenter de manière simple leur fonctionnement et les intérêts qu'elles apportent dans le cadre de notre travail.

### 4.2.1 Principes de l'ACP

L'ACP [43, 60, 77, 79] a deux objectifs majeurs. Le premier est de déterminer des représentations graphiques planes, appelées cartes, qui permettent d'analyser au mieux les proximités et les écarts entre les différentes variables. En particulier, elles doivent permettre d'isoler les individus "atypiques" et de regrouper les individus "semblables".

Le deuxième objectif de l'ACP est de fournir des "résumés linéaires" des variables, c'est-à-dire de remplacer les variables initiales par des combinaisons linéaires de celles-ci. Ces nouvelles variables, appelées composantes principales ou facteurs, ont des propriétés qu'il est important de présenter pour faciliter l'interprétation de l'analyse.

#### Cadre d'application

Les variables doivent être de type quantitatif, hétérogènes et ne doivent pas être globalement indépendantes. En effet, l'ACP procède par réduction de la dimension de l'espace des variables en tirant parti des corrélations entre les variables initiales.

#### Tableau de données

TAB. 4.1 – *Exemple de fichier de données.*

	$Y^1$	...	$Y^j$	...	$Y^p$
$X_1$			⋮		
⋮			⋮		
$X_i$	...	...	$k_{ij}$	...	...
⋮			⋮		
$X_n$			⋮		

On considère  $n$  individus  $X_1, \dots, X_n$  sur lesquels on observe les valeurs de  $p$  variables  $Y^1, \dots, Y^p$ . L'observation de la variable  $Y^j$  sur l'individu  $X_i$  est noté  $k_{ij}$ . Ces données sont présentées dans le tableau 4.1.

Pour chaque variable  $Y^j$ , on calcule sa moyenne :

$$\bar{Y}^j = \frac{1}{n} \sum_{i=1}^n k_{ij}$$

et sa variance :

$$s_j^2 = \frac{1}{n} \sum_{i=1}^n (k_{ij} - \bar{Y}^j)^2$$

Afin d'éliminer les effets dus aux différences des moyennes, on centre les variables  $Y^j$ . On travaillera donc avec un tableau centré  $Z$  de terme général  $z_{ij} = k_{ij} - \bar{Y}^j$ . Souvent, on réduit aussi les variables, afin d'éliminer les effets dus aux choix des unités. Dans notre cas, ceci n'a pas été nécessaire car les intervalles interquartiles s'expriment tous en proportions, donc dans la même unité.

### Nuage des individus

Considérons le nuage de points  $\mathcal{N}$  composé de  $n$  points  $X_i$  dans  $\mathbb{R}^p$ ,  $G$  étant le centre de gravité. Sa coordonnée représente la valeur de la variable  $Y^j$  sur cet individu.

Nous cherchons à mesurer la distance entre deux individus. Comme nos données s'expriment dans la même unité, nous avons choisi une métrique diagonale,  $M$ . Ainsi, la distance entre deux individus  $X_i, i \in \{1, \dots, n\}$  et  $X_{i'}, i' \in \{1, \dots, n\}$  est définie par la forme quadratique suivante :

$$d^2(X_i; X_{i'}) = (X_i - X_{i'})^t M (X_i - X_{i'}) = \sum_{i=1}^n m_i (X_i - X_{i'})^2$$

Dans la plupart des cas, les  $n$  individus ont tous le même poids,  $1/n$ , dans le calcul des caractéristiques de l'échantillon. Il n'en est pas toujours ainsi et il est utile, dans certains cas, de travailler avec des poids  $m_i$  éventuellement différents d'un individu à l'autre. Ces poids, qui sont des nombres positifs de somme 1, comparables à des fréquences, sont regroupés dans une matrice diagonale  $M$  de taille  $n$  :

$$M = \begin{bmatrix} m_1 & & & & 0 \\ & \ddots & & & \\ & & m_i & & \\ & & & \ddots & \\ 0 & & & & m_n \end{bmatrix}$$

Dans le cas le plus usuel de poids égaux, on a :

$$M = \frac{1}{n} I.$$

Le vecteur  $G$  des moyennes arithmétiques de chaque variable définit le point moyen :

$$G^t = (\bar{z}_{.1}, \dots, \bar{z}_{.p}).$$

On a  $G = Z^t M 1$ , où  $1$  désigne le vecteur de  $\mathbb{R}^p$  dont toutes les composantes sont égales à 1. Les variables étant centrées, le centre de gravité  $G$  du nuage se trouve donc en 0 et :

$$G^t = (0, \dots, 0).$$

On effectue alors une analyse factorielle du nuage  $\mathcal{N}$  en  $G = 0$ .

La matrice d'inertie est :

$$V = Z^t Z.$$

Les  $z_{ij}$  étant centrés,  $V$  correspond à la matrice des variances-covariances,  $s_{ij}$ , des  $Y^j$ . Par contre, si les  $z_{ij}$  avaient été réduits,  $V$  serait la matrice des corrélations  $(\rho_{j,j'})$ .

L'inertie totale du nuage de points  $\mathcal{N}$  correspond à la moyenne pondérée des carrés des distances des points au centre de gravité. En  $G = 0$ , elle est donc :

$$In = \text{trace}(VM) = p$$

L'inertie est donc égale au nombre de variables et ne dépend pas de leurs valeurs.

L'inertie expliquée par un axe  $e_j$  de la base canonique, c'est-à-dire par une variable  $Y^j$ , est :

$$In(e_j) = e_j^t V e_j = \text{cov}(e_j^t, e_j) = s_j^2$$

La diagonalisation de  $V$  permet d'obtenir ses  $p$  valeurs propres  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . Les axes associés,  $U_1, U_2, \dots, U_p$ , sont appelés facteurs principaux.

L'inertie expliquée par le premier facteur principal est  $\lambda_1$ . Elle est nécessairement supérieure à celle expliquée par chaque variable  $Y^j$ , d'où  $\lambda_1 \geq 1$ .

Plus généralement, on ne considère dans l'ACP que les facteurs principaux correspondant à des valeurs propres supérieures à 1.

Le nuage de points  $\mathcal{N}$  est alors projeté sur le sous-espace engendré par les facteurs principaux considérés, en suivant la méthode de l'analyse factorielle.

### Nuage des variables

On considère le nuage de points  $\mathcal{N}'$  composé de  $p$  points  $Y^j$  dans  $\mathbb{R}^n$ . Chaque variable  $Y^j$  est donc représentée dans  $\mathbb{R}^n$ . Chacune de ses coordonnées représente la valeur de la variable  $Y^j$  sur un des  $n$  individus.

Pour étudier la proximité des variables entre elles, il faut munir cet espace d'une métrique.

Les variables de notre étude étant centrées, le carré de la distance entre deux variables  $Y^j$ ,  $j \in \{1, \dots, p\}$  et  $Y^{j'}$ ,  $j' \in \{1, \dots, p\}$ , est égal à :

$$d^2(Y^j; Y^{j'}) = \text{var}(Y^j) + \text{var}(Y^{j'}) - 2\text{cov}(Y^j, Y^{j'}) = s_j^2 + s_{j'}^2 - 2s_{jj'}.$$

Si  $Y^j$  et  $Y^{j'}$  sont corrélées alors :

$$\text{cov}(Y^j, Y^{j'}) > 0 \quad \Rightarrow \quad Y^j \text{ et } Y^{j'} \text{ proches}$$

Si  $Y^j$  et  $Y^{j'}$  sont non corrélées alors :

$$\text{cov}(Y^j, Y^{j'}) = 0 \quad \Rightarrow \quad Y^j \perp Y^{j'}$$

Si  $Y^j$  et  $Y^{j'}$  sont anti-corrélées alors :

$$\text{cov}(Y^j, Y^{j'}) < 0 \quad \Rightarrow \quad Y^j \text{ et } Y^{j'} \text{ opposées}$$

Le carré de la distance entre deux variables s'exprime en terme de covariance et augmente avec les variances. Elle diminue si la liaison est positive et augmente si la liaison est négative.

Le carré de la distance d'une variable à l'origine des axes est sa variance :

$$d^2(0, Y^j) = \text{var}(Y^j) = s_j^2$$

On effectue maintenant une analyse factorielle (duale) du nuage de points  $\mathcal{N}'$  en 0. La matrice d'inertie du nuage  $\mathcal{N}'$  est :

$$\Gamma = ZZ^t$$

L'inertie du nuage  $\mathcal{N}'$  en 0 est :

$$In = \text{trace}(\Gamma) = p$$

Les valeurs propres (non nulles) de  $\Gamma$  sont les mêmes que celles de  $V$  :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p.$$

Les axes propres associés correspondent aux vecteurs :

$$W_1 = \frac{1}{\sqrt{\lambda_1}} ZU_1, \dots, W_p = \frac{1}{\sqrt{\lambda_p}} ZU_p,$$

Ce sont les composantes principales, elles peuvent être considérées comme de nouvelles variables :

- combinaisons linéaires des variables  $Y^j$ ,
- non corrélées entre elles,
- de variance maximale.

La composante principale  $W_1$  est la variable la plus liée aux  $Y^j$ , dans le sens où la somme des carrés des coefficients de corrélation,  $\sum_{j=1}^p \rho^2(W_1, Y^j)$ , est maximale. Il en est de même, en décroissant, pour  $W_2, \dots, W_p$ , avec  $W_2$  non corrélée avec  $W_1$ , etc.



### Interprétation d'une ACP

Comme nous l'avons vu précédemment, les valeurs propres associées à la diagonalisation de la matrice d'inertie des nuages de points  $\mathcal{N}$  et  $\mathcal{N}'$  sont identiques. Il est donc inutile de procéder à deux diagonalisations. Les logiciels de statistiques travaillent directement sur le nuage de points  $\mathcal{N}'$  et diagonalisent donc la matrice  $\Gamma = ZZ^t$ .

En général, on ne retient que les valeurs propres  $\lambda_k$  supérieures à 1 ; les composantes principales  $W$  correspondantes auront alors une variance supérieure à celle des variables initiales  $Y^j$ . Ce nombre  $k$  de composantes principales retenues dépend du pourcentage d'inertie totale expliquée. On le calcule avec l'indicateur  $F_k$  égal à :

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\lambda_1 + \lambda_2 + \dots + \dots + \lambda_p}.$$

On retient souvent les  $k$  composantes principales qui vérifient  $F_k = 80\%$ . L'interprétation se déroule en deux étapes :

- on considère d'abord le nuage de points  $\mathcal{N}'$  des variables, et on en interprète les axes : les positions des variables initiales  $Y^j$  permettent de donner un sens aux axes ;
- on passe ensuite au nuage de points  $\mathcal{N}$  des individus ; son interprétation utilise celle des axes faite à l'étape 1.

### Interprétation du nuage des variables

Le nuage  $\mathcal{N}'$  est projeté sur le plan factoriel  $(W_1, W_2)$ . Comme les variables  $Y^j$  de notre fichier sont simplement centrées, il est impossible de lire directement sur le graphique la qualité de la représentation d'une variable ; il faut donc utiliser la valeur du cosinus carré de l'angle entre l'axe principal  $W_1$  et le vecteur reliant le centre de gravité du nuage à cette variable. Ainsi, plus cette valeur est élevée, meilleure est la représentation de la variable  $Y^j$  par le plan factoriel et inversement.

Les coordonnées de  $Y^j$  dans le plan factoriel  $(W_1, W_2)$  sont égales aux covariances des  $Y^j$  avec chacune des composantes principales :

$$\frac{\text{cov}(Y^j, W_1)}{s_j} \quad \text{et} \quad \frac{\text{cov}(Y^j, W_2)}{s_j}$$

La contribution d'une variable  $Y^j$  à l'inertie expliquée par une composante principale  $k$  est proportionnelle au carré de sa coordonnée sur cette composante :

$$\text{CTR}_{\lambda_k}(j) = \frac{\text{cov}(Y^j, W_{\lambda_k})^2}{s_j^2 \lambda_k}.$$

Cette contribution est confondue avec la covariance de la variable avec la composante.

### Interprétation du nuage des individus

Le sens donné aux axes factoriels lors de l'interprétation du nuage de points  $\mathcal{N}'$  permet une interprétation du nuage de points  $\mathcal{N}$  des individus.

Comme dans l'analyse factorielle générale, le nuage  $\mathcal{N}$  est projeté sur le plan factoriel  $(U_1, U_2)$ .

La contribution d'un individu  $X_i$  à l'inertie expliquée par un facteur principal est proportionnelle au carré de sa coordonnée sur cette composante :

$$\text{CTR}_{\lambda_k}(i) = \frac{W_{ki}^2}{\lambda_k},$$

où  $W_{ki}$  est la valeur de la  $k^{\text{e}}$  composante pour l'individu concerné. Plus  $X_i$  est loin de l'axe, plus sa contribution à celui-ci est élevée et inversement. Cependant, il n'est pas souhaitable qu'un individu ait une contribution excessive, surtout pour les premières composantes, car cela serait un facteur d'instabilité, le fait de retirer cet individu modifiant souvent profondément le résultat de l'analyse. Si c'est le cas, il faut refaire l'analyse en éliminant cet individu, puis le porter en élément supplémentaire s'il ne s'agit pas d'une donnée erronée ainsi mise en évidence.

Par contre, il est impossible de lire la qualité de la représentation d'un individu  $X_i$  dans le plan factoriel, il est nécessaire de faire appel à la valeur de ses cosinus carrés,  $CO2_i$  des angles qu'ils forment avec le plan factoriel  $(U_1, U_2)$ ,  $QLT_i = CO2_i^1 + CO2_i^2$ .  $QLT_i$  désigne la qualité de la représentation de l'individu  $X_i$  sur le plan principal. Plus les valeurs de ces indicateurs sont élevées, meilleures sont les représentations des individus par le plan factoriel et inversement.

### Intérêt de l'ACP

Dans le cadre de notre étude, ce sont plutôt les aspects descriptifs de l'ACP qui nous ont intéressés. En effet, un des objectifs de cette méthode est la représentation des données, qui a un caractère optimal selon certains critères algébriques et géométriques [60]. En changeant les coordonnées, elle privilégie les combinaisons linéaires entre des variables quantitatives (notre cas), afin de trouver une ou plusieurs composantes qui résument ces variables. D'autre part, elle a aussi l'avantage de ne faire référence à aucune hypothèse statistique ni à aucun modèle quelconque. En représentant nos données symbolisant les profils de distributions de DS des GHM grenoblois sans transformation géométrique, aucune classification ne ressortait, alors que nous avons vu que l'approche empirique suggérait 9 profils caractéristiques du CHU Grenoble. C'est pourquoi, en utilisant cette méthode, nous allons pouvoir améliorer la vue d'ensemble des différents profils (surtout s'il n'apparaît pas de classification de ces profils, comme l'a montré l'approche empirique). Les représentations graphiques permettent d'analyser au mieux les proximités et les écarts entre les différentes variables. Elles ont aussi l'avantage de permettre d'isoler les individus "atypiques", de regrouper ceux qui sont "semblables", mais aussi de les caractériser. C'est exactement ce que l'on cherche à faire : voir s'il

existe, à Grenoble, des GHM “atypiques” en terme de profils de distribution de DS, quelles sont leurs formes et quelles sont les pathologies concernées.

D'autre part, avec cette méthode, il nous est possible de résumer nos 3 indicateurs (les 3 premiers intervalles interquartiles de DS en proportions cumulées) par une ou deux combinaisons linéaires de ceux-ci, non corrélées entre-elles. Ainsi, nous pouvons réduire le nombre de variables caractérisant les profils grenoblois sans perdre d'information, en obtenant des représentations graphiques de très bonne qualité.

Nous allons maintenant appliquer cette méthode factorielle à notre échantillon de profils de distribution de DS.

### **Application à la classification des profils de distribution des DS des GHM grenoblois**

Notre but est de comparer les profils des distributions des DS des 257 GHM grenoblois selon trois indicateurs : les proportions cumulées de patients grenoblois dans les 3 premiers intervalles interquartiles de DS des mêmes GHM de la BDD nationale.

Comme nous l'avons vu dans la section 2.2, notre fichier d'analyse se compose de 257 individus (les GHM) décrits par 3 variables actives de type continu (les proportions cumulées de patients grenoblois dans les 3 premiers intervalles interquartiles) et 5 variables de type indicateur (le numéro de GHM, les 3 listes de référence et le type des profils empiriques de distribution de DS).

Nous avons réalisé une ACP centrée sur les 3 variables et les 257 individus ; chaque GHM est alors défini par un point qui identifie le profil de distribution de ses DS. L'ensemble des GHM est représenté par un ensemble de points qui forme le nuage des profils de distribution des DS.

Cette analyse a été réalisée avec le logiciel SPAD version 5.3.

#### **4.2.2 Principe de la classification hiérarchique**

Les techniques de classification hiérarchique ascendante ne requièrent aucune condition de validité [43, 60, 79]. Elles vont nous permettre de découper notre échantillon de GHM en groupes homogènes en termes de profils. Ainsi, nous allons pouvoir examiner si les profils des GHM de notre échantillon se structurent effectivement en groupes, comme l'ont laissé présager les approches empirique et factorielle, mais également de déterminer le nombre de ces groupes.

Ce découpage en groupe est effectué à partir de la structure même des données. Ils nous aide à visualiser les différences entre les GHM en termes de distance.

Il faut pour cela définir au préalable :

- les variables qui vont servir de base à la formation des groupes
- la mesure de la distance entre les profils
- les critères utilisés pour agréger les profils dans les groupes

Nous allons effectuer ce découpage à partir des GHM. Les variables de l'analyse seront les GHM de notre échantillon, ceux que l'on a déjà utilisés dans l'ACP.

La distance que nous avons choisie pour notre analyse est la distance euclidienne non normée :

$$d(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$$

d'autres distances peuvent être utilisées, comme la distance de Manhattan (city-block), qui utilise la somme des valeurs absolues des différences, ou la distance de Chebyshev, qui prend le maximum de la valeur absolue des différences.

Plusieurs hiérarchies seront effectuées avec différents critères d'agrégation, c'est-à-dire de calculs de distance entre les groupes :

- critère du voisin le plus éloigné ; dans cette méthode la distance entre deux groupes,  $g_1$  et  $g_2$ , est calculée comme étant la distance entre les deux points,  $x_i$  et  $y_j$ , les plus éloignés (“complex linkage”) :

$$d(g_1, g_2) = \max[d(x_i, y_j)],$$

- critère du voisin le plus proche ; la distance entre deux groupes,  $g_1$  et  $g_2$ , est calculée comme étant la distance entre les deux points,  $x_i$  et  $y_j$ , les plus proches (“single linkage”) :

$$d(g_1, g_2) = \min[d(x_i, y_j)],$$

- critère barycentrique ; la distance entre 2 groupes,  $g_1$  et  $g_2$ , est calculée comme étant la distance entre les deux centres de gravité des groupes, centre( $c_1$ ) et centre( $c_2$ ) :

$$d(g_1, g_2) = d[\text{centre}(c_1), \text{centre}(c_2)],$$

- critère de la moyenne ; la distance entre 2 groupes,  $g_1$  et  $g_2$ , est calculée comme étant la distance moyenne entre les points  $x_i$  et les points  $y_j$  :

$$d(g_1, g_2) = \text{moyenne}[d(x_i, y_j)],$$

- critère de la variation minimale d'inertie (méthode de Ward) ; sur un ensemble  $G_1$  et  $G_2$  de centre  $g_1$  et  $g_2$ , on calcule :

$$\frac{m_1 m_2}{m_1 + m_2} \|g_1 - g_2\|^2$$

où  $m_i = (\text{Card } G_i)/N$ , avec  $N$  le nombre total d'individus.

Il est nécessaire de réaliser plusieurs hiérarchies selon différents critères d'agrégation, afin de déterminer les groupes stables, c'est-à-dire, pour notre étude, les GHM qui s'agrègent le plus souvent ensemble dans la même classe, quel que soit le critère d'agrégation utilisé. Nous les appellerons "noyaux stables".

On visualise les distances entre les groupes de GHM par des traits matérialisés sous forme d'un arbre, où chaque feuille est un GHM, et la longueur de chaque branche, une distance entre deux GHM ou groupes de GHM; l'ensemble forme ce que l'on appelle un arbre hiérarchique ou dendrogramme. L'agrégation des GHM est dessinée en fonction de la distance et du critère utilisé. L'intérêt de cet arbre est qu'il peut donner une idée du nombre de classes existant effectivement au sein des GHM. Le nombre de classes sera choisi de façon pragmatique.

Nous avons exclu de notre analyse les "bons E", GHM ne créant pas de problème en terme de distribution de DS (profils empiriques de type *(h)* et *(neutre)*). Ainsi, étudier uniquement les "mauvais E" nous permet de mieux séparer les formes des profils, un peu comme un "zoom" en optique le ferait pour affiner le détail d'un ensemble.

Une fois les noyaux stables identifiés, nous nous sommes intéressés à décrire physiquement (forme de leurs profils), économiquement (déficit en point ISA) et médicalement (type de GHM : médical ou chirurgical) chacun des groupes issus de la classification. Nous avons regardé pour chacun d'entre-eux :

- le profil moyen (valeurs prises par les 3 intervalles interquartiles de DS),
- le type de profil empirique,
- la différence d'aire moyenne entre Grenoble et la référence nationale de chaque groupe,
- le bénéfice ou le déficit moyen en nombre de points ISA,
- le pourcentage de GHM chirurgicaux et médicaux.

La mise en évidence d'un lien possible entre les caractéristiques médico-économiques des GHM et les groupes issus de la classification a été faite par une analyse de variance entre les différents groupes au seuil  $\alpha = 5\%$  si la variable était de type quantitatif (différence d'aire entre un profil grenoblois et la référence nationale, et bénéfice ou déficit de points ISA), ou par un test du  $\chi^2$  si la variable était de type qualitatif (type de GHM : médical ou chirurgical).

Toutes ces analyses ont été réalisées avec le logiciel Matlab version 6.0.

### 4.2.3 Principe de l'analyse discriminante

L'analyse factorielle discriminante selon Fisher [43, 51, 60, 77, 79] a pour principe de sélectionner les variables les plus pertinentes pour la séparation des groupes. Les facteurs discriminants trouvés, appelés fonctions discriminantes, sont des combinaisons linéaires de ces variables. Comme pour l'ACP et la classification hiérarchique ascendante, l'avantage de

cette méthode est qu'elle ne requiert aucune condition de validité. De plus, comme nous avons observé empiriquement, par l'ACP et par la classification hiérarchique, que les profils des 257 GHM semblent se structurer en plus de 2 classes, l'analyse discriminante est intéressante. Cette méthode va principalement nous permettre de valider la répartition issue de la classification hiérarchique ascendante, en regardant le taux de GHM bien classés dans chacun des différents groupes, et de vérifier si ces groupes sont pertinents ou non du point de vue médico-économique. Cette analyse va aussi nous permettre de rechercher une fonction de classement automatique des GHM dans les différents groupes, pour de nouveaux GHM.

La règle est la maximisation des variances inter-classes et la minimisation des variances intra-classes. De bonnes fonctions discriminantes impliquent un faible pourcentage de sujets mal classés. Les variables sont incluses dans l'équation des combinaisons linéaires par le calcul du  $V$  de Rao [76]. Le calcul du  $V$  permet de connaître la contribution d'une variable à la séparation des groupes. Une manière d'évaluer la contribution d'une variable dans l'équation est de voir de combien augmente le  $V$  quand elle y est introduite. On teste alors la signification de ces changements par un test qui suit une loi du  $\chi^2$ . Le  $V$  de Rao est défini comme suit :

$$V = (n - g) \sum_{i=1}^p \sum_{j=1}^p W_{ij}^* \sum_{k=1}^g (\bar{X}_{ik} - \bar{X}_i)(\bar{X}_{jk} - \bar{X}_j),$$

où  $p$  est le nombre de variables dans l'équation,  $g$  est le nombre de groupes,  $n$  est la taille de l'échantillon,  $\bar{X}_{ik}$  est la moyenne de la  $i^e$  variable pour le  $k^e$  groupe,  $\bar{X}_i$  est la moyenne de la  $i^e$  variable pour l'ensemble des groupes et  $W_{ij}^*$  est un élément de l'inverse de la matrice variance-covariance intra-groupe. Plus la différence entre la moyenne des groupes est importante, plus le  $V$  de Rao est élevé.

Nous avons réalisé deux analyses discriminantes sur un échantillon aléatoire représentant 60% des données de chaque groupe : la première avec les deux facteurs de l'ACP et le déficit en point ISA, la deuxième avec les variables d'origine, *i.e.* les proportions cumulées de patients grenoblois dans les 3 premiers intervalles interquartiles. Ces variables ont été introduites par une méthode de pas à pas pour la construction des fonctions discriminantes. Ces fonctions ont ensuite été validées avec les données restantes (40%). Nous avons présenté le pourcentage de "bien classés" pour chacune des classes, qui est un indice de la qualité de la discrimination pour le classement de GHM issus de l'échantillon du CHU de Grenoble qui n'auront pas servi à l'analyse.

Nous avons testé la robustesse de la meilleure fonction discriminante avec 10 échantillons tests, construits de la même manière. La sensibilité et la spécificité de ces classements seront aussi évaluées pour chacune des classes.

L'analyse a été réalisée avec les logiciels de statistiques SPSS version 6.1 et Matlab version 6.0.

## 4.3 Résultats

### 4.3.1 ACP sur les 257 GHM grenoblois

#### Description générale des données

TAB. 4.2 – *Moyennes et écart-types des proportions cumulées de patients grenoblois dans les 3 intervalles interquartiles nationaux de DS (n=257).*

Variables	Moyenne (en %)	Écart-type (en %)
Premier intervalle	30.1	14.4
Deuxième intervalle	52.6	15.3
Troisième intervalle	75.8	10.2

La description univariée de l'ensemble des trois intervalles interquartiles de DS nationaux en proportions cumulées montre que le profil "moyen" des distributions des DS des 257 GHM grenoblois est comparable à celui de la base nationale (0.25%, 0.50% et 0.75%) et même meilleur sur le premier intervalle interquartile. En effet, les GHM grenoblois comptent en moyenne 30.1%+/-14.4% de patients sur le premier intervalle interquartile, 52.6%+/-15.3% de patients sur le deuxième et 75.8% +/- 10.2% sur le troisième (voir tableau 4.2).

TAB. 4.3 – *Matrice des corrélations entre les trois intervalles interquartiles.*

	$Ic_{(1)}$	$Ic_{(2)}$	$Ic_{(3)}$
$Ic_{(1)}$	1.0000		
$Ic_{(2)}$	0.7409	1.0000	
$Ic_{(3)}$	0.4813	0.7527	1.0000

La matrice des corrélations, présentée dans le tableau 4.3, montre que les trois intervalles interquartiles de DS caractérisant la forme des profils grenoblois sont très corrélés. Cependant, on remarquera que les corrélations vont en décroissant : la corrélation entre le premier intervalle interquartile et le troisième est moindre qu'entre le premier et le deuxième, et qu'entre le deuxième et le troisième.

#### Interprétation et description des composantes principales

Deux axes principaux ont pu être extraits avec l'ACP. L'inertie cumulée expliquée par les deux axes est de 95.0% (voir tableau 4.4) : ils expliquent donc une part importante du

TAB. 4.4 – *Pourcentage de variance expliquée par chacune des composantes principales et inertie cumulée.*

Facteur	Valeur propre	Inertie expliquée	Inertie cumulée
1	0.0438	80.5	80.5
2	0.0079	14.5	95.0
3	0.0027	5.0	100.0

TAB. 4.5 – *Matrice des vecteurs propres.*

Variables	Facteur 1	Facteur 2
Premier intervalle	-0.13	-0.07
Deuxième intervalle	-0.15	0.03
Troisième intervalle	-0.08	0.05

nuage des profils des distributions des DS des GHM grenoblois.

La première composante principale ou facteur (80.5% de l’inertie) s’exprime comme la somme des proportions cumulées de patients dans les trois intervalles interquartiles (voir tableau 4.5). C’est un facteur de taille, car les 3 variables se trouvent dans le demi-plan négatif de la première composante (voir figure 4.1) et sont toutes très corrélées avec l’axe (voir tableau 4.6). Elles contribuent donc toutes à la définition de ce premier facteur. Ainsi, si un GHM a un pourcentage important de patients sur l’un des 3 intervalles, les deux autres intervalles compteront également des pourcentages élevés et inversement. Cet axe oppose les GHM qui ont des proportions cumulées de patients élevées dans les trois intervalles (les “bons E”, côté positif de l’axe), à ceux ayant de faibles proportions (les “mauvais E”, côté négatif de l’axe).

La deuxième composante (14.5% de l’inertie) est un facteur de forme qui oppose le premier intervalle au troisième (voir figure 4.1). Seules ces deux variables participent à la construction de ce facteur ; en effet, elles sont bien corrélées avec celui-ci par rapport au deuxième intervalle interquartile, respectivement -0.45, 0.53 contre 0.18 (voir tableau 4.6). Ainsi, sur le côté positif de l’axe se trouvent les GHM peu performants, en termes de DS, sur les courts séjours (proportions de patients grenoblois sur le premier intervalle interquartile inférieures à 25%) et efficaces sur les séjours de longue durée (proportions de patients grenoblois sur le troisième intervalle interquartile inférieures à 75%) ; sur le côté négatif figurent les GHM bons sur les courts séjours et médiocres sur les longs séjours (proportions de patients grenoblois supérieure à 75%), [voir figure 4.2].



TAB. 4.6 – *Corrélations des variables avec les 2 premières composantes principales retenues.*

Variabes	Facteur 1	Facteur 2
Premier intervalle	-0.88	-0.45
Deuxième intervalle	-0.96	0.18
Troisième intervalle	-0.77	0.53

### Projections des profils des GHM grenoblois

La projection des profils des GHM grenoblois dans le plan factoriel,  $F1 \times F2$  montre un nuage de point centré autour de l'origine des axes en un noyau assez dispersé (voir figure 4.2). Plus les GHM sont excentrés, plus ils ont tendance à être atypiques. Nous nous intéressons aux GHM identifiés comme "mauvais", en terme de profil de distribution. Ils se trouvent pour la plupart dans la partie droite du nuage (voir figure 4.3). Plus ils sont situés à l'extrémité droite de celle-ci, moins ils sont performants, beaucoup de profils empiriques de type (a) sont dans cette partie. Dans le quadrant supérieur droit, on a tendance à retrouver les GHM bons sur les séjours de longues durées (proportions cumulées de patients dans le troisième intervalle interquartile supérieures à 75%) et médiocres sur les séjours de courtes durées (proportions cumulées de patients dans premier intervalle interquartile  $< 25\%$ ), profils empiriques de types (b) et (e). On observe l'inverse dans le quadrant inférieur droit, profils empiriques de types (c) et (a).

Cependant, deux GHM se distinguent, l'un se situe à l'extrême droite du premier axe (9.1% de sa variance), l'autre dans la partie supérieure du deuxième axe (4.6% de sa variance) : il s'agit respectivement du n°48 : interventions sur la rétine et du n°230 : interventions réparatrices pour hernies inguinales et crurales, âge de 18 à 69 ans sans CMA (voir figure 4.2). Ces GHM ont des distances à l'origine des axes relativement élevées (0.63 et 0.26), ce qui confirme leur caractère atypique. En conséquence, ils ont tendance à tirer respectivement les première et deuxième composantes à eux. C'est pourquoi nous avons refait la même analyse en enlevant les données de ces deux GHM afin de vérifier si les résultats ne diffèrent pas.

#### 4.3.2 ACP sans les GHM grenoblois n°48 et n°230

L'inertie cumulée expliquée par les deux axes principaux retenus est de 94.1% (voir figure 4.4) : l'inertie du premier facteur est égale à 81.9% et celle du deuxième facteur à 13.4%. Elles ne changent pas de manière significative de celles de l'analyse précédente. L'interprétation des deux facteurs reste similaire avec celle de l'ACP incluant ces 2 GHM (voir figures 4.5 4.2). Nous pouvons donc conserver ces deux GHM pour la suite de l'analyse, car ils ne biaisent pas les résultats.

### Visualisation des listes de référence dans le plan factoriel

La figure 4.6 montre que la majorité des GHM appartenant à la liste du CHU se situent sur le côté positif du premier axe. Par contre, le deuxième axe ne semble pas discriminant pour les GHM de cette liste. Cela se vérifie sur la figure 4.9, les GHM appartenant à la liste du CHU ayant tendance à avoir des coordonnées sur le côté positif du premier facteur.

Les listes fondées sur les quartiles et le  $\chi^2$  se trouvent sur la partie positive du premier axe, mais plutôt du côté négatif du deuxième axe (voir figures 4.7 et 4.8). Ce résultat se confirme sur les figures 4.10 et 4.11, les GHM faisant partie des listes des quartiles et du  $\chi^2$  ayant plutôt des coordonnées du côté positif du premier facteur. Nous pouvons aussi remarquer que la dispersion des coordonnées des GHM de ces deux listes est moindre que celles des GHM n'appartenant pas à ces listes. Cela s'explique par le faible effectif de ces listes.

L'analyse des coordonnées sur le premier facteur selon les 8 profils types issus de l'analyse empirique montre que plus un profil est "mauvais", plus ses coordonnées se trouvent du côté positif de l'axe et inversement (voir figure 4.12). Si l'on s'intéresse au deuxième facteur (voir figure 4.13), on voit que les profils empiriques de types *(b)* et *(e)* s'opposent aux profils empiriques de types *(c)* et *(g)*. Les GHM ayant une valeur faible sur le premier intervalle interquartile et élevée sur le troisième intervalle (GHM peu performants sur les séjours de courte durée mais efficaces sur les séjours de longue durée), profils empiriques de types *(b)* et *(e)*, ont tendance à se situer sur le côté négatif du deuxième axe. Les GHM étant efficaces sur les séjours de courte durée mais peu performants sur les séjours de longue durée, profils empiriques de types *(c)* et *(g)*, se trouvent plutôt du côté positif de l'axe.

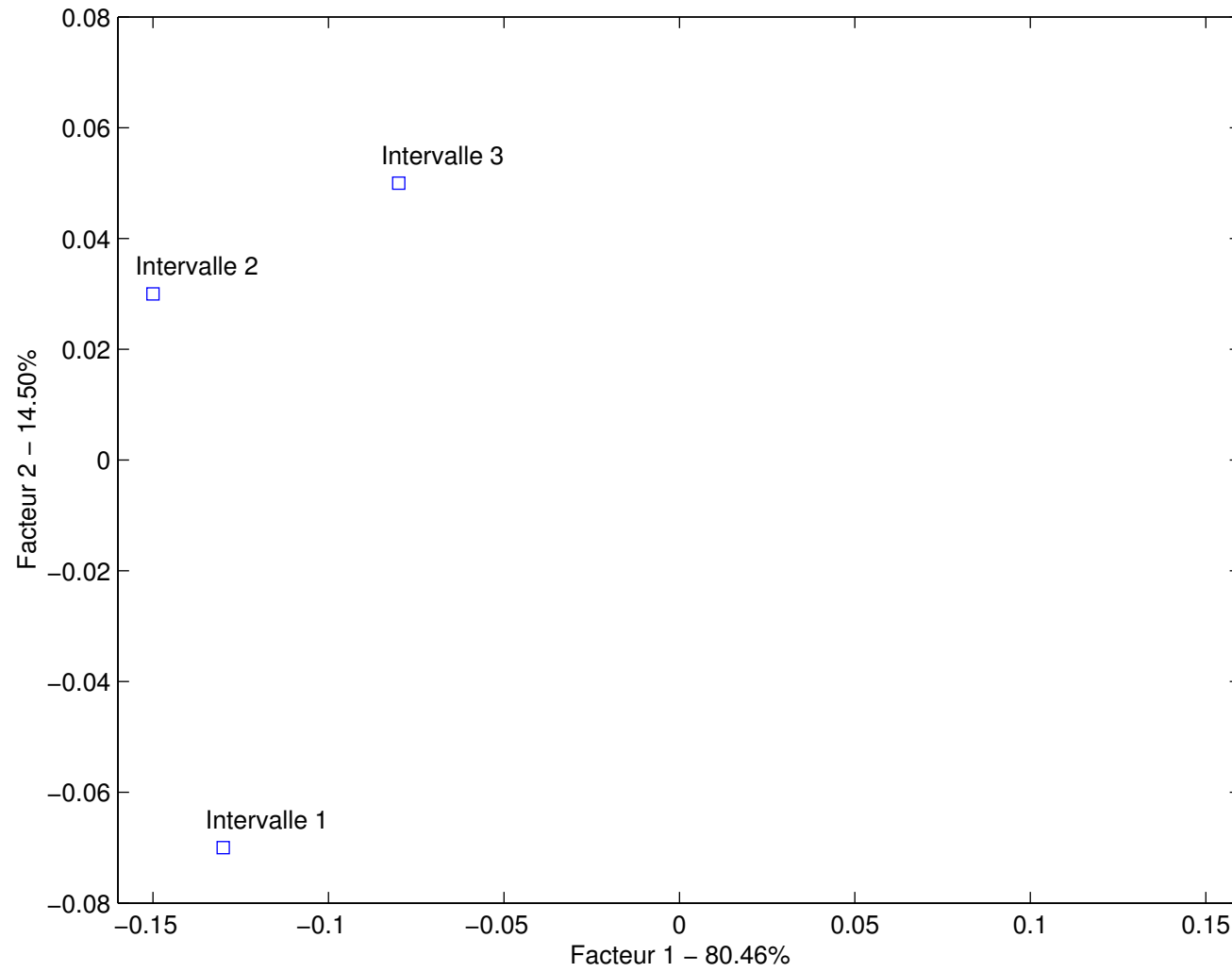


FIG. 4.1 – Projection des variables dans le plan factoriel  $F1 \times F2$  (ACP réalisée sur les 257 GHM).

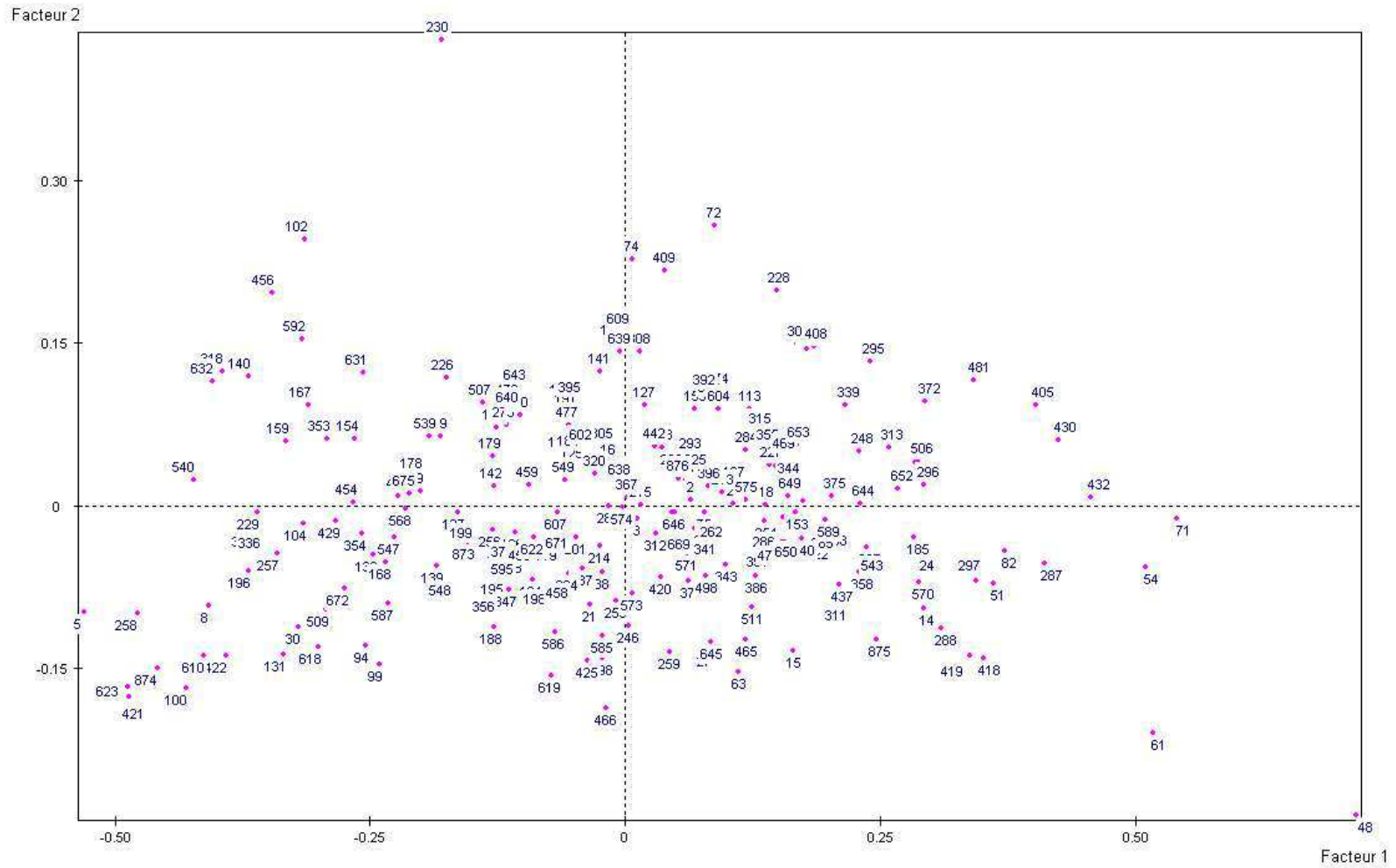


FIG. 4.2 – Projection des individus dans le plan F1 et F2 (ACP réalisée sur les 257 GHM).

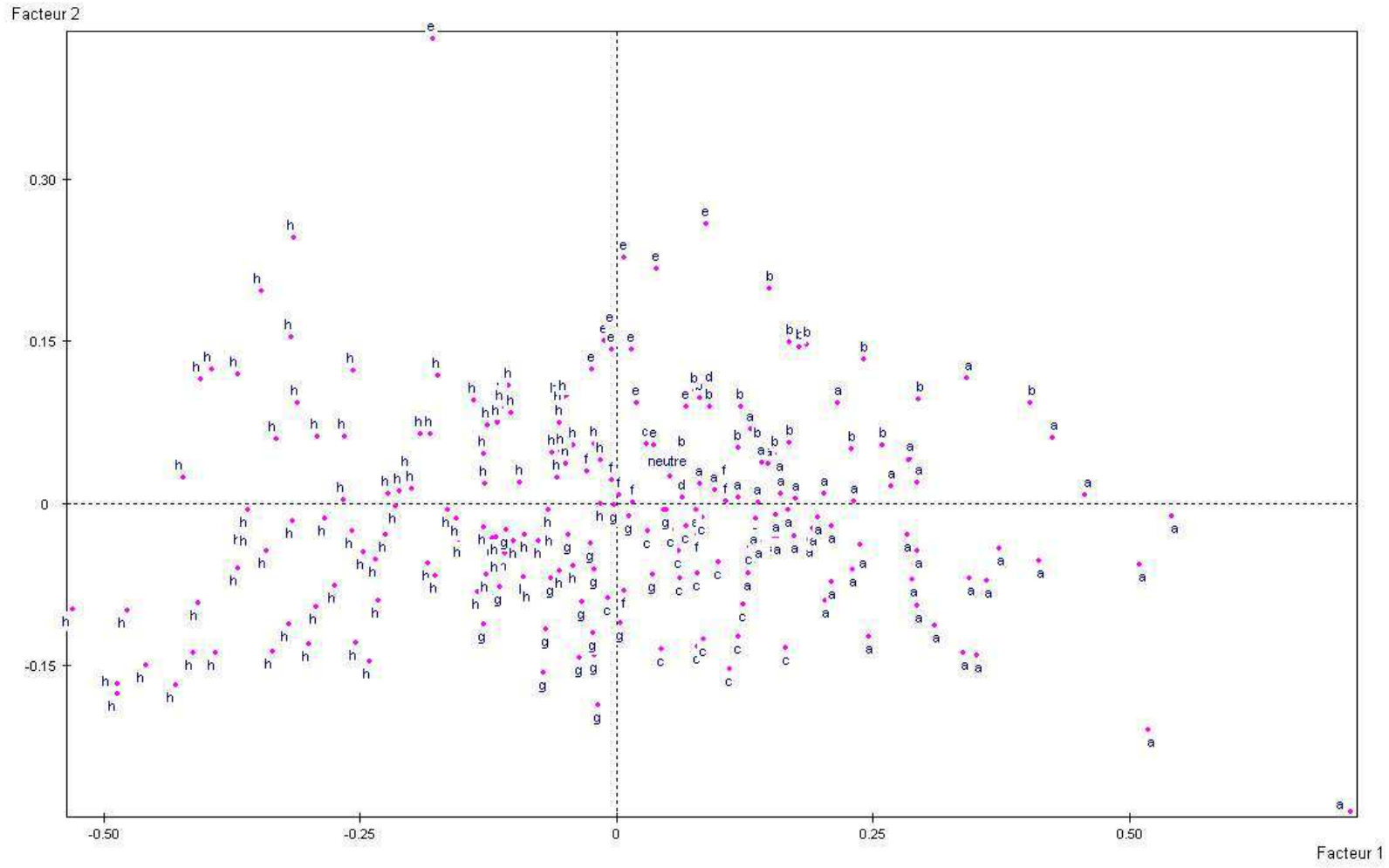


FIG. 4.3 – Identification du nuage des individus avec les 9 profils caractéristiques du CHU de Grenoble.

