

ASPECTS DE LA SÉCURITÉ SOCIALE

*Exploitation de la base de données:
de la LAMal:*

Rapport relatif à un mandat

Rapport de recherche n° 18/03



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Bundesamt für Sozialversicherungen BSV
Office fédéral des assurances sociales OFAS

L'Office fédéral des assurances sociales publie dans sa série „Aspects de la sécurité sociale“ des articles de fond et des rapports de recherches sur des sujets d'actualité dans le domaine de la sécurité sociale pour les rendre accessibles au grand public et encourager la discussion. Les analyses présentées par les auteurs ne reflètent pas forcément l'opinion de l'Office fédéral des assurances sociales.

Auteur: Prof. Antoine Gualtierotti
Institut de Hautes Etudes en Administration
Publique (IDHEAP)
Route de la Maladière 21
1022 Chavannes-près-Renens
Tél. 021 694 06 20
E-mail: antoine.gualtierotti@idheap.unil.ch

Renseignements: Herbert Känzig
Office fédéral des assurances sociales
Effingerstrasse 20
3003 Berne
Tél. 031 322 91 48
Fax 031 324 06 87
E-mail: herbert.kaenzig@bsv.admin.ch

ISBN: 3-905340-83-6

Copyright: Office fédéral des assurances sociales
CH – 3003 Berne
Reproduction d'extraits autorisée – excepté à des fins
commerciales – avec mention de la source; copie à l'Office
fédéral des assurances sociales.

Diffusion: OFCL, Diffusion publications, CH-3003 Bern
www.bbl.admin.ch/bundespublikationen

Numéro de commande: 318.010.18/03 f 8.03 250

Exploitation de la base de données de la LAMal: rapport relatif à un mandat

A.F. Gualtierotti, IDHEAP, CH-1022 Chavannes-près-Renens

Décembre 2001

Préface

Le présent rapport illustre les possibilités d'application de méthodes statistiques et de solutions informatiques récentes pour procéder à des analyses longitudinales. Il est particulièrement intéressant de voir comment ces méthodes peuvent être utilisées, en relation avec l'analyse de grandes quantités de données, pour la classification des cas.

Les statistiques disponibles dans le domaine de la santé publique sont en général des analyses transversales descriptives, se rapportant à une année donnée. La dimension temporelle est introduite lorsque les enquêtes sont effectuées périodiquement et que leurs résultats sont réunis dans des séries chronologiques. Mais comme ces statistiques reposent en général sur des données agrégées, on n'y trouve pas d'analyses longitudinales à proprement parler.

Or, pour réaliser une analyse longitudinale, il faut que les données présentent un degré de précision qui permette de suivre les individus dans la dimension temporelle ; en d'autres termes, une identification est indispensable pour pouvoir relier les indications concernant un individu en une année statistique donnée aux indications concernant le même individu les années suivantes.

C'est là une exigence à laquelle satisfont les données recueillies pour l'évaluation des différences de coûts entre les modèles d'assurance-maladie : nous disposons, sur une période de quatre ans, des données concernant environ cinq millions d'assurés. Des informations socio-démographiques, le modèle d'assurance choisi et les coûts occasionnés ont été relevés sous une forme qui permet d'opérer une distinction entre les groupes de fournisseurs de prestations (p. ex. médecins selon la spécialité, hôpitaux selon le type, pharmacies, etc.).

Il est ainsi possible de voir, pour chaque individu, à quels services il a recouru durant chacune des quatre périodes de relevé, pour quel coût et à quel moment (tant le moment du traitement que l'année du remboursement par l'assureur sont connus).

Afin de pouvoir tirer des conclusions de cette immense quantité de données, il faut classer les diverses carrières de patients par catégories. On peut le faire en fixant, à partir des expériences tirées d'une analyse statistique détaillée des données, des critères arbitraires par rapport aux questions permettant de répartir les divers modes d'évolution des maladies en catégories qui fassent sens.

Le présent travail propose une nouvelle approche, qui utilise pour la classification non plus des critères arbitraires choisis a priori, mais qui forme les catégories uniquement à partir des informations contenues dans les données.

Cette approche ne voit pas seulement l'évolution de la maladie comme l'établissement d'une chronologie des prestations consommées (classées selon la date du début du traitement). La période de quatre ans considérée est divisée en laps de temps de durée égale. Les prestations sont attribuées au laps de temps correspondant sur la base du début du traitement. De cette manière, les laps de temps durant lesquels aucune prestation n'a été consommée sont également pris en compte dans l'analyse.

La base de données utilisée se compose des informations livrées par une quinzaine de fournisseurs de données utilisant différents systèmes informatiques et n'ayant pas tous le même degré de précision dans le relevé. De ce fait, à côté de ses considérations théoriques, le rapport illustre aussi les difficultés qui surgissent au moment d'appliquer pratiquement des modèles théoriquement fondés. En ce sens, la présente étude ne constitue pas une présentation définitive de cette nouvelle méthode, mais un premier pas vers la réalisation d'autres études plus poussées.

En publiant ce volume méthodologique, l'OFAS entend apporter sa contribution à l'élargissement de la discussion qui s'impose d'urgence pour une meilleure analyse des données de registre.

Herbert Känzig

Analyses fondamentales, Office fédéral des assurances sociales

Vorwort

Der vorliegende Bericht soll die Verwendungsmöglichkeiten neuerer statistischer Methoden und Informatiklösungen zur Durchführung von Längsschnittanalysen aufzeigen. Insbesondere interessiert, wie sie im Zusammenhang mit der Analyse grosser Datenmengen zu Fallklassifikationen genutzt werden können.

Die Statistiken, die im Bereich des Gesundheitswesens vorliegen, sind in der Regel deskriptive Querschnittsanalysen, die sich auf ein bestimmtes Jahr beziehen. Die zeitliche Dimension wird eingebracht, indem die Erhebungen periodisch durchgeführt werden und die Resultate zu Zeitreihen zusammengestellt werden. Eigentliche Längsschnittanalysen liegen nicht vor, da für die Statistiken in der Regel aggregierte Daten erhoben werden.

Um Längsschnittanalysen durchführen zu können, müssen jedoch die Daten in einem Detaillierungsgrad vorliegen, der es ermöglicht, die einzelnen Subjekte in der zeitlichen Dimension verfolgen zu können, das heisst, dass eine Identifikation vorhanden sein muss, aufgrund der die Angaben zu einem Individuum von einem Statistikjahr mit denjenigen desselben Individuums der folgenden Jahre verknüpft werden kann.

Diese Forderung erfüllen die Daten, die im Rahmen der Evaluation der Kostenunterschiede zwischen den Versicherungsmodellen in der Krankenversicherung gesammelt wurden: Über eine Periode von vier Jahren liegen die Angaben zu ca. fünf Millionen Versicherten vor. Erhoben wurden sozio-demographische Informationen, gewähltes Versicherungsmodell und die verursachten Kosten in einer Form, die die Unterscheidung nach Leistungserbringergruppen (z.B. Ärzte nach Spezialität, Spitäler nach Art, Apotheken usw.) zulässt.

Für jedes Individuum ist damit ersichtlich, welche Dienste in jeder der vier Erhebungsperioden in Anspruch genommen wurden, zu welchen Kosten und zu welchem Zeitpunkt (sowohl der Zeitpunkt der Behandlung und das Jahr der Rückerstattung durch den Versicherer sind bekannt).