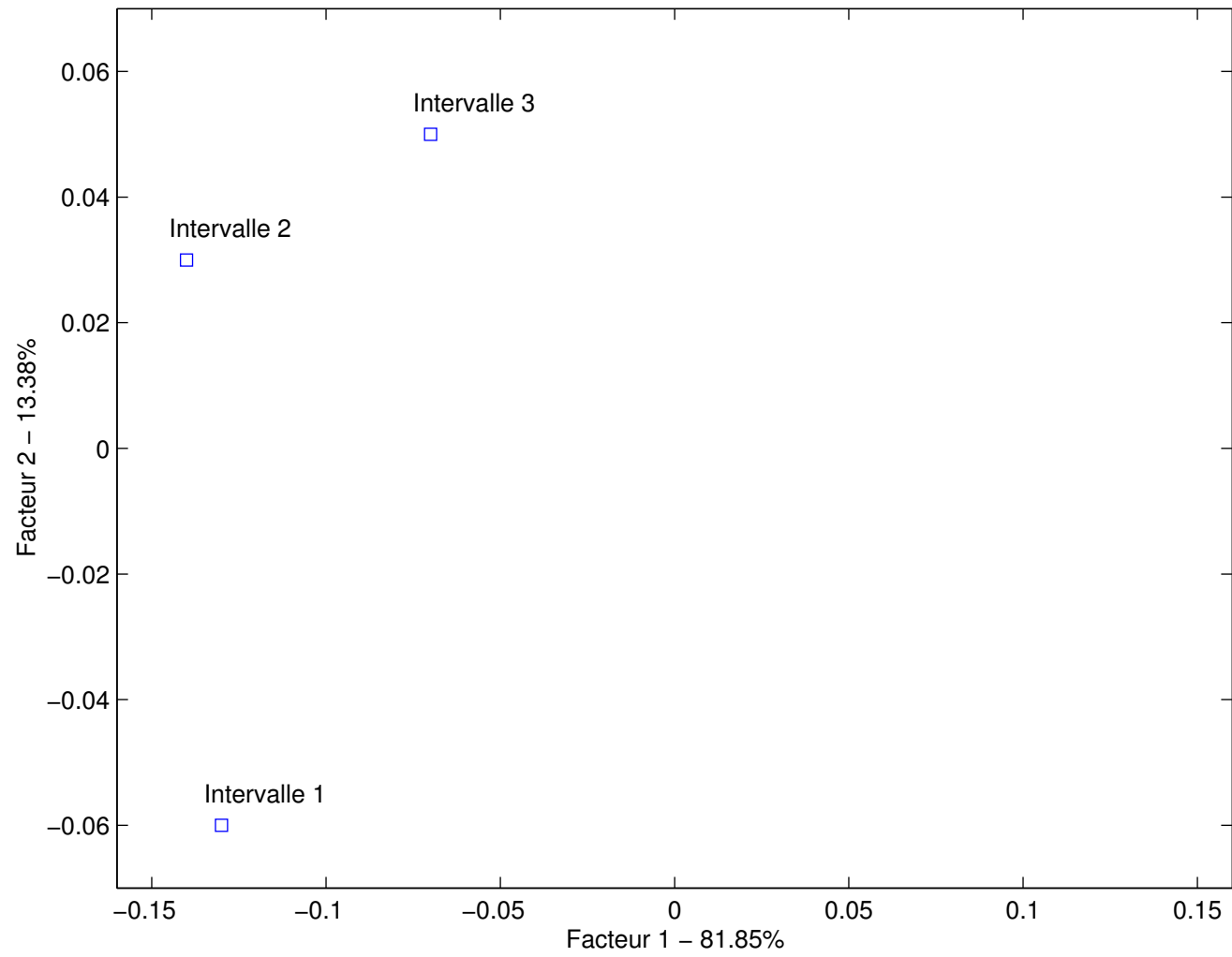


FIG. 4.4 – Projection des variables dans le plan factoriel  $F1 \times F2$  (ACP réalisée sans les GHM n°48 et n°230).

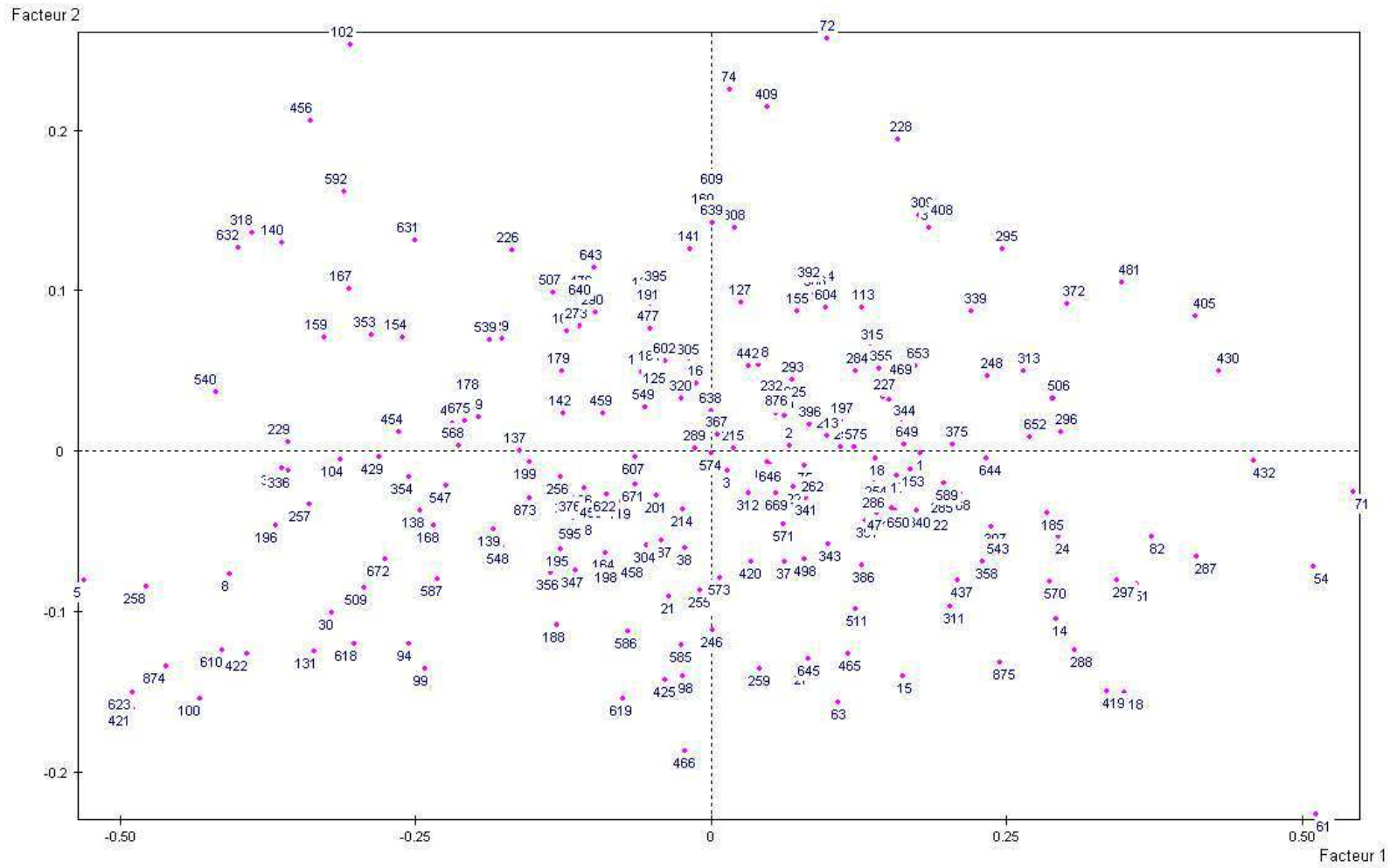


FIG. 4.5 – Représentation des individus (ACP réalisée sans les GHM n° 48 et n° 230).

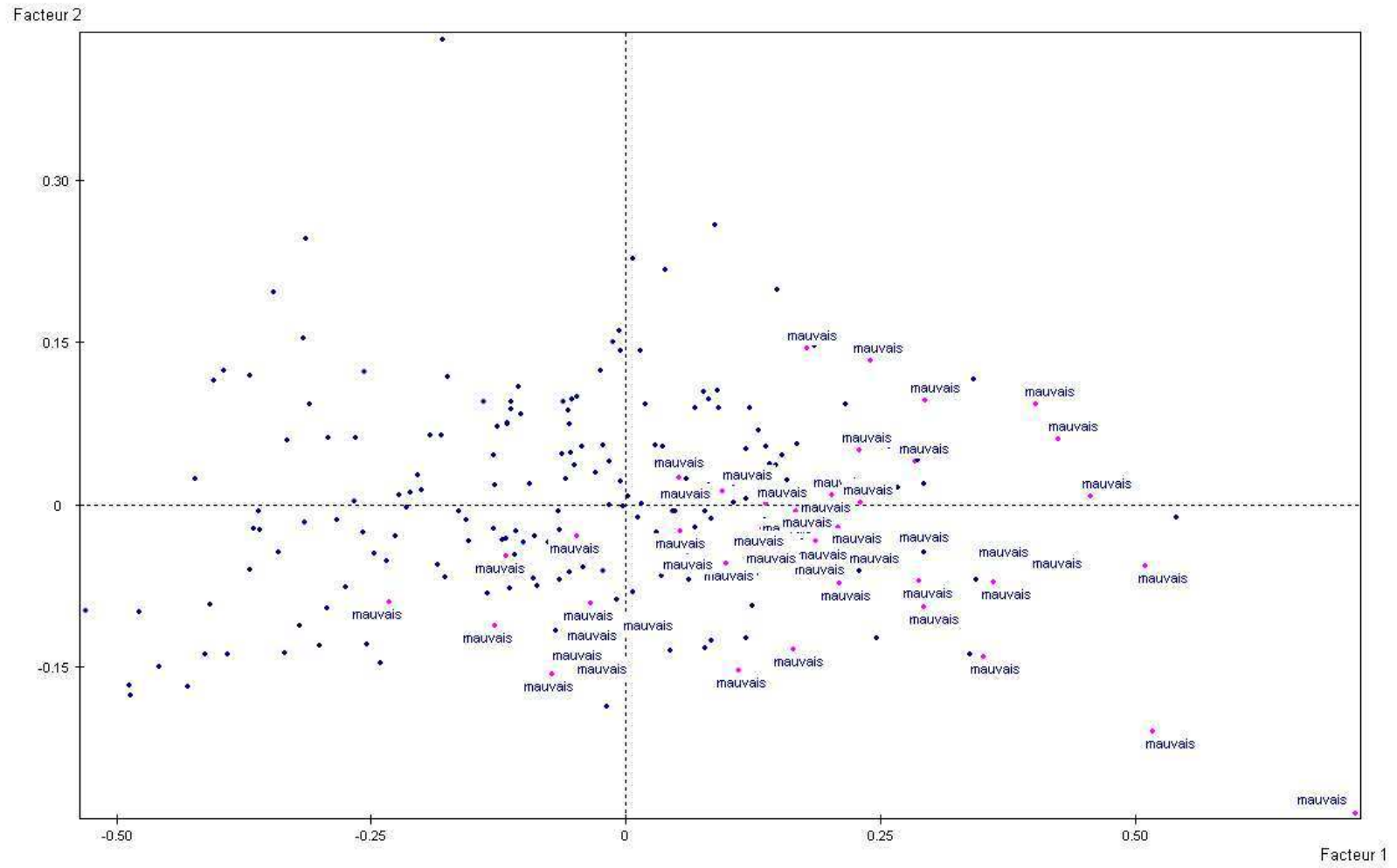


FIG. 4.6 – Identification du nuage des individus avec la liste du CHU.



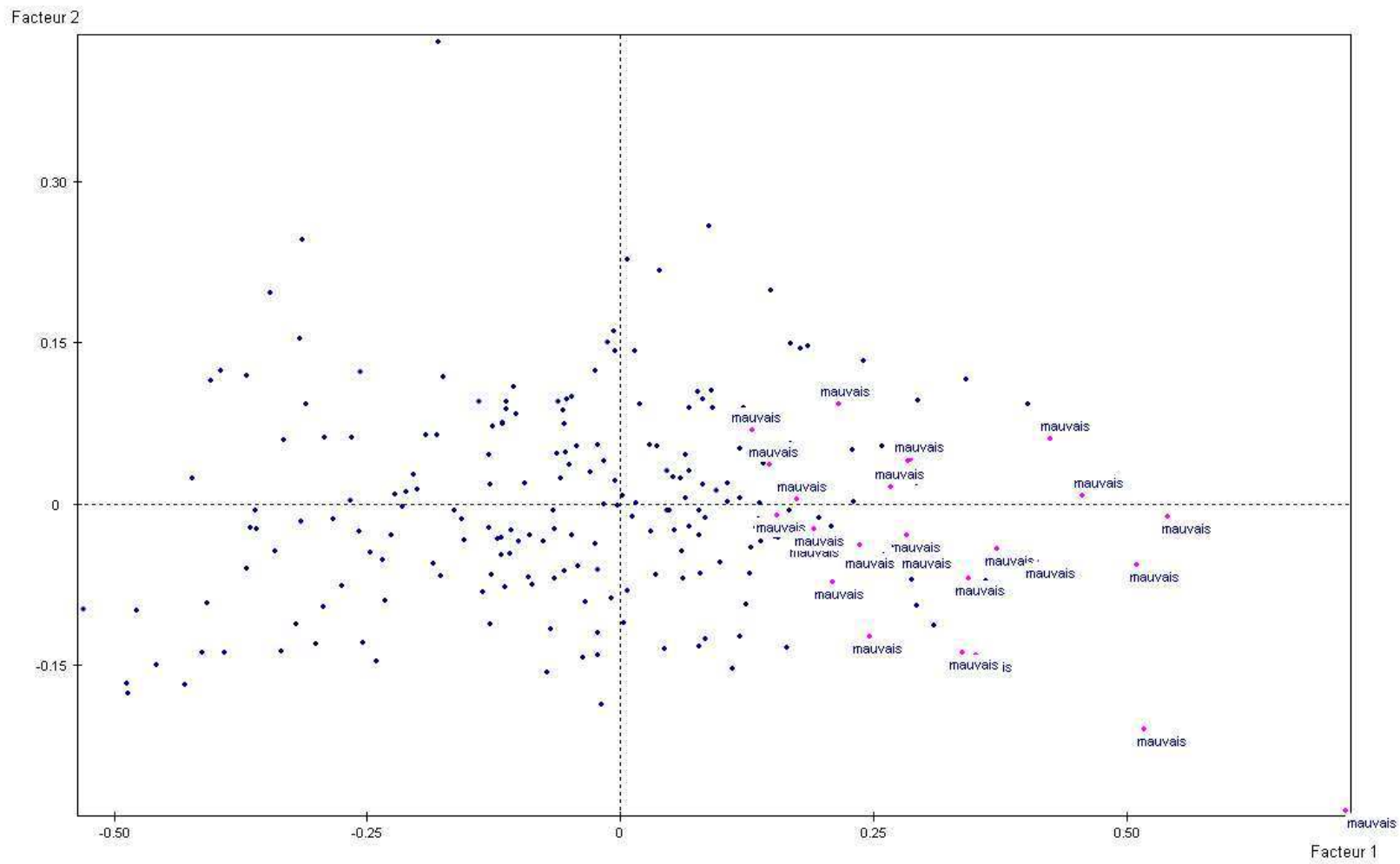


FIG. 4.7 – Identification du nuage des individus avec la liste fondée sur les quartiles.

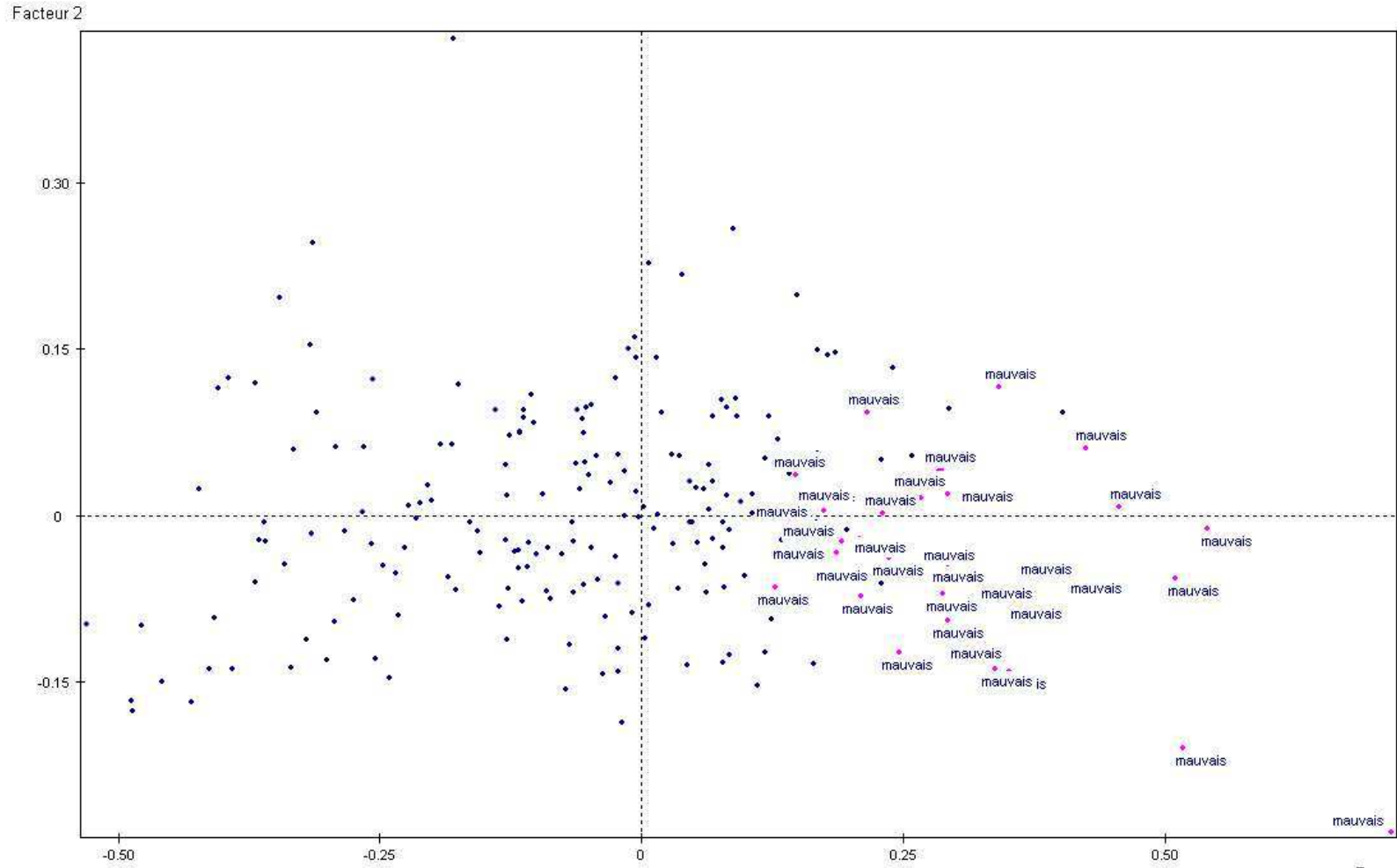


FIG. 4.8 – Identification du nuage des individus avec la liste fondée sur le test du  $\chi^2$ .

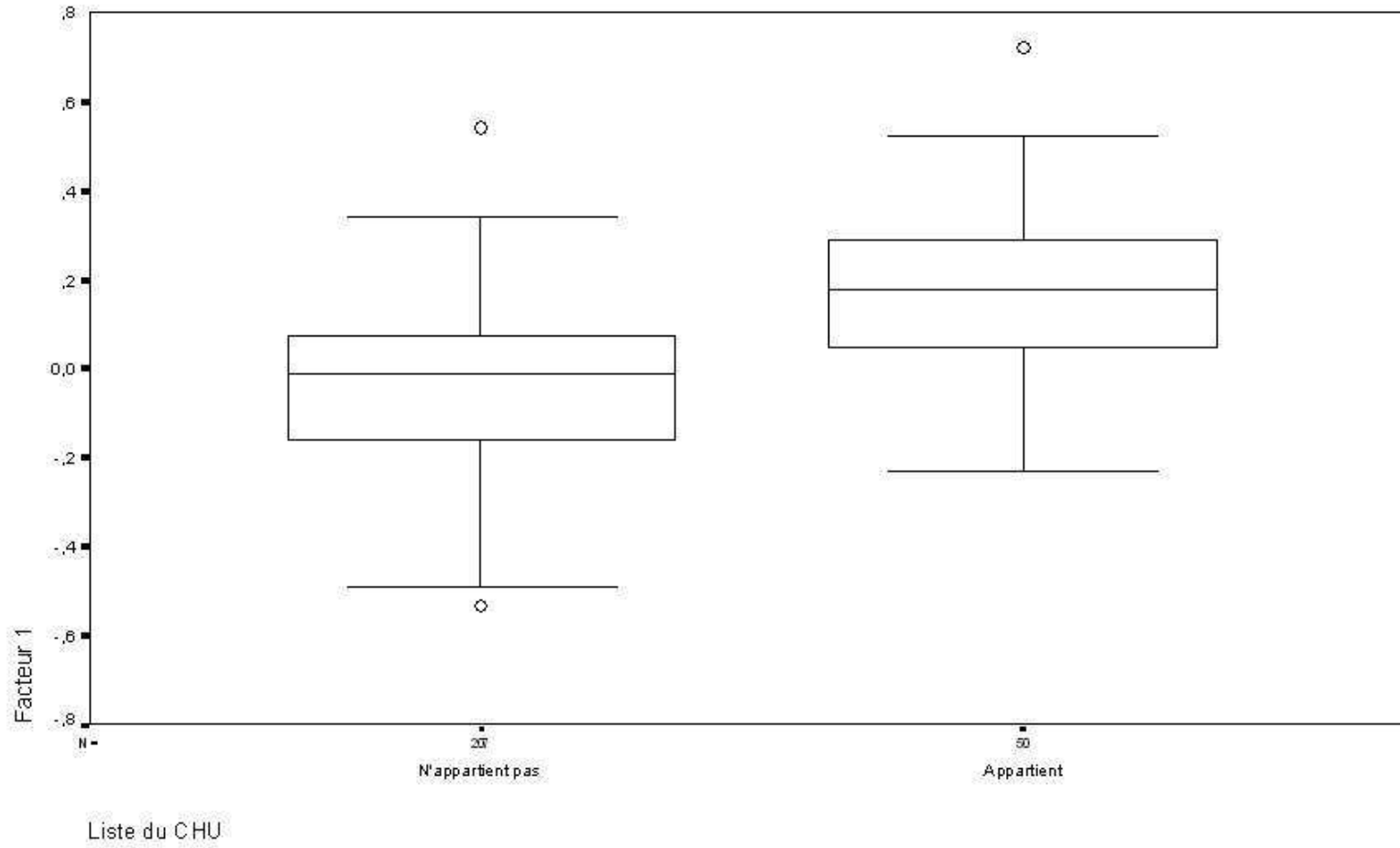


FIG. 4.9 – *Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste CHU.*

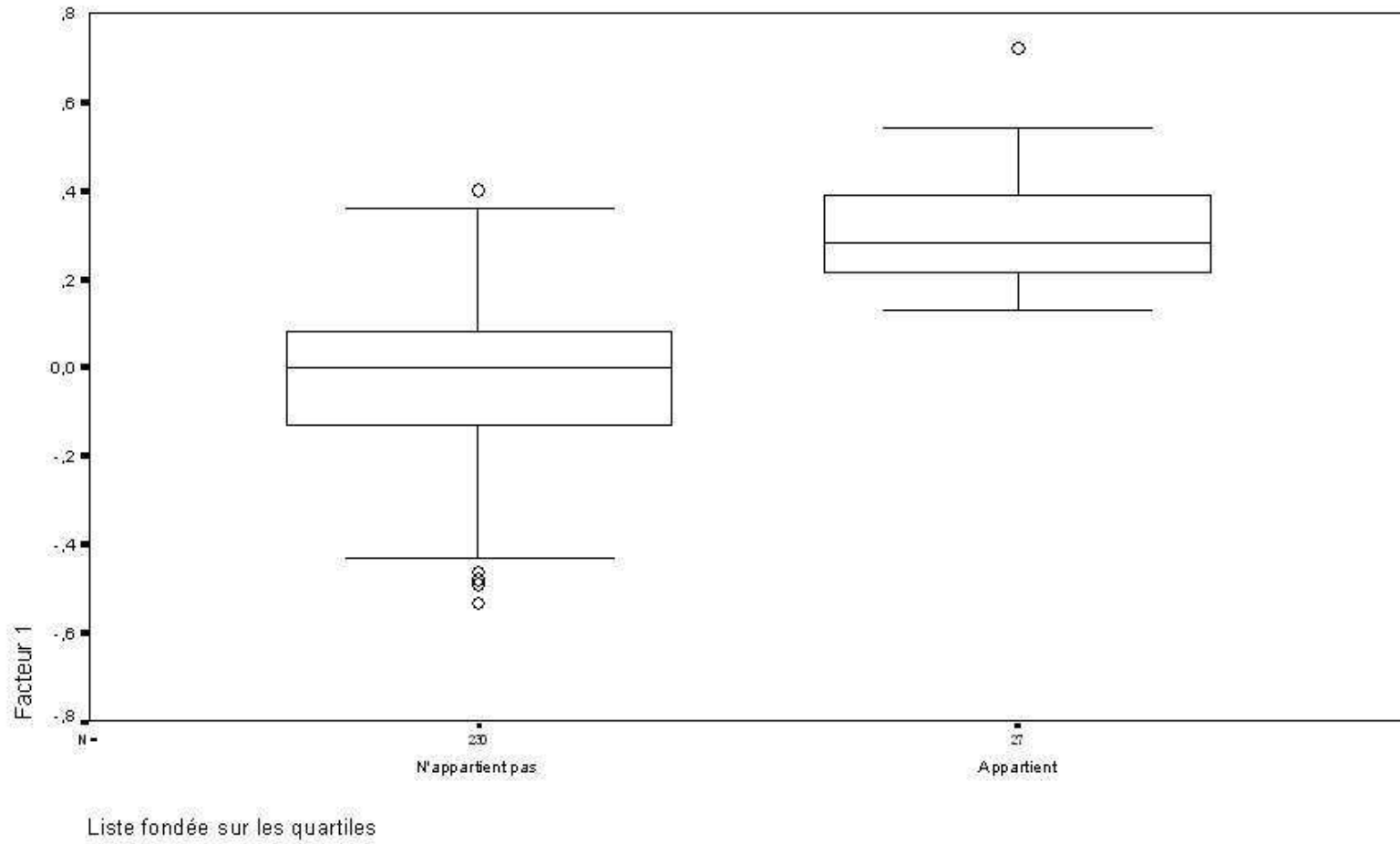


FIG. 4.10 – *Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste des quartiles.*

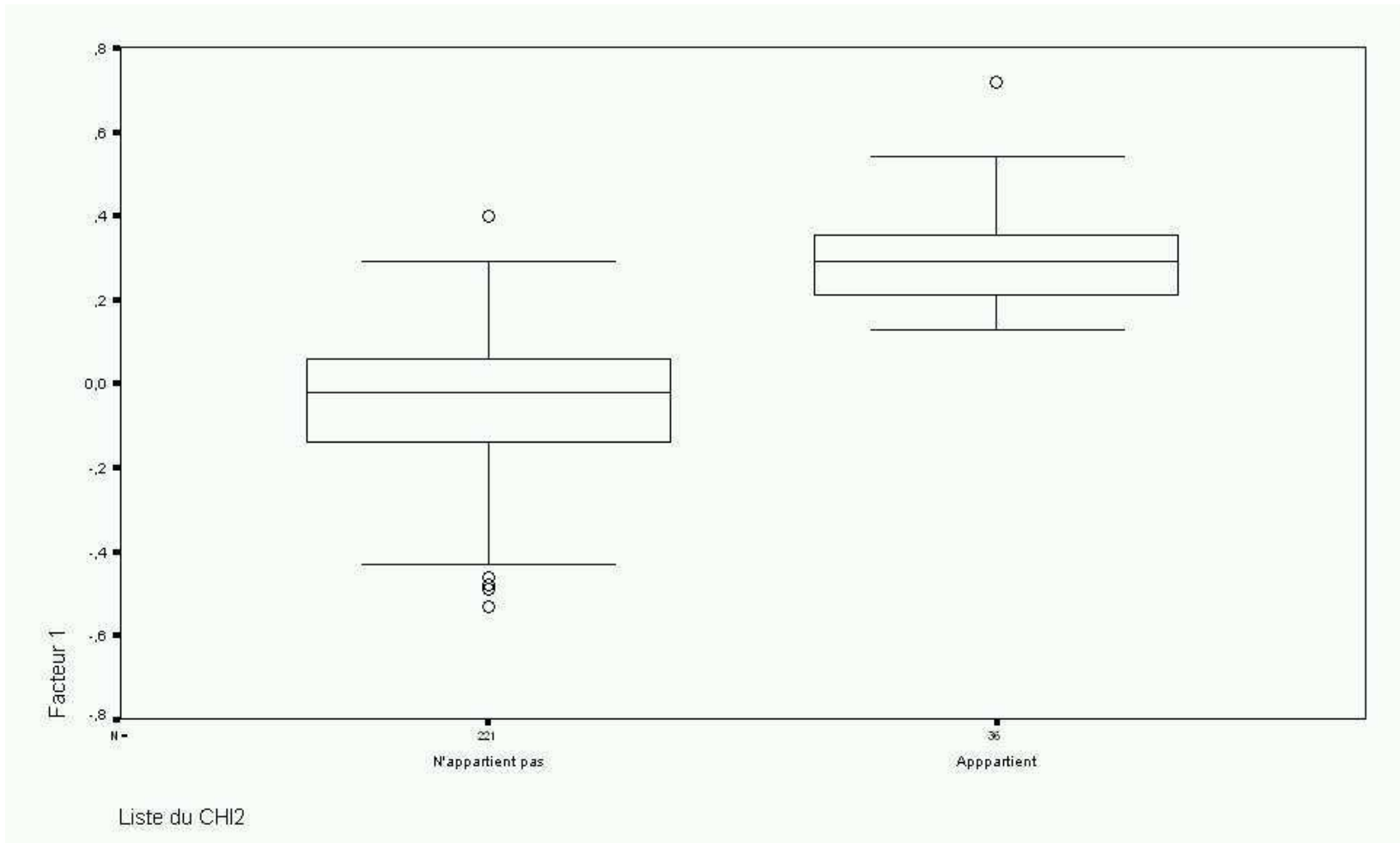


FIG. 4.11 – *Box plot des coordonnées sur le premier facteur de l'ACP pour les groupes de GHM de la liste du  $\chi^2$ .*

### Corrélations des facteurs avec les variables décrivant les profils

L'étude des corrélations des facteurs avec les variables suivantes :

- la différence entre les aires grenobloise et nationale quantifiant l'éloignement global d'un profil grenoblois par rapport à la norme nationale,
- le déficit en point ISA,

montre que le premier facteur est fortement corrélé négativement avec ces deux variables (voir tableau 4.7). Plus on va vers le côté positif du premier axe, plus on trouve les GHM avec un profil éloigné de la référence nationale et avec un déficit en points ISA important. La première composante principale peut être considérée comme un indicateur général sur le type de profil de GHM (“bon” ou “mauvais” GHM en terme de profil de distribution de DS par rapport à la référence nationale). Le deuxième facteur est très peu corrélé avec ces deux variables (voir tableau 4.7). Ce facteur est plutôt un indicateur de forme des profils.

TAB. 4.7 – *Corrélations des facteurs de l'ACP avec les variables indiquant les différences entre les aires grenobloise et nationale et le déficit en points ISA.*

Variables	Facteur 1	Facteur 2
Indicateurs des profils :		
Différence d'aire	-0.99*	0.04
Déficit en points ISA	-0.56*	-0.09*

\*  $p < 0.05$

#### 4.3.3 Classification hiérarchique ascendante

Parmi les 257 GHM, 137 ont été identifiés comme “mauvais E” avec l'approche empirique. Plusieurs classifications hiérarchiques ascendantes avec les différents critères d'agrégation présentés précédemment ont été effectuées sur les 137 GHM identifiés, en utilisant comme variables les proportions cumulées de patients sur chacun de leurs trois intervalles interquartiles.

Plusieurs dendrogrammes ont été obtenus et comparés, afin d'observer les noyaux stables, c'est-à-dire les GHM qui s'agrègent toujours ensemble. Nous présentons ici le dendrogramme de la classification (voir figure 4.14) qui utilise comme critère d'agrégation la méthode de Ward, car la séparation en grands groupes y est plus nette visuellement que dans les autres classifications.

Sur le dendrogramme, on observe principalement 3 groupes de GHM atypiques :

- un premier groupe de 97 GHM, dont le noyau stable est constitué par 75 GHM,
- un deuxième groupe constitué de 24 GHM sans noyau stable,

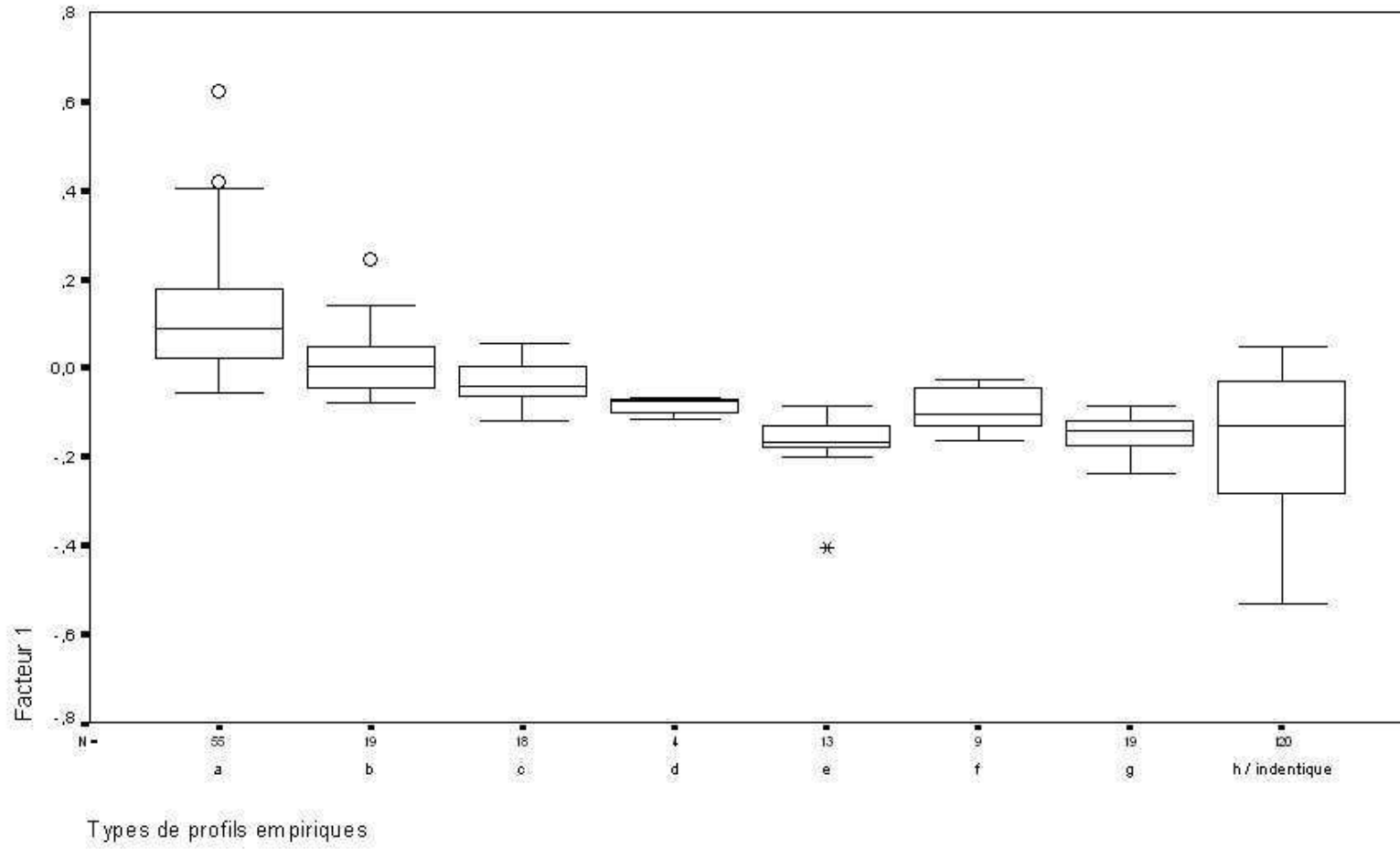


FIG. 4.12 – Box-plot des coordonnées sur le premier axe de l'ACP selon les profils empiriques.

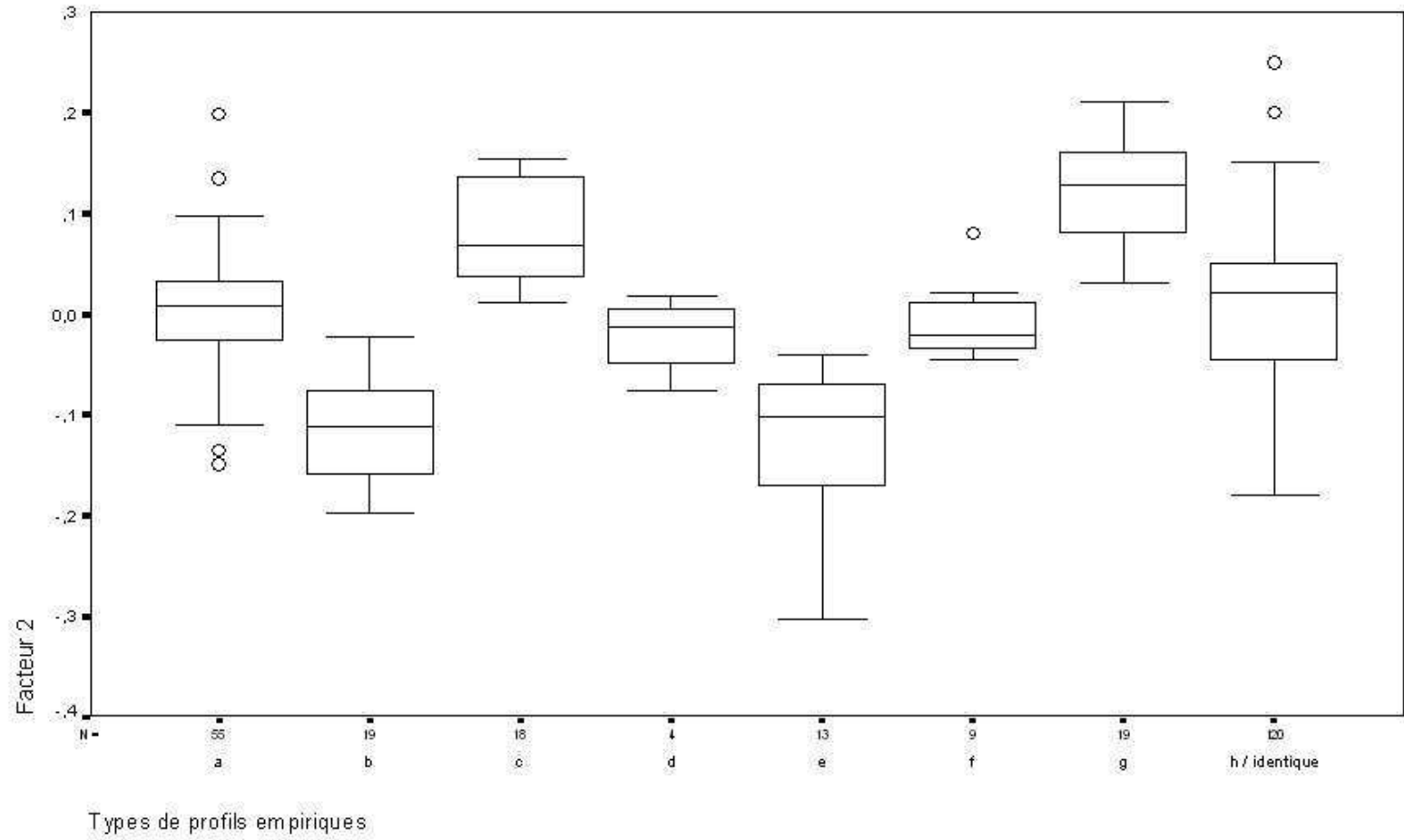


FIG. 4.13 – *Box-plot des coordonnées sur le deuxième axe de l'ACP selon les profils empiriques.*



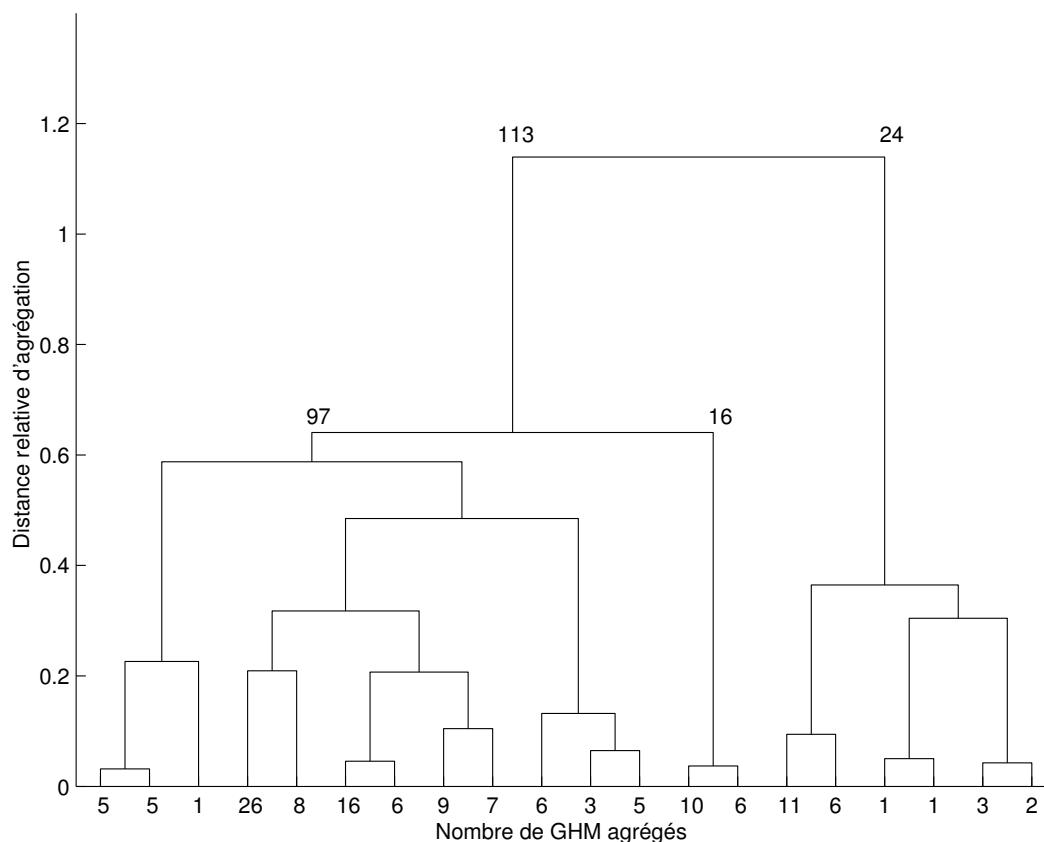


FIG. 4.14 – Dendrogramme de la classification hiérarchique de 137 GHM identifiés empiriquement comme “mauvais”. La distance utilisée est la distance euclidienne et le critère d’agrégation est celui de la variation minimale d’inertie (méthode de Ward).

– un troisième groupe de 16 GHM, sans noyau stable lui aussi.

En traçant le profil moyen des GHM de chacun des groupes, nous avons pu identifier les caractéristiques de chacun. La figure 4.15 montre les proportions cumulées moyennes de patients dans chaque intervalle interquartile pour chacun des trois groupes de GHM trouvés ci-dessus, ainsi que celles du groupe des “bons E”.

Le premier groupe réunit 24 GHM qui ont en grande majorité un profil empirique de type (a) (23/24); ces GHM seront nommés “**mauvais FA**” (voir annexe I). Ces GHM sont identifiés comme peu performants, quelle que soit la longueur du séjour ( $Ic_{(1)} < 0.25$ ,  $Ic_{(2)} < 0.50$  et  $Ic_{(3)} < 0.75$ ) [profil en pointillés de couleur bleue sur la figure 4.15]. Il ne diffère pas du CHU de Grenoble (les 257 GHM) en terme de type de GHM. Ces GHM entraînent un déficit moyen de -96 865 +/- 104 509 points ISA au CHU de Grenoble, et la différence moyenne d’aire entre le profil grenoblois et la norme nationale est de -13.4% +/-

5.4% (voir tableau 4.8).

Le deuxième groupe rassemble 16 GHM qui ont en majorité un profil empirique de type *(g)* (15/16); ces GHM sont appelés “**mauvais FG**” (voir annexe J). Ils ont tendance à être peu performants sur les séjours de longue durée uniquement ( $Ic_{(3)} < 0.75$ ) [profil en pointillés de couleur bleu ciel sur la figure 4.15]. Ce groupe compte significativement moins de GHM de type chirurgical que dans l'échantillon regroupant les 257 GHM (6% contre 33%,  $p = 0.0235$ ). Les GHM sont responsables d'un déficit moyen de -44 350 +/- 49 507 points ISA au CHU de Grenoble, et la différence moyenne d'aire entre le profil grenoblois et la norme nationale est de 3.5% +/- 1.7% (voir tableau 4.8).

Le troisième groupe (n=97) correspond aux GHM identifiés comme “mauvais”, mais sans profil caractéristique; on les a appelés les “**autres F**” (voir annexe K). Il compte principalement des profils empiriques de types *(a)*, *(b)*, *(c)* et *(e)*, respectivement 33%, 19%, 18% et 13%. Il regroupe 44% de GHM chirurgicaux. Le profil moyen de ces GHM est proche de celui de la référence nationale, tout en ayant des valeurs légèrement inférieures aux proportions de référence [profil en trait continu de couleur rouge sur la figure 4.15]. Ils occasionnent un déficit moyen de -28 955 +/- 70 885 points ISA au CHU de Grenoble, et la différence moyenne relative d'aire entre le profil grenoblois et la norme nationale est de -2.2% +/- 3.7% (voir tableau 4.8). Ce groupe compte sensiblement moins de GHM chirurgicaux par rapport à l'ensemble des 257 GHM (26% contre 33%, à la limite de la significativité  $p = 0.059$ ).

En dehors des groupes issus de la classification, il existe un quatrième groupe, c'est celui des GHM identifiés comme bons ou normaux avec l'approche empirique, les “bons E” (voir annexe L) [profil en pointillés de couleur verte sur la figure 4.15]. Il compte légèrement moins de GHM chirurgicaux que sur l'ensemble de l'échantillon (26% contre 33%). Il amène un bénéfice moyen de 39 288 +/- 63 311 points ISA au CHU de Grenoble, et la différence moyenne d'aire entre le profil grenoblois et la norme nationale est de 8.6% +/- 6.7% (voir tableau 4.8). Il compte de manière significative, plus de GHM chirurgicaux que dans l'échantillon des 257 GHM (44% contre 33%,  $p = 0.0175$ ).

TAB. 4.8 – Moyennes et écarts-types de la différence d'aire et du déficit en points ISA des groupes de GHM issus de la classification hiérarchique.

Variabes	Autres F $\bar{X}$ (E.T.)	Mauvais FA $\bar{X}$ (E.T.)	Mauvais FG $\bar{X}$ (E.T.)	Bons E $\bar{X}$ (E.T.)	p_valeurs
Différence d'aire (en %)	-2.2 (3.7)	-13.4 (5.4)	3.5 (1.7)	8.6 (6.7)	<0.0001
Bénéfice ou déficit en points ISA	-28 955 (70 885)	-96 865 (104 509)	-44 350 (49 507)	39 288 (63 311)	<0.0001
% GHM de type chirurgical	44.3	45.8	6.3	25.8	<0.0001

Dans le tableau 4.8, nous voyons qu'il existe entre les groupes une liaison statistique significative entre les différences d'aire grenobloise et nationale. Plus le groupe compte de mauvais GHM, plus cette dernière va être négative, ainsi plus le profil grenoblois sera éloigné de la référence nationale et inversement. Les “mauvais FG” ont une différence d'aire positive

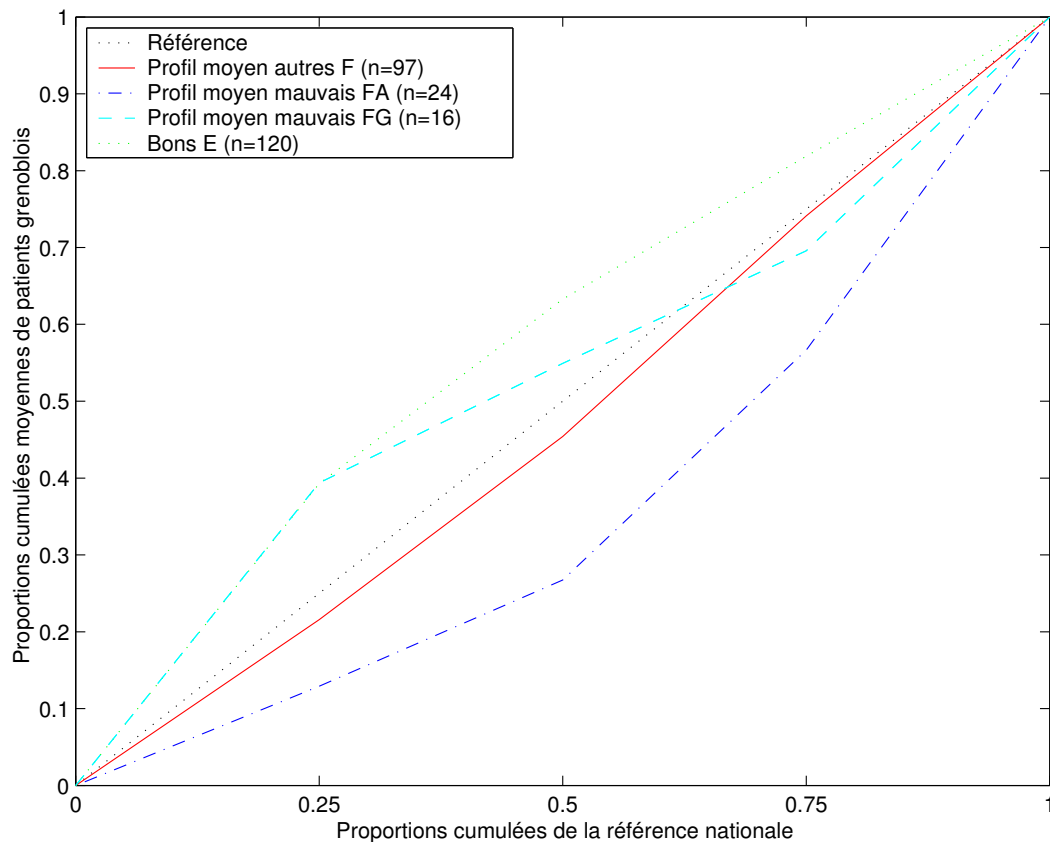


FIG. 4.15 – Proportions cumulées moyennes dans chacun des 3 intervalles interquartiles pour chacun des groupes de GHM: “bons E”, “mauvais FA”, “mauvais FG” et “autres F”.

car ils ne sont mauvais que sur le dernier tiers de la distribution de leur DS, correspondant aux séjours de longue durée.

On observe le même phénomène pour le déficit ou le bénéfice en point ISA : plus le groupe réunit de mauvais GHM, plus le déficit en point ISA est important. Il est possible que cette variable soit discriminante pour l'appartenance des GHM aux différents groupes. Cependant, on remarque que les écarts-types sont importants, cela étant synonyme d'une hétérogénéité au sein de chacun des groupes.

En représentant les trois groupes issus de la classification hiérarchique et celui réunissant les “bons E” dans le plan formé par les deux variables : différence d'aire et déficit ou bénéfice en points ISA, on s'aperçoit que les barycentres (carrés étoilés de couleur noire) des groupes des “bons E” GHM (carrés de couleur verte), des “mauvais FA” (étoiles de couleur rouge) et des profils de type “autres F” (points de couleur bleue) sont alignés entre eux, alors que celui correspondant aux “mauvais FG” (cercles de couleur bleu ciel) est légèrement décalé sur la gauche (voir figure 4.16). Ce groupe de 16 GHM semble avoir un profil particulier qui diffère de celui des 3 autres.

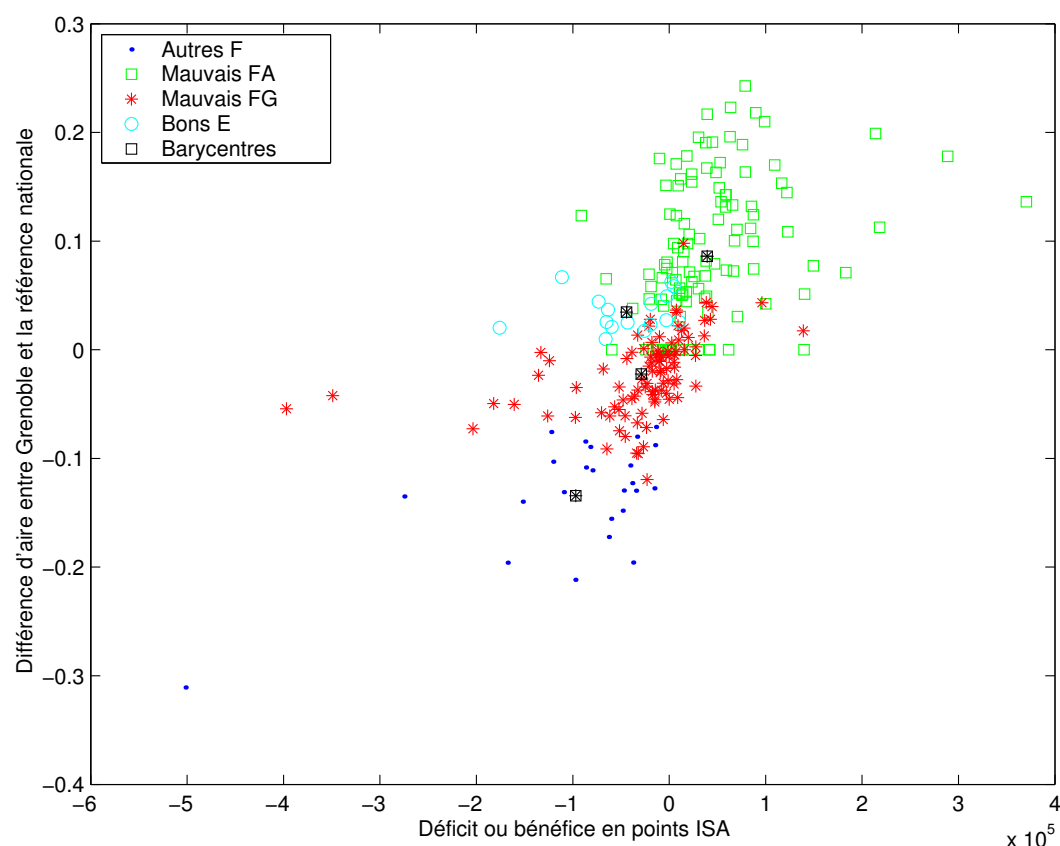


FIG. 4.16 – Représentation du déficit ou bénéfice en points ISA par rapport à la différence d’aire entre le profil grenoblois et la norme nationale pour la classification issue de l’approche factorielle.

Ce graphique représente ce que l’on peut couramment observer dans un établissement : un groupe important de bons GHM, quelques-uns très mauvais et un “marais” au milieu du graphique représentant les GHM autres, ni bons, ni mauvais ... En revanche, ce qui a l’air particulier au CHU de Grenoble est ce groupe rassemblant 16 GHM qui se situe légèrement décalé sur la gauche : tous les GHM de ce groupe ont pour caractéristique commune d’être peu performants sur les séjours de longue durée uniquement. On remarque que ces GHM ont un déficit moyen non négligeable de  $-44\,350 \pm 49\,507$  points ISA. Ce déficit est plus important que celui causé par les GHM “autres F”.

Ce groupe de “mauvais FG” a la particularité de rassembler en majorité des pathologies d’origine neurologique ou gastro-entérologique, avec comme début d’intitulé “Autres ...”, souvent synonymes d’absence de diagnostic précis (voir annexe D). L’allongement des longs séjours est-il lié aux patients, aux variations de pratique médicale, ou à la structure? Nous tenterons de répondre à ce problème dans la discussion.

### Confrontation avec les listes de référence

Maintenant que nous avons présenté les caractéristiques des 4 groupes de la classification, nous allons maintenant les confronter aux 3 listes de référence afin de voir la pertinence de ce découpage. Nous allons nous intéresser au 3 groupes : “les mauvais FA”, les “mauvais FG” et les “autres F” que nous opposerons aux “bons F”.

Les 50 GHM faisant partie de la liste CHU des “mauvais E” ont 79% de risque d'appartenir au groupe des GHM “mauvais FA” et 85% de risque d'être dans celui des “autres F” (voir tableau 4.9). Parallèlement, un “mauvais FA” aura 63% de risque de faire partie de la liste CHU, contre 50% pour un “mauvais FG” (voir tableau 4.9). 100% des GHM de la liste des quartiles et du  $\chi^2$  se répartissent entre les groupes des “autres F” et des “mauvais FA”. Si un GHM appartient au groupe des “mauvais FA”, il a 71% de risque d'appartenir à la liste des quartiles et 83% de risque d'appartenir à la liste du  $\chi^2$  (voir tableaux 4.10 et 4.11).

TAB. 4.9 – *Sensibilité et valeur prédictive positive entre la liste CHU (référence) et la classification issue de la classification hiérarchique.*

	Autres F	Mauvais FA	Mauvais FG
Sensibilité	85%	79%	66%
Valeur prédictive positive	24%	63%	50%

TAB. 4.10 – *Sensibilité et valeur prédictive positive entre la liste des quartiles (référence) et la classification issue de la classification hiérarchique.*

	Autres F	Mauvais FA	Mauvais FG
Sensibilité	100%	100%	0%
Valeur prédictive positive	10%	71%	0%

Le groupe des “mauvais FA” repère le mieux les GHM faisant partie des listes des quartiles et du  $\chi^2$ . Cela semble normal étant donné que ce sont les mêmes variables qui ont servi à construire le modèle de mélange et ces 2 listes de référence. Concernant les GHM de la liste CHU, ils sont identifiés en grande partie par le groupe des “autres F” et par celui des “mauvais FA”. Cependant, si un GHM est un “mauvais FG”, il a 50% de risque de faire partie de la liste CHU. Aucun des GHM du groupe des “mauvais FG” ne se trouve dans la liste du  $\chi^2$  ou des quartiles.

TAB. 4.11 – *Sensibilité et valeur prédictive positive entre la liste du  $\chi^2$  (référence) et la classification issue de la classification hiérarchique.*

	Autres F	Mauvais FA	Mauvais FG
Sensibilité	100%	100%	0%
Valeur prédictive positive	17%	83%	0%

#### 4.3.4 Analyse discriminante

Nous avons tiré au hasard 60% des GHM de chaque groupe pour la recherche des fonctions discriminantes. Nous avons donc à notre disposition 154 GHM répartis de la manière suivante :

- 14 GHM “mauvais FA”,
- 10 GHM “mauvais FG”,
- 58 GHM “autres F”,
- 72 GHM “bons E”.

Les caractéristiques de chacun des groupes se trouvent dans le tableau 4.12.

TAB. 4.12 – *Moyennes et écarts-types des facteurs issus de l’ACP et du déficit en points ISA pour les 154 GHM utilisés pour la discrimination.*

Groupes	Effectif	Facteur 1 $\bar{X}$ (E.T.)	Facteur 2 $\bar{X}$ (E.T.)	Déficit points ISA $\bar{X}$ (E.T.)
Mauvais FA	14	0.248 (0.142)	0.013 (0.100)	-99 084 (126 465)
Mauvais FG	10	-0.169 (0.050)	0.136 (0.040)	-33 147 (40 374)
Autres F	58	-0.029 (0.100)	-0.027 (0.094)	-24 324 (65 845)
Bons E	72	-0.167 (0.166)	-0.010 (0.075)	39 698 (68 808)

La première analyse pas à pas retient les trois variables que nous avons décidé d’inclure dans l’analyse : les deux facteurs de l’ACP et le déficit en points ISA. Elles augmentent toutes significativement le  $V$  de Rao (voir tableau 4.13). Les deux axes discriminants significatifs (voir tableau 4.14) expliquent respectivement 77% et 21% de la variance. La figure 4.17 présente la projection des GHM dans le plan de ces deux axes discriminants : facteur 1 et facteur 2. Ces axes, comme ceux de l’ACP, déterminent des facteurs ou fonctions discriminantes qui permettent de classer, c’est-à-dire d’obtenir une fonction de classement pour les GHM de notre échantillon. Néanmoins, on ne trouve que 62.3% de bien classés. Nous n’avons donc pas une très bonne fonction de classement pour nos groupes. Toutefois, les séparations entre les “mauvais FA”, les “mauvais FG” et les “autres F” sont très nettes, mais la séparation entre

les "bons E" et les trois autres groupes paraît moins franche. Cela semble être la conséquence que ces GHM n'ont pas été affectés à un groupe par la classification hiérarchique ascendante.

TAB. 4.13 – *Sélection des variables avec une analyse discriminante pas à pas avec la méthode du V de Rao.*

Pas	Variable entrée	Lambda de Wilks	p_valeur	V de Rao	Variation de V	p_valeur
1	Facteur 1	0.55	<0.00001	121,03	121,03	<0.00001
2	Facteur 2	0.43	<0.00001	171.00	49.95	<0.00001
3	Déficit en points ISA	0.39	<0.00001	187.70	16.71	0.0008

TAB. 4.14 – *Test de signification (\*) des fonctions discriminantes de la première analyse.*

Fonctions	Valeurs	% Variance propre	% Variance	Coefficient cumulée	Lambda de corrélation	$\chi^2$ de Wilks	DDL	p_valeurs
1*	0.96	76.9	76.9	0.70	0.39	139.6	9	<0.00001
2*	0.26	20.9	97.8	0.46	0.77	38.8	4	<0.00001
3*	0.03	2.2	100.0	0.16	0.97	4.0	1	0.0446

\* Le coefficient de corrélation représente ici le degré d'association entre les scores discriminants et les groupes. Le lambda de Wilks est la proportion de variance non expliquée par les différences parmi les groupes. Lambda est transformé en une variable qui suit approximativement une loi du  $\chi^2$ .

Afin de voir si les facteurs de l'ACP apportent une différence pour la discrimination des groupes, une autre analyse a été réalisée avec les variables d'origine : les pourcentages cumulés de patients dans les trois intervalles interquartiles. L'analyse pas à pas les a toutes les trois sélectionnées (elles augmentent significativement le V de Rao). Les deux axes discriminants significatifs trouvés expliquent respectivement 89% et 11% de la variance (voir tableaux 4.15 et 4.16). Cela semble indiquer que ces 3 variables sont importantes pour décrire la variété des profils des GHM. La figure 4.18 montre la projection des GHM dans le plan de ces axes. On trouve 82% de bien classés. Cela est supérieur à la première analyse, réalisée sur les facteurs issus de l'ACP et le déficit en points ISA. Cette deuxième analyse nous permet donc de trouver une meilleure fonction de classement pour nos GHM. Elle nous indique aussi que les facteurs trouvés par l'ACP sont moins discriminants que les variables d'origines.

TAB. 4.15 – *Sélection des variables avec une analyse discriminante pas à pas avec la méthode du V de Rao.*

Pas	Variable entrée	Lambda de Wilks	p_valeur	V de Rao	Variation de V	p_valeur
1	% premier intervalle	0.42	<0.00001	205.64	205.64	<0.00001
2	% deuxième intervalle	0.34	<0.00001	250.25	44.61	<0.00001
3	% troisième intervalle	0.30	<0.00001	296.90	46.66	<0.00001

TAB. 4.16 – *Test de signification (\*) des fonctions discriminantes de la deuxième analyse.*

Fonctions	Valeurs	% Variance propre	% Variance	Coefficient cumulée	Lambda de corrélation	$\chi^2$ de Wilks	DDL	p-valeurs
1*	1.75	88.5	88.5	0.79	0.30	182.2	9	<0.00001
2*	0.22	11.1	99.6	0.42	0.81	30.9	4	<0.00001
3	0.01	0.4	100.0	0.09	0.99	1.3	1	0.2620

Pour une première validation de la fonction de classement, nous avons classé les 103 GHM restants (48 considérés comme “bons E” et 55 identifiés comme “mauvais E”). Cela n’est certes pas le meilleur test pour valider ce classement, car les “mauvais” GHM ont obtenu une affectation à un groupe par la classification hiérarchique et, comme nous l’avons vu ci-avant, seuls les bons GHM n’ont pas été affectés par une classification au préalable. Toutefois, le résultat de ce classement donnera une idée des classements futurs surtout en ce qui concerne les “bons” GHM. Pour les “mauvais”, la concordance “classement/groupe prédéfini” devrait être évidemment assez bonne.

D’après le tableau 4.17 qui compare le nombre de bien classés parmi les 103 GHM entre les deux analyses, c’est la deuxième analyse, réalisée avec les variables d’origine, qui donne le meilleur taux de bien classés : 82% contre 66.0%. C’est celle-ci qui permet aussi d’obtenir un pourcentage de bien classés supérieur, que ce soit pour les GHM identifiés comme “mauvais” ou “bons”. Le tableau 4.18 donne une description du classement des 103 GHM issus de la deuxième analyse. Pour tester la robustesse du taux de bien classés dans notre deuxième analyse, nous avons refait 10 analyses incluant les mêmes variables avec 10 échantillons différents tirés de manière aléatoire (60% des GHM de chaque groupe pour calculer les fonctions discriminantes et 40% pour valider le classement). Le taux de bien classés reste très proche de celui observé dans notre deuxième analyse, 81% +/- 8% versus 82% (voir tableau 4.19).

TAB. 4.17 – *Comparaison des taux de bien classés pour deux analyses discriminantes, l’une réalisée avec les facteurs issus de l’ACP et le déficit en points ISA, l’autre avec les variables d’origine : pourcentages cumulés de patients dans les différents intervalles interquartiles.*

	Ensemble des GHM	Autres F	Mauvais FA	Mauvais FG	Bons E
Effectifs	103	39	10	6	48
<b>Première analyse :</b> F1, F2, Déficit points ISA	66%	77%	80%	100%	50%
<b>Deuxième analyse :</b> 1 <sup>er</sup> , 2 <sup>e</sup> , 3 <sup>e</sup> intervalles	82%	90%	100%	100%	71%

La sensibilité correspond à la probabilité qu’à un profil de GHM donné d’être affecté à sa classe. La valeur prédictive positive est la probabilité qu’un GHM appartenant à une classe



TAB. 4.18 – *Tableau de contingence entre les profils et les classes des 103 GHM utilisés pour construire la fonction de classement de l'analyse discriminante. Variables discriminantes utilisées : les 3 pourcentages cumulés de patients sur les intervalles interquartiles de DS. 82% de bien classés.*

Profils	Classes				Total par profil
	Autres F	Mauvais FA	Mauvais FG	Bons E	
Autres F	35	0	3	1	39
Mauvais FA	0	10	0	0	10
Mauvais FG	0	0	6	0	6
Bons E	13	0	1	34	48
Total colonne	48	10	10	35	103

TAB. 4.19 – *Analyse de la robustesse des taux de bien classés dans les différentes classes à partir de 10 nouvelles analyses des variables d'origine : proportions cumulées de patients dans les différents intervalles interquartiles.*

	Ensemble des GHM	Autres F	Mauvais FA	Mauvais FG	Bons E
	$\bar{X}[E.T.]$ (en %)				
Effectifs	103	39	10	6	48
Analyses :					
1 <sup>er</sup> , 2 <sup>e</sup> , 3 <sup>e</sup> intervalles	80.7 (4.8)	80.8 (7.6)	92.0 (6.3)	98.3 (5.3)	76.0 (6.0)

donnée ait effectivement le profil de cette classe (voir tableau 4.20). Concernant les groupes de GHM appartenant aux groupes des “mauvais FA” et “mauvais FG”, leur sensibilité et leur valeur prédictive positive sont égales à 100%. Ce chiffre est à relativiser, vu le peu d'effectif de ces deux classes. La sensibilité du groupe des “autres F” est égale à 74%, ainsi les “autres F” ont 74% de chance d'être affectés à leur classe.

Les coefficients des fonctions discriminantes et les fonctions de classement sont présentés dans l'annexe M

## 4.4 Discussion

### 4.4.1 ACP

L'ACP nous a permis de déterminer qu'une classification se dessinait sur les profils grenoblois. Le premier facteur est plutôt un indicateur sur le type de GHM : “bon F” ou “mauvais F”. Le deuxième indicateur est plutôt un indicateur de forme, il donne une idée de l'endroit de la distribution où les GHM sont peu performants ou très performants (séjours de courte ou de longue durée).

TAB. 4.20 – *Sensibilité et valeur prédictive positive de la discrimination (deuxième analyse).*

	Autres F	Mauvais FA	Mauvais FG
Sensibilité	74%	100%	100%
Valeur prédictive positive	96%	100%	100%

L’orthogonalité entre les facteurs présente un avantage certain, l’absence de corrélations évite les redondances d’informations et permet ainsi une description assez complète des profils.

Cependant, il est intéressant de décrire plus finement la classification qui se dessine sur les profils grenoblois.

#### 4.4.2 Classification des profils des 257 GHM

La classification hiérarchique ascendante nous a permis d’isoler trois groupes parmi les GHM identifiés comme mauvais à l’issue de l’approche empirique. Les “bons E” constituent le quatrième groupe que l’on n’a pas étudié en détail, puisque notre intérêt se porte sur les “mauvais E”. L’avantage de cette classification est qu’elle identifie un groupe de 16 GHM pouvant être la cause d’une partie importante du déficit du CHU de Grenoble.

##### Groupe des “mauvais FG”

Ce groupe de 16 GHM (6% de notre échantillon) a comme caractéristique commune d’être peu performant sur les séjours de longue durée uniquement ( $DS > q_{(0.75)}^N$ ) [profil empirique de type ( $g$ )]. La majorité des GHM appartenant à ce groupe ont comme début d’intitulé: “Autres ...”, synonyme d’absence de diagnostic précis. Ces GHM sont en majorité de type médical (94%), signifiant souvent une certaine hétérogénéité médicale des patients. Il est donc important de s’intéresser aux RSS de ces patients responsables de la dérive en termes de DS, afin d’en expliquer les causes. Ces patients ont sans doute dû avoir un grand nombre d’examens médicaux, car on n’arrivait pas à poser un diagnostic précis, cela s’est traduit par des DS plus longues et des coûts plus élevés que la normale pour le CHU. Cependant, cela peut aussi s’expliquer par des erreurs de classification dans les GHM. En effet, la personne responsable du codage des RSS a pu omettre des co-morbidités importantes ou des actes qu’auraient eu ces patients, l’algorithme classant alors le patient dans un GHM qui n’aurait pas dû être le sien.

Pour ce faire, il faudrait étudier plus finement les caractéristiques (diagnostics principal et associés et actes) de ces patients ayant une  $DS > q_{(0.75)}^N$  en consultant leur dossier médical, soit 709 dossiers sur les 2 359 de ces 16 GHM, si on voulait être exhaustif. Une alternative

serait de faire un tirage aléatoire de 10% de ces patients, soit environ 71 dossiers, ce qui est faisable à l'échelle du CHU. Cette démarche serait rentable pour l'hôpital, car on a réussi à isoler une sous-population de patients qui ont un allongement significatif de leur DS et qui semblent être à l'origine d'une partie du déficit en points ISA du CHU de Grenoble. De plus, le nombre de dossier médicaux à contrôler (71) est raisonnable. Cela permettrait de connaître les causes exactes (médical ou administrative) de l'allongement de leurs séjours.

Cette recherche pourrait aboutir à la mise en place d'actions (meilleure démarche diagnostic, sensibilisation à l'importance du codage des RSS, ...) auprès des médecins des services concernés, visant à résoudre les problèmes soulevés par l'analyse détaillée des 71 dossiers. La conséquence principale de cette action pourrait être économique, avec une augmentation du nombre de points ISA au CHU de Grenoble.

En effet, ce groupe de GHM est responsable d'un déficit de 709 595 points ISA, si on se fonde sur la méthode de calcul utilisée pour construire la liste CHU (voir paragraphe 2.3.1), soit 10% du déficit total du CHU de Grenoble. Pour ces 16 GHM, l'hôpital soigne au total 1 467 patients en moins par rapport à la base nationale.

Comme nous l'avons déjà identifié, la cause de la dérive en matière de points ISA est sans doute le groupe de patients ayant une  $DS > q_{(0.75)}^N$  (soit environs un tiers des patients). Il ne faudrait pas grand chose pour que ce groupe de GHM puisse regagner le groupe des bons GHM, ou au moins le groupe des GHM "autres F". Nous allons quantifier quel serait l'impact d'un tel changement en termes de points ISA.

Pour ce faire, nous avons choisi arbitrairement d'enlever, à l'effectif de chacun des GHM, les patients grenoblois ayant une DS strictement supérieure à la DS correspondant à 75% des patients nationaux. Notre hypothèse est que ces patients n'auraient pas dû se trouver classés dans ces GHM. Nous avons donc éliminé entre 25.4% (GHM n°586) et 36.7% (GHM n°466) de patients, ce qui représente 710 patients au total (31%). Nous avons ensuite recalculé l'apport relatif en points ISA par rapport à la base nationale avec la méthode fournie par le CHU. Les GHM obtiennent alors un bénéfice de 863 818 points ISA (voir tableau 4.21). En recalculant la somme des points ISA totaux grenoblois relative à la DS moyenne base nationale, on s'aperçoit que le CHU, qui était déficitaire de 1 128 368 points ISA, devient bénéficiaire de 445 045 points ISA, et se retrouve ainsi en équilibre avec la norme nationale. Nous voyons donc qu'avec cette approche, le CHU de Grenoble est gagnant en termes de points ISA.

### Groupe des "mauvais FA"

Pour décrire plus finement le groupe des 24 GHM identifiés comme "mauvais FA" (9% de notre échantillon), nous avons étudié à nouveau le dendrogramme (voir figure 4.14) à un niveau de segmentation plus petit (0.3). Il montre globalement trois sous-groupes, qui représentent différents niveaux d'éloignement du profil grenoblois à la référence nationale mesuré par la différence d'aire entre le profil grenoblois et la référence nationale. La caractéristique commune de ces GHM est qu'ils sont peu performants quelle que soit la longueur de leur séjour.

TAB. 4.21 – Calcul des points ISA totaux du CHU de Grenoble avant et après avoir enlevé le groupe des 16 GHM.

Groupes	N patients		Déficit ou bénéfice total de points ISA		Déficit ou bénéfice total de patients	
	Avant*	Après**	Avant*	Après**	Avant*	Après**
GHM “mauvais FG”	2359	1650	-709 595	863 818	-467	811
GHM “mauvais FA”	3084	3084	-2 324 771	-2 324 771	-1 467	-1 467
GHM “autres F”	15779	15779	-2 808 597	-2 808 597	-1 820	-1 820
GHM “bons E”	33428	33428	4 714 595	4 714 595	3 558	3 558
Total	54 650	53 941	-1 128 368	445 045	-277	1 082

(\* analyse avec tous les patients de la base grenobloise, \*\* analyse sans les patients appartenant au groupe des 16 GHM et ayant une  $DS > q_{(0.75)}^N$ )

Un premier sous-groupe rassemble les GHM n°48 et 61 (CMD affections de l’oeil), ils correspondent aux deux profils les plus éloignés de la norme nationale (différence d’aire la plus élevée), particulièrement sur les séjours de courte et de longue durées. Un deuxième rassemble les GHM n°54 (CMD affections de l’oeil), 71 (CMD affections des oreilles, du nez, de la gorges, de la bouche et des dents), 405, 430 et 432 (CMD affections du rein et des voies urinaires), ils ont comme caractéristique commune d’être très éloignés de la référence nationale particulièrement sur les séjours de courte et de moyenne durées. Un troisième groupe rassemble les 17 GHM restants, il sont un peu moins éloignés de la référence nationale, mais restent tout de même peu performants. Près de la moitié de ces GHM sont de type chirurgical, cela signifiant une certaine homogénéité médicale de la population.

Ces 24 GHM occasionnent un déficit important de 2 324 771 points ISA au CHU de Grenoble, si l’on se réfère à la liste CHU, qui correspond à 33% du déficit du CHU en points ISA. Cela signifie aussi que par rapport à la base nationale, le CHU de Grenoble a tendance à soigner au total 1 467 patients de moins dans les GHM de ce groupe.

Le contrôle des dossiers médicaux des patients appartenant à ce groupe pour tenter d’expliquer l’allongement des séjours risque d’être difficile à organiser. En effet, à la différence des patients faisant partie du groupe des 16 “mauvais FG”, ces 24 GHM sont identifiés comme mauvais quelle que soit la durée de leur séjour (courte, moyenne ou longue), du coup on ne sait pas exactement quels patients en sont la cause ... La seule solution serait de vérifier le RSS de tous les patients, mais cela prendrait beaucoup de temps, il y en a 3 804 ... Une alternative consisterait à faire un tirage aléatoire de 10% des RSS des patients appartenant à chacun des GHM de ce groupe, ce qui représenterait environ 308 dossiers (nombre encore élevé par rapport aux 71 dossiers médicaux correspondant aux groupe des 16 GHM). Néanmoins, cela permettrait de comprendre en partie les raisons de l’allongement de la DS à Grenoble pour les patients de ces GHM : médicale (recrutement particulier de patients à Grenoble, variation des pratiques médicales) ou administrative (erreurs de codage dans les RSS, problème de structure). L’inconvénient de cette démarche est qu’elle n’est pas

ciblée sur un sous-groupe de patients comme pour le groupe des 16 GHM, mais réalisée à l'aveugle sur tous les patients de ce groupe ... En conséquence, les caractéristiques des RSS tirés au sort n'expliqueront peut-être qu'une infime partie du déficit en points ISA; de plus, l'action qui en découlerait risquerait de ne pas être très rentable pour le CHU de Grenoble, car les problèmes révélés par l'analyse des 308 GHM peuvent être très divers, donc difficile à exploiter.

### Groupe des “autres F”

La forme des profils des GHM appartenant à ce groupe varie entre les profils empiriques de types  $(a)$ ,  $(b)$ ,  $(c)$  et  $(e)$ .

Cependant il est important d'affiner la description du groupe des 97 “autres F”. Pour ce faire, nous avons regardé le dendrogramme à un niveau de segmentation plus petit (0.3), le même que pour les 24 “mauvais FA”. Il montre trois sous-groupes, qui correspondent à différentes formes de profils empiriques. Le premier sous-groupe rassemble 11 GHM, qui ont pour la majorité un profil empirique de type  $(e)$ , ils correspondent aux GHM qui sont peu performants sur les séjours de courte durée uniquement. Le deuxième sous-groupe rassemble 14 GHM, dont huit ont un profil empirique de type  $(b)$ : GHM peu performants sur les séjours de courte et de moyenne durées, et six de type  $(a)$ : GHM peu performants quelle que soit la longueur de leur séjour. Le troisième sous-groupe réunit les 52 GHM restants, qui ont des profils divers.

Le groupe des “autres F” est responsable d'un déficit de 2 808 597 points ISA pour le CHU, soit un total de 1 820 patients. Compte tenu du nombre de GHM et de leur diversité en termes de profils de distribution, il ne semble pas concevable d'aller vérifier les dossiers médicaux de ces patients, même 10% d'entre eux (soit 1 578), ce qui ne serait qu'une perte de temps pour le CHU. Une solution serait de croiser ces GHM avec la classification empirique, afin d'en repérer les plus particuliers et d'ainsi vérifier les dossiers des quelques patients concernés. Cette démarche se ferait plutôt à titre indicatif.

### Comparatif du nombre de RSS à contrôler

Il est intéressant de comparer le nombre de RSS ou de dossier médical (sur la base d'un tirage aléatoire de 10% d'entre eux) à contrôler pour tenter d'expliquer l'allongement des DS à Grenoble, selon les listes de référence et les trois groupes de mauvais GHM de notre classification. Il faudrait vérifier :

- 1 172 dossiers de la liste CHU (50 GHM représentant 80% du déficit total du CHU),
- 660 dossiers de la liste du  $\chi^2$  (36 GHM représentant 10% du déficit),
- 422 dossiers de la liste des quartiles (27 GHM représentant 35% du déficit),
- 308 dossiers du groupe des GHM ayant un profil empirique de type  $(a)$  (24 GHM représentant 33% du déficit),

- 71 dossiers du groupe des GHM ayant un profil empirique de type (*g*) (16 GHM représentant 10% du déficit).

Pour d'être rentable en termes de points ISA, il faudrait vérifier un grand nombre de dossiers, en ciblant sur les groupes ou les listes qui ont tendance à causer une grande partie du déficit total en points ISA au CHU (liste CHU, liste des quartiles). Cependant, moins il y aura de dossiers médicaux à contrôler, meilleures en seront les analyses et la mise en place d'actions ciblées et rentables pour le CHU.

Il est ainsi important, dans un premier temps, de repérer les GHM qui ont des profils empiriques identifiés comme mauvais. Cela permet de cibler l'intervalle dans la distribution des DS où les patients ont des séjours plus longs que la norme nationale. Ensuite, on tire au sort 10% des dossiers médicaux dans cet intervalle, et on les analyse en détail. Pour finir, il est possible de mettre en œuvre une action auprès des services concernés, visant à résoudre les problèmes mis en exergue par cette catégorie de patients. C'est cette démarche qui a été utilisée pour le groupe des "mauvais FG". Nous prendrons comme exemple 8 GHM communs au groupe des 16 GHM "mauvais FG" et à la liste CHU (GHM n°21, 188, 201, 246, 398, 425, 585 et 619). Sans l'information de la classification sur la forme des profils, on aurait dû tirer au sort 124 dossiers, alors que 39 auraient suffi, soit un gain de 30% en charge de travail.

D'autre part, la méthode ayant servi à calculer le déficit en points ISA se fonde sur la DS moyenne nationale, indicateur peu pertinent d'un point de vue statistique pour les distributions de DS, souvent bi-modales ... C'est pourquoi, aujourd'hui, le déficit en points ISA calculé de cette manière n'est qu'indicatif. Pour être plus juste, il faudrait coupler cette méthode de calcul à la forme des profils empiriques en se fondant sur les quartiles de DS et non sur la moyenne.

### 4.4.3 Les facteurs de l'ACP dans l'analyse discriminante

Pour le classement des GHM, nous avons vu que les pourcentages cumulés de patients dans les trois premiers intervalles interquartiles étaient les variables les plus adaptées, même par rapport à l'analyse utilisant les deux facteurs de l'ACP et le déficit ou le bénéfice en points ISA, ceci pour trois raisons :

- les indices interquartiles classent mieux les profils des GHM (81% contre 66%),
- l'analyse de robustesse montre qu'après 10 analyses sur 10 échantillons différents, le taux de bon classement est égal à 81% +/-7.6,
- ces indices permettent une interprétation plus fine de la classification des profils des GHM : le pourcentage sur le premier intervalle interquartile détecte les GHM peu performants sur les séjours de courte durée, celui sur le deuxième intervalle décèle les GHM peu performants sur les séjours de durée moyenne et celui sur le troisième, l'allongement sur les séjours de longue durée.

### Sensibilité des classes et validation du classement

La sensibilité est la plus élevée (100%) pour les “mauvais FA” et les “mauvais FG”. Cependant, ces résultats sont à relativiser, les effectifs de ces deux groupes étant faibles, respectivement 10 et 6 (voir tableau 4.20). Ils sont donc dans la totalité des cas décelés comme GHM “atypiques”. Concernant le groupe des GHM “autres F”, 90% des GHM sont bien classés, les 10% restant se retrouvent classés dans les “mauvais FG” [3 GHM] et dans le groupe des “bons E” GHM [1 GHM] (voir tableau 4.18). Ils sont ainsi en quasi totalité détectés comme “atypiques”.

#### 4.4.4 Limites

##### Limites statistiques

La classification que nous avons présentée ne s'est pas révélée très stable, particulièrement pour les groupes de GHM ayant des profils empiriques de type  $(a)$  et  $(g)$ , où il n'y avait pas de noyau stable. C'est pourquoi il ne serait pas raisonnable d'extrapoler ces résultats à des bases PMSI plus récentes. Cependant, l'analyse discriminante portant sur cette classification a donné des résultats très robustes (un peu plus de 80% de bien classés).

La classification hiérarchique nous a permis de voir qu'il existait sans doute une classification des profils des distributions de DS grenobloises, mais que cette dernière n'était peut-être pas optimale.

Concernant l'analyse discriminante, nous n'avons pas pu valider notre classification avec des GHM d'une base PMSI plus récente, n'en n'ayant pas à disposition. La seule validation que nous avons faite consiste en 10 analyses sur 10 échantillons différents construits à partir d'un tirage aléatoire de nos données, afin de voir comment évoluent les taux de bien classés dans les différents groupes. Ces derniers sont proches de ceux de l'analyse témoin que nous avons présentée (81% contre 82%), et les écart-types sont petits, donc synonymes d'une certaine robustesse de notre classification (voir tableaux 4.17 et 4.19).

Il serait intéressant de valider ce modèle avec des GHM appartenant une base PMSI plus récente; les résultats risquent cependant d'être médiocres, notre classification n'étant pas très stable du fait du faible effectif des groupes  $(a)$  et  $(g)$ .

##### Limites économiques

La règle du cut-off (éliminer les patients ayant des  $DS > q_{(0.75)}^N$ ) que nous avons utilisée pour le groupe des “mauvais FG”, afin de recalculer la somme des points ISA relative à la DS moyenne de la base nationale, est un peu abrupte. Une alternative consisterait à modéliser les distributions de DS des 16 GHM grenoblois par des mélanges de lois, afin de tenir compte de la bi-modalité des distributions de DS mise en évidence par leur profil, puis de repérer de

manière plus précise les “outliers”, et alors éliminer seulement un certain pourcentage (5% ou 10%) de ceux qui se trouvent le plus à droite de la deuxième distribution du mélange.

#### 4.4.5 Synthèse

La classification issue de l’approche factorielle nous a permis de mettre en évidence une catégorie de 16 GHM, dans lesquels un tiers des patients semble être à l’origine de la dérive en points ISA du CHU de Grenoble. La classification nous a aussi montré que les profils des GHM grenoblois semblaient se structurer en quatre groupes, dont trois étaient identifiés comme mauvais. Cette classification a l’avantage d’être en cohérence avec celle d’origine empirique (8 groupes), plus fine que la liste du CHU, et robuste (plusieurs échantillons ont été testés avec l’analyse discriminante et ils donnent un taux moyen de 81% de bien classés). D’autre part, la classification issue de l’approche factorielle apporte des informations supplémentaires sur le type des profils des GHM du CHU de Grenoble, mais également une aide pour les gestionnaires de l’hôpital, toujours en quête de points ISA SUPPLEMENTAIRES ... faciles à obtenir.

Dans le chapitre suivant, nous allons proposer une nouvelle classification des profils des 257 GHM grenoblois, cette fois-ci issue d’une méthode différente, utilisant les modèles de mélange, pour valider la classification que nous venons d’étudier.



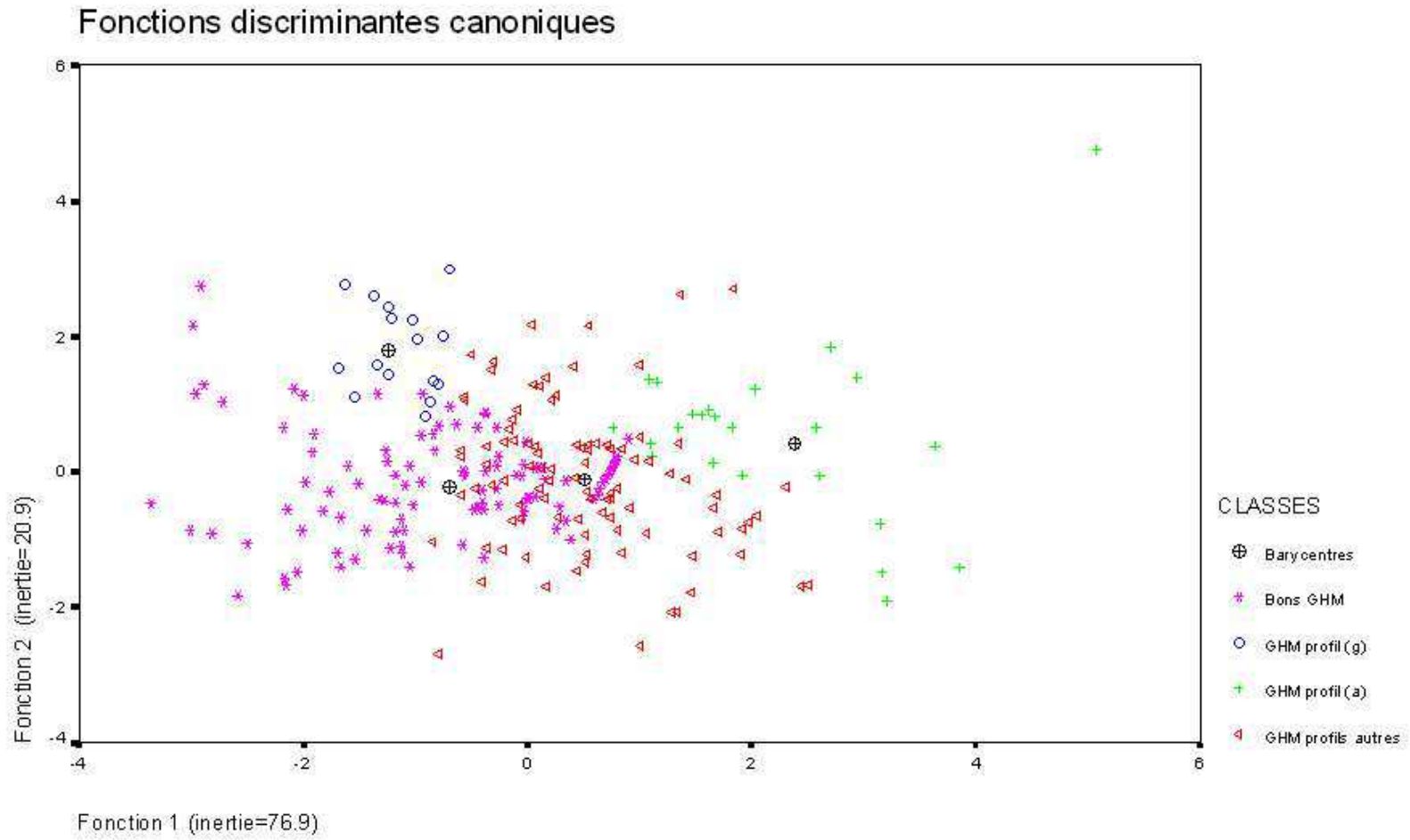


FIG. 4.17 – Première analyse : projection des 154 GHM dans le plan des 2 facteurs discriminants significatifs (62% de bien classés).