Um angesichts dieser immensen Datenmenge Aussagen machen zu können, müssen die verschiedenen Patientenlaufbahnen in Kategorien eingeteilt werden. Das kann geschehen, indem man auf der Basis der Erfahrungen einer eingehenden statistischen Datenanalyse im Hinblick auf die zu beantwortenden Fragen willkürliche Kriterien setzt, um die verschiedenen Krankheitsverläufe in sinnvolle Kategorien einzuteilen.

In der vorliegenden Arbeit wird ein neuer Ansatz vorgestellt, der für die Klassifizierung der Krankheitsverläufe nicht willkürliche Apriori-Zuordnungskriterien benutzt, sondern die Kategorien nur mit den in den Daten vorhandenen Informationen bildet.

Der Ansatz sieht den Krankheitsverlauf nicht nur als chronologische Aneinanderreihung der bezogenen Leistungen (geordnet nach dem Datum des Behandlungsbeginns). Die Periode der vier betrachteten Jahre wird in eine Anzahl gleichlanger Zeitabschnitte eingeteilt. Die Leistungen werden aufgrund des Behandlungsbeginns dem jeweiligen Zeitabschnitt zugeteilt. Mit diesem Vorgehen gehen auch diejenigen Zeitabschnitte in die Analyse ein, in denen keine Leistungen konsumiert wurden.

Die benutzte Datenbasis setzt sich zusammen aus Abgaben von ca. 15 Datenlieferanten mit unterschiedlichen EDV-Systemen und unterschiedlichem Detaillierungsgrad bei der Erfassung. So dokumentiert der Bericht neben den theoretischen Überlegungen auch die Schwierigkeiten, die bei der Umsetzung von theoretisch fundierten Modellen in die Praxis auftauchen. In diesem Sinne stellt die Untersuchung nicht eine abschliessende Abhandlung der Methode dar, sondern einen ersten Schritt zur Realisierung von weitergehenden Studien.

Das BSV möchte mit der Publikation des vorliegenden Methodenbandes einen Beitrag zu einer erweiterten, dringend notwendigen Diskussion für eine bessere Analyse von Registerdaten leisten.

Herbert Känzig

Kompetenzzentrum Grundlagen, Bundesamt für Sozialversicherung

Prefazione

Il presente rapporto intende evidenziare le modalità atte a destinare i metodi statistici e le soluzioni informatiche più recenti all'esecuzione di analisi longitudinali. Ciò che soprattutto interessa è sapere come queste modalità possono essere utilizzate, al fine di classificare singoli casi, nell'analisi di grandi quantità di dati.

Le statistiche nell'ambito della sanità pubblica sono di regola analisi trasversali descrittive relative ad un anno particolare. La dimensione temporale viene apportata eseguendo le rilevazioni periodicamente, ordinandone i risultati in una serie temporale. Non vi sono vere e proprie analisi longitudinali, dal momento che a fini statistici vengono di regola rilevati dati aggregati.

Per servire ad analisi longitudinali i dati devono tuttavia aver raggiunto un grado di dettaglio tale da consentire di seguire i singoli soggetti in una dimensione temporale. Questo significa che bisogna disporre di elementi d'identificazione di una persona che permettano di associarne i dati riferiti ad un anno statistico particolare a quelli degli anni successivi.

I dati raccolti nell'ambito della valutazione delle differenze dei costi tra i modelli assicurativi presenti nell'assicurazione malattie soddisfano questa esigenza: sono disponibili dati relativi a circa cinque milioni di assicurati, distribuiti su un periodo di quattro anni. La raccolta d'informazioni su aspetti sociodemografici, sul modello assicurativo scelto e sui costi cagionati è stata condotta in modo da permettere la differenziazione in base ai gruppi dei fornitori di prestazioni (p. es. i medici in base alla specializzazione, gli ospedali in base al genere, le farmacie ecc.).

Viene così evidenziato, per ogni persona e per ognuno dei quattro periodi di rilevazione, quali sono stati i servizi ottenuti, i loro costi ed il momento in cui sono stati richiesti (essendo noti sia il periodo di cura, sia l'anno del rimborso da parte dell'assicuratore).

Data l'immensa quantità di dati, sarà possibile esprimere un parere dopo aver suddiviso i decorsi delle malattie dei pazienti in categorie. Questa suddivisione in categorie appropriate potrà avvenire fissando criteri arbitrari

maturati da esperienze ottenute nell'ambito di un'analisi statistica e approfondita dei dati, operata in vista delle domande cui si dovrà rispondere.

In questo rapporto viene proposta una nuova impostazione che non utilizza a priori principi di attribuzione arbitrari per la classificazione dei decorsi delle malattie in categorie, ma le distingue unicamente attraverso le informazioni contenute nei dati.

Il decorso della malattia non viene qui inteso esclusivamente come sequenza cronologica delle prestazioni ottenute (ordinate secondo la data d'inizio della cura). Un periodo di osservazione di quattro anni viene suddiviso in un numero d'intervalli di uguale lunghezza. In base alla data d'inizio della cura, le prestazioni vengono distribuite sull'arco temporale corrispondente. Questa procedura permette di comprendere nell'analisi anche i periodi in cui non sono state ottenute prestazioni.

La base di dati utilizzata è composta dalle indicazioni di ca. 15 fornitori di dati con sistemi informatici diversi e un grado di dettaglio diverso al momento della rilevazione. Il rapporto documenta quindi, oltre alle considerazioni teoriche, anche le difficoltà che s'incontrano nella prassi all'applicazione di modelli aventi una base teorica. Perciò questa ricerca non rappresenta tanto una relazione conclusiva sul metodo, quanto un primo passo verso la realizzazione degli studi in corso.

Con la pubblicazione della presente raccolta di metodi l'UFAS vuole contribuire ad una discussione approfondita e assolutamente necessaria per una migliore analisi dei dati registrati.

Herbert Känzig

Centro di competenza Analisi fondamentali, Ufficio federale delle assicurazioni sociali

Foreword

The present report documents the possible applications of new statistical methods and data processing solutions to longitudinal studies, in particular for the analysis of extensive data sets for case classification.

As a rule, health care statistics are presented in the form of descriptive cross-sectional analyses for a particular year. Evolution over time is accounted for by periodic surveys, the results of which are expressed as temporal series. These are not real longitudinal studies however, since they are usually based on aggregated data.

Real longitudinal studies require data detailed enough to track particular subjects over a certain period of time. It must be possible to identify them, so as to connect data on a given individual from one statistical year to the next.

Data collected for the evaluation of cost differences between health insurance models fulfil this requirement; they provide information on approximately five million policyholders for a four-year period: socio-demographic information, the insurance model, and costs generated. They also distinguish between service providers (e.g. physicians according to their specialisation, hospitals according to type, pharmacies, etc.).

They document the services required by each policyholder during the periods surveyed, their costs, the time of treatment, and the year of compensation.

Due to the volume of the data set, patient pathways had to be categorised to allow for any meaningful readings. This was done by conducting an exhaustive statistical data analysis and establishing random criteria for the specific query, so as to divide courses of illness into meaningful categories.

The present study proposes a new approach: instead of using random a priori criteria to categorise courses of illness, it defines criteria on the basis of information provided by the data.