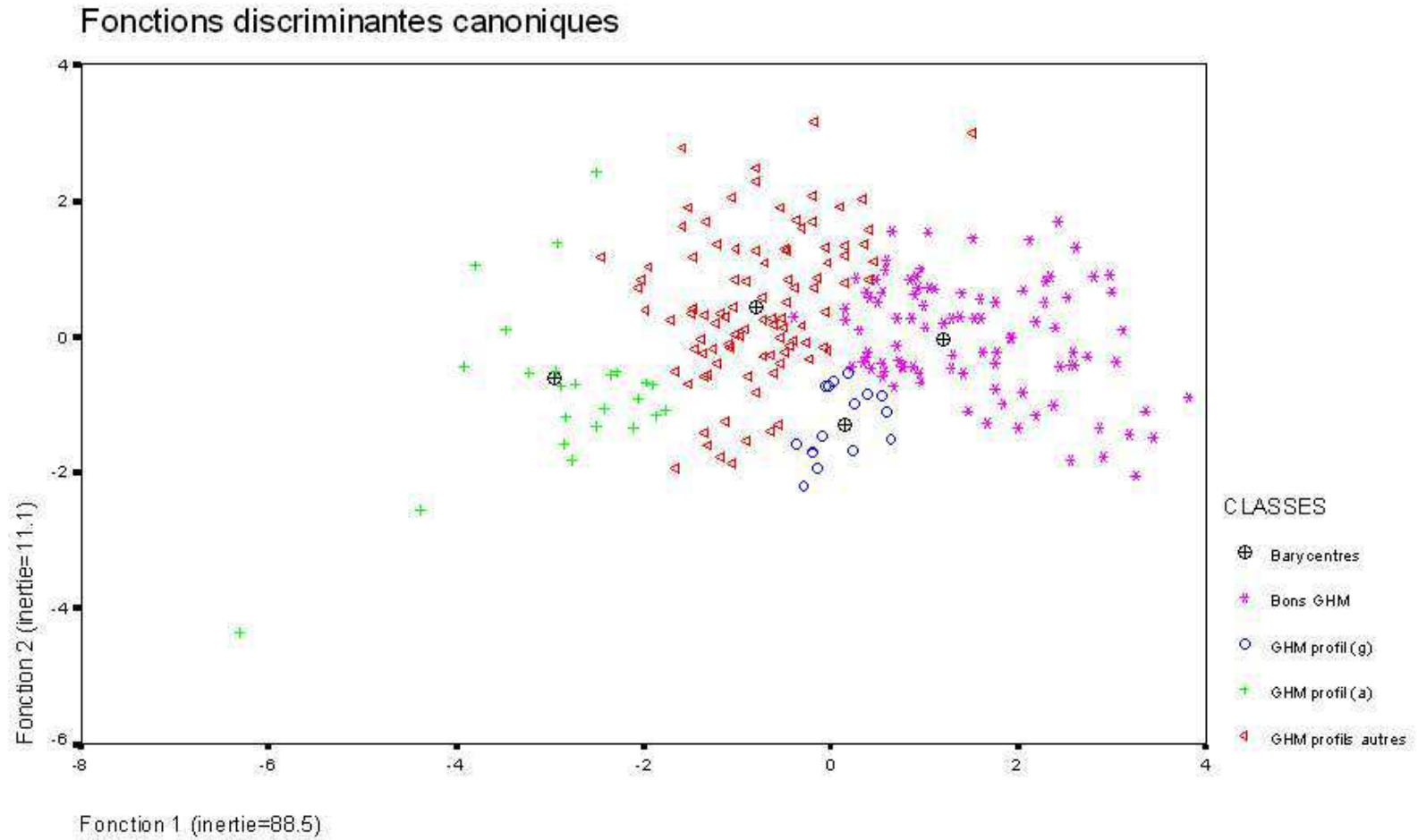


FIG. 4.18 – *Deuxième analyse : projection des 154 GHM dans le plan des 2 facteurs discriminants significatifs (82% de bien classés).*



# Chapitre 5

## Classification des profils issue des modèles de mélange de lois

Après avoir étudié, dans le chapitre précédent, la classification des profils issue d'une approche factorielle, ce chapitre s'intéresse à une approche différente, utilisant les modèles de mélange de lois, pour établir une nouvelle classification des profils des 257 GHM. Les modèles de mélange ont l'intérêt d'analyser si ceux-ci se structurent en classes, mais aussi de pouvoir inférer sur le nombre de classes qui les décrivent, en tenant compte de l'hétérogénéité observée dans nos données (profils de différents types). L'objectif de cette nouvelle classification est de valider la classification issue de l'approche factorielle, mais aussi de repérer les GHM atypiques.

Pour commencer, nous allons définir les principes mathématiques sur lesquels reposent les modèles de mélange. Ensuite, nous présenterons et discuterons les résultats des analyses.

### 5.1 Méthode

L'objectif de ce paragraphe est de présenter de manière simple le fonctionnement des modèles de mélange.

#### 5.1.1 Principe des modèles de mélange de lois normales

Les modèles de mélange [26, 39, 42, 83] sont un moyen d'obtenir des distributions complexes, tout en s'appuyant sur des lois très simples, comme les lois normales. Leur utilisation est très classique dans les problèmes de classification. Comme les profils des 257 GHM sont caractérisés par 3 variables, nous avons utilisé des modèles de mélange multivariés. D'autre part, nous avons fait l'hypothèse que l'on se trouvait dans le cas gaussien.

##### **Mélange gaussien multivarié :**

Soit  $\mathbf{X}$  une variable aléatoire dans  $\mathbb{R}^d$ , ( $d=3$ ). La variable aléatoire  $\mathbf{X}$  suit une loi de mélange gaussien à  $K$  composantes si sa densité est donnée par la fonction :

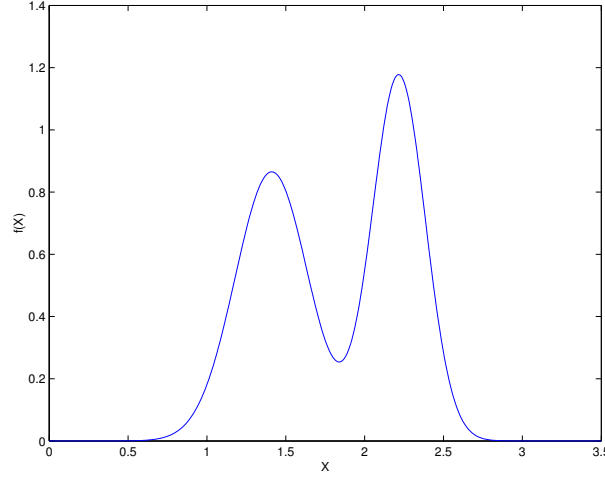


FIG. 5.1 – Exemple de densité d'un modèle de mélange de deux lois normales univariées.

$$f(\mathbf{x}; \boldsymbol{\theta}) = \sum_{k=1}^K p_k \phi(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k); \quad (5.1)$$

où

$$\boldsymbol{\theta} = ((p_k)_{k=1, \dots, K}, (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_{k=1, \dots, K}),$$

avec  $p_k$ ,  $k = 1, \dots, K$ , telles que  $p_k \in [0, 1]$  et  $\sum_{k=1}^K p_k = 1$  et  $\phi(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  est une loi normale de moyenne  $\boldsymbol{\mu}_k$  et de matrice de variance-covariance  $\boldsymbol{\Sigma}_k$ ,

$$\phi(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right).$$

La décomposition spectrale de la matrice variance-covariance proposée par Celeux et Govaert [31] permet de préciser la structure de cette matrice dans les différentes classes, à l'aide des paramètres interprétables d'un point de vue statistique et géométrique. Ainsi, en présence de liaisons entre les paramètres initiaux des lois gaussiennes, cette décomposition permet de réduire le nombre de paramètres à estimer pour la densité. La composition est la suivante :

$$\forall k \in (1, 2, \dots, K), \boldsymbol{\Sigma}_k = \lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k',$$

où  $\mathbf{D}_k$  est la matrice orthogonale des vecteurs propres,  $\mathbf{A}_k$  est une matrice diagonale constituée des valeurs propres de  $\boldsymbol{\Sigma}_k$  normalisées de sorte que leur produit fasse 1, et  $\lambda_k$  un scalaire

(moyenne géométrique des valeurs propres de  $\Sigma_k$ ).

Une description géométrique de chacun des termes de la décomposition est la suivante :

- $D_k$  détermine l'orientation de l'ellipsoïde de dispersion de la classe  $k$ ,
- $A_k$  donne une information sur l'allongement relatif de l'ellipsoïde de dispersion de la classe  $k$  le long de ses axes,
- $\lambda_k$  est associé au volume de la classe  $k$ ,  $\lambda_k |A_k|$  étant proportionnelle au volume de l'ellipsoïde correspondant.

On dénombre 28 modèles possibles pour la décomposition de la matrice variance-covariance. Ne disposant pas d'information sur la structure de nos données, nous avons adopté le modèle le plus général : les proportions de chaque groupe  $k$  du mélange sont libres et leurs ellipses peuvent être de volumes, de formes et d'orientations différentes (voir figure 5.2).

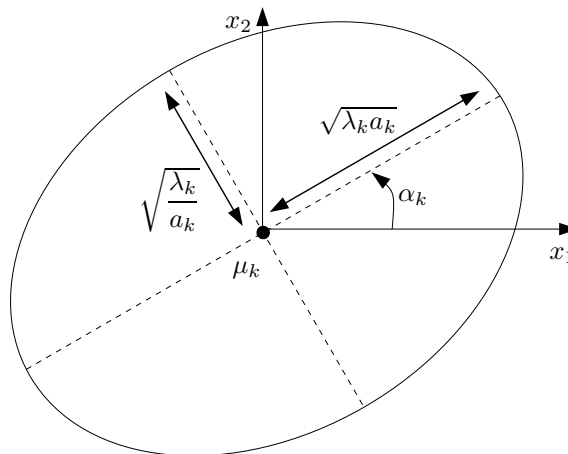


FIG. 5.2 – *Interprétation géométrique de la décomposition spectrale. La quantité  $\lambda_k$  représente le volume de l'ellipse de concentration,  $a_k = A_k(1,1)$  détermine sa forme et l'angle  $\alpha_k$  de la matrice de rotation  $D_k$  détermine son orientation.*

### Estimation des paramètres d'un mélange gaussien

L'estimation des paramètres du mélange consiste à estimer  $\theta$  à partir d'un échantillon  $\mathcal{X}=(\mathbf{x}_1, \dots, \mathbf{x}_n)$ , dont chaque élément  $\mathbf{x}_i$  de  $\mathbb{R}^d$ , ( $1 \leq i \leq n$ ) est une réalisation de la loi de mélange  $f(\mathbf{x}; \theta)$ . L'algorithme EM permet de déterminer  $\hat{\theta}$ , l'estimateur du maximum de vraisemblance de  $\theta$ . L'algorithme EM a été étudié dans un cadre très général par Dempster, Laird et Rubin en 1977 [36, 39, 65, 66]. C'est une méthode itérative très connue et largement utilisée pour l'identification de mélanges (voir annexe N).

## Choix du nombre de composantes du modèle

Le choix du nombre de composantes du modèle est un problème ancien et délicat [26, 39, 42, 83]. Dans notre étude, ce choix s’est fait de façon pragmatique : nous avons choisi le modèle le plus pertinent du point de vue de la signification médico-économique. Cependant, nous avons tout de même étudié les valeurs de deux critères statistiques couramment utilisés (BIC [*Bayes Information Criterion*] [80] et ICL [*Integrated Completed Likelihood*] [28]), afin de déterminer si le nombre de composantes du modèle choisi de manière pragmatique était proche de celui “élu” par ces 2 critères. Le nombre de composantes choisies dépend de la configuration qui maximise la valeur de ces critères.

### 5.1.2 Application à la classification des profils

Disposant de trois variables pour spécifier les profils des 257 GHM de notre échantillon (voir section 2.2), nous notons  $\mathbf{x}_i$ , pour  $i = 1, \dots, 257$ , le vecteur à 3 composantes décrivant le profil de la distribution des DS du GHM<sub>*i*</sub>. La première composante correspond à la proportion cumulée de patients sur le premier intervalle interquartile, la deuxième à la proportion cumulée sur le deuxième intervalle et la troisième à la proportion cumulée sur le troisième intervalle. Par exemple, nous avons :

$$\mathbf{x}_1 = \begin{pmatrix} 0.10 \\ 0.20 \\ 0.80 \end{pmatrix}$$

Les vecteurs  $\mathbf{x}_i$  de  $\mathbb{R}^3$  sont rassemblés dans l’ensemble  $\mathcal{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{257})$ . L’objectif est de trouver une partition en  $K$  classes de cet ensemble. On cherche donc à décrire l’appartenance des  $\mathbf{x}_i$  de  $\mathcal{X}$  aux classes. Il s’agit d’estimer les labels  $\mathbf{Z}_i \in \{1, \dots, K\}$  de chaque point  $\mathbf{x}_i$ , donc d’obtenir une classification estimée  $\hat{\mathbf{z}}_i = (\hat{z}_1, \dots, \hat{z}_{257})$  où  $\hat{z}_i = k$  si  $\mathbf{x}_i$  est dans la classe  $k$ . Dans notre travail, nous avons utilisé la méthode du maximum *a posteriori* ou méthode du MAP, pour déduire la classification  $\hat{\mathbf{z}}$ . Cette méthode consiste simplement à affecter le profil  $\mathbf{x}$  à la classe  $k$  la plus probable *a posteriori* qui vérifie  $\mathbf{z} = \arg \max_k t_k(\mathbf{x}; \boldsymbol{\theta})$ , la probabilité *a posteriori*  $t_k(\mathbf{x}; \boldsymbol{\theta})$  que  $\mathbf{x}$  appartienne à la classe  $k$  se calcule directement par le théorème de Bayes (voir équation (N.1) en annexe N).

Pour interpréter les classes du mélange obtenu, nous avons étudié les caractéristiques médico-économiques de chacun des groupes  $k$  ( $k = (1, \dots, K)$ ) du mélange selon les variables suivantes : le déficit moyen en points ISA, la différence d’aire entre les profils grenoblois, et le type de GHM (médical ou chirurgical). Nous avons testé le lien possible entre les groupes et les différentes caractéristiques par une analyse de variance si la variable était quantitative, et par un test du  $\chi^2$  si la variable était qualitative au seuil  $\alpha = 5\%$ .

### 5.1.3 Stratégie de modélisation

Pour proposer une classification des 257 profils et un découpage en classes, nous avons estimé, à 3 reprises, des modèles de mélanges gaussiens multivariés sur les trois intervalles interquartiles en proportions cumulées, avec le logiciel *XEMgaus* [27]. L'estimation de ces modèles a été faite en utilisant l'algorithme EM. L'algorithme EM a été initialisé par un tirage aléatoire des centres de classes. À chaque fois, cinq cent essais ont été réalisés pour chacun des modèles. Si l'algorithme ne convergait pas sur un essai, il avait droit à 499 essais supplémentaires au maximum pour converger. Nous avons estimé des modèles de mélange allant de 2 à 5 classes. Nous avons retenu le modèle de mélange pertinent d'un point de vue médico-économique (listes de références, déficit en points ISA) et statistique (valeur des critères BIC et ICL). Ensuite, nous avons décrit les caractéristiques médico-économiques des groupes du modèle de mélange retenu. Nous avons confronté ces groupes aux différentes listes de référence et à ceux de la classification issue de l'approche factorielle pour voir s'ils étaient proches.

## 5.2 Résultats

### 5.2.1 Classification résultant des modèles de mélange

#### Choix du meilleur modèle de mélange

Le mélange gaussien multivarié que nous avons retenu est le modèle le plus explicatif des points de vue médico-économique (déficit en point ISA et profil de distribution empirique) et statistique (valeurs des critères BIC et ICL).

D'après les figures 5.3, 5.4, 5.5, les modèles de mélanges retenus sur la base des critères statistiques varient entre 2 et 3 classes. Le critère BIC choisit toujours le modèle de mélange à 3 classes. Par contre, si on se base sur le critère ICL, il hésite entre 2 et 3 classes, cependant, dans 2 des 3 analyses, il choisit 3 classes.

Le modèle de mélange de 2 lois normales distingue 2 groupes de GHM grenoblois. L'un d'eux correspond aux GHM ayant plutôt un profil de distribution de DS proche de la base nationale, mais tout de même légèrement en dessous des proportions de référence (N=151). L'autre groupe a tendance à réunir les bons GHM, en terme de profils de distributions de DS (N=106). Ce modèle n'est donc pas convaincant d'un point de vue médico-économique puisqu'il n'arrive pas à distinguer les mauvais GHM de Grenoble.

En revanche, nous retenons les mélanges de 3 lois normales, qui sont pertinents des points de vue médico-économique et statistique tout en restant simples. Le découpage des GHM grenoblois en 3 groupes à l'aide des modèles de mélange, pourrait représenter correctement la réalité médicale : un petit groupe de mauvais GHM, un petit groupe de bons GHM et un



groupe important de GHM autres, qui semblent être ni vraiment mauvais, ni vraiment bons.

D'autre part, les paramètres trouvés pour ces modèles sont semblables dans les 3 analyses. Cela confirme donc une certaine robustesse des modèles de mélange de 3 lois normales pour donner une classification des profils de distributions DS des GHM grenoblois. Intéressons-nous maintenant au modèle détaillé.

## 5.2.2 Description et interprétation du mélange de 3 lois normales

Le modèle sélectionné pour le décrire plus précisément correspond à celui résultant de la première analyse. En effet, les valeurs des critères BIC et ICL concluent toutes les deux pour ce modèle à 3 classes. De plus, ce modèle a la plus grande vraisemblance parmi les 3 analyses, à nombre de composantes du modèle égal.

En traçant le profil moyen des GHM de chacun des groupes du mélange, nous avons pu analyser les caractéristiques de chacun. La figure 5.6 montre les proportions cumulées moyennes de patients dans chacun des intervalles interquartiles pour les 3 groupes.

Le premier groupe rassemble 52 GHM, on les a nommé les “**mauvais M**” (voir annexe O). Il compte principalement des profils empiriques de type  $(a)$  mais aussi de types  $(b)$  et  $(f)$ , respectivement 62%, 21% et 13%. Ces GHM ont tendance à être peu performants quelle que soit la longueur du séjour ( $Ic_{(1)} < 0.25$ ,  $Ic_{(2)} < 0.50$  et  $Ic_{(3)} < 0.75$ ) [profil en pointillés de couleur bleue dans la figure 5.6 et groupe de points de couleur bleue dans les figures 5.7, 5.8, 5.10 et 5.9]. Ils occasionnent un déficit moyen de 56150 +/- 84634 points ISA et la différence moyenne d'aire entre le profil grenoblois des GHM appartenant à cette classe et la norme nationale est de -8.1% +/- 6.8% (voir tableau 5.1. Ce groupe compte un pourcentage équivalent de GHM chirurgicaux par rapport à l'échantillon regroupant les 257 GHM (48% contre 33%).

Le deuxième groupe (n=127) correspond aux GHM qui ne sont, ni classés dans les “mauvais M”, ni dans les “bons M”, on les a appelé les “**autres M**” (voir annexe P). Il compte principalement des profils empiriques de types  $(neutre)$  et  $(h)$  (profils regroupés ensemble pour l'analyse), mais aussi des profils empiriques de type  $(g)$  et  $(c)$ , 21% pour chacun des profils (42%), 18%, 14% et 12%. Le profil moyen de ces GHM ne diffère pas de celui de la référence nationale [profil en trait continu de couleur rouge sur la figure 5.6 et groupe de points, croix de couleur rouge, dans les figures 5.7, 5.8, 5.10 et 5.9]. Ils causent un déficit moyen de -18 190 +/- 69 931 points ISA au CHU de Grenoble et la différence moyenne d'aire entre le profil grenoblois et la norme nationale est égale à 0.4% +/- 3.8% (voir tableau 5.1). Cette classe rassemble le même pourcentage de GHM chirurgicaux par rapport à l'échantillon des 257 GHM (27.6% contre 33%).

Le troisième groupe réunit 78 GHM qui ont pour la grande majorité des profils empiriques de type  $(h)$  (86%), on les a donc nommé les “**bons M**” (voir annexe Q). Ces GHM

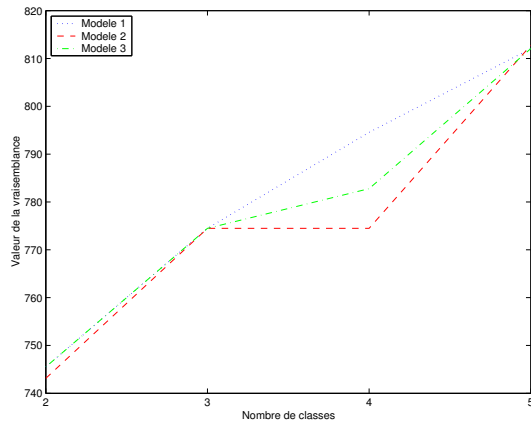


FIG. 5.3 – Valeurs de la vraisemblance pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses.

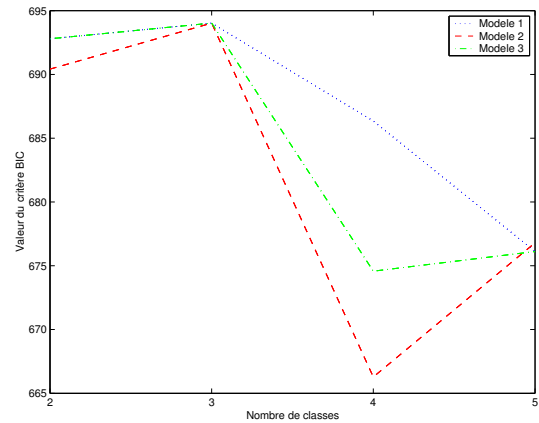


FIG. 5.4 – Valeurs du critère BIC pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses.

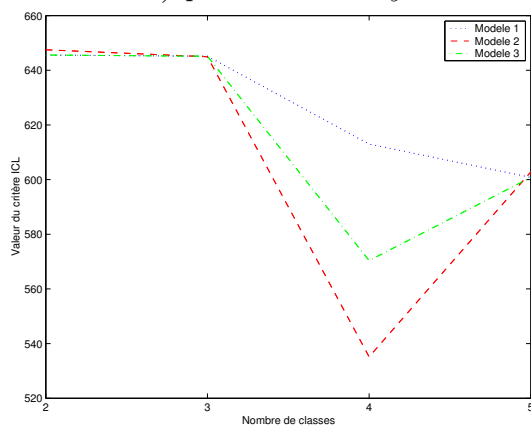


FIG. 5.5 – Valeurs du critère ICL pour les différents modèles de mélange (de 2 à 5 classes) pour les 3 analyses.

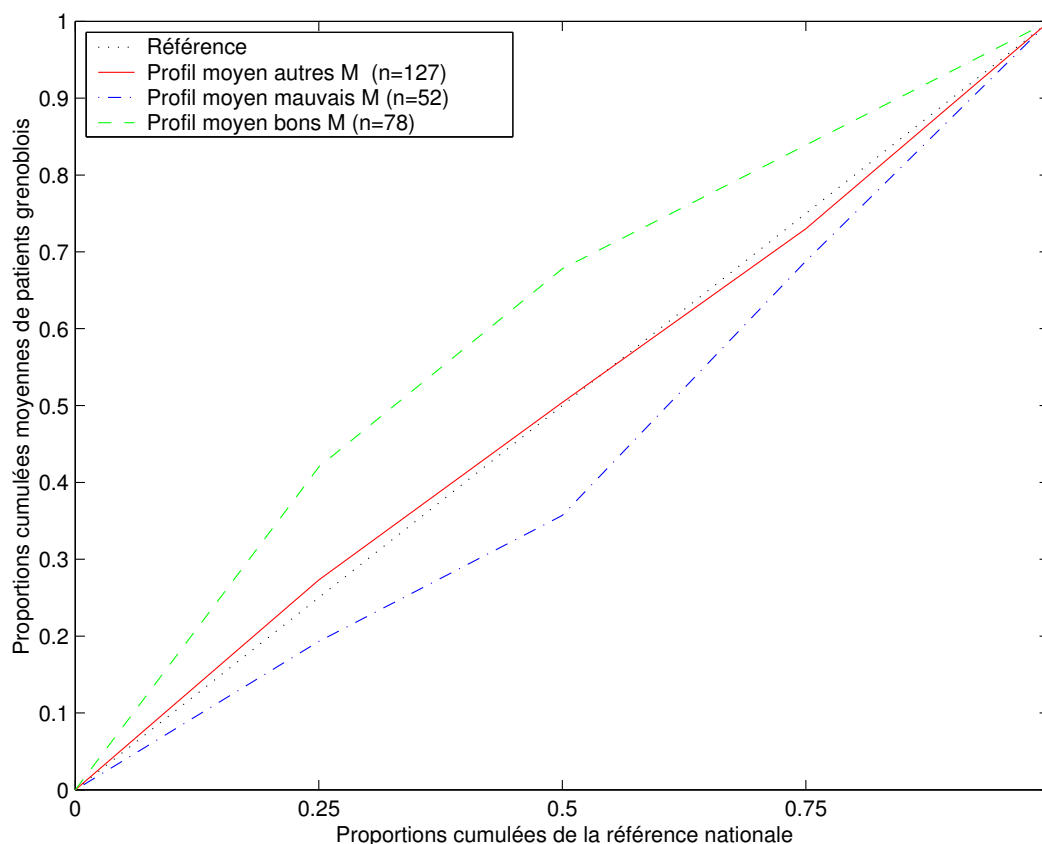


FIG. 5.6 – *Proportions cumulées moyennes dans chacun des 3 intervalles interquartiles pour chacun des groupes de GHM: “bons M”, “mauvais M” et “autres M”.*

sont performants quelques soient la longueur des séjours [profil en pointillés de couleur verte sur la figure 5.6 et groupe de points, étoiles de couleur verte, dans les figures 5.7, 5.8, 5.10 et 5.9]. Ils amènent un bénéfice moyen de 52585 +/- 69074 points ISA au CHU de Grenoble et la différence moyenne d’aire entre le profil grenoblois et la norme nationale est de 11.8% +/- 5.9% (voir tableau 5.1). Ce groupe compte sensiblement plus de GHM chirurgicaux par rapport à l’ensemble des 257 GHM, mais ce n’est pas significatif (39% contre 33%).

Si nous représentons cette classification en 3 groupes selon le bénéfice ou le déficit en points ISA et la différence d’aire entre le profil empirique grenoblois et la référence nationale, nous remarquons que les barycentres de chacun des groupes sont globalement alignés (voir tableau 5.11). Le barycentre correspondant aux groupes des “autres M” est légèrement décalé sur la gauche. Cela s’explique du fait que le groupe des “mauvais FG”, formant un ensemble homogène dans la classification issue de l’approche factorielle, a été englobé dans le groupe des “autres M”.

Maintenant que nous avons décrit les 3 groupes de la classification issue des modèles

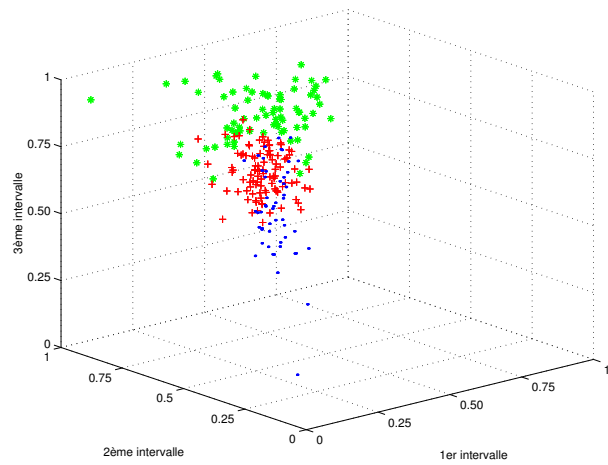


FIG. 5.7 – Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur les trois intervalles interquartiles nationaux.

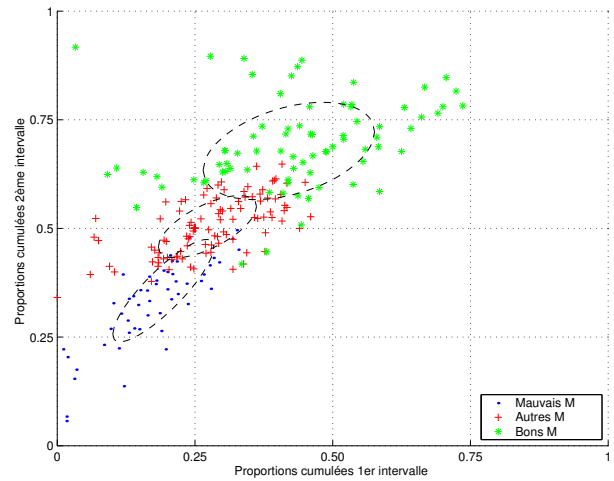


FIG. 5.8 – Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le premier et le deuxième intervalle interquartile national.

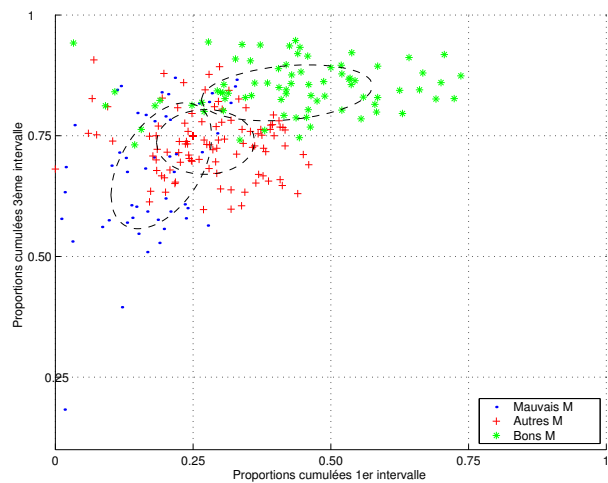


FIG. 5.9 – Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le premier et le troisième intervalle interquartile national.

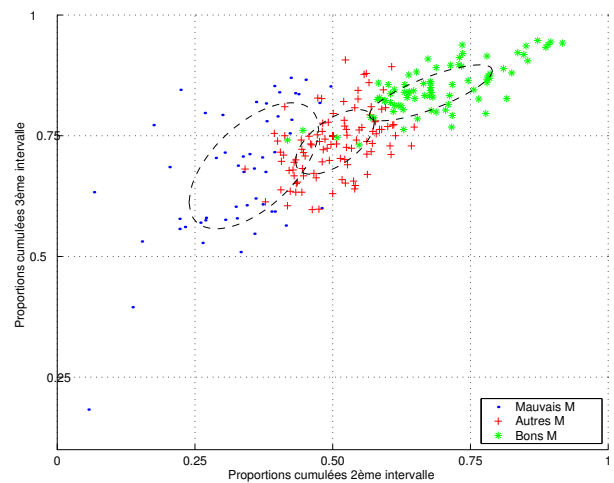


FIG. 5.10 – Représentation des groupes de GHM issus du modèle de mélange de 3 lois normales sur le deuxième et le troisième intervalle interquartile national.

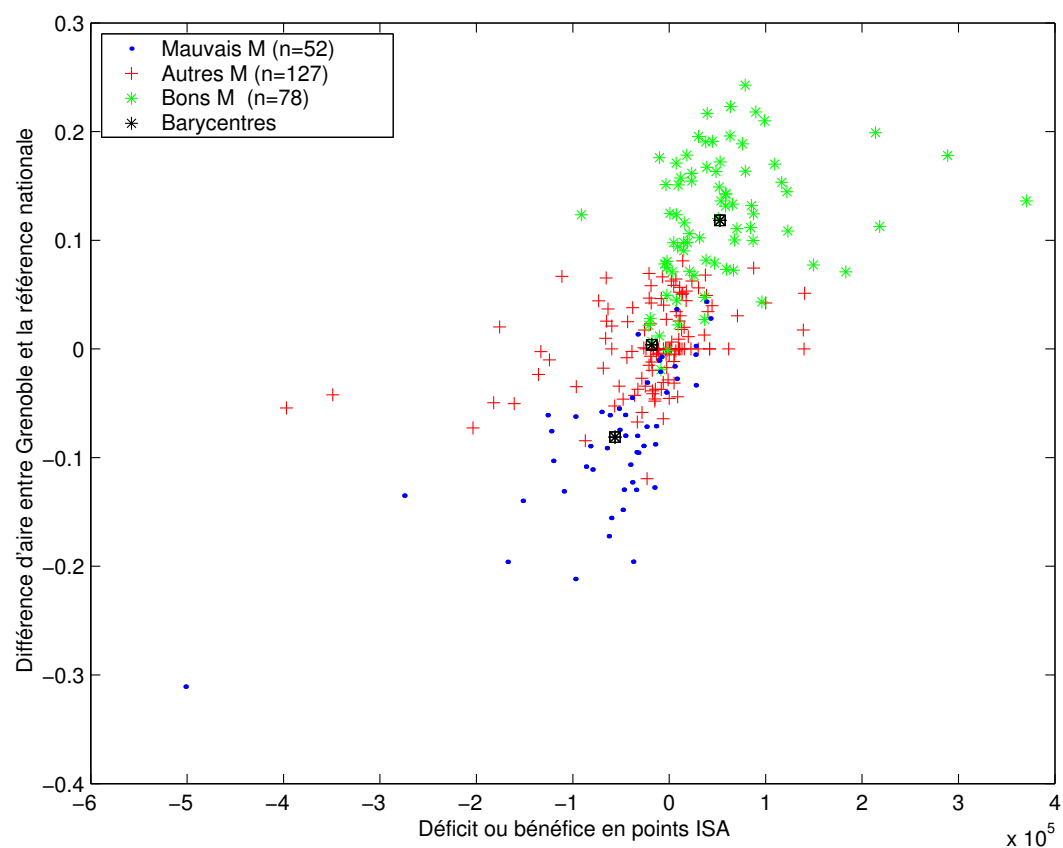


FIG. 5.11 – Représentation des déficit ou bénéfice en points ISA par rapport à la différence d'aire entre le profil grenoblois et la norme nationale pour la classification issue des modèles de mélange.

TAB. 5.1 – *Moyennes et écarts-types de la différence d'aire et du déficit en points ISA des groupes de GHM issus du modèle de mélange en 3 classes.*

GHM	Mauvais M $\bar{X}$ (E.T.)	Autres M $\bar{X}$ (E.T.)	Bons M $\bar{X}$ (E.T.)	p_valeurs
Différence d'aire (en %)	-8.1 (6.8)	0.4 (3.8)	11.8 (5.9)	<0.0001
Bénéfice ou déficit en points ISA	-56 150 (84 634)	-18 190 (69 931)	52 585(69 074)	
% GHM chirurgicaux	48.1	27.6	39.3	<0.0001

de mélange, nous allons la confronter aux 3 listes de référence pour voir si ce découpage est pertinent du point de vue médico-économique et à la classification issue de l'approche factorielle pour regarder si celle-ci est validée.

### 5.2.3 Confrontation de la classification avec les listes de référence

TAB. 5.2 – *Sensibilités et valeurs prédictives positives entre les listes de référence : CHU, quartile et  $\chi^2$ , et le groupe des "mauvais" GHM issu du modèle de mélange.*

	Liste CHU	Liste quartile	Liste $\chi^2$
Sensibilité	46%	78%	78%
Valeur prédictive positive	44%	40%	54%

Nous allons principalement nous intéresser au groupe des "mauvais M". Pour ce faire, nous avons regroupé les GHM "autres M" avec les "bons M" et nous les avons opposés au groupe des "mauvais M", notre centre d'intérêt.

En confrontant les GHM appartenant aux 3 groupes du modèle de mélange avec les 3 listes de référence : CHU, quartiles et  $\chi^2$ , nous voyons que le groupe des "mauvais M" (N=52) compte :

- 21 des 50 GHM de la liste CHU
- 21 des 27 GHM de la liste des quartiles,
- 28 des 36 GHM de la liste du  $\chi^2$ .

Les GHM appartenant à la liste des quartiles ou du  $\chi^2$  ont autant de risque de se trouver dans le groupe des "mauvais M" [78%] (voir tableau 5.2). Les GHM faisant partie de la liste CHU ont 46% de risque de se retrouver dans ce groupe.

Un GHM appartenant au groupe des "mauvais M" a 54% de risque de faire partie de la liste du  $\chi^2$ . La valeur prédictive positive pour les GHM appartenant à la liste CHU est égale

à 44 %.

Le groupe des “mauvais M” semble mieux repérer les GHM appartenant aux listes du  $\chi^2$  et du CHU. Cela paraît normal étant donné que ce sont les mêmes variables qui ont servi à construire le modèle de mélange et ces 2 listes de référence.

## 5.2.4 Confrontation de la classification avec celle issue l’approche factorielle

TAB. 5.3 – *Intersection entre les classes du modèle de mélange et celles issues de l’approche factorielle.*

Classes de l’approche factorielle	Classes du modèle de mélange			
	Autres M n(%)	Mauvais M n(%)	Bons M n(%)	Total ligne n(%)
Autres F	58(45.7)	29(55.8)	10(12.8)	97(37.7)
Mauvais FA	1(0.8)	23(44.2)	0(0.0)	24(9.3)
Mauvais FG	15(11.8)	0(0.0)	1(1.3)	16(6.2)
Bons F	53(41.7)	0(0.0)	67(85.9)	120(46.7)
Total colonne	127(100.0)	52(100.0)	78(100.0)	257(100.0)

Le modèle de mélange sépare les 257 profils des GHM en 3 groupes alors que la classification issue de l’approche factorielle les divise en 4 : le groupe des “mauvais FG” se retrouve englobé dans le groupe des “autres M”.

Le groupe des “mauvais M” compte exclusivement des GHM appartenant aux groupes des “autres F” (56%) et “mauvais FA” (44%), voir tableau 5.3.

Le groupe des GHM “autres M” compte près de la moitié de GHM appartenant au groupe des “autres F” (46%), 42% de “bons F” et 12% de “mauvais FG”.

Le groupe des “bons M” du modèle de mélange compte principalement des GHM appartenant au groupe des “bons F” (86%), mais également 13% de GHM appartenant au groupe des “autres F”.

De manière générale, les “bons F” ne sont jamais affectés dans le groupe des “mauvais M” et inversement. On constate donc un fort regroupement entre les “mauvais” GHM des 2 classifications. Cela renforce la pertinence de la classification issue de l’approche factorielle.

D’autre part, en comparant les caractéristiques économiques des 3 groupes de ces 2 classifications (le groupe des 16 “mauvais FG” étant mis à part), on s’aperçoit qu’il n’y a pas de différence significative sur le bénéfice ou le déficit en points ISA, voir tableau 5.4. Cela renforce encore plus la première classification.

TAB. 5.4 – *Comparaison des caractéristiques médico-économiques entre les 3 groupes des classifications issues de l’approche factorielle et des modèles de mélange.*

Groupes	Classification issue des modèles de mélanges (M)		Classification issue de l’approche factorielle (F)		p-valeurs
	N GHM	$\bar{X}(E.T)$	N	$\bar{X}(E.T)$	
“Mauvais M” ou “Mauvais FA”	52	-56150 (84634)	24	-96865 (104509)	0.075
GHM “Autres M” ou “Autres F”	127	-18190 (69931)	97	-28955 (70885)	0.090
“Bons” GHM	78	52585 (69074)	120	39288 (63311)	0.165

## 5.3 Discussion

### 5.3.1 Apport de la classification en trois groupes

Cette classification nous a permis d’isoler 3 groupes : les “bons M” (N=127), les “mauvais M” (N=52) et les “autres M” (N=78) , qui semblent être représentatif du CHU de Grenoble. Cette nouvelle classification confirme la cohérence de la classification issue de l’approche factorielle en quatre groupes. En effet, en utilisant une méthode complètement différente, on retrouve des classes équivalentes en terme de profils de distributions et de caractéristiques économiques. D’autre part, les “mauvais F” ne sont jamais affectés au groupe des “bons M” et inversement. On remarque donc un fort regroupement entre les GHM des groupes extrêmes des 2 classifications (les “bons M” avec les “bons F”, et les “mauvais M” ou les “mauvais FA”). Cela valorise encore la classification issue de l’approche factorielle. De plus, la classification issue des mélanges est en cohérence avec la classification empirique, le groupe des “mauvais M” regroupant principalement des GHM ayant un profil empirique de type  $(a)$ , mauvais sur toute la distribution des DS et celui des “bons M” ayant des profils empiriques de type  $(h)$  ou  $(neutre)$ , bons sur la totalité de la distribution. Les groupes des GHM “autres M” ou “autres F” correspondent au reliquat.

### 5.3.2 Groupe des “mauvais M”

Le groupe des “mauvais M” compte principalement des profils empiriques de type  $(a)$ , mais aussi de types empiriques  $(b)$  et  $(f)$ , respectivement 62%, 21% et 13%. Ces GHM ont tendance à être peu performants, quelle que soit la longueur du séjour ( $Ic_{(1)} < 0.25$ ,  $Ic_{(2)} < 0.50$  et  $Ic_{(3)} < 0.75$ ) Ce groupe compte un pourcentage plus important de GHM chirurgicaux par rapport à l’ensemble des 257 GHM.

Ce groupe est responsable d’un déficit de 2 919 802 points ISA au CHU de Grenoble, soit 42% du déficit total. Cela signifie que, par rapport à la base nationale, le CHU soigne l’équivalent de 1 836 patients de moins.

Le contrôle des dossiers médicaux des patients appartenant à ce groupe effectué pour tenter d’expliquer l’allongement des séjours, risque d’être difficile à organiser pour le CHU, comme pour le groupe des “mauvais FA”. En effet, la majorité de ces 52 GHM (60%) sont identifiés comme mauvais sur la totalité de leur distribution de DS (profil empirique de type



(a)). Une solution serait le contrôle exhaustif de tous les dossiers de ce groupe, mais c'est irréalisable, il y en a 13 169 ... Une alternative pourrait être un tirage aléatoire de 10% d'entre-eux, soit environ 1 317 dossiers, cependant ce nombre est encore important. Pour le diminuer un peu, on pourrait s'aider de la classification empirique qui nous donnerait une indication sur l'endroit des distributions de DS où les 40% de GHM restant sont mauvais. Comme pour le groupe des "mauvais FA", l'inconvénient de cette vérification est qu'elle comporte beaucoup de dossiers, elle n'est pas ciblée sur un sous-groupe de patients précis, mais réalisée à l'aveugle sur tous les patients de ce groupe. En conséquence, les caractéristiques des RSS tirés au sort n'expliqueront peut-être qu'une infime partie du déficit en points ISA. L'action qui pourrait en découler risque de ne pas être très rentable pour le CHU de Grenoble car les causes des problèmes révélés par l'analyse de 1 317 dossiers risquent d'être multiples et de ce fait difficiles à exploiter.

Un atout supplémentaire de la classification issue de l'approche factorielle est qu'elle isole un groupe homogène de 16 GHM "mauvais FG" (voir section 4.4.2), ce qui est beaucoup plus simple pour la mise en place d'une action visant à rapporter des points au CHU de Grenoble.

### 5.3.3 Limites de l'analyse

Ces résultats sont très dépendants de l'hypothèse que l'on a faite, à savoir : la normalité des 3 variables décrivant les profils, les conditions de départ de l'algorithme EM et son nombre d'itérations. Cependant, la classification est proche de celle issue de l'approche factorielle. L'hypothèse de normalité ne semble donc pas illogique.

### 5.3.4 Synthèse

La classification issue du modèle de mélange renforce la justesse de la classification issue de l'approche factorielle. En utilisant deux méthodes différentes, nous obtenons deux classifications similaires. On retrouve les mêmes résultats dans les groupes extrêmes (les "bons" et les "mauvais" GHM). Elle est cohérente avec la classification empirique. Elle semble séparer les profils en 3 groupes : 52 GHM identifiés comme "mauvais M", 78 comme "bons M" et 127 "autres M" correspondant au reliquat. Cependant, elle ne trouve pas de structure particulière aux profils identifiés comme "mauvais" par l'approche empirique, à l'inverse de la classification issue de l'approche factorielle qui identifiait 2 groupes de "mauvais F", les "mauvais FA" et les "mauvais FB".

## Chapitre 6

# Modélisation des DS du GHM n°18 : accidents vasculaires cérébraux non transitoires

Dans les chapitres précédents, nous avons proposé deux classifications des profils des 257 GHM grenoblois, qui nous ont permis d'identifier un certain nombre de GHM "atypiques". Nous nous intéressons, à travers ce chapitre, à l'un d'entre eux, le GHM n°18 : accidents vasculaires cérébraux non transitoires (AVC), afin d'analyser plus spécifiquement les causes de ce mauvais classement économique.

L'AVC touche entre 120 000 et 140 000 Français chaque année [56]. C'est la troisième cause de décès en France, après les cardiopathies ischémiques et les cancers, mais surtout la première cause de handicap acquis de l'adulte et la première cause de mortalité chez les femmes. C'est une maladie où les gens meurent immédiatement ou meurent de complications. Elle survient plutôt chez le sujet âgé. Les taux de mortalité à cinq ans sont élevés : de l'ordre de 50% pour les infarctus cérébraux et de 70% pour les hémorragies cérébrales [59]. Cependant ces taux de mortalité sont en train de baisser de manière phénoménale. Dans notre département, il est passé de 170 pour 100 000 à 70 pour 100 000 par an. Notre société évolue avec des sujets de plus en plus âgés qui sont relativement seuls dans la vie. Si ces personnes vivent en couple c'est avec quelqu'un du même âge. Ainsi, si un AVC survient, soit c'est un homme qui est parfois pris en charge par une femme qui est aussi âgée, soit c'est une femme dont le mari est déjà mort et dont les enfants travaillent. Cela implique donc des problèmes de prise en charge pouvant se répercuter sur la DS à l'hôpital. En effet, les séquelles de cette pathologie lourde subsistent à long terme lorsque l'AVC est non transitoire (GHM n°18). Elles exigent la prise en charge des personnes qui survivent à un AVC par toutes les structures du système de santé français (hôpitaux de court et moyen séjour, soins à domicile, établissements d'hébergement pour les personnes âgées, soins hospitaliers de long séjour, soins à domicile) [57]. Les AVC non transitoires concernent un peu moins de 100 000 personnes en France. Il est possible de considérer qu'en moyenne, 80% des AVC sont des infarctus cérébraux, 15% des hémorragies cérébrales, 5% des hémorragies méningées.

L’objectif de l’analyse des DS des patients atteints d’accidents vasculaires cérébraux (AVC) non transitoires est de trouver des éléments de réponse à la question suivante : pourquoi les patients grenoblois appartenant à ce GHM ont-ils des DS plus élevées que les patients de la base nationale, à pathologie égale ? Sont-elles dues à un problème de structure (peu de places en établissements hospitaliers de long séjour,...), à un recrutement de patients plus complexes à Grenoble (proportion importante d’“outliers”), à des variations de pratique médicale, à des erreurs de classifications ?

L’utilisation de modèles de mélange permet de le faire. Ces modèles ont l’avantage de prendre en compte l’hétérogénéité de la distribution des DS (présence d’“outliers”), mais aussi de les repérer. Dans cet objectif, le professeur Quantin et ses collaborateurs les ont appliqués aux distributions de DS et de coûts de quelques GHM [52, 72, 73], et ont mis en place une méthode intitulé “Justification clinique des séjours coûteux d’un GHM” dont nous nous sommes inspirés (voir section 1.4.1) [41]. Au Royaume-Uni, le Professeur Millard et son équipe se sont aussi beaucoup intéressés aux modèles de mélanges, pour comprendre comment se structuraient les DS dans les services de gériatrie [63].

Pour ce faire, nous allons décrire la population étudiée et les variables d’intérêt pour l’analyse. Puis, nous donnerons la stratégie de modélisation que nous avons employée. Nous exposerons ensuite les différentes méthodes statistiques utilisées pour modéliser les distributions de DS des bases nationale et grenobloise de ce GHM et les expliquer. Pour finir, nous exposerons les résultats de nos analyses et nous en discuterons la pertinence d’un point de vue médico-économique et méthodologique.

## 6.1 Population et méthodes

### 6.1.1 Population et variables d’intérêt

Nous avons choisi ce GHM car il représente une pathologie traçante, est cohérent du point de vue médical et coûte cher au CHU de Grenoble en termes de points ISA (déficit de 349 030 points ISA, 3<sup>e</sup> au classement de la liste CHU, voir annexe F). Ce GHM a un profil empirique de type  $(a)$ , significativement différent de la référence nationale ( $p = 0.004$ ). Il fait aussi partie de la liste fondée sur le  $\chi^2$  (voir annexe H).

Notre échantillon est composé de 617 RSA pour le CHU de Grenoble et de 950 pour la base nationale.

Les variables que nous avons retenues pour l’analyse sont une partie de celles contenues dans les RSA (voir annexe C) :

variable à expliquer : la DS du patient (en jours),

variables explicatives :

- le nombre de RUM,
- l’âge du patient,

- le sexe du patient,
- le diagnostic principal regroupé en classes :
  - \* hémorragies,
  - \* infarctus,
  - \* malformation,
  - \* non précisé ;
- le nombre d’actes,
- le nombre de diagnostics associés ;
- le fait d’avoir eu ou non un acte classant et sa valeur moyenne ICR,
- le mode d’entrée :
  - \* entrée directe,
  - \* transfert,
  - \* transfert long et moyen séjour ;
- le mode de sortie :
  - \* transfert,
  - \* transfert long et moyen séjour,
  - \* domicile,
  - \* décès.

### 6.1.2 Démarche suivie

Nous avons commencé par décrire la distribution de DS du GHM n°18 (minimum, quartiles, moyennes, maximum, type de profil) pour les bases nationale et grenobloise. Ensuite, nous avons cherché la forme globale de la distribution des DS de la base nationale. Pour ce faire, nous leur avons ajusté des lois univariées normales, log-normales, exponentielles, gammas et de Weibull. Les paramètres de ces lois ont été estimés par la méthode du maximum de vraisemblance. Nous avons mesuré la qualité des ajustements à ces lois avec le test de Kolmogorov-Smirnov (KS) au seuil  $\alpha = 5\%$  [79].

#### Description de notre hypothèse de travail

Nous sommes partis de l’hypothèse générale suivante, applicable à tous les GHM : la distribution des DS pour un  $\text{GHM}_i$  de la base nationale suit une loi univariée. En revanche, le  $\text{GHM}_i$  grenoblois, “jumeau” de celui de la base nationale, ne suit pas une loi univariée (présence de sous-groupes de patients, synonyme d’hétérogénéité). Nous allons donc estimer sur les DS du  $\text{GHM}_i$  grenoblois un modèle de mélange de 2 ou 3 lois du même type que celle qui ajuste la distribution du  $\text{GHM}_i$  de la base nationale. Nous aurons comme seule contrainte que les paramètres de l’un des groupes du modèle de mélange soient les mêmes que ceux estimés sur la distribution de DS du  $\text{GHM}_i$  de la base nationale. L’hypothèse de travail est qu’à Grenoble, nous devrions retrouver au moins la population standard (celle de

la base des CHU français).

Si cela se vérifie, nous avons donc affaire à deux populations différentes pour le GHM<sub>*i*</sub> grenoblois; nous obtenons alors :

- une certaine proportion  $P_N$  d’une population que nous pouvons nommer  $N$ . Elle a des paramètres proches de ou identiques à ceux estimés sur la distribution des DS du GHM<sub>*i*</sub> “jumeau” de la base nationale. Cette population de patients est donc la même en terme de DS que celle de la base française.
- La proportion complémentaire  $(1 - P_N)$  d’une autre population que nous pouvons appeler  $\bar{N}$ . Elle a des paramètres différents de ceux estimés sur la distribution des DS du GHM<sub>*i*</sub> “jumeau” de la base nationale. Ce groupe peut correspondre aux patients inhabituels qui ne se trouvent pas dans de telles proportions dans la BDD nationale. L’allongement ou le raccourcissement de la DS pour ce GHM peut alors peut-être s’expliquer par un recrutement de patients plus compliqués à Grenoble par rapport aux CHU de la base nationale.

Cependant, deux cas peuvent se présenter :

- **Cas 1** : une des 5 lois unimodales ajuste correctement les distributions des DS d’un GHM de la base nationale (résultat du test de KS). Nous testerons alors, sur la distribution des DS de son homologue grenoblois, l’adéquation avec la loi de même famille (paramètres estimés de manière empirique) :
  - \* **Cas n°1A** : si ce type de loi est acceptable pour le GHM grenoblois (résultat du test de KS), on peut penser que l’allongement des DS de ce GHM peut s’expliquer par une variation de la pratique médicale à Grenoble par rapport aux autres CHU nationaux.
  - \* **Cas n°1B** : si ce type de loi n’est pas acceptable, nous testerons notre hypothèse de travail : “nous pouvons avoir affaire, pour le GHM grenoblois, à différents groupes de patients en termes de distribution de DS”. Nous utiliserons alors un modèle de mélange de la loi qui ajuste le mieux les DS de la base nationale pour identifier ces groupes de patients. Nous initialiserons les paramètres d’une des deux classes du modèle avec les mêmes que ceux de la base nationale.
- **Cas n°2** : aucune des 5 lois unimodales n’ajuste la distribution des DS d’un GHM de la base nationale et grenobloise ; nous faisons à nouveau la même hypothèse : présence de sous-groupes de patients pour les GHM concernés des bases nationale et grenobloise. Nous choisirons alors un modèle de mélange de 2 ou 3 lois log-normales. C’est à dire que nous allons estimer un modèle de mélange de lois normales sur le logarithme des DS de la base nationale. Nous ferons de même pour leurs homologues grenoblois, en initialisant les paramètres des modèles de mélange avec ceux de la base nationale.

Nous avons appliqué cette méthodologie aux DS du GHM n°18. Pour estimer les paramètres d’un modèle de mélange de lois normales sur le logarithme des DS des GHM (voir

section 5.1.1), nous avons utilisé le logiciel *XEM gauss* [27].

L'estimation des paramètres des modèles a été faite avec l'algorithme EM dans les mêmes conditions que pour les mélanges utilisés pour la classification des profils (voir paragraphe 5.1.3). Le choix du nombre de composants des modèles (2 ou 3 lois) s'est fait de façon pragmatique, en choisissant le modèle le plus pertinent du point de vue de la signification médico-sociale, mais aussi sur la base des critères BIC [80] et ICL [28] afin de déterminer si le nombre de composants du modèle choisi de manière pragmatique était proche de celui "élu" par ces 2 critères.

### 6.1.3 Interprétation du modèle de mélange

Pour interpréter les différentes composantes des modèles, nous avons considéré deux analyses différentes, qui utilisent les probabilités conditionnelles  $t_k(\mathbf{x}_i; \boldsymbol{\theta})$  d'appartenance aux composantes  $k$  des modèles (voir équation (N.1)).

La première méthode nous a permis de savoir de quel composante du modèle se rapprochait le plus un patient ayant une certaine caractéristique médico-sociale [39]. La deuxième méthode nous a servi à décrire les caractéristiques médico-sociales de chacun des sous-groupes de patients et, par conséquent à analyser les causes de l'allongement des DS du GHM étudié.

Nous avons réalisé ces analyses pour chacune des variables d'intérêt du GHM étudié pour les bases nationale et grenobloise.

#### Première analyse

Pour savoir de quel sous-groupe se rapprochait le plus un patient ayant une certaine caractéristique  $j$ , nous avons appliqué la méthode suivante.

Considérons une variable qualitative  $\mathbf{S}$  à deux modalités représentant, par exemple, le sexe du patient. Nous notons  $\mathbf{s}_i$  le sexe du patient  $i$ . Nous choisissons pour homme " $\mathbf{s}_i = 0$ " et pour femme " $\mathbf{s}_i = 1$ ". Nous voulons savoir de quel composante  $k$  ( $k = 1, \dots, K$ ) du modèle les hommes et les femmes auraient plus le profil. Pour ce faire, nous avons calculé la probabilité moyenne d'appartenance,  $P_{kj}$ , à l'une des composantes  $k$  du mélange pour la modalité  $j$ ,  $j = (0,1)$  de la variable qualitative  $\mathbf{S}$ ,

$$P_{kj} = \frac{\sum_{i=1}^n t_k(\mathbf{x}_i; \boldsymbol{\theta}) \delta\{\mathbf{s}_i = j\}}{\sum_{i=1}^n \delta\{\mathbf{s}_i = j\}},$$

où  
 $\delta\{\mathbf{s}_i = j\} = 1$  si  $\mathbf{s}_i = j$  et 0 sinon, et  $n$  est le nombre de RSA dans l'échantillon.

Une fois la probabilité moyenne  $P_{kj}$  calculée, nous la comparons avec la proportion  $P_k$  estimée dans le modèle de mélange de référence.

Par exemple, considérons un mélange à 2 composantes, tel que la proportion de la première composante  $P_1 < 0.5$ . De manière intuitive, si :

- $P_{1j} > P_1$   
 $\Rightarrow$  la modalité  $j$  de la variable  $\mathbf{S}$  a plus de chance d'appartenir à la première composante du modèle qu'à la seconde,
- $P_{1j} < P_1$   
 $\Rightarrow$  la modalité  $j$  de la variable  $\mathbf{S}$  a plus de chance d'appartenir à la deuxième composante du modèle qu'à la première,
- $P_{1j} \approx P_1$   
 $\Rightarrow$  la modalité  $j$  de la variable  $\mathbf{S}$  a autant de chance d'appartenir à l'une ou l'autre des composantes du modèle.

Nous avons réalisé ce calcul pour chacune des modalités des variables qualitatives et dans les bases grenobloise et nationale.

Ensuite, pour regarder quelles sont les variables étudiées qui diffèrent de manière significative entre les composantes  $k$  des modèles pour la base nationale et grenobloise, nous avons utilisé un test du  $\chi^2$  (effectifs attendus supérieurs à 5) ou de Fisher (effectifs attendus inférieurs à 5), les variables étant qualitatives. Nous avons choisi un seuil  $\alpha = 5\%$ .

## Deuxième analyse

Pour comprendre quelles étaient les causes de l'allongement de la DS des GHM grenoblois par rapport à la base nationale : variation des pratiques médicales, problème de structure ou recrutement de patients plus complexes, nous avons décrit les caractéristiques médico-sociales des patients de chaque sous-groupe en utilisant les variables explicatives que nous avons sélectionnées.

Considérons de nouveau la variable qualitative  $\mathbf{S}$ . Nous pouvons estimer le pourcentage  $PO_{kj}$ , ( $j = 0,1$ ), d'hommes et de femmes dans chacune des composantes  $k$  du modèle ( $k = 1, \dots, K$ ),

$$PO_{kj} = \frac{\sum_{i=1}^n t_k(\mathbf{x}_i; \boldsymbol{\theta}) \delta\{\mathbf{s}_i = j\}}{\sum_{i=1}^n t_k(\mathbf{x}_i; \boldsymbol{\theta})}.$$

Pour une variable quantitative  $\mathbf{D}$ , par exemple le nombre de diagnostics associés d'un patient, nous notons  $\mathbf{d}_i$  sa réalisation pour le patient  $i$ . Nous voulons regarder si sa moyenne  $\bar{\mathbf{d}}_k$  est différente selon les composantes  $k$  du modèle avec  $k = 1, \dots, K$ . Pour ce faire, nous avons calculé pour chacune composante  $k$  :

$$\bar{\mathbf{d}}_k = \frac{\sum_{i=1}^n t_k(\mathbf{x}_i; \boldsymbol{\theta}) \mathbf{d}_i}{\sum_{i=1}^n t_k(\mathbf{x}_i; \boldsymbol{\theta})}.$$

Ensuite, pour trouver quelles sont les variables étudiées qui diffèrent de manière significative entre les composantes  $k$  des bases nationale et grenobloise, à indice  $k$  égal, nous avons utilisé un test du  $\chi^2$  (effectifs attendus supérieurs à 5) ou de Fisher (effectifs attendus inférieurs à 5) si la variable était qualitative, et un test de Student si la variable était quantitative. Nous avons choisi un seuil  $\alpha = 5\%$ .

### Variables explicatives de la DS selon les composantes

Pour identifier de manière précise les variables explicatives des DS de chaque composante  $k$  du modèle dans une base donnée (nationale ou grenobloise), nous avons ajusté un modèle multivarié de régression log-normal sur les DS de chacune des composantes  $k$ , où chaque patient est muni de son poids d'appartenance  $t_k(\mathbf{x}_i; \boldsymbol{\theta})$  à la composante  $k$ . Les modèles ont été ajustés sur les variables sexe et âge.

Nous avons introduit dans les modèles toutes les variables du PMSI que nous avons sélectionnées pour cette étude. Pour sélectionner les variables explicatives du modèle multivarié, la méthode pas à pas descendante a été utilisée. Nous sommes partis du modèle le plus complet, c'est-à-dire du modèle contenant toutes les variables sélectionnées. Nous avons ensuite retiré successivement du modèle multivarié toutes les variables qui n'étaient pas significatives (test de Student au seuil de 5%), sans tenir compte des interactions. Le choix du modèle multivarié définitif s'est fait sur la base de la statistique du rapport de vraisemblance au seuil  $\alpha$  de 5%, qui teste l'adéquation des modèles emboîtés. À partir de ce dernier modèle, toutes les interactions deux à deux des variables qualitatives ont été testées avec le test du rapport de vraisemblance au seuil  $\alpha$  de 5%.



TAB. 6.1 – Description de la distribution des DS des bases grenobloise et nationale.

N° GHM	Base	Minimum	$q(0.25)$	$q(0.50)$	$X$	$q(0.75)$	Maximum
18	grenobloise	1	5	10	14.7	20	110
18	nationale	1	3	9	11.6	15	92

## 6.2 Résultats

### 6.2.1 Description des distributions des DS

On constate un allongement des DS des patients grenoblois ayant eu un AVC non transitoire par rapport à la référence nationale, quelle que soit la longueur du séjour (voir tableau 6.1). En effet, quel que soit l'indicateur que l'on regarde (quartiles ou DS moyenne), on remarque qu'il est supérieur à la base nationale.

### 6.2.2 Description médico-sociale

Les patients grenoblois semblent être comparables à ceux de la base nationale par leur sexe, leur âge, la présence ou non d'un acte classant et le nombre d'ICR pour les actes classants. Cependant, le nombre de RUM est plus important à Grenoble (1.8 +/- 0.7 contre 1.3 +/- 0.6,  $p < 0.0001$ ), ainsi les patients grenoblois semblent passer en moyenne dans deux unités médicales lors de leur hospitalisation. Il y a plus d'AVC ayant comme origine un infarctus au CHU Grenoble que sur la base nationale (40.4% contre 30.8%,  $p < 0.0001$ ). De même, les causes de l'AVC sont plus souvent inconnues à Grenoble que sur la base nationale (39.2% contre 26.2%,  $p < 0.0001$ ). Le nombre de diagnostics associés est plus important au CHU de Grenoble (2.7 +/- 2.5 contre 1.6 +/- 1.8,  $p < 0.0001$ ), par contre le nombre d'actes est inférieur (0.2 +/- 0.9 contre 1.6 +/- 2.3,  $p < 0.0001$ ). Les patients arrivent plus rarement d'un autre service quand ils ont eu un AVC au CHU de Grenoble (5.4% contre 11.8%,  $p < 0.0001$ ). Près de 95% des patients grenoblois arrivent directement au CHU, contre 87.6% sur la base nationale. Près d'un quart des patients grenoblois et nationaux sont transférés dans un autre service après avoir fait un AVC, respectivement 24.3% et 27.7%, et près de 11% d'entre eux décèdent. Plus de patients grenoblois retournent chez eux après un AVC que sur la base nationale (63.7% contre 57.2%,  $p < 0.0001$ ).

Pour chacune des deux bases, les femmes ayant un AVC non transitoire sont plus âgées que pour les hommes (71 ans contre 64 ans pour Grenoble et 69 contre 64 pour la base interchu,  $p < 0.0001$ ), cette variation étant de même ordre entre les 2 bases.

### 6.2.3 Modélisation des distributions de DS

#### Lois de probabilités ajustant les DS nationales

D'après les résultats du test de KS à 5%, aucune des lois unimodales (normale, log-normale, exponentielle, gamma ou Weibull) ne semble ajuster les DS du GHM n°18 de la base nationale,  $p < 0.0001$  (voir tableaux 6.3, 6.4, 6.5, 6.6 et graphiques 6.2 et 6.2)

Nous faisons donc l'hypothèse que les distributions du logarithme des DS des bases grenobloise et nationale suivent un mélange de 2 ou 3 lois normales.

#### Modélisation des DS par des mélanges de lois

À travers ce paragraphe, nous nous intéressons aux résultats de la comparaison des modèles de mélange obtenus sur le logarithme des DS de la base nationale avec ceux estimés sur le logarithme des DS de la base grenobloise.

Quelle que soit la base (nationale ou grenobloise), les valeurs des critères BIC et ICL sont maximums pour 2 classes (voir tableau 6.8).

Ainsi, les DS des patients hospitalisés suite à des AVC non transitoires se répartissent comme suit :

- 84.1% des patients grenoblois sont hospitalisés en moyenne durant 15 +/- 11 jours, et 15.9% ne restent que 1 ou 2 jours.
- 74.7% patients de la base nationale demeurent en moyenne 13 +/- 6 jours dans les CHU, et 25.3% ne restent que 1 ou 2 jours.

Le groupe de patients grenoblois restant hospitalisés 2 semaines en moyenne est plus important que sur la base nationale (84% contre 75%), de plus ils demeurent en moyenne 2 jours de plus que les patients faisant partie des autres CHU nationaux. Les patients qui restent 1 ou 2 jours sont moins nombreux à Grenoble que sur la base nationale (16% contre 25%).

Il ne semble donc pas y avoir la présence un groupe supplémentaire d'“outlier” à Grenoble pour ce GHM; en revanche, il est intéressant de constater que le modèle de mélange isole comme deuxième composante un groupe notable de patients restant 1 ou 2 jours, que ce soit dans la base nationale ou grenobloise. Plusieurs questions découlent de cette classification en 2 groupes :

- Pourquoi les patients de Grenoble sont-ils plus nombreux à rester hospitalisés 2 semaines?
- Est-ce dû à un problème de fonctionnement des services ou à un recrutement différent des patients à Grenoble?
- Pourquoi ces patients restent-ils en moyenne 2 jours de plus que ceux de la base nationale?
- Ces patients sont-ils plus complexes que ceux de la base nationale?

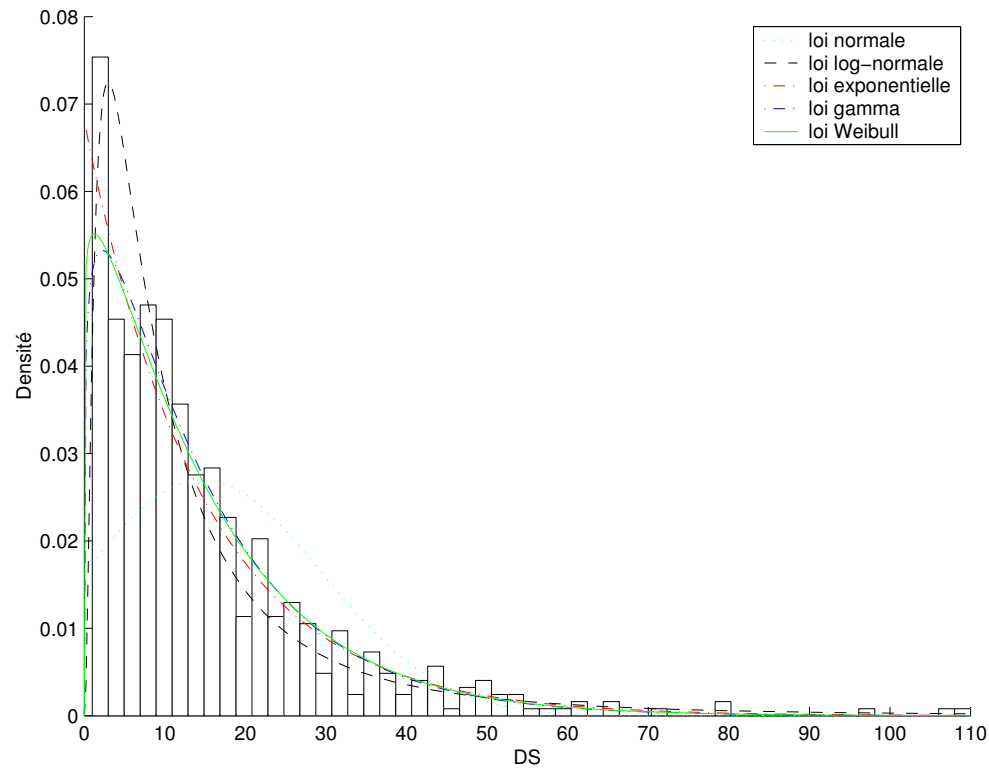


FIG. 6.1 – Ajustement de la distribution des DS du GHM n° 18 de la base grenobloise par différentes lois.

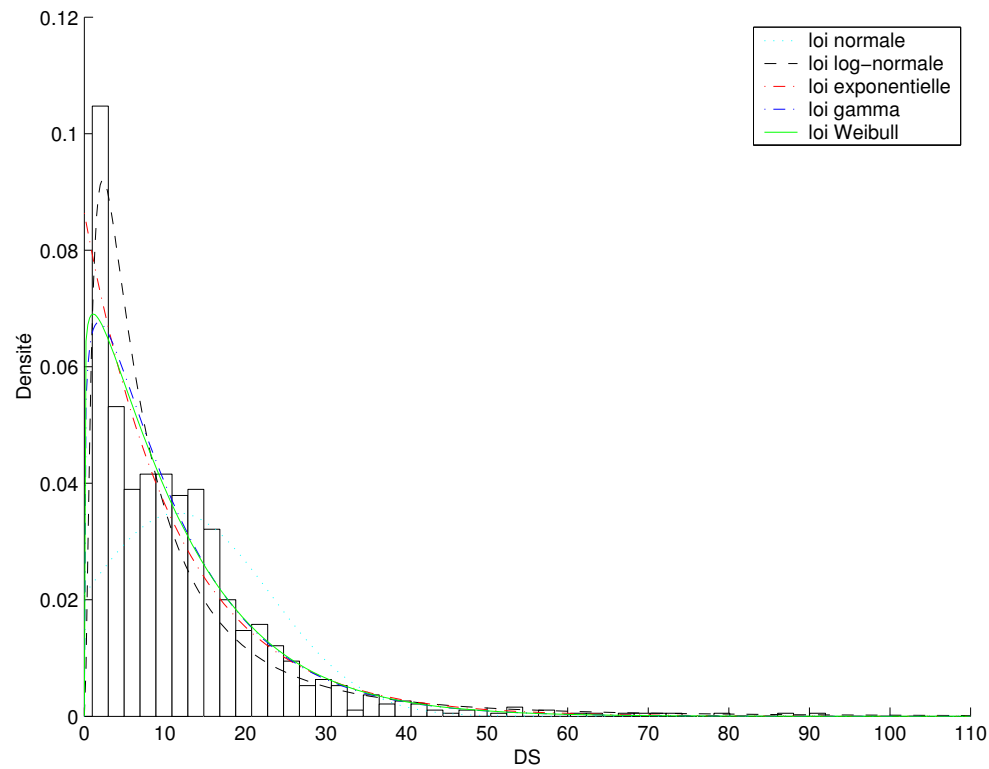


FIG. 6.2 – Ajustement de la distribution des DS du GHM n° 18 de la base nationale par différentes lois.

- Est-ce dû à des erreurs de classification?
- Pourquoi aussi peu de patients restent 1 ou 2 jours à Grenoble?
- Est-ce dû au fait que plus décèdent ou qu'ils sont transférés?

Nous avons essayé de répondre à ces questions en décrivant les caractéristiques médico-sociales de chacun des sous-groupes de patients des bases nationale et grenobloise.

## 6.2.4 Description médico-sociale des composantes

### Première analyse

Les variables qui diffèrent significativement entre les composantes des modèles de mélange sont :

- pour la base grenobloise, le nombre de RUM ( $p < 0.0001$ ), le type de diagnostic ( $p < 0.0001$ ), le nombre de diagnostic associé ( $p < 0.0001$ ) et le mode de sortie ( $p = 0.001$ ),
- pour la base nationale, le nombre de RUM ( $p < 0.0001$ ), le type de diagnostic ( $p < 0.0001$ ), le nombre de diagnostic associé ( $p < 0.0001$ ), le nombre d'actes ( $p = 0.003$ ), le mode d'entrée ( $p = 0.017$ ) et le mode de sortie ( $p = 0.001$ ).

Les modalités distinguant plutôt les patients appartenant plutôt à la première composante du modèle de mélange de la base grenobloise (DS moyenne de 15 jours) sont le fait d'être passé dans plusieurs unités médicales et d'avoir plus de deux diagnostics associés, voir tableau 6.10.

Les modalités des variables décrivant plutôt les patients faisant plutôt partie de la deuxième composante du modèle de mélange de la base grenobloise (DS de un ou deux jours) sont (voir tableau 6.10) :

- un séjour dans une seule unité médicale,
- un diagnostic principal d'hémorragie ou de malformation,
- aucun diagnostic associé,
- un décès.

Les caractéristiques décrivant plutôt les patients appartenant plutôt à la première composante du modèle de mélange de la base nationale (DS moyenne 13 jours) sont (voir tableau 6.11) :

- un séjour dans plusieurs unités médicales,
- un diagnostic principal d'infarctus,
- deux diagnostics associés ou plus,
- aucun acte associé,
- un mode d'entrée en transfert interne,
- un mode de sortie en transfert interne.

Les caractéristiques des patients faisant plutôt partie de la deuxième composante du modèle de mélange de la base nationale (DS de un ou deux jours) sont (voir tableau 6.11) :

- un séjour dans une seule unité médicale,

- un diagnostic principal d’hémorragie ou de malformation,
- aucun diagnostic associé,
- un mode d’entrée en transfert de long et moyen séjour,
- un décès.

## Deuxième analyse

Les patients grenoblois restent 15 jours en moyenne au CHU (première composante) différent de ceux appartenant à la base nationale, qui restent 13 jours en moyenne, par (voir tableau 6.12) :

- le type des diagnostics principaux ( $p < 0.0001$ ),
- le mode d’entrée ( $p < 0.0001$ ),
- le mode de sortie ( $p < 0.0001$ ).

Ils ont moins fréquemment un diagnostic non précisé ou peu de diagnostic d’infarctus. Ils entrent souvent directement au CHU (94.2% contre 86.0%) et retournent la plupart du temps directement au domicile (64.3% contre 57.7%).

D’après le tableau 6.12, les variables décrivant les patients grenoblois restant 1 ou 2 jours au CHU (deuxième composante du modèle) qui diffèrent significativement de celles de la base nationale restant la même durée sont les suivantes :

- le sexe ( $p = 0.009$ ),
- le type des diagnostics principaux ( $p < 0.0001$ ),
- le mode d’entrée ( $p < 0.0001$ ),
- le mode de sortie ( $p < 0.0001$ ).

En effet, nous constatons une majorité d’hommes hospitalisé au CHU de Grenoble (51.4% contre 43.9%). Ils ont plus fréquemment un diagnostic d’infarctus (26.7% contre 21.4%) et moins de diagnostic non précisé (25.5% contre 29.2%). Ils entrent moins souvent au CHU directement (92.3% contre 96.9%), mais plus par transfert interne (6.9% contre 3.1%). Ils semblent décéder moins fréquemment (18.2% contre 22.8%) et sont transférés davantage en long et moyen séjour (21% contre 17%).

## Variables explicatives de la DS selon les composantes

Le modèle de régression log-normal qui ajustait le mieux les DS de la première composante du mélange de la base grenobloise (DS moyenne de 15 jours) est présenté dans le tableau 6.13. Plus l’âge augmente, plus le patient séjourne dans des unités médicales différentes, et plus sa DS aura tendance à être longue. Les patients ayant fait un infarctus ont tendance à rester moins longtemps que ceux ayant fait une hémorragie. On constate l’inverse chez les patients ayant un diagnostic de malformation. Nous avons testé l’interaction entre l’âge et le sexe des patients, mais celle-ci n’est pas significative.

Aucune des variables que nous avons retenues dans le PMSI n'expliquait les DS de la deuxième composante du mélange du CHU de Grenoble.

Les variables qui semblent expliquer les DS de la première composante du mélange (DS moyenne de 13 jours) de la base nationale sont (voir tableau 6.14) : l'âge, le diagnostic, le nombre d'unités médicales où le patient a séjourné, et le nombre de diagnostics associés. Plus un patient est âgé, plus il a séjourné dans des unités médicales différentes, plus il a eu d'actes associés et plus sa DS aura tendance à être longue. Le fait d'avoir un diagnostic de malformation a tendance à diminuer la DS moyenne par rapport au fait d'avoir un diagnostic d'hémorragie. Nous avons testé l'interaction entre l'âge et le sexe des patients, mais celle-ci n'est pas significative.

D'après le tableau 6.15, les DS des patients nationaux appartenant à la deuxième composante du modèle ont tendance à augmenter avec l'âge et le nombre d'unités médicales où le patient a séjourné. Cependant, les DS ont tendance à diminuer si le patient est transféré dans un autre service ou dans des structures de moyen et long séjours, par rapport aux patients qui rentrent au domicile. Comme pour la première composante, l'interaction entre l'âge et le sexe des patients n'est pas significative.

## 6.3 Discussion

### 6.3.1 Principaux résultats

#### Structure des distribution de DS en deux groupes

Les DS des patients ayant fait un AVC non transitoire semblent se structurer en deux groupes, quelle que soit la base (nationale ou grenobloise). Le groupe de patients grenoblois restant 2 semaines en moyenne au CHU est plus important que dans la base nationale (84% contre 75%) ; ils passent en moyenne 2 jours de plus que les patients de la base nationale. Les patients qui restent 1 ou 2 jours (première composante du modèle) sont moins nombreux à Grenoble (16% contre 25%). Il ne semble donc pas y avoir présence d'un groupe supplémentaire d'"outliers" à Grenoble pour ce GHM, mais seulement de deux groupes de patients, l'un avec de faibles et l'autre avec de fortes DS.

D'après les tableaux 6.10 et 6.11), les patients qui ont la plus grande probabilité d'appartenir à la deuxième composante des modèles (1 ou 2 jours de DS moyenne), quelle que soit la base (nationale ou grenobloise), sont plutôt ceux qui ont eu un diagnostic principal de malformation ou d'hémorragie et ceux qui décèdent (près d'un quart d'entre-eux, voir tableau 6.12), ces deux variables étant très liées. Ce sont des patients qui sont décéder de l'AVC et seulement de cela, ce qui explique le fait qu'ils n'aient pas de diagnostics associés, et qu'ils ne séjournent que dans une seule unité médicale. Cependant, un nombre important de ces patients rentrent directement au domicile (plus de la moitié, voir tableau 6.12). Cela est difficilement explicable avec les données sur lesquelles nous avons travaillé. La solution

idéale serait d'étudier en détail quelques RSS de ses patients afin de connaître leurs DGA et leurs actes associés,...

D'autre part, Grenoble a une proportion de patients deux fois moins importante dans ce groupe, qui s'explique par le fait que le CHU de Grenoble est généraliste (un seul établissement), alors que dans les villes importantes, il y a plusieurs hôpitaux spécialisés (par exemple, les Hospices Civils de Lyon). Ainsi, un patient étant admis en urgence dans l'un d'entre eux pourra être transféré vers l'hôpital spécialisé qui traite les AVC (par exemple, l'hôpital de neurologie à Lyon). Cela est confirmé par le modèle log-normal : la DS a tendance à diminuer si le patient est transféré vers un autre hôpital ou vers une structure de moyen ou de long séjour (voir tableau 6.15).

Les patients ayant la plus grande probabilité d'appartenir à la première composante du mélange (15 ou 13 jours en moyenne, respectivement sur la base grenobloise et nationale), sont plutôt ceux qui ont eu un diagnostic d'infarctus, la présence de plusieurs diagnostics associés et des séjours dans plusieurs unités médicales, ces deux dernières variables étant liées (voir tableaux 6.10 et 6.11). Ces caractéristiques s'expliquent aussi par le fait que le CHU de Grenoble est un établissement généraliste. Ainsi, si un patient est admis dans le service de neurologie générale et que les médecins diagnostiquent un AVC, il pourra se voir transféré dans l'unité spécialisée pour les AVC du CHU. On remarque aussi qu'à Grenoble, il y a un pourcentage important de diagnostics non précisés (41%, voir tableau 6.12). Cela peut être synonyme d'une démarche diagnostique plus longue que la moyenne, et donc d'un allongement de la DS qui s'avère problématique du point de vue économique. Il faut savoir aussi qu'il est possible de faire des soins de bonne qualité dans les 10-12 jours suivant l'AVC. Ainsi, toutes les DS qui dépassent 12 jours ne correspondent plus à des actes de soin, mais à des consommations de ressource pour le CHU. Concernant la base nationale, elle est proche de cette DS moyenne, mais à Grenoble, on la dépasse de 3 jours. Cela pourrait être une raison du mauvais classement économique de ce GHM.

D'autre part, cette pathologie lourde nécessite un accompagnement et une rééducation importants, d'où l'orientation de ces patients vers des structures de soins de suite et de réadaptation (SSR) ou de soins de longue durée (SLD) et le besoin de lits hospitaliers pour les nouveaux patients. Cependant, il y a peu d'établissements de ces types, particulièrement dans la région grenobloise, donc un manque de places important. C'est pourquoi, les patients qui le peuvent (contexte social favorable à un retour à domicile), rentrent chez eux, même s'ils seraient préférables qu'ils intègrent des structures de SSR ou de SLD. Cependant, le CHU a besoin de ces lits pour les nouveaux patients atteints d'AVC. Il aurait été intéressant de sélectionner la variable du PMSI indiquant la destination des patients à leur sortie d'hôpital, elle aurait permis de valider cette caractéristique grenobloise. Concernant, les patients plus compliqués d'un point de vue social (entourage peu favorable à un retour à domicile), ils sont obligés d'attendre au CHU qu'une place se libère dans ces structures pour pouvoir partir. Ce sont ce type de patients qui reviennent cher au CHU. En effet, ils restent au CHU à cause de problèmes sociaux et non médicaux.

Ces deux raisons (démarche diagnostique plus longue et manque de place dans les struc-

tures de SSR et de SLD) pourraient expliquer en partie l'allongement de 2 jours de la DS moyenne à Grenoble. Cependant il peut aussi s'expliquer par des erreurs de classification : en effet, si la personne responsable du codage des RSS a oublié une co-morbidité importante, cela pourra entraîner le classement du patient dans un GHM qui n'aurait pas dû être le sien. Néanmoins, ce type d'erreur n'est sans doute pas propre au CHU de Grenoble.

### 6.3.2 Variables explicatives de la DS des composantes

Concernant les premières composantes du modèles, trois variables sont communes à la base de Grenoble et à la base nationale, pour expliquer la DS. Il s'agit de l'âge, du nombre de RUM et du diagnostic principal. En effet, plus l'âge augmente, plus le risque de faire un AVC non transitoire est élevé. Une des raisons qui pourrait expliquer que le nombre de RUM augmente avec la DS est la structure du CHU de Grenoble : c'est le seul établissement de la ville, l'hôpital sud jouant le rôle d'hôpital de proximité, et c'est un établissement généraliste comme nous l'avons expliqué plus haut. Les patients ayant fait un infarctus à Grenoble restent moins longtemps à l'hôpital que ceux ayant eu une hémorragie. Comme la DS moyenne est comparée aux patients ayant fait une hémorragie (degré de complication élevé), il paraît normal que la DS soit plus courte. De plus, une hémorragie est synonyme de complication et donc d'un allongement des DS. Un AVC avec un diagnostic de malformation est lui aussi synonyme de complication, et donc d'un allongement des DS.

Dans la base nationale, deux autres variables expliquent aussi les DS de cette composante, il s'agit du nombre de diagnostics et d'actes associés. Le fait d'avoir de multiples pathologies est reconnu pour allonger les DS.

Seules les DS de la base nationale de la deuxième composante peuvent être expliquées avec les variables du PMSI avec lesquelles nous avons travaillé. Comme pour le premier groupe, l'âge a tendance à allonger les DS. La relation entre le nombre d'unités médicales où le patient a séjourné et l'allongement est logique, car plus un patient change d'unité médicale, plus il aura tendance à rester au CHU, cependant ceci concerne le groupe ayant des DS comprises entre un et deux jours. La cause est donc sans doute à chercher ailleurs. Comme nous l'avons vu avant, ces faibles DS peuvent s'expliquer par des transferts, dont le pourcentage est important dans ce groupe.

Les patients grenoblois ne diffèrent donc pas de manière significative d'un point de vue médical des patients de la base nationale. Le problème de l'allongement des DS des patients appartenant à la première composante semble plutôt venir d'un problème de structure. Les modèles ne mettent pas en évidence l'indicateur de fragilité sociale concernant les femmes âgées.



### 6.3.3 Validité interne de l'étude

#### Choix du modèle

En étudiant la forme des histogrammes de DS des patients ayant eu un AVC non transitoire dans les deux bases nationale et grenobloise, nous voyons que leurs distributions ne sont approximées par aucune des lois simples que nous avons testées (normale, log-normale, exponentielle, gamma et Weibull). Cela est synonyme d'une hétérogénéité des DS au sein de ces distributions, comme l'ont déjà montré des auteurs ayant étudié les distributions de DS de certains GHM [52, 72, 73, 74, 75]. Pour la prendre en compte, nous avons réalisé des modèles de mélange de lois log-normales sur les DS de chacune des bases, montrant que c'est bien des modèles de mélange de deux lois normales qui ajustent le mieux le logarithme des DS. Les critères BIC et ICL sont maximums pour ce modèle, quelle que soit la base.

#### Interprétation des composantes du mélange

Les deux analyses utilisées nous ont permis de décrire les composantes du mélange sous deux angles : l'une plutôt prédictive (première analyse), l'autre plutôt descriptive (deuxième analyse). La première analyse nous a permis de donner la probabilité moyenne d'appartenance à l'une des composantes, connaissant la caractéristique d'un patient. C'est une démarche pronostique. La deuxième nous a servi à décrire les caractéristiques des composantes  $k$  des deux modèles. Cette démarche a été renforcée par la modélisation des DS de chacune des composantes du modèle, afin d'identifier plus précisément les variables du PMSI influençant les DS de ces différents groupes. On trouve des résultats similaires, ce qui renforce la validité du nombre de classes des modèles de mélange.

### 6.3.4 Validité externe de l'étude

Concernant la validité externe de l'étude sur le plan statistique, des analyses du Professeur Quantin et de son équipe ont aussi montré qu'en étudiant les DS de certains GHM (n°128, n°589, n°590, et du n°873 au n°876), celles-ci suivaient plutôt des modèles de mélange de trois lois de Weibull [52, 72, 73, 74, 75]. En appliquant une méthodologie proche de la leur, nous voyons que les distributions de la base grenobloise et de la base nationale, pour les patients atteints d'un AVC non transitoires, suivent un modèle de mélange de deux lois log-normales. Il n'y a donc pas présence d'"outliers" dans ces distributions, ou les méthodes que nous avons appliquées ne le permettent pas. Le choix d'approximation de la distribution des DS de ce GHM diffère de celui fait par l'équipe du Professeur Quantin. En effet, nous avons choisi des mélanges de lois log-normales car elles ont une forme simple, elles sont robustes d'un point de vue statistique, et sont souvent utilisées en étude unimodale. Pour renforcer la validité de notre modèle, il aurait été intéressant d'étudier les mêmes GHM que le Prof. Quantin, afin d'analyser si les DS grenobloises se répartissaient elles-aussi en 3 sous-groupes et avaient des caractéristiques proches, en termes de DS et de proportions de patients.

Le but de notre étude n'était pas de mettre en évidence des variables pouvant expliquer les DS de patients ayant eu un AVC non transitoire, mais de repérer quelles étaient les variables du PMSI qui différaient entre la base nationale et la base grenobloise afin d'expliquer l'allongement de 2 jours en moyenne des patients appartenant à la deuxième composante du modèle. Nous n'avons pas rencontrée d'étude dans la littérature ayant fait ce type de comparaison sur ce type de patients. Néanmoins, les variables ayant une influence sur les DS dans notre étude sont connus de la littérature [59, 57].

### 6.3.5 Limites de l'étude

#### Limites statistiques

Du point de vue statistique, les résultats trouvés sont très dépendants de l'hypothèse faite, à savoir la log-normalité des DS. La comparaison graphique de la densité des modèles à l'histogramme du logarithme des DS permet de s'apercevoir que pour les bases nationales, elle a tendance à surestimer la proportion de patients restant entre 7 et 15 jours en moyenne, ce phénomène est moins prononcé dans la base grenobloise. Concernant la première composante du modèle pour la base nationale, la loi log-normale n'est peut-être pas optimale. Il aurait été intéressant de modéliser les DS de la base nationale par un modèle de mélange de lois exponentielles, qui auraient l'avantage de mieux épouser le début de la distribution de DS où se trouve un fort pourcentage de patients restant un ou deux jours (près d'un cinquième des patients).