The course of an illness is not seen as a simple chronological listing of services provided (by beginning of treatment). The four years under investigation are subdivided into periods of equal length, and services are

listed under the relevant period by beginning of treatment. This ensures that periods in which no services were provided are also accounted for.

The data base was compiled from approximately 15 sources, in varying detail, and using different data processing systems. The report therefore documents not only theoretical considerations, but also the difficulties of implementing theoretical models in practice. In this sense, the study does not provide a final evaluation, and is only the first of an ongoing series of studies.

With its publication, the FSIO would like to contribute to a comprehensive and indispensable debate on ways to improve the analysis of register data.

Herbert Känzig

Bases (competence centre), Federal Office for Social Security

Synthèse

Ce qui suit est un rapport qui décrit les activités et consigne les réflexions issues d'un mandat de l'OFAS relatif à l'exploitation de la base de données de la LAMal. Ce rapport a deux parties distinctes (ce qui induit certaines répétitions, mais on veut qu'il soit possible de lire les deux parties indépendamment l'une de l'autre), la première, intitulée "Etat des lieux, travaux à faire," présente une version "synthétique et littéraire" de ce qui est décrit de manière plus technique dans la seconde. Celle-ci comporte trois parties: la description d'une méthode, son illustration, et enfin son évaluation telle qu'elle avait été prévue par le mandat. On trouvera en annexe une description de certains algorithmes nécessaires à la mise en oeuvre de la méthode que l'on décrit.

Zusammenfassung

Der vorliegende Bericht beschreibt die Aktivitäten und methodischen Überlegungen im Zusammenhang mit der vom BSV in Auftrag gegebenen Evaluierung der KVG-Datenbank. Der Bericht ist zweigegliedert (die zwei Teile sollen unabhängig voneinander gelesen werden können, weshalb es zu Wiederholungen kommen kann). Der erste, deskriptive Teil ist eine Zusammenfassung der Informationen, die im zweiten, technischen Teil dann näher erläutert werden. Der zweite Teil umfasst die drei folgenden Schritte: Beschreibung einer Methode, Darstellung der Methode und deren Evaluierung auf der Basis des Auftrags. Die Beilage enthält eine Beschreibung gewisser Algorithmen, die für die Anwendung der beschriebenen Methode erforderlich waren.

Riassunto

Il rapporto descrive le attività e le riflessioni scaturite da un mandato dell'UFAS relativo all'utilizzazione della banca dati della LAMal. Esso è suddiviso in due parti ben distinte: la prima, intitolata "Etat des lieux, travaux à faire", presenta in modo "sintetico e letterario" ciò che viene descritto in maniera più tecnica nella seconda. Quest'ultima è a sua volta costituita da tre parti: la descrizione di un metodo, la sua illustrazione e, infine, una sua valutazione come previsto dal mandato. La struttura particolare di questo rapporto implica certe ripetizioni. Ciò è voluto affinché le due parti summenzionate possano essere lette indipendentemente l'una dall'altra. In un allegato vengono descritti alcuni algoritmi necessari all'applicazione del metodo in questione.

Summary

The present document is a report which describes the activities and indicates the considerations which resulted from a mandate issued by the Swiss Federal Office of Social Insurance relating to the use of the health insurance law data base. The report comprises two separate parts, which has led to a certain degree of repetition, but the intention is that each part can be read as an independent whole. The first is entitled “Current situation – Work to be done” and presents a “synthetic and literary” version of what is described more technically in the second part. The latter is divided into three sections: the description of a method, its illustration and finally its evaluation as foreseen under the terms of the mandate. A description of certain algorithms necessary for implementing the method described can be found in the appendix.

Contents

1	Etat des lieux, travaux à faire	1
1.1	Objectifs	2
1.2	Etat des lieux	3
1.3	Travaux à faire	5
1.3.1	Problème N° 2	5
1.3.2	Problème N° 3	6
1.4	Conclusions	6
2	Les faits premiers	7
3	De la méthode	9
3.1	Usage actuel de la BDL	9
3.2	Une autre forme de description des histoires des remboursements	11
3.3	Des volumes	15
3.4	Les données de la méthode	17
3.4.1	Le fichier des trajectoires	17
3.4.2	Le fichier des segments	17
4	Illustration de la méthode: la psychiatrie en cabinet privé	18
4.1	Notation	18
4.2	Quelques repères concernant la psychiatrie en cabinet privé . .	18
4.3	La psychiatrie en cabinet privé sporadique et la psychiatrie en cabinet privé régulière	24
4.3.1	Effectifs de la PCPS et de la PCPR dans le corpus utilisé	25
4.3.2	Domaines médicaux présents dans les remboursements relatifs à la PCPR et à la PCPS	26
4.3.3	Comparaison des durées des périodes selon la catégorie (régulière ou sporadique) et le domaine médical	30
4.4	Un examen plus détaillé des données de la classe PCPR	33
4.4.1	La répartition des remboursements pour les soins en psychiatrie	33
4.4.2	Comparaison des répartitions des remboursements, dans la classe PCPR, pour les divers domaines médicaux . .	34

4.4.3	Assurés de la PCPR pour lesquels il y a exactement une période de remboursements dans le domaine de la psychiatrie	39
4.4.4	Assurés de la PCPR pour lesquels il y a exactement deux périodes de remboursements dans le domaine de la psychiatrie	41
4.4.5	Assurés de la PCPR pour lesquels, au cours d'au moins une période, la psychiatrie est à l'origine des remboursements les plus importants	44
4.5	Les profils psychiatriques	44
5	Evaluation de la méthode selon les normes du mandat	45
5.1	<i>Ausgangslage</i>	46
5.1.1	<i>Die Daten mit Hilfe von Kohorten- und Längsschnittansätzen aufzubereiten und zu analysieren:</i>	46
5.1.2	<i>Die verwendeten Methoden sowie die verwendeten Kategorisierungs- und Klassifikationssysteme allgemein verständlich darzustellen und überprüfen:</i>	46
5.1.3	<i>Die Programme so zu verallgemeinern, dass sie auf unterschiedliche Fragestellungen und Datenbestände angewendet und durch beliebige - fachkundige - Benutzer verwendet werden können. Voraussetzung dafür ist eine entsprechende Anleitung und eine detaillierte Dokumentation:</i>	49
5.2	<i>Auftrag</i>	49
5.2.1	<i>... erarbeitet einen Bericht über die Verwendungsmöglichkeiten neuester statistischen Methoden und Informationslösungen zur Analyse von Längsschnitt und Kohortendaten. Der Bericht hat insbesondere aufzuzeigen wie sich im Rahmen der Administrativdatenstatistik stellenden Probleme in Zusammenhang mit der Analyse grosser Datenmengen und zur Fallklassifikation nutzen lassen.</i>	49