#### Limites épidémiologiques

Il est dommage que nous n'ayons pas pu travailler sur les diagnostics et les actes associés, mais seulement considérer leur nombre. En effet, ces derniers sont très liés à la DS. Il aurait été intéressant de pouvoir les regrouper en classes, comme nous l'avons fait avec les diagnostics principaux. Cependant, étant donné leur grand nombre, on peut s'interroger si un regroupement en classes des actes associés et médicaux serait faisable, et si cela serait pertinent du point de vue médical. Néanmoins, il n'est pas sûr que ces variables affinent de manière notable la description des sous-groupes de patients en termes de DS, car les variables du fichier PMSI n'ont pas été créées pour cet objectif. Elles permettent tout de même, au terme de cette analyse, de pointer vers des problèmes de structure pour le CHU de Grenoble.

### 6.3.6 Synthèse

Les DS des patients hospitalisés suite à des AVC non transitoires semblent se structurer en deux groupes, que ce soit dans la base nationale ou grenobloise, sous l'hypothèse que les distributions des logarithmes de DS suivent des mélange de lois normales. Le premier groupe de patients grenoblois restant hospitalisés 2 semaines en moyenne est plus important que sur la base nationale (84 % contre 75%). Les patients de ce groupe demeurent en moyenne 2 jours de plus que les patients du même groupe dans les autres CHU nationaux. Les patients

du second groupe, restant 1 ou 2 jours à l'hôpital, sont quasiment deux fois moins nombreux à Grenoble (16% contre 25%).

Il ne semble pas y avoir présence d'un groupe d'"outliers" à Grenoble pour ce GHM, le mauvais classement de celui-ci semble s'expliquer seulement par l'allongement global des séjours de 2 jours, observés dans le premier groupe de patients. Deux raisons pourraient expliquer cet allongement : d'une part, une démarche diagnostique plus longue que la moyenne et d'autre part, un manque de place dans les structures de SSR ou de SLD, donc plutôt des problèmes liés au fonctionnement des services et non au recrutement des patients. La faible DS du deuxième groupe pourrait s'expliquer par un taux de décès des patients plus important.

TAB. 6.2 – *Caractéristiques du GHM n°18: accidents vasculaires cérébraux non transitoires, pour les bases grenobloise et nationale.*

Variables	Base grenobloise	Base nationale	p_valeur
	$n = 617$ $n$ (%)	$n = 950$ $n$ (%)	
Sexe			0.455
Homme	323(52.4)	479(50.4)	
Femme	294(47.6)	471(49.6)	
Âge [ $\bar{X}(ET)$ ]	66.9(17.7)	66.2(18.1)	0.451
Nombre de RUM [ $\bar{X}(ET)$ ]	1.8(0.7)	1.3(0.6)	<0.0001
<b>Diagnostic principal regroupé en classes</b>			<0.0001
Hémorragie	166(26.9)	279(29.4)	
Infarctus	190(30.8)	384(40.4)	
Malformation	19(3.1)	38(4.0)	
Non précisé	242(39.2)	249(26.2)	
Nombre de diagnostics associés [ $\bar{X}(ET)$ ]	2.7(2.5)	1.6(1.8)	<0.0001
Nombre d'actes [ $\bar{X}(ET)$ ]	0.2(0.9)	1.6(2.3)	<0.0001
Présence d'acte classant			0.308
Non	611(99.0)	935(98.4)	
Oui	6(1.0)	15(1.6)	
ICR (actes classants) [ $\bar{X}(ET)$ ]	25.7(14.3)	33.1(33.1)	0.605
<b>Mode d'entrée</b>			<0.0001
Transfert long et moyen séjour	0(0)	6(0.6)	
Transfert interne	33(5.4)	112(11.8)	
Entrée directe	584(94.6)	832(87.6)	
<b>Mode de sortie</b>			<0.0001
Transfert long et moyen séjour	0(0)	38(4.0)	
Transfert interne	150(24.3)	263(27.7)	
Domicile	393(63.7)	543(57.2)	
Décès	74(12.0)	106(11.2)	

TAB. 6.3 – Paramètres de la lois normale estimés sur les DS de la base nationale et résultats du test d'adéquation de KS.

N° GHM	$\mu^{\text{NAT}}$	$\sigma^{\text{NAT}}$	$KS^{\text{NAT}}$	Valeur de $p^{\text{NAT}}$
18	11.5758	130.666	0.1774	<0.0001

TAB. 6.4 – Paramètres de la lois log-normale estimés sur les DS de la base nationale et résultats du test d'adéquation de KS.

N° GHM	$\mu^{\text{NAT}}$	$\sigma^{\text{NAT}}$	$KS^{\text{NAT}}$	Valeur de $p^{\text{NAT}}$
18	1.9700	1.1374	0.1023	<0.0001

TAB. 6.5 – Paramètre de la lois exponentielle estimé sur les DS de la base nationale et résultats du test d'adéquation de KS.

N° GHM	$\lambda^{\text{NAT}}$	$KS^{\text{NAT}}$	Valeur de $p^{\text{NAT}}$
18	11.5758	0.0864	<0.0001

TAB. 6.6 – Paramètres de la lois gamma estimés sur les DS de la base nationale et résultats du test d'adéquation de KS.

N° GHM	$\alpha^{\text{NAT}}$	$\beta^{\text{NAT}}$	$KS^{\text{NAT}}$	Valeur de $p^{\text{NAT}}$
18	1.1829	9.7856	0.3529	<0.0001

TAB. 6.7 – Paramètres de la lois de Weibull estimé sur les DS de la base nationale et résultats du test d'adéquation de KS.

N° GHM	$\alpha^{\text{NAT}}$	$\beta^{\text{NAT}}$	$KS^{\text{NAT}}$	Valeur de $p^{\text{NAT}}$
18	0.0683	1.0819	0.6301	<0.0001

TAB. 6.8 – Valeurs des critères ICL et BIC pour les modèles de mélange de 2 ou 3 lois log-normales ajustés sur les DS de la base nationale et grenobloise.

GHM	BDD	ICL (2 classes)	ICL (3 classes)	BIC (2 classes)	BIC (3 classes)
18	nationale	-1396.50	-1889.90	-1347.60	-1357.90
18	grenobloise	-929.96	-1042.40	-898.45	-907.30

TAB. 6.9 – Paramètres estimés des modèles de mélange de 2 lois normales sur le logarithme de DS des GHM de la base nationale et grenobloise.

N° GHM	Base	Nombre de composantes	$p_1^{(1)}$ (%)	$n_1^{(2)}$	$\bar{X}(DS_1)(E.T.(DS_1))^{(3)}$	$p_2$ (%)	$n_2$	$\bar{X}(DS_2)(E.T.(DS_2))$
18	grenobloise	2	84.1	524	15.4 (10.8)	15.9	93	1.7 (0.3)
18	nationale	2	74.7	697	13.1 (6.4)	25.3	253	1.7 (0.4)

<sup>(1)</sup>  $p_1$  : proportion de la première composante (idem pour la 2<sup>e</sup> composante),

<sup>(2)</sup>  $n_1$  : effectif de la première classe (idem pour la 2<sup>e</sup> classe),

<sup>(3)</sup>  $\bar{X}(DS_1)(E.T.(DS_1))$  : DS moyenne et écarts-types estimés avec les paramètres de la loi normale pour la 1<sup>ère</sup> composante, (idem pour la 2<sup>e</sup> composante), avec  $\bar{X}(DS_1) = E(DS_1) = \exp(\mu_1 + \frac{\sigma_1^2}{2})$  et  $E.T.(DS_1) = \sqrt{V(\bar{X})} = \sqrt{(\exp(2\mu_1 + \sigma_1^2))(\exp \sigma_1^2 - 1)}$  où  $\mu_1$  correspond à la moyenne des logarithmes de DS de la 1<sup>ère</sup> composante et  $\sigma_1^2$  à la variance.

TAB. 6.10 – GHM n°18, accidents vasculaires cérébraux non transitoires (base grenobloise) : probabilités conditionnelles d'appartenance à l'une ou à l'autre des classes du modèle selon les modalités des variables explicatives.

Variables	Effectif	Probabilité d'appartenance à la :		p.valeurs	Proche du profil de la :	
		1 <sup>ère</sup> composante	2 <sup>e</sup> composante		1 <sup>ère</sup> comp.	2 <sup>e</sup> comp.
Sexe						
Homme	323	83.0	17.0			
Femme	294	85.4	14.6			
Nombre de RUM				<0.0001		
1 RUM	175	63.1	36.9			+
2 RUM et plus	442	92.4	7.6		+	
Diagnostic principal regroupé en classes				<0.0001		
Hémorragie	166	77.4	22.6			+
Infarctus	190	89.0	11.0			
Malformation	19	42.7	57.3			+
Non précisé	242	88.2	11.8			
Nombre de diagnostics associés				<0.0001		
Aucun DGA	124	68.1	31.9			+
1 DGA	123	82.4	17.6			
2 DGA et plus	370	90.1	9.9		+	
Nombre d'actes						
Aucun acte	566	84.3	15.7			
1 acte et plus	51	81.9	18.1			
Présence d'acte classant						
Non	611	84.0	16.0			
Oui	6	98.1	1.9			
Mode d'entrée						
Transfert long et moyen séjour	0					
Transfert interne	33	90.9	9.1			
Entrée directe	584	83.7	16.3			
Mode de sortie				0.001		
Transfert long et moyen séjour	0					
Transfert interne	150	88.9	11.1			
Domicile	393	85.0	15.0			
Décès	74	69.9	30.1			+

Légende : nous avons noté par “+” les modalités des variables dont les probabilités moyennes sont supérieures ou inférieures de manière significative à l'une des proportions estimées (84.1% ou 15.9%) des 2 composantes du modèle de mélange gaussien ajustant le logarithme des DS de la base grenobloise.

TAB. 6.11 – *GHM n° 18, accidents vasculaires cérébraux non transitoires (base nationale) : probabilités conditionnelles d'appartenance à l'une ou à l'autre des classes du modèle selon les modalités des variables explicatives.*

Variables	Effectif	Probabilité d'appartenance à la :		p.valeurs	Proche du profil de la :	
		1 <sup>ère</sup> composante	2 <sup>e</sup> composante		1 <sup>ère</sup> comp.	2 <sup>e</sup> comp.
Sexe						
Homme	479	75.6	24.4			
Femme	471	73.8	26.2			
Nombre de RUM				<0.0001		
1 RUM	691	68.2	31.8			+
2 RUM et plus	259	92.3	7.7		+	
Diagnostic principal regroupé en classes				<0.0001		
Hémorragie	279	67.1	32.9			+
Infarctus	384	83.3	16.7		+	
Malformation	38	40.0	60.0			+
Non précisé	249	75.4	24.6			
Nombre de diagnostics associés				<0.0001		
Aucun DGA	320	60.2	39.8			+
1 DGA	220	73.8	26.2			
2 DGA et plus	410	86.6	13.4		+	
Nombre de actes				0.003		
Aucun acte	464	84.3	15.7		+	
1 acte et plus	486	76.6	23.4			
Présence d'acte classant						
Non	935	74.7	25.3			
Oui	15	76.3	23.7			
Mode d'entrée				0.017		
Transfert long et moyen séjour	6	69.0	31.0			+
Transfert interne	112	85.2	14.8		+	
Entrée directe	832	73.3	26.7			
Mode de sortie				<0.0001		
Transfert long et moyen séjour	38	77.7	22.3			
Transfert interne	263	80.8	19.2		+	
Domicile	543	74.7	25.3			
Décès	106	58.7	41.3			+

Légende : nous avons noté par “+” les modalités des variables dont les probabilités moyennes sont supérieures ou inférieures de manière significative à l'une des proportions estimées (74.7% ou 25.3%) des 2 composantes du modèle de mélange gaussien ajustant le logarithme des DS de la base nationale.



TAB. 6.12 – *GHM n°18 : caractéristiques médico-sociales des deux classes du modèle pour chacune des bases.*

Variables	1 <sup>ère</sup> composante		p_valeurs	2 <sup>e</sup> composante		p_valeurs
	nationale %	grenobloise %		nationale %	grenobloise %	
Homme	49.0	48.4		43.9	51.4	0.009
Âge ( $\bar{X}$ )	67.4	67.9		61.7	62.8	
Nombre de RUM ( $\bar{X}$ )	1.4	1.9		1.4	1.1	
Diagnostic principal regroupé en classes			<0.0001			<0.0001
Hémorragie	26.4	24.7		38.3	38.3	
Infarctus	45.1	32.6		21.4	26.7	
Malformation	2.1	1.6		11.1	9.5	
Non précisé	26.4	41.1		29.2	25.5	
Nombre de diagnostics associés ( $\bar{X}$ )	1.9	2.9		1.6	0.9	
Nombre d'actes ( $\bar{X}$ )	1.8	0.2		0.4	1.1	
Présence d'acte classant						
Oui	1.0	1.0		1.0	1.0	
Mode d'entrée			<0.0001			<0.0001
Transfert long et moyen séjour	0.6	0		0	0.7	
Transfert interne	13.4	5.8		3.1	6.9	
Entrée directe	86.0	94.2		96.9	92.3	
Mode de sortie			<0.0001			<0.0001
Transfert long et moyen séjour	4.2	0		0	3.5	
Transfert interne	29.9	25.7		17.0	21.0	
Domicile	57.1	64.3		60.3	57.2	
Décès	8.8	10.0		22.8	18.2	

TAB. 6.13 – *Résultat de la modélisation de la première composante du mélange pour la base grenobloise par une régression log-normale.*

Variables	$\hat{\beta}$	$\hat{\beta}/\sigma(\hat{\beta})$	$t$	p_valeur
Intercept	1.229	0.174	7.047	<0.0001
Sexe				
Femme*	0.117	0.067	1.747	0.081
<b>Âge</b>	0.005	0.002	2.301	<b>0.022</b>
<b>Nombre de RUM</b>	0.347	0.050	6.865	<b>&lt;0.0001</b>
Diagnostic principal :				
<b>Infarctus**</b>	0.295	0.087	3.379	<b>0.001</b>
<b>Malformation**</b>	-0.671	0.268	-2.502	<b>0.013</b>
Non précisé**	0.153	0.085	1.814	0.070

En gras : les variables ayant une influence sur les DS,

\* versus Homme,

\*\* versus Hémorragie.

TAB. 6.14 – *Résultat de la modélisation de la première composante du mélange pour la base nationale par une régression log-normale.*

Variabiles	$\hat{\beta}$	$\hat{\beta}/\sigma(\hat{\beta})$	$t$	p_valeur
Intercept	1.707	0.129	13.273	<0.0001
Sexe :				
Femme*	0.087	0.050	1.750	0.081
<b>Âge</b>	0.005	0.001	3.394	<b>0.001</b>
<b>Nombre de RUM</b>	0.130	0.043	3.023	<b>0.003</b>
<b>Nombre de d'actes</b>	0.037	0.011	3.489	<b>0.001</b>
<b>Nombre de diagnostics associés</b>	0.034	0.015	2.276	<b>0.023</b>
Diagnostic principal :				
Infarctus**	-0.005	0.630	-0.075	0.941
<b>Malformation**</b>	-0.433	0.174	-2.487	<b>0.013</b>
Non précisé**	-0.033	0.072	-0.455	0.649

En gras : les variables ayant une influence sur les DS,

\* versus Homme,

\*\* versus Hémorragie.

TAB. 6.15 – *Résultat de la modélisation de la deuxième composante du mélange pour la base nationale par une régression log-normale.*

Variabiles	$\hat{\beta}$	$\hat{\beta}/\sigma(\hat{\beta})$	$t$	p_valeur
Intercept	0.132	0.182	0.725	0.469
Sexe :				
Femme* :	0.117	0.061	1.923	0.056
<b>Âge :</b>	0.001	0.002	0.033	0.973
<b>Nombre de RUM</b>	0.263	0.099	2.654	<b>0.009</b>
Mode de sortie :				
<b>Transfert long et moyen séjour**</b>	-0.356	0.166	-2.143	<b>0.033</b>
<b>Transfert**</b>	-0.293	0.076	-3.836	<b>&lt;0.0001</b>
Décès**	-0.155	0.087	-1.791	0.075

En gras : les variables ayant une influence sur les DS,

\* versus Homme,

\*\* versus Domicile.

# Conclusion

Ce travail de thèse a permis tout d’abord de proposer une méthodologie statistique rigoureuse permettant de pointer des GHM pouvant être à l’origine du fort coût du point ISA. Ces méthodes ont l’avantage d’être facilement automatisables, et ainsi généralisables à l’échelle nationale. Ensuite, d’un point de vue médico-économique, il apparaît que la classification la plus pertinente est celle issue de la classification hiérarchique ascendante. Les 257 GHM grenoblois de la base PMSI de 1997 semblent se structurer en 4 catégories de GHM, dont 2 sont identifiées comme mauvaises. Cette classification présente l’avantage d’être cohérente avec celle d’origine empirique, qui laissait apparaître la présence de 9 catégories. Elle est robuste, car elle est en accord avec la classification issue des modèles de mélange, et est plus précise que la liste du CHU. Elle a également permis d’isoler un groupe de 16 GHM, les “mauvais FG”, très homogènes du point de vue des profils de distribution de DS et des caractéristiques médico-économiques. Un tiers des patients de ce groupe participe fortement à la dérive du coût du point ISA à Grenoble. Comme on l’a vu, la mise en place d’une action dans les services hospitaliers concernés pourrait s’avérer rentable en termes de points ISA pour le CHU. De plus, le fait d’avoir identifié, grâce à nos profils, l’endroit des distributions de DS où se trouvent les anomalies va aider les médecins pour la vérification des dossiers médicaux. En effet, ils pourront cibler à partir de quelle DS ils doivent les vérifier. Ainsi, cette classification pourrait être une aide pour les gestionnaires des hôpitaux qui sont toujours en quête de points ISA supplémentaires, facile à obtenir.

## Classification des profils des 257 GHM

Du point de vue empirique, les profils des GHM grenoblois semblent se structurer en huit groupes (les profils empirique de types *(h)* et *(neutre)* étant regroupés ensemble). Six d’entre eux, les “mauvais E”, sont considérés comme mauvais de manière significative, de manière diverse : sur l’ensemble de la distribution des DS, sur les deux tiers ou sur un tiers seulement. Cette classification a l’avantage d’indiquer l’endroit de la distribution DS où les GHM grenoblois sont moins performants que la base nationale.

La classification issue de l’approche factorielle indique que les profils des GHM grenoblois semblent se structurer en quatre groupes, dont deux peuvent être identifiés comme mauvais (FA et FG). Cette classification a l’avantage d’être en cohérence avec celle d’origine empirique (huit groupes). Elle est plus précise que la liste du CHU, constituée par les 50

GHM causant 80% du déficit du CHU, et robuste (plusieurs échantillons ont été testés avec l'analyse discriminante, et ils donnent un taux moyen de 81% de bien classés).

La classification issue du modèle de mélange confirme la justesse de la classification issue de l'approche factorielle. En utilisant deux méthodes différentes, nous obtenons deux classifications similaires. De plus, la classification issue du modèle de mélange est elle aussi cohérente avec la classification empirique. Elle semble seulement séparer les profils des 257 GHM en trois groupes, un seul d'entre eux pouvant être considéré comme mauvais, les "mauvais M". D'autre part, elle ne trouve pas de structure particulière au groupe des "mauvais E", à l'inverse de la classification issue de l'approche factorielle qui identifiait deux sous-groupes.

Il semble donc que la classification issue de l'approche factorielle soit la plus pertinente d'un point de vue médico-économique. En effet, elle est cohérente avec la classification d'origine empirique, elle est robuste et est confirmée par une autre classification utilisant une méthode différente. Par rapport à la liste CHU, cette approche est plus rigoureuse et apporte des informations supplémentaires sur le type des profils des GHM.

### Analyse intra-GHM

Les DS des patients hospitalisés suite à des AVC non transitoires semblent se structurer en deux groupes, que ce soit dans la base nationale ou grenobloise. Le premier groupe de patients grenoblois, restant hospitalisés deux semaines en moyenne, est plus important que celui de la base nationale (84 % contre 75%); les patients de ce groupe demeurent en moyenne deux jours de plus (15 jours contre 13) que dans les autres CHU. Les patients du second groupe, restant entre un et deux jours à l'hôpital, sont deux fois moins nombreux à Grenoble (16% contre 29%) que dans la base nationale.

Il ne semble donc pas y avoir présence d'un groupe supplémentaire d'"outliers" à Grenoble pour ce GHM, mais seulement de deux groupes de patients, l'un de faible et l'autre de forte DS. Son mauvais classement s'explique seulement par l'allongement global des séjours des patients de son premier groupe. Deux raisons pourraient expliquer cet allongement : d'une part, une démarche diagnostique plus longue que la moyenne (pourcentage important de diagnostics non précisés [41.1%]), et, d'autre part, un manque de place dans les structures de soins de suite et de réadaptation. La faible DS du deuxième groupe pourrait s'expliquer en grande partie par un taux de décès plus important des patients.

### Complexité du travail de vérification des RSS

Avant de mettre en place une action visant à augmenter les points ISA, il faut, d'abord, trouver une explication au classement des GHM appartenant aux listes de référence ou aux groupes des "mauvais" dans les classifications (problème de structure, variation des pratiques médicales, erreur de classification, etc...). Comme nous l'avons vu dans les chapitres 4 et 5, la solution idéale serait le contrôle exhaustif des RSS de tous les patients correspondant à

ces groupes. Cependant, ce nombre est beaucoup trop élevé pour réaliser cette vérification. Une alternative serait le tirage au sort de 10% des RSS de ces groupes. Cela correspondrait alors au contrôle de (par ordre décroissant) :

- 1 172 dossiers de la liste CHU (issus des 50 GHM représentant 80% du déficit total du CHU),
- 1 317 dossiers du groupe des “mauvais M” (issus des 52 GHM représentant 42% du déficit total du CHU),
- 660 dossiers de la liste du  $\chi^2$  (issus des 36 GHM représentant 10% du déficit),
- 422 dossiers de la liste des quartiles (issus des 27 GHM représentant 35% du déficit),
- 308 dossiers du groupe des “mauvais FA” (issus des 24 GHM représentant 33% du déficit),
- 71 dossiers du groupe des “mauvais FG” (issus des 16 GHM représentant 10% du déficit).

Ainsi, en tenant compte simplement du déficit en points ISA calculé par rapport à la DS moyenne nationale, il faudrait vérifier un nombre important de RSS ( $n=1\ 172$ ), alors qu’en se focalisant d’abord sur la forme des profils des GHM, ce nombre serait significativement plus faible (groupe des “mauvais FG”,  $n=71$ ), tout en restant rentable pour le CHU en termes de points ISA, même si ce groupe ne représente que 10% du déficit. Le point le plus intéressant de cette thèse a consisté à montrer que le groupe, bien identifié, des 16 GHM responsables de 10% du déficit, comportait un ensemble de patients, bien identifié également, de DS supérieure à  $q_{(0.75)}^N$ ; s’ils n’avaient pas été classés dans ces GHM (souvent qualifiés d’“autres...” par suite d’absence de diagnostic précis), ils auraient conduit à un équilibre du CHU de Grenoble par rapport à la base nationale en terme de points ISA.

Ainsi, plutôt que de fonder une vérification des RSS sur le déficit en points ISA de tous les GHM (liste CHU), il serait souhaitable de repérer d’abord les “mauvais” GHM en termes de profils, afin d’identifier l’endroit de la distribution des DS où ces GHM sont peu performants (intervalle de DS dans lequel il faut vérifier les RSS), et de quantifier, ensuite, leur déficit en points ISA. Comme nous l’avons vu, cette démarche permettrait de diminuer le nombre de RSS à contrôler, et donc de les étudier plus précisément. Après quoi il serait possible de mettre en œuvre une action ciblée auprès des services concernés, visant à résoudre les problèmes mis en évidence avec cette catégorie de patients. À moyen terme, cette action pourrait rapporter des points ISA au CHU. Cependant, cette méthodologie montre des limites quand les GHM sont identifiés comme mauvais sur la totalité de leurs distributions de DS (profil empirique de type *(a)* et groupe des “mauvais FA”): dans ce cas, le calcul du déficit en points ISA pourrait être utile pour choisir les GHM dont les RSS sont à contrôler. L’inconvénient de cette démarche est qu’elle n’est pas ciblée sur un sous-groupe de patients comme elle l’était pour les “mauvais FG”, mais réalisée à l’aveugle sur tous les patients de ce groupe. En conséquence, les caractéristiques des RSS tirés au sort n’expliqueront peut-être qu’une infime partie du déficit en points ISA, et l’action qui en découlera ne sera peut-être

pas rentable pour le CHU.

Nous voyons donc que travailler uniquement avec le déficit en points ISA par rapport à la DS moyenne nationale n'est pas suffisant pour repérer les GHM qui coûtent cher en points ISA au CHU de Grenoble. Il faut coupler cette méthode avec des techniques de classification, donnant une indication de la forme des profils des distributions de DS, puis calculer le déficit en points ISA à partir des quartiles de DS et non à partir de la moyenne. En conséquence, le meilleur choix réside dans l'examen des 71 RSS des "mauvais FG", qui présentent toutes les caractéristiques des mauvais profils "rentables" en terme de points ISA s'ils sont supprimés, pour rééquilibrer le CHU de Grenoble par rapport à la norme nationale.

Enfin, la modélisation des DS d'un GHM repéré comme "mauvais" par un modèle de mélange (comme le GHM 18 pris en exemple) permet d'étudier les caractéristiques médico-sociales de ses patients et donc les causes précises de l'allongement de ses DS.

## Perspectives

Les perspectives que nous pouvons envisager à partir de ce travail sont de différents ordres.

Tout d'abord, nous pourrions améliorer la méthodologie proposée en prenant en compte le statut du patient à la sortie de l'hôpital (vivant guéri, vivant non guéri ou décédé), cela aurait une grande importance d'un point de vue médical en terme de qualité des soins, et nous ne l'avons pas considéré dans cette étude.

Ensuite, nous pourrions encore affiner l'étude des distributions de DS des GHM du groupe des "mauvais FG", par des mélanges de lois qui permettraient de caractériser leurs outliers. Concernant le calcul du déficit en points ISA, la méthode actuelle est insuffisante, puisque ce calcul est fait à partir de la DS moyenne. Cet indicateur est à améliorer, étant donné que les distributions de DS ne suivent pas des lois normales, et sont souvent bi-modales. Il serait intéressant de développer une nouvelle méthode de calcul, qui pourrait être fondée sur les quartiles des DS nationales. Cela permettrait de prendre en compte la forme des profils grenoblois.

Concernant le groupe des 16 GHM appelés "mauvais FG", il serait intéressant de modéliser leurs distributions de DS par des modèles de mélange, qui permettraient de caractériser leurs "outliers" (patients ayant une  $DS > q_{(0.75)}^N$ , environ le tiers) de manière plus précise. D'autre part, ce modèle affinerait le choix des RSS à vérifier, par exemple pour sélectionner leurs patients excentrés (au sens des grandes déviations).

La mise en exergue de ces profils a nécessité de nombreuses et laborieuses étapes, qui ne pourraient pas être aisément reproduites. Ainsi, la rédaction d'un logiciel automatisant ces processus permettrait de fournir un outil facilement exploitable dans les DIM des CHU.

D'autre part, maintenant que les informations sur les coûts par pathologies sont disponibles de manière courante dans les CHU, il serait très intéressant de transposer cette méthodologie à la distribution des coûts et pas seulement à celles des DS comme nous l'avons fait. Cela serait plus informatif d'un point de vue économique.

Il serait aussi pertinent d'être capable, en utilisant cette méthodologie, d'anticiper sur le nombre de lits à fermer dans les services selon les périodes de l'année.

De plus, d'un point de vue plus technique, on pourrait améliorer la méthode de calcul du déficit en points ISA en prenant en compte la forme des profils, et pas uniquement la durée moyenne de séjour.

Pour terminer, il serait intéressant d'intégrer les notions d'origine géographique et de longévité, pour qualifier plus précisément le recrutement des patients des GHM identifiés comme principaux responsables du déficit du CHU. Des cartes du bassin de recrutement et des pyramides des âges des patients à problèmes pourraient être ainsi élaborées et utilisées par les décideurs des systèmes de soin.





# Annexe A

## Liste des CMD

Vingt cinq Catégories Majeures de Diagnostic (CMD) et deux Catégories Majeures (CM) ont à l'heure actuelle été définies :

- CMD n°01 : affections du système nerveux ;
- CMD n°02 : affections de l'oeil ;
- CMD n°03 : affections des oreilles, du nez, de la gorge, de la bouche et des dents ;
- CMD n°04 : affections de l'appareil respiratoire ;
- CMD n°05 : affections de l'appareil circulatoire ;
- CMD n°06 : affections du tube digestif ;
- CMD n°07 : affections du système hépato-biliaire et du pancréas ;
- CMD n°08 : affections et traumatismes de l'appareil musculo-squelettique et du tissu conjonctif ;
- CMD n°09 : affections de la peau, des tissus sous-cutanés et des seins ;
- CMD n°10 : affections endocriniennes, métaboliques et nutritionnelles ;
- CMD n°11 : affections du rein et des voies urinaires ;
- CMD n°12 : affections de l'appareil génital masculin ;
- CMD n°13 : affections de l'appareil génital féminin ;
- CMD n°14 : grossesses pathologiques, accouchements et affections du post partum ;
- CMD n°15 : nouveaux-nés, prématurés et affections de la période périnatale ;
- CMD n°16 : affections du sang et des organes hématopoïétiques ;
- CMD n°17 : affections myéloprolifératives et tumeurs de siège imprécis ou diffus ;
- CMD n°18 : maladies infectieuses et parasitaires ;
- CMD n°19 : maladies et troubles mentaux ;
- CMD n°20 : troubles mentaux organiques liés à l'absorption de drogues ou induits par celles-ci ;
- CMD n°21 : traumatismes, allergies et empoisonnements ;
- CMD n°22 : brûlures ;

- CMD n°23 : facteurs influant sur l'état de santé et autres motifs de recours aux services de santé ;
- CMD n°25 : maladies dues à une infection à VIH (non encore disponible) ;
- CMD n°26 : traumatismes multiples graves (non encore disponible) ;
- CM n°24 : séances et séjours de moins de 24 heures ;
- CM n°90 : erreurs et autres séjours inclassables.

## Annexe B

### Exemple : liste des GHM appartenant à la CMD 02

Ci-après, un exemple avec la liste des 13 GHM appartenant à la CMD n°02 “Affections de l’oeil” :

- GHM n°48 : Interventions sur la rétine ;
- GHM n°49 : Interventions sur l’orbite ;
- GHM n°50 : Interventions primaires sur l’iris ;
- GHM n°51 : Interventions sur le cristallin avec ou sans vitrectomie ;
- GHM n°52 : Autres interventions extra-oculaires, âge supérieur à 17 ans ;
- GHM n°53 : Autres interventions extra-oculaires, âge inférieur à 18 ans ;
- GHM n°54 : Autres interventions intra-oculaires ;
- GHM n°59 : Hyphéma
- GHM n°60 : Infections oculaires aiguës sévères ;
- GHM n°61 : Affections oculaires d’origine neurologique ;
- GHM n°62 : Autres affections oculaires, âge supérieur à 17 ans avec CMA ;
- GHM n°63 : Autres affections oculaires, âge supérieur à 17 ans sans CMA ;
- GHM n°64 : Autres affections oculaires, âge inférieur à 18 ans.



## Annexe C

# Informations contenues dans le RUM

Le **RUM** contient les informations administratives et médicales suivantes :

- Numéro FINESS ;
- Numéro de RSS ;
- Date de naissance ;
- Sexe ;
- Code postal du lieu de résidence du patient ;
- Numéro de l'Unité Médicale ;
- Date d'entrée dans l'Unité Médicale ;
- Mode d'entrée dans l'Unité Médicale ;
- Provenance ;
- Date de sortie de l'Unité Médicale ;
- Mode de sortie de l'Unité Médicale ;
- Destination ;
- Hospitalisation d'une durée totale inférieure à 24 heures, le cas échéant ;
- Hospitalisation au cours des 30 jours précédents ;
- Nombre de séances, équivalant au nombre de présences dans l'établissement d'une durée inférieure à 24 heures impliquant habituellement sa fréquentation itérative dans un but thérapeutique, quel que soit le tarif appliqué à ce passage ;
- Diagnostic principal ;
- Diagnostics associés, s'il y a lieu et jusqu'à 99 ;
- Actes médicaux et chirurgicaux, s'il y a lieu et jusqu'à 99 ;
- Poids de naissance, pour les nouveau-nés (<28 jours) ;
- Score Indice de Gravité Simplifié (Version 2).

Le **RSS** est une synthèse des informations qui se trouvent dans le(s) **RUM**.



## Annexe D

# Numéros, effectifs et libellés des GHM

Les cinq tableaux ci-après donnent les numéros, indices et libellés des 257 GHM retenus pour l'analyse.



N° GHM	N <sub>Gre</sub>	N <sub>Nat</sub>	Libellés des GHM
1	423	641	Craniotomies en dehors de tout traumatisme, âge supérieur à 17 ans
2	74	140	Craniotomies pour traumatisme, âge supérieur à 17 ans
3	58	96	Craniotomies, âge inférieur à 18 ans
5	85	164	Interventions sur le système vasculaire précérébral
8	159	217	Interventions sur les nerfs crâniens ou périphériques et autres interventions sur le système nerveux, âge inférieur à 70 ans sans CMA
14	59	95	Tumeurs du système nerveux, âge supérieur à 69 ans et/ou CMA
15	94	171	Tumeurs du système nerveux, âge inférieur à 70 ans sans CMA
16	425	698	Maladies dégénératives du système nerveux
17	159	325	Sclérose en plaques et ataxie cérébelleuse
18	617	950	Accidents vasculaires cérébraux non transitoires
19	177	337	Accidents ischémiques transitoires et occlusions des artères précérébrales
21	103	196	Autres affections cérébro-vasculaires sans CMA
22	93	171	Affections des nerfs crâniens et rachidiens, âge supérieur à 69 ans et/ou CMA
23	179	398	Affections des nerfs crâniens et rachidiens, âge inférieur à 70 ans sans CMA
24	68	136	Infections du système nerveux à l'exception des méningites virales
25	61	74	Méningites virales
27	90	164	Troubles de la conscience et comas d'origine non traumatique
28	188	299	Convulsions, épilepsies et céphalées, âge supérieur à 69 ans et/ou CMA
29	339	649	Convulsions, épilepsies et céphalées, âge de 18 à 69 ans sans CMA
30	162	441	Convulsions, épilepsies et céphalées, âge inférieur à 18 ans
37	158	192	Autres affections du système nerveux, âge supérieur à 69 ans et/ou CMA
38	179	408	Autres affections du système nerveux, âge inférieur à 70 ans sans CMA
40	418	595	Affections du système nerveux, avec CMAS
45	204	165	Commotions cérébrales, âge supérieur à 69 ans et/ou CMA
46	714	854	Commotions cérébrales, âge inférieur à 70 ans sans CMA
48	229	395	Interventions sur la rétine
51	593	1157	Interventions sur le cristallin avec ou sans vitrectomie
52	142	216	Autres interventions extra-oculaires, âge supérieur à 17 ans
53	129	168	Autres interventions extra-oculaires, âge inférieur à 18 ans
54	228	422	Autres interventions intra-oculaires
61	124	84	Affections oculaires d'origine neurologique
63	299	363	Autres affections oculaires, âge supérieur à 17 ans sans CMA
68	92	139	Interventions majeures sur la tête et le cou
71	60	74	Réparations de fissures labiales et palatines
72	86	230	Interventions sur les sinus et l'apophyse mastoïde, âge supérieur à 17 ans
74	352	550	Interventions diverses sur les oreilles, le nez, la gorge ou le cou
75	83	165	Rhinoplasties
82	186	287	Autres interventions chirurgicales portant sur les oreilles, le nez, la gorge ou le cou
94	191	458	Tumeurs malignes des oreilles, du nez, de la gorge ou de la bouche
95	105	189	Troubles de l'équilibre
99	82	120	Otites moyennes et infections des voies aériennes supérieures, âge de 18 à 69 ans sans CMA
100	149	257	Otites moyennes et infections des voies aériennes supérieures, âge inférieur à 18 ans
102	144	94	Traumatismes et déformations du nez
103	131	184	Affections de la bouche et des dents sans certaines extractions, réparations ou prothèses dentaires, âge supérieur à 17 ans
104	77	113	Affections de la bouche et des dents sans certaines extractions, réparations ou prothèses dentaires, âge inférieur à 18 ans
105	135	323	Affections de la bouche et des dents avec certaines extractions, réparations et prothèses dentaires
106	199	338	Autres diagnostics portant sur les oreilles, le nez, la gorge ou la bouche, âge supérieur à 17 ans
113	181	201	Interventions majeures sur le thorax
118	203	265	Embolies pulmonaires
119	57	162	Infections et inflammations respiratoires, âge supérieur à 69 ans et/ou CMA
122	209	691	Tumeurs de l'appareil respiratoire
125	65	69	Épanchements pleuraux, âge supérieur à 69 ans et/ou CMA
127	198	718	Œdème pulmonaire et détresse respiratoire
128	455	623	Bronchopneumopathies chroniques
129	298	543	Pneumonies et pleurésies banales, âge supérieur à 69 ans et/ou CMA
130	103	162	Pneumonies et pleurésies banales, âge de 18 à 69 ans sans CMA
131	96	165	Pneumonies et pleurésies banales, âge inférieur à 18 ans

N° GHM	N <sub>Gre</sub>	N <sub>Nat</sub>	Libellés des GHM
136	231	279	Bronchites et asthme, âge supérieur à 69 ans et/ou CMA
137	129	231	Bronchites et asthme, âge de 18 à 69 ans sans CMA
138	376	812	Bronchites et asthme, âge inférieur à 18 ans
139	296	366	Signes et symptômes respiratoires, âge supérieur à 69 ans et/ou CMA
140	843	497	Signes et symptômes respiratoires, âge inférieur à 70 ans sans CMA
141	129	118	Autres diagnostics portant sur le système respiratoire, âge supérieur à 69 ans et/ou CMA
142	156	108	Autres diagnostics portant sur le système respiratoire, âge inférieur à 70 ans sans CMA
153	114	300	Chirurgie de remplacement valvulaire avec circulation extracorporelle, sans cathétérisme cardiaque, ni coronarographie
154	155	143	Pontages aorto-coronariens avec cathétérisme cardiaque ou coronarographie
155	209	388	Pontages aorto-coronariens sans cathétérisme cardiaque, ni coronarographie
158	185	217	Chirurgie majeure de revascularisation, âge supérieur à 69 ans et/ou CMA
159	59	113	Chirurgie majeure de revascularisation, âge inférieur à 70 ans sans CMA
160	160	177	Autres interventions de chirurgie vasculaire
164	150	264	Pose d'un stimulateur cardiaque permanent sans infarctus aigu du myocarde, ni insuffisance cardiaque congestive, ni état de choc
167	153	270	Ligatures de veines et éveinages
168	95	73	Autres interventions sur le système circulatoire
178	180	167	Infarctus aigu du myocarde avec complications cardio-vasculaires et sortie en vie de l'établissement
179	438	632	Infarctus aigu du myocarde sans complications cardio-vasculaires
181	214	393	Cathétérismes cardiaques ou coronarographies avec pathologie cardiaque sévère en dehors d'un infarctus aigu du myocarde
182	846	2034	Cathétérismes cardiaques ou coronarographies pour une pathologie autre
184	467	1032	Insuffisances cardiaques et états de choc circulatoire
185	160	201	Thrombophlébites veineuses profondes
187	293	522	Troubles vasculaires périphériques, âge supérieur à 69 ans et/ou CMA
188	218	343	Troubles vasculaires périphériques, âge inférieur à 70 ans sans CMA
191	248	428	Hypertension artérielle
195	174	478	Arythmies et troubles de la conduction cardiaque, âge supérieur à 69 ans et/ou CMA
196	94	257	Arythmies et troubles de la conduction cardiaque, âge inférieur à 70 ans sans CMA
197	153	368	Angine de poitrine
198	168	282	Syncopes et lipothymies, âge supérieur à 69 ans et/ou CMA
199	86	204	Syncopes et lipothymies, âge inférieur à 70 ans sans CMA
200	214	395	Douleurs thoraciques
201	173	190	Autres affections de l'appareil circulatoire avec CMA
202	198	266	Autres affections de l'appareil circulatoire sans CMA
213	161	337	Interventions majeures sur l'intestin grêle et le côlon, âge supérieur à 69 ans et/ou CMA
214	73	206	Interventions majeures sur l'intestin grêle et le côlon, âge inférieur à 70 ans sans CMA
215	63	134	Interventions sur l'œsophage, l'estomac et le duodénum, âge supérieur à 69 ans et/ou CMA
225	192	368	Appendicectomies non compliquées, âge inférieur à 70 ans sans CMA
226	55	82	Interventions réparatrices pour hernies et éventrations à l'exception des hernies inguinales et crurales, âge supérieur à 69 ans et/ou CMA
227	54	136	Interventions réparatrices pour hernies et éventrations à l'exception des hernies inguinales et crurales, âge de 18 à 69 ans sans CMA
228	75	136	Interventions réparatrices pour hernies inguinales et crurales, âge supérieur à 69 ans et/ou CMA
229	98	260	Interventions réparatrices pour hernies inguinales et crurales, âge de 18 à 69 ans sans CMA
230	120	167	Interventions réparatrices pour hernies et éventrations, âge inférieur à 18 ans
232	74	180	Interventions sur le rectum et l'anus autres que les résections rectales, âge inférieur à 70 ans sans CMA
246	132	360	Tumeurs malignes du tube digestif, âge supérieur à 69 ans et/ou CMA
248	140	226	Hémorragies digestives, âge supérieur à 69 ans et/ou CMA
254	128	200	Occlusions intestinales non dues à une hernie, âge supérieur à 69 ans et/ou CMA
255	98	143	Occlusions intestinales non dues à une hernie, âge inférieur à 70 ans sans CMA
256	451	622	Gastro-entérites et maladies diverses du tube digestif, âge supérieur à 69 ans et/ou CMA
257	566	965	Gastro-entérites et maladies diverses du tube digestif, âge de 18 à 69 ans sans CMA
258	550	1229	Gastro-entérites et maladies diverses du tube digestif, âge inférieur à 18 ans
259	228	231	Autres affections digestives, âge supérieur à 69 ans et/ou CMA
260	222	238	Autres affections digestives, âge de 18 à 69 ans sans CMA
262	174	257	Affections du tube digestif avec CMAS
273	105	141	Cholécystectomies sans exploration de la voie biliaire principale, âge supérieur à 69 ans et/ou CMA

N° GHM	N <sub>Gre</sub>	N <sub>Nat</sub>	Libellés des GHM
274	104	260	Cholécystectomies sans exploration de la voie biliaire principale, âge inférieur à 70 ans sans CMA
284	106	531	Hépatites alcooliques et cirrhoses
285	146	516	Affections malignes du système hépato-biliaire ou du pancréas
286	124	266	Affections non malignes du pancréas
287	82	234	Affections hépatiques à l'exception des tumeurs malignes, des cirrhoses, des hépatites alcooliques, âge supérieur à 69 ans et/ou CMA
288	57	488	Affections hépatiques à l'exception des tumeurs malignes, des cirrhoses, des hépatites alcooliques, âge inférieur à 70 ans sans CMA
289	119	165	Affections des voies biliaires, âge supérieur à 69 ans et/ou CMA
290	100	153	Affections des voies biliaires, âge inférieur à 70 ans sans CMA
293	85	156	Affections du système hépato-biliaire et du pancréas avec CMAS
295	568	778	Interventions majeures sur les articulations et greffes de membres
296	125	344	Interventions sur la hanche et le fémur, autres que les interventions majeures, âge supérieur à 69 ans et/ou CMA
297	100	160	Interventions sur la hanche et le fémur, autres que les interventions majeures, âge de 18 à 69 ans sans CMA
301	75	204	Interventions sur le rachis, âge supérieur à 69 ans et/ou CMA
302	411	870	Interventions sur le rachis, âge inférieur à 70 ans sans CMA
304	59	114	Interventions sur l'humérus ou sur le membre inférieur à l'exception de la hanche, du fémur et du pied, âge supérieur à 69 ans et/ou CMA
305	534	432	Interventions sur l'humérus ou sur le membre inférieur à l'exception de la hanche, du fémur et du pied, âge de 18 à 69 ans sans CMA
306	92	131	Interventions sur l'humérus ou sur le membre inférieur à l'exception de la hanche, du fémur et du pied, âge inférieur à 18 ans
308	334	196	Interventions sur le genou, âge inférieur à 70 ans sans CMA
309	109	144	Résections osseuses localisées et/ou ablation de matériel de fixation interne au niveau de la hanche et du fémur
310	716	624	Résections osseuses localisées et/ou ablation de matériel de fixation interne au niveau d'une localisation autre que la hanche et le fémur
311	51	92	Interventions sur le membre supérieur à l'exception de l'humérus et de la main, âge supérieur à 69 ans et/ou CMA
312	280	371	Interventions sur le membre supérieur à l'exception de l'humérus et de la main, âge inférieur à 70 ans sans CMA
313	276	323	Interventions sur le pied
315	175	206	Interventions sur les tissus mous, âge inférieur à 70 ans sans CMA
317	396	580	Interventions sur la main et le poignet pour une affection autre qu'un kyste synovial
318	438	277	Arthroscopies
320	135	160	Autres interventions portant sur l'appareil musculo-squelettique et le tissu conjonctif, âge inférieur à 70 ans sans CMA
336	150	80	Fractures de la hanche et du bassin
339	115	259	Fractures pathologiques et affections malignes de l'appareil musculo-squelettique et du tissu conjonctif
340	60	249	Affections du tissu conjonctif, âge supérieur à 69 ans et/ou CMA
341	102	456	Affections du tissu conjonctif, âge inférieur à 70 ans sans CMA
343	777	1510	Pathologies rachidiennes relevant d'un traitement médical
344	60	194	Maladies osseuses et arthropathies spécifiques, âge supérieur à 69 ans et/ou CMA
347	96	160	Signes et symptômes concernant l'appareil musculo-squelettique et le tissu conjonctif
348	64	189	Tendinites, myosites et bursites
349	183	178	Suites de traitement après une affection de l'appareil musculo-squelettique ou du tissu conjonctif

N° GHM	N <sub>Gre</sub>	N <sub>Nat</sub>	Libellés des GHM
351	133	91	Fractures, entorses, luxations et dislocations de l'avant-bras, de la main, du pied ou multiples ou mal précisées, âge de 18 à 69 ans sans CMA
352	225	208	Fractures, entorses, luxations et dislocations de l'avant-bras, de la main, du pied ou multiples ou mal précisées, âge inférieur à 18 ans
353	125	113	Fractures, entorses, luxations et dislocations du bras, de la jambe à l'exception du pied, âge supérieur à 69 ans et/ou CMA
354	269	143	Fractures, entorses, luxations et dislocations du bras, de la jambe à l'exception du pied, âge de 18 à 69 ans sans CMA
355	105	111	Fractures, entorses, luxations et dislocations du bras, de la jambe à l'exception du pied, âge inférieur à 18 ans
356	160	413	Autres pathologies de l'appareil musculo-squelettique et du tissu conjonctif
358	50	119	Affections de l'appareil musculo-squelettique et du tissu conjonctif, avec CMAS
367	170	261	Greffes de peau et/ou parages de plaie à l'exception des ulcères cutanés et cellulites sans CMA
372	161	172	Interventions sur le sein pour des affections non malignes autres que les actes de biopsie et d'excision locale
375	78	95	Interventions plastiques
376	92	88	Autres interventions sur la peau, les tissus sous-cutanés ou les seins, âge supérieur à 69 ans et/ou CMA
377	197	289	Autres interventions sur la peau, les tissus sous-cutanés ou les seins, âge inférieur à 70 ans sans CMA
386	50	108	Affections dermatologiques sévères, âge supérieur à 69 ans et/ou CMA
391	109	177	Lésions, infections et inflammations de la peau et des tissus sous-cutanés, âge supérieur à 69 ans et/ou CMA
392	93	233	Lésions, infections et inflammations de la peau et des tissus sous-cutanés, âge de 18 à 69 ans sans CMA
394	177	142	Traumatismes de la peau et des tissus sous-cutanés, âge supérieur à 69 ans et/ou CMA
395	253	304	Traumatismes de la peau et des tissus sous-cutanés, âge de 18 à 69 ans sans CMA
396	103	178	Traumatismes de la peau et des tissus sous-cutanés, âge inférieur à 18 ans
397	111	193	Autres affections dermatologiques, âge supérieur à 69 ans et/ou CMA
398	211	467	Autres affections dermatologiques, âge inférieur à 70 ans sans CMA
405	57	81	Interventions sur l'hypophyse ou les glandes surrénales
408	63	76	Interventions sur les parathyroïdes
409	165	554	Interventions sur la thyroïde
418	203	1162	Diabète, âge supérieur à 35 ans
419	53	272	Diabète, âge inférieur à 36 ans
420	100	232	Troubles métaboliques ou nutritionnels divers, âge supérieur à 69 ans et/ou CMA
421	98	320	Troubles métaboliques ou nutritionnels divers, âge de 18 à 69 ans sans CMA
422	54	221	Troubles métaboliques ou nutritionnels divers, âge inférieur à 18 ans
425	85	477	Autres troubles endocriniens, âge inférieur à 70 ans sans CMA
429	64	83	Transplantations rénales
430	54	109	Interventions sur les reins et les uretères et chirurgie majeure de la vessie pour une affection tumorale
432	90	197	Interventions sur les reins et les uretères et chirurgie majeure de la vessie pour une affection non tumorale, âge inférieur à 70 ans sans CMA
437	172	216	Interventions transurétrales, âge supérieur à 69 ans et/ou CMA
438	207	272	Interventions transurétrales, âge inférieur à 70 ans sans CMA
442	80	94	Autres interventions sur les reins et les voies urinaires
454	187	193	Infections des reins et des voies urinaires, âge supérieur à 69 ans et/ou CMA
455	111	274	Infections des reins et des voies urinaires, âge de 18 à 69 ans sans CMA
456	165	256	Infections des reins et des voies urinaires, âge inférieur à 18 ans
458	251	497	Lithiases urinaires, âge inférieur à 70 ans sans CMA
459	104	136	Signes et symptômes concernant les reins et les voies urinaires, âge supérieur à 69 ans et/ou CMA
465	69	183	Autres affections des reins et des voies urinaires, âge supérieur à 69 ans et/ou CMA
466	100	184	Autres affections des reins et des voies urinaires, âge de 18 à 69 ans sans CMA
469	104	222	Affections des reins et des voies urinaires, avec CMAS
471	145	343	Insuffisance rénale, sans dialyse
476	51	86	Interventions sur le pénis
477	100	133	Prostatectomies transurétrales, âge supérieur à 69 ans et/ou CMA
481	113	116	Interventions sur les testicules pour des affections non malignes, âge inférieur à 18 ans
498	58	141	Infections et inflammations de l'appareil génital masculin

N° GHM	N <sub>Gre</sub>	N <sub>Nat</sub>	Libellés des GHM
506	128	219	Hystérectomies, âge inférieur à 70 ans sans CMA
507	56	136	Interventions réparatrices sur l'appareil génital féminin
509	358	395	Interventions sur le système utéro-annexiel pour des affections non malignes, autres que les interruptions tubaires
511	67	109	Interventions sur la vulve, le vagin ou le col utérin
512	117	138	Laparoscopies ou cœlioscopies diagnostiques
515	136	139	Dilatations et curetages, conisations pour des affections non malignes
527	149	353	Autres affections de l'appareil génital féminin
531	308	661	Césariennes sans CMA
539	323	529	Accouchements par voie basse avec complications
540	852	3158	Accouchements par voie basse sans complication
543	54	92	Grossesses ectopiques
545	259	582	Avortements avec aspiration ou curetage ou hystérotomie
546	79	107	Avortements sans aspiration, ni curetage, ni hystérotomie
547	138	257	Faux travail
548	101	162	Affections de l'ante partum, avec ou sans intervention chirurgicale, avec complications
549	207	297	Affections de l'ante partum, avec ou sans intervention chirurgicale, sans complications
568	53	55	Autres interventions pour affections du sang et des organes hématopoïétiques
570	53	86	Affections du sang et des organes hématopoïétiques, avec CMAS
571	158	356	Troubles de la lignée érythrocytaire, âge supérieur à 17 ans
573	92	179	Troubles de la coagulation
574	81	70	Affections du système réticulo-endothélial ou immunitaire, âge supérieur à 69 ans et/ou CMA
575	123	104	Affections du système réticulo-endothélial ou immunitaire, âge inférieur à 70 ans sans CMA
585	93	141	Chimiothérapie pour leucémie aiguë
586	122	157	Chimiothérapie pour autre tumeur, avec CMA
587	683	1924	Chimiothérapie pour autre tumeur, sans CMA
589	52	353	Lymphomes ou leucémies, âge supérieur à 69 ans et/ou CMA
592	294	156	Radiothérapie et surveillance
595	184	137	Antécédents ou suspicion d'affections tumorales, sans exploration endoscopique
602	121	199	Maladies infectieuses et parasitaires, avec CMAS, avec ou sans acte
604	73	117	Septicémies, âge supérieur à 17 ans
607	118	106	Fièvres d'étiologie indéterminée, âge supérieur à 69 ans et/ou CMA
609	66	113	Maladies virales, âge supérieur à 17 ans
610	168	288	Maladies virales et fièvres d'étiologie indéterminée, âge inférieur à 18 ans
611	153	341	Autres maladies infectieuses ou parasitaires
618	199	228	Troubles aigus de l'adaptation et du fonctionnement psycho-social
619	226	395	Névroses dépressives
622	236	413	Troubles mentaux d'origine organique et retard mental
623	87	105	Psychoses
625	54	96	Autres troubles mentaux
631	113	323	Éthylisme avec dépendance
632	239	337	Éthylisme sans dépendance
638	82	82	Interventions sur la main ou le poignet à la suite de blessures
639	79	84	Autres interventions pour blessures ou complications d'acte, âge supérieur à 69 ans et/ou CMA
640	111	154	Autres interventions pour blessures ou complications d'acte, âge inférieur à 70 ans sans CMA
643	84	87	Traumatismes, allergies et empoisonnements, avec CMAS
644	129	80	Traumatismes complexes, âge supérieur à 69 ans et/ou CMA
645	276	141	Traumatismes complexes, âge de 18 à 69 ans sans CMA
646	67	89	Traumatismes complexes, âge inférieur à 18 ans
649	296	285	Effets toxiques des médicaments et autres produits, âge supérieur à 69 ans et/ou CMA
650	587	854	Effets toxiques des médicaments et autres produits, âge de 18 à 69 ans sans CMA
651	115	237	Effets toxiques des médicaments et autres produits, âge inférieur à 18 ans
652	77	103	Complications iatrogéniques non classées ailleurs, âge supérieur à 69 ans et/ou CMA
653	100	125	Complications iatrogéniques non classées ailleurs, âge inférieur à 70 ans sans CMA
669	269	235	Interventions chirurgicales avec autres motifs de recours aux services de santé
671	316	338	Signes et symptômes avec CMA
672	36	711	Signes et symptômes sans CMA
674	199	1124	Suivi thérapeutique sans antécédent d'affections malignes
675	204	2288	Autres facteurs influant sur l'état de santé
873	125	214	Nouveau-nés de 2000 g à 2499 g, sans problème majeur
874	899	2690	Nouveau-nés de 2500 g et plus, sans problème significatif
875	190	433	Nouveau-nés de 2500 g et plus, avec problème majeur
876	367	956	Nouveau-nés de 2500 g et plus, avec autre problème significatif

## Annexe E

# Comparaison des quartiles de DS entre Grenoble et la base nationale

En comparant les quartiles de DS entre les GHM du CHU de Grenoble et ceux de la base nationale, nous pouvons rencontrer 31 configurations différentes, présentées ci-dessous. Dans notre fichier, nous avons rencontré 19 de ces 31 configurations et nous avons indiqué le nombre de GHM concernés (voir tableau ??). Les voici de manière détaillée:

Considérons les configurations suivantes :

- **1** :  $q_{(0.25)}^N \neq q_{(0.50)}^N \neq q_{(0.75)}^N$ ,
- **2** :  $q_{(0.25)}^N = q_{(0.50)}^N \neq q_{(0.75)}^N$ ,
- **3** :  $q_{(0.25)}^N \neq q_{(0.50)}^N = q_{(0.75)}^N$ ,
- **4** :  $q_{(0.25)}^N = q_{(0.50)}^N = q_{(0.75)}^N$ .

Considérons, pour chacune des configurations **1**, **2**, **3** et **4**, les cas suivants :

**A** :

- $q_{(0.25)}^N \neq q_{(0.25)}^G$ ,
- $q_{(0.50)}^N \neq q_{(0.50)}^G$  et
- $q_{(0.75)}^N \neq q_{(0.75)}^G$ ,

**B** :

- $q_{(0.25)}^N = q_{(0.25)}^G$ ,
- $q_{(0.50)}^N \neq q_{(0.50)}^G$  et
- $q_{(0.75)}^N \neq q_{(0.75)}^G$ ,

**C** :

- $q_{(0.25)}^N \neq q_{(0.25)}^G$ ,
- $q_{(0.50)}^N = q_{(0.50)}^G$  et

$$- q_{(0.75)}^N \neq q_{(0.75)}^G,$$

**D :**

$$- q_{(0.25)}^N \neq q_{(0.25)}^G,$$

$$- q_{(0.50)}^N \neq q_{(0.50)}^G \text{ et}$$

$$- q_{(0.75)}^N = q_{(0.75)}^G,$$

**E :**

$$- q_{(0.25)}^N = q_{(0.25)}^G,$$

$$- q_{(0.50)}^N = q_{(0.50)}^G \text{ et}$$

$$- q_{(0.75)}^N \neq q_{(0.75)}^G,$$

**F :**

$$- q_{(0.25)}^N \neq q_{(0.25)}^G,$$

$$- q_{(0.50)}^N = q_{(0.50)}^G \text{ et}$$

$$- q_{(0.75)}^N = q_{(0.75)}^G,$$

**G :**

$$- q_{(0.25)}^N = q_{(0.25)}^G,$$

$$- q_{(0.50)}^N \neq q_{(0.50)}^G \text{ et}$$

$$- q_{(0.75)}^N = q_{(0.75)}^G,$$

**H :**

$$- q_{(0.25)}^N = q_{(0.25)}^G,$$

$$- q_{(0.50)}^N = q_{(0.50)}^G \text{ et}$$

$$- q_{(0.75)}^N = q_{(0.75)}^G.$$

TAB. E.1 – Récapitulatif des 31 configurations possibles avec leurs effectifs.

	Configurations				Total (% colonne)
	1 $n_1$	2 $n_2$	3 $n_3$	4 $n_4$	
Cas					
A	91	1	0	0	92 (35.8)
B	46	6	0	0	52 (20.2)
C	11	0	1	0	12 (4.7)
D	14	0	0	0	14 (5.4)
E	30	5	0	2	37 (14.4)
F	7	1	0	0	8 (3.1)
G	13	1	1	Impossible	15 (5.8)
H	18	5	2	2	27 (10.5)
Total (% ligne)	230 (89.5)	19 (7.4)	4 (1.55)	4 (1.55)	257 (100)





## Annexe F

# Liste de référence fondée sur les points ISA

Le tableau ci-après regroupe les 50 GHM qui causent 80% du déficit en points ISA du CHU de Grenoble (données PMSI 1997).

N° GHM	Déficit en points ISA	Classement
1	-396896	2
2	-124131	14
14	-86258	24
15	-126116	13
17	-47748	45
18	-349030	3
21	-65009	32
48	-501480	1
51	-274630	4
54	-167506	8
61	-97278	20
63	-96370	21
68	-97290	19
82	-151869	10
122	-70148	28
153	-56580	40
182	-59544	39
185	-120267	16
188	-111294	17
201	-63490	34
213	-68388	29
246	-65969	30
248	-45780	47
285	-52041	41
286	-51974	42
287	-60210	37
295	-203439	5
302	-160860	9
343	-135641	11
372	-45535	48
375	-61540	36
397	-87080	23
398	-59616	38
405	-47073	46
418	-109224	18
425	-43128	50
430	-48257	44
432	-62582	35
437	-122382	15
438	-64666	33
570	-79489	26
571	-43865	49
585	-175924	7
587	-91140	22
595	-65296	31
619	-73139	27
644	-51510	43
650	-182016	6
669	-133155	12
875	-81780	25

## Annexe G

### Liste de référence fondée sur le “test des quantiles”

Le tableau ci-après donne la liste des 26 GHM grenoblois dont les quartiles de DS,  $q_{(0.25)}^G$ ,  $q_{(0.50)}^G$  et  $q_{(0.75)}^G$  sont strictement supérieurs, de manière significative, à ceux de la base nationale,  $q_{(0.25)}^N$ ,  $q_{(0.50)}^N$  et  $q_{(0.75)}^N$ , au seuil du  $\chi^2$  de 5% :

N° GHM	$q_{(0.25)}^N$	$q_{(0.25)}^G$	$p_{q(0.25)}$	$q_{(0.50)}^N$	$q_{(0.50)}^G$	$p_{q(0.50)}$	$q_{(0.75)}^N$	$q_{(0.75)}^G$	$p_{q(0.75)}$
1	7	8	0.013	12	14	<0.0001	21	25	<0.0001
17	2	3	<0.0001	4	5	<0.0001	6	7.5	<0.0001
18	3	5	0.002	9	10	0.037	15	20	0.037
24	3	4.75	0.026	6.5	9.5	0.025	12.25	15.75	0.025
40	4	6	<0.0001	10	12	0.013	19	21	0.013
48	2	7	<0.0001	4	9	<0.0001	6	12	0.000
54	2	6	<0.0001	4	7	<0.0001	6.75	9	0.000
61	2	5	<0.0001	3	7	<0.0001	6	10	0.000
71	2	6	<0.0001	4	7	<0.0001	6.75	8	0.000
82	2	4	<0.0001	4	6	<0.0001	7	10.75	0.000
122	2	3	<0.0001	4	8	<0.0001	12	15	<0.0001
185	4	7	0.007	8	11	<0.0001	12	16	<0.0001
285	2	3	0.031	4	7	<0.0001	10	11.75	<0.0001
287	1	3	<0.0001	2	7	<0.0001	7	12	0.000
296	10	13	<0.0001	13	15	<0.0001	18	19	<0.0001
297	7	9	0.007	9	13	<0.0001	13	15	<0.0001
315	2	3	0.012	4	5	<0.0001	6	6.5	<0.0001
339	2	5	<0.0001	7	9	0.017	14	15	0.017
397	3	6	0.023	7	9	0.013	11	15	0.013
418	4	7	<0.0001	5	10	0.000	10	13.5	0.000
419	3	4	0.047	4	8	<0.0001	8	12	<0.0001
430	8	11	<0.0001	10	13	<0.0001	15	16	<0.0001
432	4	7	<0.0001	6	8	<0.0001	9	11	<0.0001
437	3	4	<0.0001	4	6	<0.0001	6	11	<0.0001
438	2	3	0.000	3	4	<0.0001	5	6	0.000
506	5	6	<0.0001	6	7	<0.0001	8	9	<0.0001
543	2	3	0.031	3	4	0.005	4	5.75	0.005
652	2	4	0.002	4	7	0.027	10	14	0.027
875	5	6	<0.0001	6	9	0.000	9	13	0.000



## Annexe H

### Liste de référence fondée sur le “test du $\chi^2$ ”

Le tableau ci-après donne la liste des 36 GHM grenoblois ayant des effectifs de patients sur les 4 intervalles interquartiles,  $I_{(1)}$ ,  $I_{(2)}$ ,  $I_{(3)}$  et  $I_{(4)}$ , significativement différentes de  $n/4$ ,  $n$  correspondant à l'effectif total des patients grenoblois pour chaque GHM.

N° GHM	$I_{(1)}^*$	$I_{(2)}^*$	$I_{(3)}^*$	$I_{(4)}^*$	$n/4$	$\chi^2$	$p$
1	78	100	109	136	16.35	105.75	<0.0001
14	11	7	16	25	12.25	14.75	0.007
18	122	168	119	208	34.75	154.25	<0.0001
24	10	12	19	27	10.47	17.00	0.015
40	74	114	108	122	12.81	104.50	0.005
48	4	9	29	187	398.20	57.25	<0.0001
51	77	77	184	255	153.97	148.25	<0.0001
54	7	28	86	107	117.23	57.00	<0.0001
61	15	2	32	75	97.87	31.00	<0.0001
68	19	12	34	27	11.91	23.00	0.008
71	1	3	34	22	50.00	15.00	<0.0001
82	18	32	57	79	47.08	46.50	<0.0001
122	45	34	62	68	13.95	52.25	0.003
185	22	33	42	63	22.65	40.00	<0.0001
285	32	19	53	42	17.23	36.50	<0.0001
287	7	12	27	36	26.20	20.50	<0.0001
288	9.5	9.5	10	28	17.70	14.25	<0.0001
296	16	20	52	37	26.33	31.25	<0.0001
297	14	13	31	42	23.60	25.00	<0.0001
339	12	34	39	30	14.43	28.75	0.002
386	12	12	6	20	7.92	12.50	0.048
397	19	23	26	43	12.06	27.75	0.007
418	40	5	68	90	79.74	50.75	<0.0001
419	10	4	14	25	17.72	13.25	<0.0001
430	1	10	26	17	24.96	13.50	<0.0001
432	1	19	32	38	35.78	22.50	<0.0001
437	36	32	34	70	22.79	43.00	<0.0001
438	24	39	85	59	40.40	51.75	<0.0001
481	0	38.5	38.5	36	37.81	28.25	<0.0001
506	13	29	46	40	19.69	32.00	<0.0001
570	8	11	10	24	11.98	13.25	0.007
644	21	25	42	41	10.88	32.25	0.012
649	60	60	92	84	11.03	74.00	0.012
650	127.3	127.3	127.3	205	30.85	146.75	<0.0001
652	10	16	26	25	9.08	19.25	0.028
875	45	17	48	80	41.96	47.50	<0.0001

\* effectif de patient grenoblois dans chaque intervalle interquartile

# Annexe I

## Groupe des “mauvais FA”

Le tableau ci-dessous donne la liste détaillée (numéro de GHM, proportions cumulées de patients dans les 3 intervalles interquartiles, déficit ou bénéfice en points ISA et différence d’aire entre Grenoble et la référence nationale) des 24 GHM grenoblois faisant partie d’un groupe de “mauvais FA”, ayant un profil empirique de type (a) et issu de la classification hiérarchique ascendante.

N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d’aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
14	0.186	0.305	0.576	-86258	-0.108
24	0.147	0.324	0.603	-40420	-0.107
48	0.017	0.057	0.183	-501480	-0.311
51	0.130	0.260	0.570	-274630	-0.135
54	0.031	0.154	0.531	-167506	-0.196
61	0.121	0.137	0.395	-97278	-0.212
71	0.017	0.067	0.633	-37281	-0.196
82	0.097	0.269	0.575	-151869	-0.140
185	0.138	0.344	0.606	-120267	-0.103
287	0.085	0.232	0.561	-60210	-0.156
288	0.167	0.333	0.509	-38318	-0.123
297	0.140	0.270	0.580	-15340	-0.128
311	0.235	0.373	0.608	-13730	-0.071
358	0.200	0.360	0.620	-33300	-0.080
397	0.171	0.378	0.613	-87080	-0.085
405	0.035	0.175	0.772	-47073	-0.130
418	0.197	0.222	0.557	-109224	-0.131
419	0.189	0.264	0.528	-34392	-0.130
430	0.019	0.204	0.685	-48257	-0.148
432	0.011	0.222	0.578	-62582	-0.172
437	0.209	0.395	0.593	-122382	-0.076
543	0.167	0.389	0.593	-14625	-0.088
570	0.151	0.358	0.547	-79489	-0.111
875	0.237	0.326	0.579	-81780	-0.090





## Annexe J

### Groupe des “mauvais FG”

Le tableau ci-dessous donne la liste détaillée des 16 GHM grenoblois faisant partie d'un groupe de “mauvais FG”, ayant un profil empirique de type  $(g)$  et issu de la classification hiérarchique ascendante.

N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d'aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
21	0.369	0.563	0.670	-65009	0.026
38	0.358	0.525	0.726	-3018	0.027
106	0.362	0.643	0.729	4605	0.059
188	0.450	0.606	0.711	-111294	0.067
201	0.341	0.572	0.734	-63490	0.037
214	0.329	0.562	0.712	10308	0.026
246	0.364	0.523	0.652	-65969	0.010
255	0.378	0.490	0.724	-19425	0.023
347	0.417	0.604	0.729	2222	0.063
398	0.408	0.517	0.659	-59616	0.021
425	0.412	0.541	0.647	-43128	0.025
458	0.382	0.570	0.717	-18600	0.042
466	0.440	0.500	0.630	-25592	0.018
585	0.387	0.538	0.656	-175924	0.020
586	0.443	0.508	0.746	-2526	0.049
619	0.460	0.527	0.690	-73139	0.044



## Annexe K

### Groupe des “autres F”

Les deux tableaux ci-après donnent la liste détaillée des 97 GHM grenoblois faisant partie d'un groupe de “autres F”, sans profil particulier et issu de la classification hiérarchique ascendante.

N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d'aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
1	0.184	0.421	0.678	-396896	-0.054
2	0.243	0.514	0.703	-124131	-0.010
3	0.293	0.534	0.724	36360	0.013
15	0.277	0.415	0.564	-126116	-0.061
17	0.208	0.428	0.679	-47748	-0.046
18	0.198	0.470	0.663	-349030	-0.042
22	0.280	0.462	0.731	-19308	-0.007
25	0.279	0.361	0.820	-10392	-0.010
27	0.344	0.444	0.633	-17556	-0.020
37	0.335	0.418	0.741	-1550	-0.002
40	0.177	0.450	0.708	-17050	-0.041
52	0.218	0.436	0.654	-15030	-0.048
63	0.338	0.418	0.605	-96370	-0.035
68	0.207	0.337	0.707	-97290	-0.062
72	0.070	0.523	0.907	15576	0.000
74	0.108	0.639	0.841	9080	0.022
75	0.247	0.494	0.699	-20940	-0.015
113	0.193	0.403	0.840	5634	-0.016
122	0.215	0.378	0.675	-70148	-0.058
127	0.222	0.540	0.808	139000	0.018
128	0.248	0.501	0.796	19700	0.011
141	0.233	0.566	0.860	44505	0.040
153	0.193	0.430	0.667	-56580	-0.053
155	0.187	0.522	0.770	0	-0.005
160	0.181	0.619	0.812	-19605	0.028
181	0.248	0.612	0.813	96110	0.043
197	0.261	0.379	0.817	-10512	-0.011
213	0.236	0.460	0.733	-68388	-0.018
215	0.302	0.492	0.778	12820	0.018
227	0.185	0.444	0.722	-8315	-0.037
228	0.067	0.480	0.827	5058	-0.032
230	0.033	0.917	0.942	14865	0.098
232	0.284	0.432	0.838	-32655	0.014
248	0.164	0.300	0.793	-45780	-0.061
254	0.227	0.430	0.695	-14355	-0.037
259	0.377	0.447	0.667	-38646	-0.002
262	0.270	0.443	0.741	4864	-0.012
274	0.144	0.548	0.731	-8515	-0.019
284	0.208	0.425	0.783	-9285	-0.021
285	0.219	0.349	0.712	-52041	-0.055
286	0.242	0.411	0.710	-51974	-0.034
293	0.235	0.482	0.776	6848	-0.002
295	0.060	0.394	0.755	-203439	-0.073
296	0.128	0.288	0.704	-33780	-0.095
301	0.240	0.507	0.733	-13194	-0.005
302	0.075	0.472	0.752	-160860	-0.050
306	0.217	0.424	0.870	27626	0.003
308	0.156	0.629	0.763	-10146	0.012

N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d'aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
309	0.119	0.394	0.853	27600	-0.034
312	0.307	0.486	0.743	9830	0.009
313	0.149	0.268	0.797	-23596	-0.072
315	0.171	0.457	0.749	-6105	-0.031
320	0.326	0.496	0.852	38522	0.044
339	0.104	0.400	0.739	-6165	-0.064
340	0.200	0.433	0.633	-28230	-0.059
341	0.294	0.422	0.755	-8100	-0.007
343	0.279	0.445	0.682	-135641	-0.024
344	0.183	0.433	0.700	-14769	-0.046
355	0.200	0.400	0.790	7748	-0.028
367	0.318	0.476	0.818	42768	0.028
372	0.112	0.224	0.845	-45535	-0.080
375	0.179	0.372	0.705	-61540	-0.061
386	0.240	0.480	0.600	-38760	-0.045
391	0.266	0.394	0.716	-23062	-0.031
392	0.194	0.473	0.828	-11869	-0.001
396	0.238	0.476	0.738	-18180	-0.012
408	0.095	0.413	0.810	0	-0.046
409	0.091	0.624	0.812	-17908	0.007
420	0.300	0.540	0.640	-3190	-0.005
438	0.116	0.304	0.715	-64666	-0.091
442	0.225	0.562	0.738	2679	0.006
465	0.319	0.406	0.638	-24890	-0.034
469	0.183	0.413	0.750	-17442	-0.039
471	0.248	0.407	0.697	-32494	-0.037
481	0.000	0.341	0.681	-22950	-0.120
498	0.293	0.466	0.672	-2697	-0.017
506	0.102	0.328	0.688	-32164	-0.096
511	0.269	0.463	0.597	-35881	-0.043
571	0.291	0.481	0.696	-43865	-0.008
573	0.380	0.446	0.761	-21097	0.022
574	0.296	0.543	0.741	15534	0.020
575	0.224	0.447	0.715	-1187	-0.029
589	0.173	0.423	0.635	-32997	-0.067
604	0.205	0.438	0.836	27269	-0.005
609	0.197	0.561	0.879	7322	0.034
611	0.275	0.497	0.732	-26298	0.001
625	0.241	0.481	0.759	5530	-0.005
638	0.329	0.451	0.866	7404	0.037
639	0.190	0.595	0.823	36808	0.027
644	0.163	0.357	0.682	-51510	-0.075
645	0.319	0.475	0.598	-28347	-0.027
646	0.261	0.522	0.701	-10676	-0.004
649	0.203	0.405	0.716	8700	-0.044
650	0.217	0.434	0.651	-182016	-0.050
652	0.130	0.338	0.675	-26660	-0.089
653	0.180	0.380	0.780	-3090	-0.040
669	0.286	0.483	0.721	-133155	-0.003



## Annexe L

### Groupe des “bons E”

Les deux tableaux ci-après donnent la liste détaillée des 120 GHM grenoblois faisant partie du groupe des “bons E” (GHM ayant des profils empiriques de types *(a)* et *(neutre)*).



N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d'aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
5	0.706	0.847	0.918	78816	0.243
8	0.642	0.730	0.912	63038	0.196
16	0.296	0.520	0.821	40125	0.034
19	0.429	0.638	0.876	70176	0.111
23	0.250	0.500	0.750	-3525	0.000
28	0.250	0.500	0.750	-3702	0.000
29	0.250	0.500	0.750	-5944	0.000
30	0.580	0.710	0.815	-3290	0.151
45	0.446	0.652	0.882	50940	0.120
46	0.250	0.500	0.750	10608	0.000
53	0.250	0.500	0.750	-16514	0.000
94	0.555	0.654	0.785	7455	0.124
95	0.250	0.500	0.750	11022	0.000
99	0.585	0.585	0.829	667	0.125
100	0.691	0.765	0.826	30368	0.196
102	0.278	0.896	0.944	23310	0.155
103	0.305	0.679	0.802	21112	0.072
104	0.519	0.714	0.896	11952	0.157
105	0.250	0.500	0.750	-16450	0.000
118	0.250	0.500	0.750	-16500	0.000
119	0.368	0.579	0.754	13216	0.050
125	0.292	0.600	0.769	-5724	0.040
129	0.362	0.678	0.859	86961	0.100
130	0.272	0.592	0.845	10664	0.052
131	0.625	0.677	0.844	23391	0.162
136	0.303	0.589	0.805	38688	0.049
137	0.419	0.612	0.845	8667	0.094
138	0.489	0.676	0.832	87374	0.124
139	0.459	0.628	0.804	18992	0.098
140	0.425	0.851	0.936	288754	0.178
142	0.385	0.583	0.859	38304	0.082
154	0.419	0.729	0.897	370328	0.136
158	0.308	0.649	0.827	183125	0.071
159	0.458	0.780	0.915	48565	0.163
164	0.413	0.553	0.767	-18848	0.058
167	0.405	0.810	0.889	9370	0.151
168	0.463	0.716	0.768	84458	0.112
178	0.406	0.678	0.850	123030	0.109
179	0.345	0.635	0.829	149568	0.077
182	0.250	0.500	0.750	-59544	0.000
184	0.315	0.546	0.844	140346	0.051
187	0.375	0.526	0.751	-37950	0.038
191	0.266	0.605	0.819	36855	0.048
195	0.437	0.575	0.787	-2616	0.075
196	0.585	0.734	0.894	18336	0.178
198	0.417	0.548	0.762	10827	0.057
199	0.419	0.605	0.837	15158	0.090
200	0.250	0.500	0.750	13275	0.000
202	0.394	0.611	0.773	-20880	0.070
225	0.250	0.500	0.750	-1223	0.000
226	0.327	0.673	0.909	31560	0.102
229	0.520	0.786	0.878	7200	0.171
256	0.410	0.581	0.825	47109	0.079
257	0.544	0.746	0.864	78890	0.164
258	0.667	0.825	0.880	89698	0.218
260	0.250	0.500	0.750	-24115	0.000
273	0.314	0.638	0.838	66550	0.073
289	0.319	0.521	0.782	11304	0.031
290	0.300	0.630	0.840	25680	0.068
304	0.390	0.525	0.763	17472	0.045

N° GHM	Proportions cumulées de patients grenoblois			Déficit ou bénéfice en points ISA	Différence d'aire
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle		
305	0.266	0.577	0.779	70668	0.031
310	0.250	0.500	0.750	41850	0.000
317	0.535	0.785	0.869	52765	0.172
318	0.436	0.872	0.947	76046	0.189
336	0.533	0.780	0.867	109327	0.170
348	0.406	0.578	0.781	-7126	0.066
349	0.250	0.500	0.750	29995	0.000
351	0.250	0.500	0.750	-11025	0.000
352	0.250	0.500	0.750	-11760	0.000
353	0.440	0.736	0.920	51842	0.149
354	0.487	0.677	0.862	58473	0.132
356	0.456	0.569	0.788	-4740	0.078
376	0.391	0.609	0.772	37584	0.068
377	0.250	0.500	0.750	41960	0.000
394	0.250	0.500	0.750	22484	0.000
395	0.279	0.557	0.877	17730	0.053
421	0.724	0.816	0.827	39368	0.217
422	0.630	0.778	0.796	-10180	0.176
429	0.500	0.688	0.891	122130	0.145
454	0.460	0.717	0.856	65736	0.133
455	0.396	0.568	0.793	6880	0.064
456	0.339	0.891	0.939	39091	0.167
459	0.346	0.596	0.808	23496	0.063
476	0.294	0.647	0.843	3110	0.071
477	0.270	0.610	0.800	7636	0.045
507	0.304	0.679	0.839	-2183	0.081
509	0.559	0.682	0.830	58995	0.143
512	0.250	0.500	0.750	-12585	0.000
515	0.250	0.500	0.750	9452	0.000
527	0.250	0.500	0.750	-11560	0.000
531	0.250	0.500	0.750	61560	0.000
539	0.356	0.712	0.833	67773	0.100
540	0.538	0.836	0.922	214000	0.199
545	0.250	0.500	0.750	384	0.000
546	0.250	0.500	0.750	1712	0.000
547	0.464	0.667	0.833	15700	0.116
548	0.475	0.594	0.822	4560	0.098
549	0.333	0.546	0.826	13962	0.051
568	0.415	0.717	0.792	20631	0.106
587	0.537	0.601	0.856	-91140	0.124
592	0.354	0.854	0.905	116615	0.153
595	0.397	0.614	0.750	-65296	0.065
602	0.289	0.570	0.810	100128	0.042
607	0.339	0.585	0.763	-20696	0.047
610	0.661	0.756	0.845	37950	0.191
618	0.583	0.688	0.799	58464	0.143
622	0.373	0.589	0.763	30416	0.056
623	0.736	0.782	0.874	63720	0.223
631	0.372	0.735	0.938	53781	0.136
632	0.444	0.887	0.933	44856	0.191
640	0.306	0.631	0.856	59436	0.073
643	0.298	0.607	0.893	87584	0.075
651	0.250	0.500	0.750	-3633	0.000
671	0.354	0.573	0.759	-8190	0.047
672	0.520	0.705	0.803	85255	0.132
674	0.250	0.500	0.750	139808	0.000
675	0.428	0.665	0.858	218250	0.113
873	0.408	0.648	0.768	13932	0.081
874	0.700	0.780	0.860	98990	0.210
876	0.250	0.500	0.750	2145	0.000



## Annexe M

# Tableaux des coefficients de classement des GHM

TAB. M.1 – *Coefficients des fonctions discriminantes (première analyse).*

Variables	Fonction 1	Fonction 2	Fonction 3
Facteur 1	6.84	-1.44	4.47
Facteur 2	-4.95	10.20	5.24
Déficit en points ISA	2.58E <sup>-06</sup>	7.41E <sup>-06</sup>	1.29E <sup>-05</sup>
Constante	0.50	-0.79	0.40

TAB. M.2 – *Coefficients de classement : fonction linéaire discriminante de Fisher (première analyse).*

Variables	Autres M	Mauvais FA	Mauvais FG	Bons E
Facteur 1	-2.57	11.70	-16.70	-9.27
Facteur 2	-3.49	-4.75	25.12	2.89
Déficit en points ISA	-0.7E <sup>-05</sup>	-0.99E <sup>-05</sup>	-0.15E <sup>-04</sup>	-0.5E <sup>-06</sup>
Constante	-1.56	-3.30	-4.76	-2.16

TAB. M.3 – *Coefficients des fonctions discriminantes (deuxième analyse).*

Variables	Fonction 1	Fonction 2	Fonction 3
Premier intervalle	4.06	-5.58	-8.31
Deuxième intervalle	4.14	-1.99	-12.42
Troisième intervalle	5.50	12.09	9.17
Constante	-7.61	-6.38	2.95

TAB. M.4 – *Coefficients de classement : fonction linéaire discriminante de Fisher, (deuxième analyse).*

Variables	Autres M	Mauvais FA	Mauvais FG	Bons E
Premier intervalle	13.74	12.66	25.79	25.39
Deuxième intervalle	-21.20	-30.77	-11.72	-13.31
Troisième intervalle	146.31	123.91	129.08	152.08
Constante	-51.95	-33.74	-48.11	-64.82

## Annexe N

# Algorithme EM

Cette annexe présente l'algorithme EM dans le cas du mélange gaussien univarié.

À partir d'une solution initiale  $\theta^0 = ((p_k^0, \mathbf{a}_k^0), k = 1, \dots, K)$ , la  $m^{\text{ième}}$  itération de l'algorithme EM est la suivante :

- **étape E (estimation)** : estimation de toutes les probabilités d'appartenance des  $\mathbf{x}_i$  aux classes, connaissant le paramètre vectoriel  $\theta^{m-1}$  ; pour  $k = 1, \dots, K$ ,  $i = 1, \dots, n$ , calcul de :

$$t_k^m(\mathbf{x}_i; \theta^{m-1}) = \frac{p_k^{m-1} f(\mathbf{x}_i; \mathbf{a}_k^{m-1})}{\sum_{k=1}^K p_k^{m-1} f(\mathbf{x}_i; \mathbf{a}_k^{m-1})}, \quad (\text{N.1})$$

où  $t_k^m(\mathbf{x}_i; \theta^{m-1})$  est, à la  $m^{\text{ième}}$  itération, la probabilité d'appartenance de  $\mathbf{x}_i$  au composant numéro  $k$  ;

- **étape M (maximisation)** : choix de l'estimateur  $\theta^m$  qui maximise la vraisemblance complétée  $L(\theta)$  ; pour  $k = 1, \dots, K$ , calcul de :

$$p_k^m = \frac{1}{n} \sum_{i=1}^n t_k^m(\mathbf{x}_i; \theta^{m-1}),$$

(estimateur du maximum de vraisemblance des proportions du mélange) et résolution des équations de vraisemblance pour le paramètre vectoriel  $\mathbf{a}_k^m = (\mathbf{a}_{kj}^m, j = 1, \dots, d) \text{ de } \mathbb{R}^d$  :

$$\forall k = 1, \dots, K \quad j = 1, \dots, d \quad \sum_{i=1}^n t_k^m(\mathbf{x}_i; \theta^{m-1}) \frac{\partial \text{Log } f(\mathbf{x}_i; \mathbf{a}_k^{m-1})}{\partial \mathbf{a}_{kj}^m} = 0.$$

L'estimation des  $\mathbf{a}_k^m$  dépend de la famille paramétrée, posée a priori, des composants du mélange. Dans le cas gaussien,  $\mathbf{a}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ , et la résolution des équations de

vraisemblance donne :

$$\boldsymbol{\mu}_k^m = \frac{\sum_{i=1}^n t_k^m(\mathbf{x}_i; \boldsymbol{\theta}^{m-1}) \mathbf{x}_i}{\sum_{i=1}^n t_k^m(\mathbf{x}_i; \boldsymbol{\theta}^{m-1})},$$

$$\boldsymbol{\Sigma}_k^m = \frac{\sum_{i=1}^n t_k^m(\mathbf{x}_i; \boldsymbol{\theta}^{m-1}) (\mathbf{x}_i - \boldsymbol{\mu}_k^m)(\mathbf{x}_i - \boldsymbol{\mu}_k^m)^t}{\sum_{i=1}^n t_k^m(\mathbf{x}_i; \boldsymbol{\theta}^{m-1})}.$$

Ces équations montrent que les probabilités conditionnelles  $t_k^m(\mathbf{x}_i; \boldsymbol{\theta}^{m-1})$  peuvent être vues comme des poids attachés aux points de l'échantillon.

Dans le cas des mélanges gaussiens multivariés contraints, il faudra, à l'étape M, estimer les centres  $\boldsymbol{\mu}_k^m$ , les proportions  $p_k^m$  et les matrices de variance-covariance  $\boldsymbol{\Sigma}_k^m$  propres au modèle gaussien choisi.

### Remarques sur l'algorithme EM

Si le nombre de classes du modèle est connu, et si celles-ci sont bien séparées d'une part, si les proportions ne sont pas trop déséquilibrées et si les valeurs initiales de  $p_k^0$  et  $\boldsymbol{\Sigma}_k^0$  ne sont pas trop éloignées de la solution optimale inconnue d'autre part, l'algorithme EM donne de bons résultats. De plus, ces résultats pouvant dépendre de la position initiale de l'algorithme, il est utile de la faire varier, pour voir si la solution reste la même. Par ailleurs, l'algorithme EM n'a pas de condition d'arrêt spécifique; en conséquence, il faut choisir un grand nombre d'itérations pour lui éviter de converger vers un maximum local dépendant lui aussi des conditions initiales de l'algorithme.

## Annexe O

### Groupe des “mauvais M”

Le tableau ci-après donne la liste détaillée (numéro de GHM et proportions cumulées de patients dans les 3 intervalles interquartiles) des 52 GHM grenoblois faisant partie du groupe des “mauvais M”, issu du modèle de mélange de 3 lois normales.



N° GHM	Proportions cumulées de patients grenoblois			$p_1^{(1)}$	$p_2^{(2)}$	$p_3^{(3)}$	Déficit en points ISA
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle				
14	0.186	0.305	0.576	0.0929	0.907	0.0001	-86258
15	0.277	0.415	0.564	0.4849	0.5151	0	-126116
24	0.147	0.324	0.603	0.2201	0.7796	0.0003	-40420
25	0.279	0.361	0.820	0	0.9982	0.0018	-10392
48	0.017	0.057	0.183	0	1	0	-501480
51	0.130	0.260	0.570	0.0191	0.9808	0	-274630
54	0.031	0.154	0.531	0.0002	0.9998	0	-167506
61	0.121	0.137	0.395	0	1	0	-97278
68	0.207	0.337	0.707	0.0234	0.9598	0.0168	-97290
71	0.017	0.067	0.633	0	1	0	-37281
82	0.097	0.269	0.575	0.0435	0.9565	0.0001	-151869
113	0.193	0.403	0.840	0.0055	0.9935	0.001	5634
122	0.215	0.378	0.675	0.4499	0.5393	0.0109	-70148
185	0.138	0.344	0.606	0.3688	0.6309	0.0003	-120267
197	0.261	0.379	0.817	0.0001	0.9974	0.0025	-10512
232	0.284	0.432	0.838	0.0018	0.9914	0.0067	-32655
248	0.164	0.300	0.793	0	0.9999	0.0001	-45780
284	0.208	0.425	0.783	0.3546	0.6041	0.0413	-9285
285	0.219	0.349	0.712	0.0344	0.942	0.0236	-52041
287	0.085	0.232	0.561	0.008	0.992	0	-60210
288	0.167	0.333	0.509	0.0244	0.9756	0	-38318
296	0.128	0.288	0.704	0.0052	0.9913	0.0035	-33780
297	0.140	0.270	0.580	0.0275	0.9724	0.0001	-15340
306	0.217	0.424	0.870	0.0013	0.9985	0.0002	27626
309	0.119	0.394	0.853	0.063	0.9361	0.0009	27600
311	0.235	0.373	0.608	0.4906	0.5091	0.0003	-13730
313	0.149	0.268	0.797	0	1	0	-23596
320	0.326	0.496	0.852	0.0386	0.902	0.0594	38522
341	0.294	0.422	0.755	0.1219	0.7135	0.1646	-8100
355	0.200	0.400	0.790	0.0724	0.9115	0.0161	7748
358	0.200	0.360	0.620	0.415	0.5844	0.0007	-33300
367	0.318	0.476	0.818	0.0902	0.7654	0.1444	42768
372	0.112	0.224	0.845	0	1	0	-45535
375	0.179	0.372	0.705	0.3839	0.596	0.0201	-61540
386	0.240	0.480	0.600	0.3259	0.674	0	-38760
391	0.266	0.394	0.716	0.2018	0.7231	0.0751	-23062
405	0.035	0.175	0.772	0	1	0	-47073
418	0.197	0.222	0.557	0.0014	0.9985	0.0001	-109224
419	0.189	0.264	0.528	0.0191	0.9809	0	-34392
430	0.019	0.204	0.685	0.0004	0.999	0.0007	-48257
432	0.011	0.222	0.578	0.0248	0.975	0.0002	-62582
437	0.209	0.395	0.593	0.4633	0.5367	0.0001	-122382
438	0.116	0.304	0.715	0.0222	0.9719	0.0059	-64666
506	0.102	0.328	0.688	0.3638	0.622	0.0142	-32164
543	0.167	0.389	0.593	0.4091	0.5908	0.0001	-14625
570	0.151	0.358	0.547	0.0944	0.9056	0	-79489
604	0.205	0.438	0.836	0.1033	0.8881	0.0086	27269
638	0.329	0.451	0.866	0.0002	0.9954	0.0044	7404
644	0.163	0.357	0.682	0.3904	0.5986	0.011	-51510
652	0.130	0.338	0.675	0.3592	0.6316	0.0092	-26660
653	0.180	0.380	0.780	0.0521	0.9362	0.0116	-3090
875	0.237	0.326	0.579	0.1967	0.8031	0.0001	-81780

(1)  $p_1$  : probabilité d'appartenir à la première composante,(2)  $p_2$  : probabilité d'appartenir à la deuxième composante,(3)  $p_3$  : probabilité d'appartenir à la troisième composante.