5.2.2	<i>Parallel zum Bericht sind die praktischen Arbeiten anhand der vorhandenen Daten zu illustrieren und die notwendigen Programme aufzubauen, damit diese für die zukünftige Administrativdatenstatistik in den produktiven Betrieb übernommen werden können:</i>	50
5.3	<i>Teilschritte</i>	52
5.3.1	<i>Aufzeigen der zusätzlichen Analysemöglichkeiten/Notwendigkeit von Kohorten-, Längsschnitt- und Pfadanalysen für die Bearbeitung von Problemstellungen, wie sie im Rahmen des Administrativdatenstatistik vorkommen:</i>	52
5.3.2	<i>Erörterung der Vor- und Nachteile des Vorgehens verglichen mit anderen Analysemethoden:</i>	53
5.3.3	<i>Allgemein verständliche Erläuterung der in der HMO/-Bonus-Evaluation erarbeiteten konkreten Lösungen sowie Illustration anhand konkreter Beispiele:</i>	54
5.3.4	<i>Programme:</i>	54
5.3.5	<i>Die theoretischen Überlegungen und die darauf aufbauende Programmierung ist anhand von Datenanalysen mit den daten des Rechnungsjahre 1991-1994 sowie 1997 und 1998 auszutesten und zu illustrieren:</i>	55
6	Description de quelques algorithmes	56
6.1	Organisation générale des algorithmes	56
6.2	Préparation des données	58
6.3	Transformation de Haar, "décimation" et transformation de Haar inverse	63
6.3.1	Transformation de Haar	63
6.3.2	Niveau des coefficients	66
6.3.3	"Décimation"	67
6.3.4	Reconstruction de (l'approximation de) la trajectoire	68
6.3.5	Réduction du nombre de valeurs de la fonction	68
6.4	Segmentation des trajectoires	69
6.4.1	Segmentation proprement dite	69
6.4.2	Périodes	73
6.4.3	Spécialités médicales et pourcentages des remboursements associés aux périodes	73

1 Etat des lieux, travaux à faire

Une exploitation de la base de données de la LAMal qui réponde aux exigences de la loi demande qu'auparavant soient nécessairement résolus trois problèmes, et cela dans l'ordre qui s'impose "naturellement" et qui est le suivant:

1. Il faut pouvoir accéder "intelligemment" aux données, c'est-à-dire pouvoir en faire la statistique exploratoire.
2. Il faut pouvoir évaluer la qualité et l'utilité de l'information contenue dans cette base, c'est-à-dire déterminer ce qu'on peut en attendre en termes d'analyse et d'exploitation.
3. Il faut enfin pouvoir maîtriser les volumes informatiques, c'est-à-dire pouvoir mener à terme analyses et exploitations.

La base de données de la LAMal résulte de la fusion et du toilettage de données administratives fournies dans le cadre de l'OAMal (Ordonnance relative à l'assurance maladie, article 28) par les différentes caisses maladie autorisées à pratiquer. De ce fait elle ne contient que des informations relatives aux caractéristiques d'assurance des assurés et aux remboursements dont ils bénéficient. La seule information médicale dont on dispose est le domaine de spécialisation du fournisseur qui établit la facture dont les éléments sont enregistrés dans la base.¹ *Toute l'utilité de la base, à des fins "managériales," va dépendre d'une part de l'étroitesse du lien que l'on saura tisser entre remboursements et spécialités médicales, et, d'autre part, de la capacité d'"expression" de ce lien, à savoir de la possibilité de l'interpréter causalement.*

La caractéristique de la base qui permet de penser que la recherche de ce lien peut être fructueuse² est la nature chronologique des observations:

¹Dans ce contexte, l'hôpital est un domaine de spécialisation! Dans la suite, tout ce qui a trait au domaine médical l'est dans ce sens restreint.

²Mais qui en fait aussi la difficulté!

pour chaque assuré, on dispose de l'historique des remboursements dont il a bénéficié et des domaines médicaux qui sont à leur source.

A l'heure actuelle, on dispose d'une méthode dont on a des raisons de croire, avec un degré de certitude suffisant pour justifier la poursuite de ce travail, qu'elle résout le premier problème énoncé plus haut. Elle est le résultat concret du mandat qui fait l'objet de ce rapport. Cette méthode doit encore être "calibrée," et cette "calibration" ne peut être précisée qu'au travers des travaux relatifs à la solution du second problème énoncé plus haut. Finalement, les trois problèmes inventoriés ne peuvent en pratique être complètement dissociés, et la méthode retenue pour la solution du premier problème ne mobilise que des instruments compatibles avec la solution du troisième problème, même si l'on n'a pas fait l'effort, dans ce cadre, de trouver les algorithmes les plus performants. On peut en effet considérer ce premier travail comme une étude de faisabilité.

1.1 Objectifs

On s'intéresse ici non pas aux objectifs généraux tels qu'ils apparaissent dans la loi, mais à la manière dont ceux-ci sont poursuivis en pratique. Et l'on choisit un exemple particulier, celui de la surmédicalisation. On cite:³

Selon la jurisprudence, un traitement contrevient à l'économie lorsque les soins donnés par un médecin pendant une longue période sont en moyenne plus chers que les soins dispensés par un groupe de médecins comparables, cela sans que des circonstances particulières ne permettent de justifier le dépassement de la moyenne.

L'approche retenue permet tout d'abord de mettre immédiatement en cause le recours aux chiffres moyens, car pratiquement toutes les répartitions de remboursements que l'on peut examiner sont fortement asymétriques si bien que les moyennes en sont de très piètres descripteurs.⁴ Elle permet ensuite

³La surmédicalisation et ses conséquences, *infosantésuisse*, 9 (2001), page 9, Critères économiques.

⁴La moyenne des valeurs 1, 2, 3, 4, 5, 6, 7, 8, 9, 105 vaut 15, ce qui ne représente adéquatement ni les valeurs de 1 à 9, ni la valeur 105. Or les données relatives aux remboursements sont toutes de cette nature!

de prendre “naturellement” en compte la nature chronologique des soins, car elle est construite sur de telles chronologies, et elle permet de comparer les pratiques des médecins *en tenant compte des caractéristiques du “portefeuille” d’assurés que soigne chacun de ceux-ci,*” les coûts étant fortement dépendants de l’état de santé du malade et les divers médecins ne soignant pas des groupes identiques de personnes.

1.2 Etat des lieux

Les méthodes traditionnelles d’accès aux données d’une base de données, comme l’âge ou le sexe, ou le type d’assurance, ne permettent pas, dans le cas présent, de comprendre la nature des données de la base car elles sont impuissantes à en saisir la dimension chronologique et évolutive, dimension qui lui confère sa seule utilité potentielle. La première tâche a donc été de construire les instruments de cet accès et de vérifier leur pertinence. Ces instruments sont au nombre de deux et constituent en fait deux bases de données dérivées des données originales:

1. une base de données de *trajectoires* de remboursements:

Une trajectoire représente l’histoire des remboursements dont un assuré a bénéficié ainsi que celle des domaines médicaux qui en sont à l’origine. Cette base de données contient des trajectoires *comparables*, ce qui n’est pas le cas pour les trajectoires telles qu’elles peuvent être extraites de la base “originelle.” Finalement, ces trajectoires sont des approximations efficaces et modulables des trajectoires “complètes,” l’approximation étant rendue nécessaire par les volumes de données à traiter.

2. une base de données de *périodes* de remboursements:

Les périodes émergent empiriquement des trajectoires des remboursements dont on exploite la “géométrie.” C’est au travers de ces périodes que se tisse le lien entre remboursements et activités de soins.⁵ C’est

⁵Pratiquement, il faut pouvoir dissocier les soins “importants” des soins “marginaux.” Ainsi une visite chez un ophtalmologue pour un contrôle de vue annuel est “marginale” chez quelqu’un qui est soigné intensivement pour des problèmes cardiaques. Si l’on veut une évaluation appropriée du coût des soins relevant de la cardiologie il faut donc “évacuer” ces interventions “marginales.” Si on ne le fait pas, le “bruit” du système augmente et les effets “significatifs” sont d’autant plus difficiles à déceler.

ici que la “calibration” dont il a été question doit se faire et elle doit résulter en un compromis qui ménage les nécessités de la connaissance et celles de l’efficacité lors du traitement des volumes: plus les approximations sont bonnes, plus le lien est étroit, mais aussi plus le traitement informatique devient lourd. Seule une exploitation fine des données peut déceler les seuils de “rendement décroissant.”