## Annexe P

### Groupe des “autres M”

Les deux tableaux ci-après donnent la liste détaillée (numéro de GHM et proportions cumulées de patients dans les 3 intervalles interquartiles) des 127 GHM grenoblois appartenant au groupe des “autres M”, issu du modèle de mélange de 3 lois normales.

N° GHM	Proportions cumulées de patients grenoblois						Déficit en points ISA
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle	$p_1^{(1)}$	$p_2^{(2)}$	$p_3^{(3)}$	
1	0.184	0.421	0.678	0.8472	0.1453	0.0075	-396896
2	0.243	0.514	0.703	0.9101	0.0656	0.0243	-124131
3	0.293	0.534	0.724	0.9125	0.0510	0.0365	36360
16	0.296	0.520	0.821	0.6278	0.2309	0.1413	40125
17	0.208	0.428	0.679	0.8354	0.1576	0.0070	-47748
18	0.198	0.470	0.663	0.8574	0.1367	0.0059	-349030
21	0.369	0.563	0.670	0.9131	0.0839	0.0030	-65009
22	0.280	0.462	0.731	0.7755	0.1707	0.0538	-19308
23	0.250	0.500	0.750	0.8936	0.0537	0.0527	-3525
27	0.344	0.444	0.633	0.9149	0.0836	0.0015	-17556
28	0.250	0.500	0.750	0.8936	0.0537	0.0527	-3702
29	0.250	0.500	0.750	0.8936	0.0537	0.0527	-5944
38	0.358	0.525	0.726	0.8960	0.0449	0.0590	-3018
40	0.177	0.450	0.708	0.9204	0.0573	0.0223	-17050
46	0.250	0.500	0.750	0.8936	0.0537	0.0527	10608
52	0.218	0.436	0.654	0.8146	0.1833	0.0021	-15030
53	0.250	0.500	0.750	0.8936	0.0537	0.0527	-16514
63	0.338	0.418	0.605	0.8790	0.1206	0.0004	-96370
72	0.070	0.523	0.907	0.9859	0.0001	0.0140	15576
75	0.247	0.494	0.699	0.9137	0.0719	0.0144	-20940
95	0.250	0.500	0.750	0.8936	0.0537	0.0527	11022
105	0.250	0.500	0.750	0.8936	0.0537	0.0527	-16450
106	0.362	0.643	0.729	0.5871	0.1547	0.2582	4605
118	0.250	0.500	0.750	0.8936	0.0537	0.0527	-16500
119	0.368	0.579	0.754	0.8350	0.0328	0.1321	13216
125	0.292	0.600	0.769	0.6105	0.0262	0.3633	-5724
127	0.222	0.540	0.808	0.8468	0.0156	0.1376	139000
128	0.248	0.501	0.796	0.8236	0.0964	0.0800	19700
130	0.272	0.592	0.845	0.7195	0.0201	0.2604	10664
136	0.303	0.589	0.805	0.6963	0.0250	0.2786	38688
141	0.233	0.566	0.860	0.8685	0.0216	0.1099	44505
153	0.193	0.430	0.667	0.8459	0.1494	0.0047	-56580
155	0.187	0.522	0.770	0.8239	0.0095	0.1666	0
164	0.413	0.553	0.767	0.5286	0.0282	0.4432	-18848
182	0.250	0.500	0.750	0.8936	0.0537	0.0527	-59544
184	0.315	0.546	0.844	0.5825	0.2465	0.1710	140346
187	0.375	0.526	0.751	0.7355	0.0541	0.2104	-37950
188	0.450	0.606	0.711	0.9421	0.0149	0.0430	-111294
198	0.417	0.548	0.762	0.5138	0.0260	0.4602	10827
200	0.250	0.500	0.750	0.8936	0.0537	0.0527	13275
201	0.341	0.572	0.734	0.8897	0.0437	0.0665	-63490
202	0.394	0.611	0.773	0.6852	0.0256	0.2892	-20880
213	0.236	0.460	0.733	0.8512	0.1141	0.0347	-68388
214	0.329	0.562	0.712	0.9101	0.0608	0.0291	10308
215	0.302	0.492	0.778	0.6461	0.2066	0.1472	12820
225	0.250	0.500	0.750	0.8936	0.0537	0.0527	-1223
227	0.185	0.444	0.722	0.9001	0.0738	0.0261	-8315
228	0.067	0.480	0.827	0.8867	0.0003	0.1130	5058
246	0.364	0.523	0.652	0.9389	0.0601	0.0010	-65969
254	0.227	0.430	0.695	0.8069	0.1799	0.0132	-14355
255	0.378	0.490	0.724	0.7141	0.0700	0.2159	-19425
259	0.377	0.447	0.667	0.8784	0.0702	0.0513	-38646
260	0.250	0.500	0.750	0.8936	0.0537	0.0527	-24115
262	0.270	0.443	0.741	0.5824	0.3394	0.0782	4864
286	0.242	0.411	0.710	0.5493	0.4164	0.0343	-51974
289	0.319	0.521	0.782	0.7350	0.1047	0.1603	11304
293	0.235	0.482	0.776	0.8372	0.0993	0.0635	6848
295	0.060	0.394	0.755	0.9375	0.0125	0.0500	-203439
301	0.240	0.507	0.733	0.9132	0.0428	0.0440	-13194
302	0.075	0.472	0.752	0.6034	0.0008	0.3958	-160860
304	0.390	0.525	0.763	0.5258	0.0571	0.4171	17472
305	0.266	0.577	0.779	0.7080	0.0190	0.2730	70668
310	0.250	0.500	0.750	0.8936	0.0537	0.0527	41850

(1)  $p_1$  : probabilité d'appartenir à la première composante,(2)  $p_2$  : probabilité d'appartenir à la deuxième composante,(3)  $p_3$  : probabilité d'appartenir à la troisième composante.

N° GHM	Proportions cumulées de patients grenoblois						Déficit en points ISA
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle	$p_1^{(1)}$	$p_2^{(2)}$	$p_3^{(3)}$	
312	0.307	0.486	0.743	0.7918	0.1245	0.0837	9830
315	0.171	0.457	0.749	0.9175	0.0385	0.0440	-6105
339	0.104	0.400	0.739	0.9135	0.0493	0.0373	-6165
340	0.200	0.433	0.633	0.7319	0.2673	0.0008	-28230
343	0.279	0.445	0.682	0.8355	0.1544	0.0101	-135641
344	0.183	0.433	0.700	0.8868	0.0979	0.0153	-14769
347	0.417	0.604	0.729	0.8937	0.0254	0.0810	2222
348	0.406	0.578	0.781	0.5775	0.0277	0.3949	-7126
349	0.250	0.500	0.750	0.8936	0.0537	0.0527	29995
351	0.250	0.500	0.750	0.8936	0.0537	0.0527	-11025
352	0.250	0.500	0.750	0.8936	0.0537	0.0527	-11760
376	0.391	0.609	0.772	0.6981	0.0263	0.2756	37584
377	0.250	0.500	0.750	0.8936	0.0537	0.0527	41960
392	0.194	0.473	0.828	0.7762	0.1893	0.0345	-11869
394	0.250	0.500	0.750	0.8936	0.0537	0.0527	22484
395	0.279	0.557	0.877	0.6935	0.2385	0.0681	17730
396	0.238	0.476	0.738	0.8850	0.0772	0.0378	-18180
397	0.171	0.378	0.613	0.5247	0.4749	0.0004	-87080
398	0.408	0.517	0.659	0.9786	0.0182	0.0032	-59616
408	0.095	0.413	0.810	0.9135	0.0656	0.0209	0
420	0.300	0.540	0.640	0.6018	0.3975	0.0007	-3190
425	0.412	0.541	0.647	0.9732	0.0260	0.0008	-43128
442	0.225	0.562	0.738	0.6661	0.0266	0.3073	2679
455	0.396	0.568	0.793	0.5049	0.0418	0.4533	6880
458	0.382	0.570	0.717	0.9303	0.0332	0.0365	-18600
459	0.346	0.596	0.808	0.6489	0.0340	0.3170	23496
465	0.319	0.406	0.638	0.7719	0.2217	0.0064	-24890
466	0.440	0.500	0.630	0.9921	0.0068	0.0011	-25592
469	0.183	0.413	0.750	0.6426	0.3191	0.0383	-17442
471	0.248	0.407	0.697	0.5623	0.4112	0.0265	-32494
481	0.000	0.341	0.681	0.9077	0.0160	0.0763	-22950
498	0.293	0.466	0.672	0.8886	0.1068	0.0047	-2697
511	0.269	0.463	0.597	0.5407	0.4592	0.0000	-35881
512	0.250	0.500	0.750	0.8936	0.0537	0.0527	-12585
515	0.250	0.500	0.750	0.8936	0.0537	0.0527	9452
527	0.250	0.500	0.750	0.8936	0.0537	0.0527	-11560
531	0.250	0.500	0.750	0.8936	0.0537	0.0527	61560
545	0.250	0.500	0.750	0.8936	0.0537	0.0527	384
546	0.250	0.500	0.750	0.8936	0.0537	0.0527	1712
549	0.333	0.546	0.826	0.5798	0.1718	0.2485	13962
571	0.291	0.481	0.696	0.9012	0.0864	0.0124	-43865
574	0.296	0.543	0.741	0.8938	0.0432	0.0630	15534
575	0.224	0.447	0.715	0.8555	0.1226	0.0219	-1187
585	0.387	0.538	0.656	0.9567	0.0420	0.0013	-175924
589	0.173	0.423	0.635	0.7446	0.2540	0.0014	-32997
595	0.397	0.614	0.750	0.7950	0.0314	0.1736	-65296
602	0.289	0.570	0.810	0.7694	0.0313	0.1992	100128
607	0.339	0.585	0.763	0.8025	0.0348	0.1627	-20696
609	0.197	0.561	0.879	0.9269	0.0105	0.0626	7322
611	0.275	0.497	0.732	0.8944	0.0682	0.0374	-26298
619	0.460	0.527	0.690	0.8938	0.0056	0.1006	-73139
622	0.373	0.589	0.763	0.7887	0.0309	0.1804	30416
625	0.241	0.481	0.759	0.8579	0.0870	0.0551	5530
643	0.298	0.607	0.893	0.7604	0.0904	0.1492	87584
645	0.319	0.475	0.598	0.7520	0.2479	0.0000	-28347
646	0.261	0.522	0.701	0.9073	0.0723	0.0205	-10676
649	0.203	0.405	0.716	0.6208	0.3523	0.0269	8700
650	0.217	0.434	0.651	0.8044	0.1938	0.0018	-182016
651	0.250	0.500	0.750	0.8936	0.0537	0.0527	-3633
669	0.286	0.483	0.721	0.8780	0.0906	0.0313	-133155
671	0.354	0.573	0.759	0.8295	0.0354	0.1351	-8190
674	0.250	0.500	0.750	0.8936	0.0537	0.0527	139808
873	0.408	0.648	0.768	0.5542	0.0304	0.4154	13932
876	0.250	0.500	0.750	0.8936	0.0537	0.0527	2145

(1)  $p_1$  : probabilité d'appartenir à la première composante,(2)  $p_2$  : probabilité d'appartenir à la deuxième composante,(3)  $p_3$  : probabilité d'appartenir à la troisième composante.



# Annexe Q

## Groupe des “bons M”

Les deux tableaux ci-après donnent la liste détaillée (numéro de GHM et proportions cumulées de patients dans les intervalles interquartiles) des 78 GHM grenoblois faisant partie du groupe des “bons M”, issu du modèle de mélange de 3 lois normales.

N° GHM	Proportions cumulées de patients grenoblois			$p_1^{(1)}$	$p_2^{(2)}$	$p_3^{(3)}$	Déficit en points ISA
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle				
5	0.706	0.847	0.918	0.0000	0.0000	1.0000	78816
8	0.642	0.730	0.912	0.0000	0.0000	1.0000	63038
19	0.429	0.638	0.876	0.1833	0.0618	0.7549	70176
30	0.580	0.710	0.815	0.0141	0.0001	0.9858	-3290
37	0.335	0.418	0.741	0.0662	0.4603	0.4735	-1550
45	0.446	0.652	0.882	0.1238	0.0428	0.8334	50940
74	0.108	0.639	0.841	0.0001	0.0000	0.9999	9080
94	0.555	0.654	0.785	0.0497	0.0002	0.9502	7455
99	0.585	0.585	0.829	0.0000	0.0000	1.0000	667
100	0.691	0.765	0.826	0.0003	0.0000	0.9997	30368
102	0.278	0.896	0.944	0.0000	0.0000	1.0000	23310
103	0.305	0.679	0.802	0.0208	0.0025	0.9767	21112
104	0.519	0.714	0.896	0.0139	0.0032	0.9829	11952
129	0.362	0.678	0.859	0.1560	0.0101	0.8340	86961

<sup>(1)</sup>  $p_1$  : probabilité d'appartenir à la première composante,

<sup>(2)</sup>  $p_2$  : probabilité d'appartenir à la deuxième composante,

<sup>(3)</sup>  $p_3$  : probabilité d'appartenir à la troisième composante.

N° GHM	Proportions cumulées de patients grenoblois			$p_1^{(1)}$	$p_2^{(2)}$	$p_3^{(3)}$	Déficit en points ISA
	1 <sup>er</sup> intervalle	2 <sup>e</sup> intervalle	3 <sup>e</sup> intervalle				
131	0.625	0.677	0.844	0.0001	0.0000	0.9999	23391
137	0.419	0.612	0.845	0.2487	0.0442	0.7072	8667
138	0.489	0.676	0.832	0.0810	0.0034	0.9157	87374
139	0.459	0.628	0.804	0.2467	0.0083	0.7451	18992
140	0.425	0.851	0.936	0.0000	0.0000	0.9999	288754
142	0.385	0.583	0.859	0.3101	0.2049	0.4850	38304
154	0.419	0.729	0.897	0.0402	0.0072	0.9526	370328
158	0.308	0.649	0.827	0.1950	0.0068	0.7982	183125
159	0.458	0.780	0.915	0.0046	0.0021	0.9933	48565
160	0.181	0.619	0.812	0.0172	0.0001	0.9828	-19605
167	0.405	0.810	0.889	0.0001	0.0001	0.9998	9370
168	0.463	0.716	0.768	0.1589	0.0343	0.8068	84458
178	0.406	0.678	0.850	0.1647	0.0128	0.8225	123030
179	0.345	0.635	0.829	0.4181	0.0196	0.5623	149568
181	0.248	0.612	0.813	0.2866	0.0041	0.7093	96110
191	0.266	0.605	0.819	0.4910	0.0091	0.4999	36855
195	0.437	0.575	0.787	0.2985	0.0163	0.6852	-2616
196	0.585	0.734	0.894	0.0013	0.0001	0.9985	18336
199	0.419	0.605	0.837	0.2571	0.0406	0.7024	15158
226	0.327	0.673	0.909	0.3875	0.0261	0.5864	31560
229	0.520	0.786	0.878	0.0026	0.0007	0.9967	7200
230	0.033	0.917	0.942	0.0000	0.0000	1.0000	14865
256	0.410	0.581	0.825	0.2766	0.0540	0.6694	47109
257	0.544	0.746	0.864	0.0077	0.0005	0.9918	78890
258	0.667	0.825	0.880	0.0001	0.0000	0.9999	89698
273	0.314	0.638	0.838	0.3513	0.0120	0.6367	66550
274	0.144	0.548	0.731	0.1448	0.0024	0.8527	-8515
290	0.300	0.630	0.840	0.4043	0.0115	0.5842	25680
308	0.156	0.629	0.763	0.0003	0.0000	0.9997	-10146
317	0.535	0.785	0.869	0.0027	0.0006	0.9968	52765
318	0.436	0.872	0.947	0.0000	0.0000	1.0000	76046
336	0.533	0.780	0.867	0.0033	0.0006	0.9961	109327
353	0.440	0.736	0.920	0.0454	0.0150	0.9397	51842
354	0.487	0.677	0.862	0.0507	0.0052	0.9441	58473
356	0.456	0.569	0.788	0.1367	0.0083	0.8549	-4740
409	0.091	0.624	0.812	0.0000	0.0000	1.0000	-17908
421	0.724	0.816	0.827	0.0001	0.0000	0.9999	39368
422	0.630	0.778	0.796	0.0093	0.0001	0.9906	-10180
429	0.500	0.688	0.891	0.0260	0.0075	0.9665	122130
454	0.460	0.717	0.856	0.0446	0.0046	0.9508	65736
456	0.339	0.891	0.939	0.0000	0.0000	1.0000	39091
476	0.294	0.647	0.843	0.2122	0.0054	0.7824	3110
477	0.270	0.610	0.800	0.3955	0.0094	0.5952	7636
507	0.304	0.679	0.839	0.0375	0.0016	0.9609	-2183
509	0.559	0.682	0.830	0.0106	0.0002	0.9892	58995
539	0.356	0.712	0.833	0.0145	0.0022	0.9833	67773
540	0.538	0.836	0.922	0.0003	0.0004	0.9994	214000
547	0.464	0.667	0.833	0.1316	0.0069	0.8615	15700
548	0.475	0.594	0.822	0.0432	0.0055	0.9513	4560
568	0.415	0.717	0.792	0.0440	0.0144	0.9416	20631
573	0.380	0.446	0.761	0.0232	0.1310	0.8458	-21097
586	0.443	0.508	0.746	0.1112	0.0094	0.8794	-2526
587	0.537	0.601	0.856	0.0005	0.0004	0.9992	-91140
592	0.354	0.854	0.905	0.0000	0.0000	1.0000	116615
610	0.661	0.756	0.845	0.0003	0.0000	0.9997	37950
618	0.583	0.688	0.799	0.0190	0.0001	0.9809	58464
623	0.736	0.782	0.874	0.0000	0.0000	1.0000	63720
631	0.372	0.735	0.938	0.0762	0.0132	0.9106	53781
632	0.444	0.887	0.933	0.0000	0.0000	1.0000	44856
639	0.190	0.595	0.823	0.2037	0.0007	0.7956	36808
640	0.306	0.631	0.856	0.4753	0.0164	0.5083	59436
672	0.520	0.705	0.803	0.0844	0.0019	0.9137	85255
675	0.428	0.665	0.858	0.1793	0.0182	0.8025	218250
874	0.700	0.780	0.860	0.0000	0.0000	1.0000	98990

(1)  $p_1$  : probabilité d'appartenir à la première composante,(2)  $p_2$  : probabilité d'appartenir à la deuxième composante,(3)  $p_3$  : probabilité d'appartenir à la troisième composante.

## Bibliographie

- [1] Circulaire n°79 du 14 février 1985 relative à la mise en place de la comptabilité analytique. *B Off*.
- [2] Arrêté du 20 septembre 1994, relatif au recueil et au traitement des données d'activité médicale et de coût, visées à l'article L.710-5 du code de santé publique, par les établissements de santé publics et privés visés aux articles L.741-1, L.715-5 du code de santé publique et aux articles L.162-23, L.162-23-1 et L.162-25 du code de la sécurité sociale et à la transmission aux services de l'Etat et aux organismes d'assurance maladie d'informations issues de ces traitements. *B Off*, 18 Octobre 1994.
- [3] Handbook of methods of applied statistics. 1:392-394, 1967.
- [4] Décret du 11 août 1983 relatif au financement des hôpitaux publics. *J Off*, 1983.
- [5] Circulaire n° 119 du 4 octobre 1985 relative à la mise en place, dans les établissements hospitaliers, des RSS. *B Off*, 1985.
- [6] Circulaire n°119 du 4 octobre 1985 relative à l'obligation de fournir des RSS. *B Off*, 1985.
- [7] Circulaire n° 160 du 5 août 1986 relative à la définition de la classification en Groupes Homogènes de Malades énonçant, CMD par CMD, le libellé et le numéro de chacun des 471 GHM. *Fascicule spécial B Off n° 85-30 bis*, 1986.
- [8] Circulaire n°160 du 5 août 1986 - la première classification en Groupes Homogènes de Malades. *B Off*, 1986.
- [9] Circulaire DH/PMSI/n° 303 du 24 juillet 1989 relative à la généralisation du Programme de Médicalisation des Systèmes d'Information (PMSI) et à l'organisation de l'information médicale dans les hôpitaux publics. *B Off*, 1989.
- [10] Circulaire n° 275 du 6 janvier 1989 relative à l'informatisation des hôpitaux publics. *B Off*, 1989.
- [11] Circulaire n° 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. *B Off*, 1990.
- [12] Circulaire DH/PMSI n° 45 du 9 juillet 1991 de mise en place des Groupes Régionaux d'Animation sur l'Information Hospitalière (GRAIH). *B Off*, 1991.
- [13] Circulaire DH/PMSI n° 56 du 16 septembre 1991 relative aux relations conventionnelles entre les représentants de l'Etat et les établissements hospitaliers. *B Off*, 1991.



- [14] Loi n° 91-748 du 31 juillet 1991 portant réforme hospitalière. *J Off*, 1991.
- [15] Classification internationale des maladies et des problèmes de santé. Dixième révision (CIM 10). *Connexes*, 1993.
- [16] Circulaire DH/DSS/n° 23 du 10 mai 1995 relative à la généralisation du dispositif technique, expérimenté en région Languedoc-Roussillon, de mesure de l'activité et des coûts des établissements hospitaliers sous compétence tarifaire de l'Etat, à l'aide des informations provenant du système d'information médicalisé. *B Off*, 1995.
- [17] Circulaire DH/PMSI/95 n°48 du 11 décembre 1995 relative à la création des comités techniques régionaux de l'information médicale (COTRIM) et à l'organisation des contrôles externes dans les établissements de santé, dans le cadre du développement du PMSI. *B Off*, 1995.
- [18] L'échelle nationale de coûts relatifs par Groupe Homogène de Malades. *J Off*, 1995.
- [19] Lettre-Circulaire du 22 août 1995 à l'intention des Préfets de région - DRASS et des Préfets de Département - DDASS confirmant la généralisation du dispositif technique de mesure de l'activité et des coûts des établissements hospitaliers sous compétence tarifaire de l'Etat. *B Off*, 1995.
- [20] Ordonnance n° 96-346 du 24 avril 1996 portant réforme de l'hospitalisation publique et privée. *J Off*, 1996.
- [21] Circulaire DH/PMSI/97 n° 251 du 3 avril 1997 relative à la mise en place, dans les établissements de santé ayant une activité de soins de suite ou de réadaptation, d'un recueil de résumés hebdomadaires standardisés (RHS) ; dispositions particulières pour les établissements sous compétence tarifaire de l'état. *J Off*, 1997.
- [22] Circulaire DH/PMSI/98 n° 153 du 9 mars 1998 relative à la généralisation, dans les établissements de santé sous dotation globale et ayant une activité de soins de suite ou de réadaptation, d'un recueil de résumés hebdomadaires standardisés (RHS) ; calendrier de montée en charge pour l'année 1998. *J Off*, 1998.
- [23] Manuel des groupes homogènes de malades, version 1. *Fascicule spécial J Off*. n° 92-9 bis, Janvier 1992.
- [24] P. Barberger-Gateau, F. Dabis, A. Moise, M. Gimbert, P. Galley et R. Salamon. Facteurs liés à la durée d'hospitalisation des personnes âgées en court séjour. *Rev Epidém et Santé Publ*, **35**:463–73, 1987.
- [25] B. Bertozzi, P. Barbisoni, S. Franzoni, R. Rozzini, GB. Frisoni et M. Trabucchi. Factors related to length of stay in a geriatric evaluation and rehabilitation unit. *Aging Clin Exp Res*, **8**:170–175, 1996.
- [26] C. Biernacki. *Choix de modèles en classification*. Thèse de doctorat, Université de Technologie de Compiègne, 1997.
- [27] C. Biernacki. *XEMgaus: Software for Model-Based Cluster and Discriminant Analysis*, 1999.
- [28] C. Biernacki, G. Celeux et G. Govaert. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**:7:719–725, 2001.

- [29] D. Blum et A. Loth. Durées de séjour par GHM : premières références. *La lettre d'Information Hospitalière*, **12–13**:7–8, 1993.
- [30] C. Catala. *La protection sociale en France, chapitre : L'offre et la demande dans le domaine de la santé*. Paris : La documentation française, 1997.
- [31] G. Celeux et G. Govaert. Gaussian parsimonious models. *Pattern Recognition*, **28**:5;781–793, 1995.
- [32] P. Cinquin. *Réflexions préliminaires sur le développement d'un système d'information hospitalier intégré*.
- [33] E. Coca. *L'ISA et le PMSI : comment mieux gérer les établissements hospitaliers?* Paris : Berger-Levrault, 1998.
- [34] E. Coca et J. Dubois-Lefrère. *Maîtriser l'évolution des dépenses hospitalières : le PMSI*. Paris : Berger-Levrault, 1992.
- [35] Service d'Analyse de Gestion. Rapport d'activité du CHU de Grenoble. Rapport interne, 1997.
- [36] A.P. Dempster, Laird M.N. et D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *J. R. Statist. Soc.*, **39**:1–38, 1977.
- [37] B. Dervaux et S. Jacobzone. *Économie de la santé : trajectoires du futur, chapitre : L'activité hospitalière*. Economica, INSEE, 1997.
- [38] Ministère des affaires sociales et de l'intégration. Manuel des groupes homogènes de malades, version 6. *Fascicule spécial J Off. n° 2000-2 bis*, Janvier 2000.
- [39] J.J. Droesbeke, B. Fichet et P. Tassi, éditeurs. *Modèles pour l'analyse des données multidimensionnelles*. Paris : Economica, 2ème édition, 1992.
- [40] A. Dussaucy, J.-F. Viel, B. Mulin et J. Euvrard. L'outil PMSI : biais, sources d'erreurs et conséquences. *Rev Epidém et Santé Publ*, **42**:345–358, 1994.
- [41] L. Dusserre, C. Quantin, P. Métral, C. Mathy et A. Lalie. Justification clinique des séjours coûteux d'un GHM. *Gestion hospitalière*, 303–308, 1995.
- [42] B.S. Everitt et D.J. Hand. *Finite mixture distributions*. London : Chapman and Hall, 1981.
- [43] B. Falissard. *Comprendre et utiliser les statistiques dans les sciences de la vie*. Paris : Masson 1996.
- [44] J. Fauconnier. Résultats PMSI 1996 et durées de séjour. Note interne du CHU de Grenoble, 1997.
- [45] R.B. Fetter. Cost Models and DRGs. An International Comparison. *Australian Health Review*, **8**:2;116–125, 1985.
- [46] R.B. Fetter. Hospital. *Journal of the Society for Health Systems*, **3**:4–15, 1992.
- [47] R.B. Fetter. Casemix classification systems. *Australian Health Review*, **22**:2;16–34, 1999.
- [48] R.B. Fetter et J.L. Freeman. Diagnosis related groups : Product line management within hospitals. *Academy of Management Review*, **11**:41–54, 1986.

- [49] R.B. Fetter, YS. Shin, JL. Freeman et al. Casemix definition by diagnosis related groups. *Med Care*, **18**:1–53, 1980.
- [50] N. Fineberg, S. Merson et T. Kurian. Length of stay of acute psychiatric inpatients: an exponential model. *Acta Psy Scand*, **92**:315–317, 1995.
- [51] R.A. Fisher. The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, **7**:179–188, 1936.
- [52] MG. Fleischmann, C. Quantin, M. Vourc’h, P. Brunet-Lecomte, C. Binquet, JB. Gouyon, G. Putet et L. Dusserre. Analyse des séjours de longue durée des groupes homogènes de malades de néonatalogie. *Journal d’Économie Médicale*, **18**:7–8;421–431, 2000.
- [53] P. Frutiger et J.M. Fessler. *La Gestion hospitalière médicalisée*. Paris: ESF éditeur, 1991.
- [54] M. Girardier, D. Blum et E. Gomez. Rôle du PERNNS dans l’évolution de la classification des groupes homogènes de malades. *Informatique et Santé*, pages 65–71, 1997.
- [55] A. Giraud et J-C. Sournia. Le “système Fetter”, ou la définition du produit hospitalier par Diagnosis-Related-Groups (D.R.G ou Groupes Homogènes de Malades (G.H.M)). Analyse critique. *Journal d’Économie Médicale*, **5**:213–225, 1985.
- [56] M Giroud et col. Incidence et mortalité des accidents vasculaires cérébraux en France. *Rev Neurol*, **43**:4;214-218, 1994.
- [57] M. Giroud et M. Lemesle. Prise en charge des accidents vasculaires cérébraux. Apport du registre dijonnais des accidents vasculaires cérébraux. *Rev Epidém Et Santé Publ*, 44 1996.
- [58] G. Haenni. *Les déterminants de la durée de séjour et des coûts à l’hôpital de Genève*. Thèse de doctorat, Université de Genève, 1996.
- [59] G.J. Hankey et Warlow C.P. Treatment and secondary prevention of stroke: evidence, costs and effects on individuals and populations. *Lancet*, **354**:9188;1457-1463, 1999.
- [60] L. Lebart, A. Morineau et M. Piron. *Statistique exploratoire multidimensionnelle*. Paris: Dunod, 1995.
- [61] A. Marazzi 1997. Estimating and testing the mean of asymmetric distributions using M-estimators. Rapport de recherche, Institut Universitaire de médecine sociale et préventive, Lausanne.
- [62] A. Marazzi, F. Paccaud, C. Ruffieux et C. Beguin. Fitting the Distributions of Length of Stay by Parametric Models. *Medical Care*, **36**:6;915–927, 1998.
- [63] SI. McClean et PH. Millard. A decision support system for bed-occupancy management and planning hospitals. *IMA Journal of Mathematics Applied in Medecine and Biology*, pages 225–234, 1995.
- [64] M. McClellan. *Économie de la santé: trajectoires du futur, chapitre: Incitations et financement des hôpitaux: le partage prospectif et rétrospectif des coûts*. Economica, INSEE, 1997.
- [65] G.J. McLachlan et Basford K.E. *Mixture Models: Inference and Application to Clustering*. New-York: Marcel Dekker, 1989.

- [66] G.J. McLachlan et T. Krishnan. *The EM Algorithm and Extensions*. New-York : Wiley-Interscience, 1997.
- [67] PH. Millard, G. Christodoulou, C. Jagger, GW. Harrison et SI. McClean. Modelling hospital and social care bed occupancy and use by elderly people in an English health district. *Health Care Management Science*, **4**:57–62, 2001.
- [68] BJ. Naughton. Reducing hospital cost for the geriatric patient admitted from the emergency department : a randomized trial. *J Am Geriatr Soc*, **42**:1045–1049, 1994.
- [69] N. Nighoghossian, P. Adeleine, L. Derex, J. Honnorat, A. Moussa, B Perel, G. Rode, D. Boisson et P. Trouillas. Facteurs influençant la durée de séjour des patients victimes d'un accident vasculaire cérébral. *Journal d'Économie Médicale*, pages 3–4;179–190, 2000.
- [70] F. Paccaud et L. Schenker. *D.R.G. (Diagnosis Related Groups)*. Éditions Alexandre Lacassagne, diffusion Masson, 1989.
- [71] S. Pérez-Hoyos, F. Ballester, JM. Tenias, A. Merelles et ML. Rivera. Length of stay in a hospital emergency room due to asthma and chronic obstructive pulmonary diseases: Implications for air pollution studies. *European Journal of Epidemiology*, **16**:455–463, 2000.
- [72] C. Quantin, F. Entezam, C. Bourdais, T. Moreau, P. Brunet-Lecomte, H. Bouzelat, C. Mathy, P. Foucher, P. Métral et L. Dusserre. Modélisation de la distribution des durées de séjour et des coûts dans deux groupes homogènes de malades d'hématologie et de pneumologie : caractérisation clinique des patients de longue durée et de coût élevé. *Rev Épidém et Santé Publ*, **45**:117–130, 1997.
- [73] C. Quantin, F. Entezam, P. Brunet-Lecomte, E. Lepage, H. Guy et L. Dusserre. High cost factors for leukaemia and lymphoma patients: a new analysis of cost within these diagnosis related groups. *J Epidemiol Community Health*, **53**:24–31, 1999.
- [74] C. Quantin, P. Métral et L. Dusserre. PMSI et contractualisation de la justification clinique des séjours coûteux d'un GHM. *Journal d'Économie Médicale*, **6**:337–339, 1995.
- [75] C. Quantin, P. Métral, André-Allaert F. et L. Dusserre. Aspects déontologiques de l'identification clinique de patients coûteux lors de l'évaluation de l'activité hospitalière. *Informatique et Santé : Springer-Verlag*, **8**:77–85, 1996.
- [76] R. Rao. *Linear statistical inference and its Application*. Ney York : édition Wiley, 1965.
- [77] C. Robert. *Analyse descriptive multivariée. Application à l'intelligence artificielle*. Paris : Flammarion, 1989.
- [78] C. Ruffieux et F. Marazzi, A. Paccaud. Exploring models for the length of stay distribution. *Soz Präventivmed*, **38**:77–82, 1993.
- [79] G. Saporta. *Probabilités, Analyse des Données et Statistique*. Éditions Technip, 1990.
- [80] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, **6**:461–464, 1978.
- [81] I. Tatsuro, I. Kai, M. Hisata, Y. Kobayashi, BA. Wakatsuki et G. Ohi. Factors influencing user's return home on discharge from a geriatric intermediate care facility in Japan. *J Am Geriatr Soc*, **43**:623–626, 1995.

- [82] G. Taylor, SI. McClean et PH. Millard. *Go with the Flow: a Systems Approach to Health Care Planning*, chapter A model to analyse the flow of elderly patients around departments of geriatric medicine and the community. London : Royal Society of Medicine Press Ltd., 1996.
- [83] D.M. Titterington, A.F.M. Smith et U.E. Markov. *Statistical analysis of mixture distribution*. London : Chapman and Hall, 1985.
- [84] M. Topp, L. Drasbek Husom, J. Langhoff-Roos, C. Delhumeau, J. L. Hutton et H. Dolk. Multiple birth and cerebral palsy in Europe - a multiple-centre study. *Paediatric and Perinatal Epidemiology*, article soumis.
- [85] IS. Udvarhelyi, L. Goldman, AL. Komaroff et TH. Lee. Determinants of resource utilization for patients admitted for evaluation of acute chest pain. *Journal of General Internal Medicine* 1992.
- [86] J. Valois, JR. Abrial et S. Cohen. Description d'un système de gestion en temps réel du dossier médical. *Grenoble Médico-Chirurgical*, 8:7;301-314, 1970.
- [87] J. Valois, P. Cohen, P. Hernicot et JR. Abrial. Application d'un système de définition et d'interrogation de données à différentes formes du dossier médical. *Rev. Informatique Méd.*, 3:2;123-129, 1972.
- [88] J. Valois, M. Girardier, Hernicot P. et C. Deyres. Dossier résumé de sortie des malades et statistiques médicales. système DOSTAM. *Bulletin statistiques et commentaires*, 4:5-22, 1976.
- [89] A. Vergnenègre, JJ. Chale, C. Grouchka, E. Séné, J. Pous et F. Bonnaud. Indicateurs explicatifs de la durée de séjour dans un service de pneumologie. *Rev Mal Respir*, 22:479-488, 1995.
- [90] G. Weil, D. Guillon et J. Demongeot. Utilisation de l'analyse des données dans l'étude des paramètres d'activité d'un hôpital. Note interne du CHU Grenoble, 1986.
- [91] S. Wyatt. The Bed Occupancy Management and Planning System (BOMPS). *The Lancet*, 345:243-245, 1995.

## Glossaire

- **BDD**: Base De Données.
- **CdAM**: Catalogue des Actes Médicaux; il est publié et mis à jour régulièrement par le Pôle d'Expertise et de Référence Nationale des Nomenclatures de Santé et il sert à coder les actes médicaux dans les RUM.
- **CIM-10**: Classification Internationale des Maladies (Dixième révision); elle est publiée par l'Organisation Mondiale de la Santé et est actualisée régulièrement; elle sert à coder le diagnostic principal (et éventuellement les diagnostics associés) dans les RUM.
- **CM**: Catégorie Majeure; au nombre de 2 (CM 24 et CM 90), elles correspondent l'une aux séances et séjours ambulatoires à l'hôpital, et l'autre aux erreurs de codages.
- **CMA**: Complication(s) et Morbidité(s) Associée(s); elle(s) correspond(ent) aux affections associées au diagnostic principal, complications de celui-ci ou de son traitement, dont la présence induit une augmentation significative du coût du séjour et l'orientation du RSS dans un GHM avec complication.
- **CMAS**: Complication(s) et Morbidité(s) Associée(s) Sévère(s); voir **CMA**.
- **CMD**: Catégorie Majeure de Diagnostic; au nombre de 23, elles regroupent les GHM d'une même spécialité médicale.
- **DA**: Diagnostics Associés; ils correspondent en général aux DP inscrits dans les RUM et non retenus comme DP.
- **DIM**: Département d'Information Médicale; structure hospitalière, créée pour les besoins de la généralisation du PMSI, où est gérée l'information médicale.
- **DRG**: Diagnosis Related Groups; classification dont sont issus les GHM français, élaborée par Robert Fetter, dans les années 70, aux Etats-Unis.
- **DP**: Diagnostic Principal; c'est celui qui a mobilisé l'essentiel de l'effort médical et soignant au cours du séjour hospitalier.
- **DS**: Durée de Séjour à l'hôpital.
- **ET**: Écart Type.
- **Fonction groupe**: partie du logiciel "GROUPEUR"; propriété de l'Etat, qui contient l'algorithme de la classification, avec tous les arbres de décision, et toutes les tables (listes des actes ou des diagnostics, utilisées à chaque nœud de décision). C'est elle qui permet d'affecter à un RSS, un numéro de GHM et un seul.
- **GHM**: Groupes Homogènes de Malades. Les patients de court séjour de Médecine, de Chirurgie ou d'Obstétrique sont regroupés dans des GHM à partir des données de leur

RSS et de la fonction groupage. Chaque patient, dont le RSS est affecté à un GHM, est supposé avoir entraîné, en standard, le même coût que tous les malades du GHM, quels que soient l'hôpital, la DS ou les examens réalisés. Les GHM sont regroupés dans une classification publiée et mise à jour régulièrement par le Pôle d'Expertise et de Référence Nationale des Nomenclatures de Santé.

- **ICR** (Indice de Coût Relatif) : unité d'œuvre des actes produits par les services médico-techniques, indiquant le degré de mobilisation de ressources humaines et matérielles directement nécessaires à leur production. L'ICR est "relatif", un acte ayant un ICR de 200 "coûte" deux fois plus cher que celui dont l'ICR vaut 100.
- **Points ISA** : Points d'Indice Synthétique d'Activité; correspondant à des points attribués à un GHM à partir de l'échelle nationale des coûts par activité médicale. item **PERNNS** : Pôle d'Expertise et de Référence Nationale des Nomenclatures de Santé.
- **PMSI** : Programme de Médicalisation des Systèmes d'Information.
- **RSA** : Résumé de Sortie Anonyme ; il permet la transmission d'informations médicales à la direction de l'établissement, ou à l'extérieur de celui-ci, par anonymisation du RSS. La production des RSA est automatisée. À partir du fichier des RSS groupés, le médecin responsable du DIM utilise le logiciel "GENRSA" (GÉNÉrateur de RSA), propriété de l'État, pour produire des RSA.
- **RSS** : Résumé de Sortie Standardisé ; il est constitué d'un ou plusieurs RUM. Tout séjour hospitalier, réalisé en court séjour de Médecine, de Chirurgie ou d'Obstétrique, fait l'objet d'un RSS.
- **RUM** : Résumé d'Unité Médicale; il est produit à la fin du séjour d'un malade dans une unité médicale assurant des soins de courte durée, quel que soit le mode de sortie de cette unité. Le RUM contient un nombre limité d'informations d'ordre administratif et médical qui doivent être systématiquement renseignées.

## CONTRIBUTION À LA MODÉLISATION DES DURÉES DE SÉJOUR DU CHU DE GRENOBLE

### RÉSUMÉ

Cette thèse propose une méthodologie permettant l'identification des groupes homogènes de malades (GHM) du Centre Hospitalier Universitaire (CHU) de Grenoble dont la durée de séjour (DS) s'écarte de la "référence nationale", mais aussi d'identifier la présence d'éventuels groupes d'"outliers" (patients avec des DS extrêmes) dans ces GHM. L'enjeu est de taille : des séjours longs entraînent une occupation des lits supérieure à ce que permet la valorisation financière correspondant à la pathologie prise en charge. Il est donc important de repérer ces GHM, responsables du coût élevé du point Indice Synthétique d'Activité (ISA), unité d'œuvre des hôpitaux utilisée par le Programme de Médicalisation des Systèmes d'Information pour connaître leur activité, leur allouer le budget correspondant, et contraindre à une productivité optimale. Les écarts entre la distribution des DS des GHM grenoblois et celle de leurs homologues de la base nationale ont été comparés. Pour chaque GHM, un profil, fondé sur la comparaison des quartiles de DS des distributions nationales et grenobloises, deux à deux, a été construit. Des méthodes statistiques de classification (analyse en composantes principales, classification hiérarchique, analyse discriminante et modèles de mélange) ont été utilisées pour repérer aisément et sûrement les GHM économiquement coûteux. De manière empirique, les profils grenoblois semblent se structurer en 9 catégories. La classification hiérarchique identifie 4 catégories de GHM, dont une particulière comptant 16 GHM, dans laquelle un tiers des patients participe fortement à la dérive en points ISA du CHU, et pour qui la mise en place d'une action économiquement rentable serait aisée. Le modèle de mélange montre que les GHM se structurent en 3 catégories et permet de valider la classification issue de l'approche multidimensionnelle.

**MOTS CLEFS :** durées de séjour (DS), points Indice Synthétique d'Activité (ISA), Groupe Homogène de Malades (GHM), analyse en composante principale, classification hiérarchique, analyse discriminante, modèle de mélange.

## CONTRIBUTION TO THE MODELLING OF LENGTHS OF STAY AT GRENOBLE UNIVERSITY HOSPITAL

### ABSTRACT

This thesis provides a methodology to identify diagnosis related groups (DRG) from Grenoble University Hospital which have a length of stay (LOS) longer than the national reference LOS, and also to identify outliers (patients with extreme length of stay) in these groups. Long LOS are critical because the hospital's rooms tend to be used too much, compared to what the financial score (value) of the corresponding DRG (pathology) would allow. It is therefore important to identify the groups that are responsible for high costs in terms of Indice Synthétique d'Activité (ISA) point. The ISA point is a unit used in the French national program settled to medicalize the information system in hospitals with the aim of optimizing hospital efficiency by measuring their activity and determining the corresponding budget to provide them with (french Prospective Payment System). We compared the differences between the length of stays distribution for the DRG in Grenoble and the LOS distribution for the corresponding DRG in the national database. For each DRG we built a profile by comparing the quartiles of the two distributions. We used statistical classification methods, such as Principal Components Analysis, Hierarchical Clustering, Discriminant Analysis and mixtures models, to identify the groups with highest costs. Empirically 9 different profile types seem to be distinguished. Hierarchical clustering methods identified a classification of DRG in 4 groups. One of them (16 DRG's) counted one third of patients who were highly responsible for costly ISA points, and for which an efficient economical action could be easily undertaken. The mixture models distinguished 3 groups of DRG, and validated the classification derived from the Hierarchical Clustering Method.

**KEY-WORDS:** length of stay (LOS), Prospective Payment System, Diagnosis Related Groups (DRG), principal components analysis, hierarchical clustering, discriminant analysis, mixture models.