Etant donné que l’approche retenue, faute d’informations relatives aux diagnostics et aux soins délivrés, est nécessairement empirique, sa validation passe par sa mise en oeuvre. On a donc procédé à des analyses statistiques élémentaires pour vérifier plusieurs points indispensables à la validation, points que l’on va aborder sous la forme d’une suite de questions auxquelles la réponse s’est toujours révélée affirmative:

1. *Les bases de données dérivées que l’on envisage de construire permettent-elles un accès “intelligent” aux données?*

Exemple: Peut-on distinguer entre les malades chroniques et les autres? Les premiers présentent souvent, dans leur trajectoire des remboursements, quand la maladie est stationnaire, des périodes longues, peu nombreuses et “linéaires.”⁶

2. *Les bases de données dérivées que l’on envisage de construire aident-elles à la mise en oeuvre de modèles statistiques explicatifs?*

Exemple: En étudiant, par exemple, la psychiatrie, on voit immédiatement qu’il faut trouver un moyen pour à la fois tenir compte des séjours hospitaliers, car ils sont coûteux, et les traiter de façon spéciale car ils “écrasent” les autres remboursements.⁷

3. *Les bases de données dérivées que l’on envisage de construire révèlent-elles des aspects inattendus des données, ou auxquels on n’a pas prêté attention?*

Exemple: Le discours relatif à la base de données de la LAMal est un discours qui privilégie le problème des volumes “excessifs.” Or l’examen

⁶Il est clair que ces études ne visent qu’une exploitation scientifique de la base de la LAMal à des fins publiques, et non pas l’identification des “bons” et “mauvais” cas.

⁷En jargon on parle de valeurs excentriques.

des séquences médicales issues des périodes montre bien que ce volume peut s'avérer insuffisant!

Ici s'arrête le travail fait dans le cadre du mandat. Les étapes suivantes (problèmes 2 et 3 de l'énoncé de départ) n'ont été prises en compte que marginalement, le but essentiel étant d'en dégager les grandes lignes (*Travaux à faire*).

1.3 Travaux à faire

1.3.1 Problème N° 2

La première tâche à accomplir est celle de reprendre les programmes qui ont servi pour la première partie de ce travail, et ce dans le but de rendre possible la "calibration" dont il a été question plus haut.

La deuxième tâche est celle qui recherche les modèles explicatifs appropriés. Il y a plusieurs approches potentiellement utiles parmi lesquelles celles qui paraissent les plus immédiatement utiles sont connues sous les noms respectifs de *event history analysis*, *generalized hidden Markov models*, *variable length Markov models*. Toutes ces techniques servent à la poursuite d'objectifs semblables: décrire et caractériser les trajectoires des remboursements et domaines médicaux associés. Toutes s'efforcent de prendre en compte le fait que l'état présent de santé d'un assuré dépend de variables contingentes comme l'âge et le sexe, mais aussi de son histoire de santé, donc de son passé. On peut ainsi s'intéresser à l'évolution des soins psychiatriques dont bénéficie un assuré: les périodes de traitement s'allongent-elles ou raccourcissent-elles, les remboursements associés augmentent-ils ou diminuent-ils, y a-t-il des différences régionales ... ? L'utilité de ces modèles est crucialement dépendante du contrôle de la variabilité que l'on peut exercer puisque la signification statistique se fait toujours à l'aulne de la variabilité intrinsèque et que celle-ci est d'autant plus grande que l'on échoue à identifier les sources extrinsèques de la variabilité globale.⁸ A cette fin les outils déjà développés joueront un rôle déterminant.

Finalement il faudra sur la base des informations révélées par les analyses

⁸Il n'y a pas à proprement parler de répétition des observations!

établir un portefeuille du souhaitable et du possible, car c'est ce qui fera l'objet des exploitations systématiques de la base, et qu'il faudra informatiser et diffuser. On a déjà constaté une difficulté réelle à donner une expression opérationnelle au souhaitable et le possible relève de l'habileté des analystes!

1.3.2 Problème N° 3

L'exploitation de la base de données de la LAMal exige un investissement informatique non négligeable en termes de calcul tout d'abord, mais surtout en termes de définition du système d'information qui va alimenter cette exploitation. Cette définition doit émerger des solutions que l'on donne au problème N° 2 et des utilisations envisagées. Il faudra en effet optimiser et l'on ne sait optimiser que ce qui est étroitement défini. Dans toute la démarche que l'on vient de décrire, il faut faire très attention à ne pas "polluer" les solutions du problème N° 2 par des excursions intempestives dans le domaine des solutions du problème N° 3!⁹

1.4 Conclusions

Dans la mesure où l'on a l'ambition de *comprendre* pour les *maîtriser* les phénomènes *réels* de l'assurance maladie, une démarche du type de celle que l'on vient de décrire est incontournable.¹⁰ C'est une démarche de longue haleine qui exige la plus grande rigueur intellectuelle et un recours à des techniques statistiques et informatiques très contemporaines. Elle doit être comprise comme une "démarche qualité" d' "amélioration continue."¹¹ A ce titre toutes les améliorations qui peuvent être apportées aux données "originelles" auront un impact dépassant de beaucoup les coûts qu'elles occasionnent.

⁹Il y a, par exemple, une possibilité, grâce aux méthodes de classification utilisées, d'évaluer les avantages et inconvénients d'un recours à des échantillonnages, ce qui ne peut pas se faire correctement sans ces méthodes. Si ce recours était possible, cela aurait pour conséquence d'alléger considérablement la charge informatique, et donc d'accélérer l'étape "recherche et développement."

¹⁰Les discussions publiques - journaux, radio, télévision - de septembre-octobre 2001 ayant pour sujet l'assurance maladie ne font que conforter cette opinion!

¹¹Voir W.E. Deming, *Out of the Crisis*, Cambridge University Press, Cambridge, UK (1982), page 465: *Efforts and methods for improvement of quality and productivity are in most companies and government agencies fragmented, with no overall competent guidance, no integrated system for continual improvement.*

2 Les faits premiers

La loi sur l'assurance maladie (LAMal dans la suite) impose le développement d'un répertoire de données (BDL dans la suite, pour **B**ase de **D**onnées de la LAMal) qui doit servir à en comprendre les effets et à en permettre l'évolution. A cette fin, l'Office fédéral des assurances sociales (OFAS dans la suite) récolte auprès des caisses maladie concernées des informations de deux types (qui doivent être complétées: on tient compte dans ce qui suit des ajouts):

- des informations de nature administrative, rendues anonymes:
 - sexe, âge, commune de domicile, couverture ... des bénéficiaires de soins,
 - canton, domaine d'activité médicale ... des prestataires de soins;
- des informations de nature financière et médicale:

pour chaque facture on dispose d'un "marquage temporel et médical" (dates liées à la facture, catégorie médicale dont relèvent les remboursements) ainsi que de diverses informations quantitatives relatives aux remboursements.

L'exploitation de ces données soulève deux problèmes de nature très différente, mais qu'on ne peut dissocier:

- la masse considérable des données (chaque assuré est potentiellement présent dans la base),
- la nature des exploitations (quelles analyses à quelles fins) qui elle exige le développement d'un appareil conceptuel qui est, pour l'instant, embryonnaire.

Ce rapport décrit et illustre une méthode potentiellement utile pour l'exploitation de la BDL. Elle a été conçue comme la première étape d'un processus

qui devrait se décomposer en trois tâches distinctes et successives qui sont:

Etape 1: Construction et évaluation d'un instrument de lecture de la BDL qui volontairement ignore la dimension "volume" du problème pour dégager les conditions *nécessaires* à une exploitation intelligente et utile de la BDL.

Il s'agit avant tout de produire des fichiers qui permettent d'exploiter en particulier la dimension temporelle des données ainsi que d'acquérir une certaine familiarité avec l'exploitation de ces fichiers.

Etape 2: Exploration d'un champ restreint de la BDL dans le but d'en évaluer le contenu en information.¹²

Cela doit permettre aussi de dégager une organisation précise de la partie informatique: qu'est-ce qui peut être obtenu par l'utilisation d'un langage d'interrogation de base de données, qu'est-ce qui doit être programmé et comment . . . Par ailleurs, la segmentation (l'un des instruments de lecture issus de la première étape) se fait à partir de seuils et d'une approximation particulière des histoires des remboursements: il faut pouvoir expérimenter avec ces seuils et ces approximations pour pouvoir faire les choix les plus appropriés. Vaut-il mieux avoir des périodes (les divers segments) plus longues ou plus courtes? Y a-t-il un lien entre la longueur d'une période et la précision de son marquage médical (les domaines médicaux qui lui sont associés)? Il est inutile d'investir dans des développements informatiques lourds avant d'avoir réglé ces questions et celles-ci ne peuvent trouver de réponse sans une expérimentation extensive!

Etape 3: Développement de l'infrastructure technique permettant la prise en compte du volume.

Il est indispensable, pour mener à bien cette information, d'avoir complété de manière exhaustive, l'Etape 2.

¹²Ce sera la psychiatrie.

Le “devrait” utilisé ci-dessus veut attirer l’attention et insister sur le fait que les problèmes relatifs à l’Etape 3 ne peuvent être ignorés lors de l’Etape 1. On s’est donc efforcé, lors de cette première étape, de ne faire appel qu’à des méthodes susceptibles d’être “portées” vers l’Etape 3.

Ce rapport est structuré de la manière suivante. Dans un premier temps on décrit et justifie les choix méthodologiques retenus. Dans un deuxième temps, on illustre à l’aide d’exemples “immédiats” ce que la méthode développée permet de faire, et ce dans le cadre de données relevant de la psychiatrie en cabinet privé. Enfin, dans un troisième temps, on évalue ce qui a été fait et ce qui reste à faire, et cela dans le cadre qui a été fixé lors de la définition du mandat dont ce rapport est l’émanation.

On n’exploite, dans ce rapport, que les données relatives aux remboursements et pas celles relatives aux prestations et aux factures. La raison en est toujours la même: l’objectif premier est ici l’évaluation du potentiel de la méthode proposée et non pas son exploitation exhaustive. Il est clair que la mise en oeuvre doit tenir compte de toute l’information disponible et le développement de la méthode de son extraction est l’objet de l’Etape 2.

3 De la méthode

3.1 Usage actuel de la BDL

L’usage qui est actuellement fait de la BDL relève de deux catégories conceptuelles qui sont largement déterminées par le problème des volumes:

- *l’agrégation*

Si l’on veut, par exemple, savoir combien “coûte” la psychiatrie, on fait la somme, “sur la BDL,” des remboursements qui sont libellés “psychiatrie.” Il est clair que cette *information*, qui n’est pas en soi inintéressante, ni inutile, n’apporte qu’une *connaissance* très limitée de ce qui se passe dans le domaine des soins psychiatriques. Une exemple

de connaissance plus utile serait une réponse à la question suivante:

Quel est le pourcentage d'assurés, qui, suite à un traitement psychiatrique, voient leur état s'améliorer?

La réponse à une telle question présuppose que l'on sache définir, à partir des données disponibles, et de manière opérationnelle deux notions: celle de "patient en psychiatrie" et celle d'"amélioration."

L'intérêt de la réponse à une telle question semble évident, mais il convient peut-être d'insister: si l'on connaît le taux d'amélioration et le taux d'"entrée en maladie," on peut commencer à étudier l'évolution des "coûts" de la psychiatrie.¹³ On verra plus loin comment la méthode proposée s'efforce de donner des réponses à de telles questions.

- *l'indexation*

Il a été tôt reconnu à l'OFAS que l'"histoire" des remboursements des assurés est probablement l'information la plus utile que contient la BDL et on s'est efforcé de l'exploiter en associant à chaque assuré, par période administrative, deux indices, celui d'"intervenant principal" (IP dans la suite) et celui de "catégorie médicale principale" (MP dans la suite). Le premier veut saisir le domaine médical des médecins les plus actifs alors que le deuxième représente le domaine médical principal dont relèvent les soins prodigués. Ces indices souffrent d'imperfections de deux types. Ils se calculent selon un "algorithme" de type majoritaire, sur des périodes typiquement administratives, avec le résultat que des prestations chevauchant les périodes administratives sont "ventilées" sur les périodes adjacentes et que des prestations coûteuses mais peu fréquentes et de courte durée sont occultées par des prestations moins importantes mais plus fréquentes.

Par ailleurs on a essayé d'"expliquer" statistiquement les indices IP et MP par les données disponibles et l'on a dû constater que le lien entre

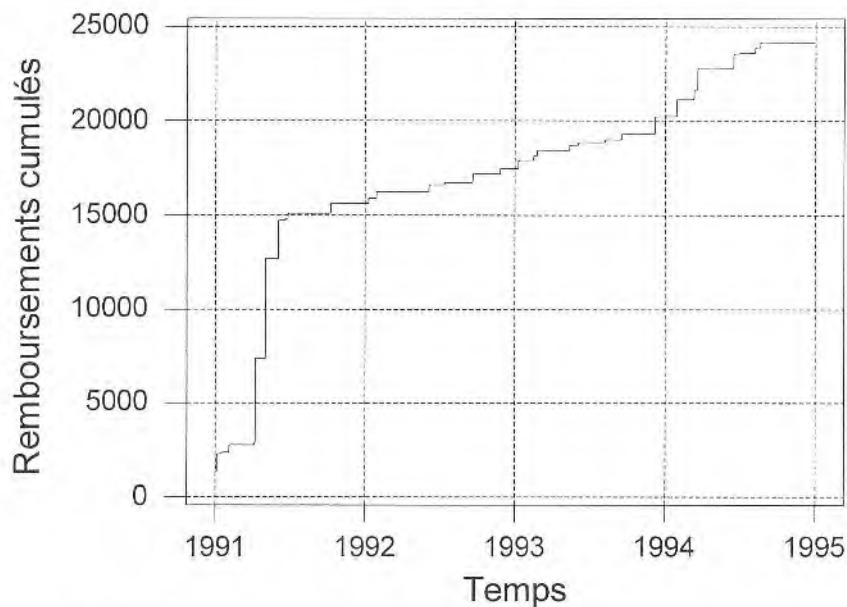
¹³Les guillemets s'efforcent d'attirer l'attention du lecteur sur le fait que de nombreux termes utilisés couramment n'ont pas de définition opérationnelle et que cette définition reste à trouver. C'est encore un travail dévolu à l'Etape 2.

données explicatives et données à expliquer est, dans le meilleur des cas, ténu.

3.2 Une autre forme de description des histoires des remboursements

On a donc recherché un moyen plus précis de rendre compte de l'évolution des remboursements dont a bénéficié un assuré. Or il se trouve que l'examen de l'historique des remboursements cumulés (c'est ce qu'on appelle dans la suite une trajectoire) conduit inmanquablement à la conclusion qu'il s'y trouve des périodes de remboursements, soit des "durées" durant lesquelles la "trajectoire" des remboursements (leur évolution dans le temps) présente des caractéristiques géométriques précises et interprétables. Voici un exemple de trajectoire:

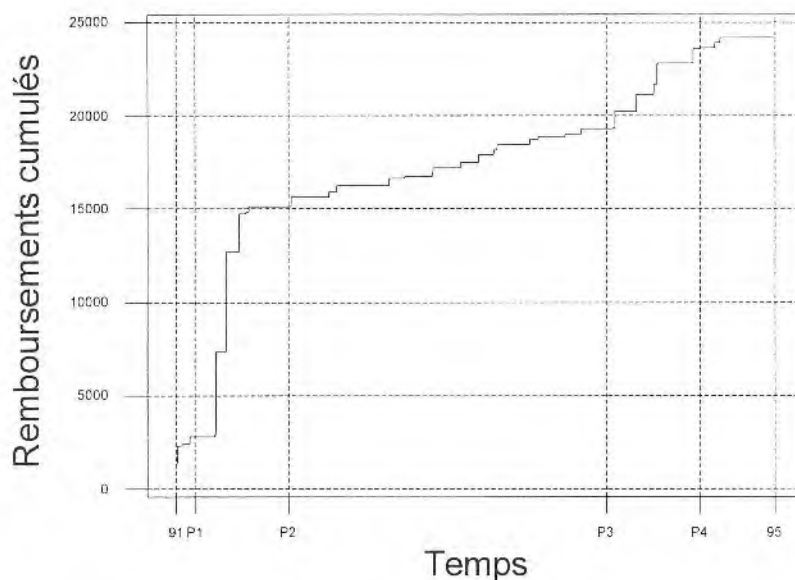
Une trajectoire



Assuré No 391

Ainsi une tendance linéaire représente des remboursements réguliers dont l'importance relative est stable. La "pente" de cette tendance est liée à l'importance des remboursements. Par contre, une portion de trajectoire qui croît très rapidement pour se tasser par la suite peut représenter une période de soins intensifs. C'est probablement par la médiation de ces périodes que la maladie manifeste sa présence dans les coûts. Voici le résultat du découpage d'une trajectoire de remboursements cumulés, découpage exploitant les propriétés géométriques de la trajectoire:

Une trajectoire et ses périodes



Assuré No 391

Supposons alors que les périodes successives des remboursements du patient A soient $P_1^A, \dots, P_{n_A}^A$.¹⁴ A toute période P_i^A , $1 \leq i \leq n_A$, on peut associer

¹⁴Il y a donc n_A périodes de remboursements pour l'assuré A .

- sa durée δ_i^A ,
- l'importance des remboursements ρ_i^A ,
- les pourcentages des remboursements imputables aux diverses catégories médicales retenues $\pi_{i,k}^A$, $1 \leq k \leq p$.¹⁵

On peut ensuite ordonner la série $\pi_{i,1}^A, \dots, \pi_{i,p}^A$ par ordre décroissant pour obtenir

$$\pi_{i,(p)}, \pi_{i,(p-1)}, \pi_{i,(p-2)}, \dots, \pi_{i,(1)}.$$

Si σ désigne un seuil prédéfini et “élevé,” et que $\pi_{i,(p)} > \sigma$, on pourra dire que le domaine médical représenté par (p) est celui qui “résume médicalement” la période de remboursements P_i^A . D'autres choix sont évidemment possibles: on peut retenir les trois “premiers” domaines médicaux (p) , $(p-1)$, et $(p-2)$, et c'est ce qui a été fait pour ce qui suit.

Voici une table contenant trois exemples qui illustrent ce qui précède:¹⁶

¹⁵L'indice k représente la catégorie médicale numéro k et le nombre de catégories est p . On a travaillé avec $p = 41$, et une définition des catégories due à l'OFAS. Si $\rho_{i,k}^A$ est l'importance du remboursement pour le patient A , durant la période numéro i , qui relève de la catégorie médicale numéro k , on calcule:

$$\pi_{i,k}^A = 100 \times \frac{\rho_{i,k}^A}{\rho_i^A}.$$

¹⁶Les commentaires sont à la suite de la table.

Segmentation de quelques trajectoires de remboursements cumulés

ID	Période	Durée en %	Domaine médical	Pct des rems	Domaine médical	Pct des rems	Domaine médical	Pct des rems	Rembt total	Total % de rems
6502	7	25	PHARM	44	GENER	30	SOINS	14	5642	88
6502	6	28	AUN+M	56	SUN-M	31	NO	4	39669	91
6502	5	16	GENER	53	PHARM	47	Néant		1134	100
6502	4	12	AUN+M	55	SUN-M	18	AUN-M	13	17265	86
6502	3	3	PHARM	100	Néant		Néant		282	100
6502	2	12	AUN-M	59	SOINS	22	GENER	9	5927	90
6502	1	3	GENER	68	PHARM	32	Néant		258	100
7603	8	16	SGN-M	100	Néant		Néant		102	100
7603	7	16	NO	91	GENER	6	CHIRG	3	2487	100
7603	6	6	Néant		Néant		Néant		Néant	Néant
7603	5	12	NO	84	PSYCH	11	CHIRG	5	5244	100
7603	4	19	NO	54	PSYCH	30	SGN-M	9	4260	93
7603	3	9	NO	72	PSYCH	15	SGN-M	8	2960	95
7603	2	12	NO	64	PSYCH	27	CHIRG	4	3497	95
7603	1	9	CHIRG	100	Néant		Néant		259	100
7772	6	31	SGN-M	48	PSYCH	13	SOINS	12	11913	73
7772	5	28	PHARM	27	OPHTL	26	GENER	24	6059	77
7772	4	16	SGN-M	73	OPHTL	8	PSYCH	8	9833	89
7772	3	12	PSYCH	34	PHARM	33	GENER	21	941	88
7772	2	9	PSYCH	58	PHARM	23	NO	11	1194	92
7772	1	3	SGN-M	96	OPHTL	4	Néant		9680	100

Remarques:

1. Tous les pourcentages ont été arrondis, ce qui explique qu'ils ne somment pas à cent dans la colonne "Durée." Le pourcentage est sur quatre années: la période 7 de l'assuré 6502 est donc une période d'une année.
2. Les périodes sont numérotées dans l'ordre inverse de leur apparition.
3. Les libellés des domaines médicaux dont le sens n'est pas évident ont les significations suivantes:
 - AUN±M := Soins ambulatoires dans un hôpital universitaire avec (+) ou sans (-) maternité
 - SUN±M := Séjour dans un hôpital universitaire avec (+) ou sans (-) maternité
 - SGN±M := Séjour dans un hôpital de soins généraux avec (+) ou sans (-) maternité
 - NO := Domaine médical inconnu

Le groupe d'assurés qui a servi à la construction de la table a été choisi "aléatoirement" parmi ceux qui, au moins une fois, au cours des quatre ans pris en compte, ont bénéficié de remboursements pour des soins relevant de la psychiatrie de ville, mais qui n'ont jamais été, au cours de ces mêmes quatre ans, au bénéfice de remboursements pour prestations en établissement psychiatrique. On a que:

1. l'assuré 6502 a bénéficié de remboursements dans le domaine de la psychiatrie mais leur montant est suffisamment petit pour ne pas apparaître dans ce décompte: on peut considérer dans ce cas que la psychiatrie n'intervient que marginalement;

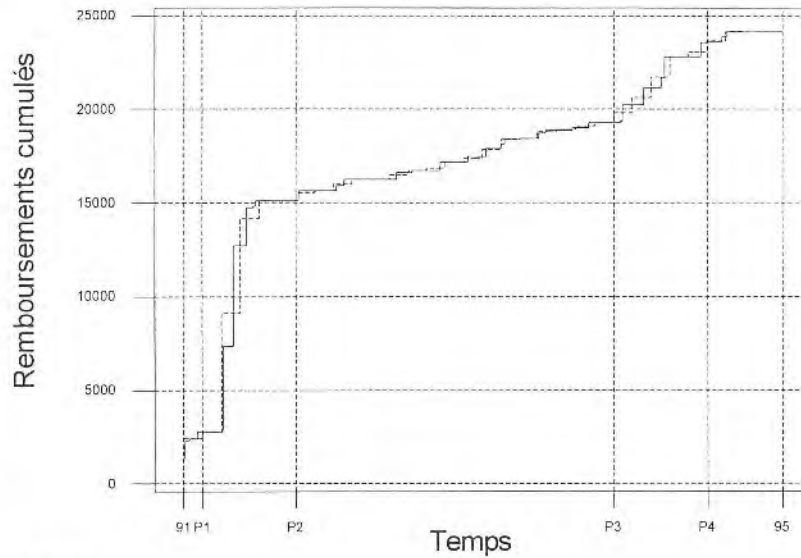
2. l'assuré 7603 a bénéficié de remboursements relativement importants dans le domaine de la psychiatrie bien que leur montant ne soit jamais le plus important;
3. l'assuré 7772 a bénéficié de remboursements relativement faibles dans le domaine de la psychiatrie bien que leur montant soit à deux reprises le plus important.

Si l'on veut "évaluer" la psychiatrie, il faudra décider si les assurés 7603 et 7772 relèvent d'une même catégorie ou de catégories différentes. La validité d'un tel choix ne peut qu'émerger à l'usage, et c'est pourquoi les trois étapes du processus décrit plus haut sont indispensables.

3.3 Des volumes

Comme on l'a mentionné, l'unité statistique de base est la trajectoire des remboursements. Il faut donc un moyen de représenter les trajectoires de manière économique et comparable, et l'on a choisi, dans un premier temps, d'utiliser des approximations des trajectoires à l'aide de 32 nombres, ce qui représente une économie non négligeable puisque le nombre de prestations dont bénéficie en moyenne un assuré est de l'ordre de la centaine. Les approximations sont faites sur une grille régulière, et c'est ce qui permet la comparaison directe des trajectoires des remboursements. Voici un exemple de trajectoire avec son approximation:

L'approximation d'une trajectoire



Assuré No 391

Remarques:

1. La durée d' "observation" va du début 1991 à fin 1994.
2. Les débuts et fins de périodes sont délimités par les libellés $P1$, $P2$, $P3$, $P4$.
3. La trajectoire "observée" est en trait continu, l'approximation en pointillé.

Cette manière de procéder est coûteuse en temps de calcul intermédiaire: c'est pourquoi il est nécessaire de bien calibrer les algorithmes et de produire des entrepôts de données dédiés à des analyses spécifiques.¹⁷

¹⁷Il faut en sus développer des instruments de manipulation des trajectoires.

3.4 Les données de la méthode

3.4.1 Le fichier des trajectoires

Une *trajectoire* désigne une suite de couples $(t_i, f(t_i))$, $1 \leq i \leq m$, où t_i est l'instant de réalisation d'un événement et $f(t_i)$ la valeur d'une variable f observée à ce même instant. Dans le contexte de la BDL, les t_i sont des dates issues des factures qui servent aux remboursements et les variables f sont, par exemple, le montant des remboursements, le nombre des prestations ou encore le nombre de factures. La méthode que l'on est en train de décrire produit et utilise un ou plusieurs fichiers de trajectoires.¹⁸ Un fichier de trajectoires contient pour chaque assuré¹⁹ une histoire, celle d'une variable observée au cours d'une période donnée,²⁰ sous forme de 16, 32, 64 ou davantage de "points d'histoire." Il y a une ligne par assuré et ces lignes constituent l' "univers statistique" qu'il faut exploiter.

Ces histoires ont trois fonctions:

1. Elles sous-tendent la classification des trajectoires selon les caractéristiques d'évolution qu'elles "incarnent" et le niveau des remboursements ou des prestations qu'elles représentent.
2. Elles servent de support à leur segmentation en "périodes de remboursements," périodes qui permettent de faire le lien avec la "maladie."
3. Elles favorisent, quand on les regroupe en trajectoires semblables, l'application des techniques statistiques indispensables à une étude "fine" des données de la BDL.²¹

3.4.2 Le fichier des segments

Ce fichier a pour l'essentiel la forme du tableau présenté en 3.2 (page 14) et constitue le premier résultat de l'utilisation de la méthode. Il permet une

¹⁸On n'a pas examiné les avantages et les inconvénients d'utiliser un fichier par caractéristique observée plutôt qu'un fichier "universel."

¹⁹Qui a bénéficié de remboursements!

²⁰1991-1994, pour les données ayant servi à l'étude.

²¹A supposer bien sûr que ces données contiennent l'information nécessaire, ce qu'il est impossible de dire sans "passer à l'acte."

analyse descriptive “immédiate” des données: les instruments privilégiés de cette analyse sont le langage d’interrogation SQL et le graphisme. La section qui suit illustre ce propos. L’objectif est de montrer ce qu’il est possible d’attendre de la méthode et non pas une étude du contenu de la BDL. Celle-ci est du ressort de l’Etape 2 (page 9).

4 Illustration de la méthode: la psychiatrie en cabinet privé

4.1 Notation

On va constamment faire, dans la suite, référence au tableau des données relatives aux périodes (section 3.2, page 14). Pour ne pas encombrer l’exposé de phrases répétitives inutiles, on va recourir aux abréviations suivantes dont la signification est celle des cases correspondantes du tableau.

ID	Période	Durée	Domaine médical	Pct des rems	Domaine médical	Pct des rems	Domaine médical	Pct des rems	Rembt total	Total % rems
i	$P_{i,j}$	$D_{i,j}$	$M_{i,j}^{(1)}$	$\Pi_{i,j}^{(1)}$	$M_{i,j}^{(2)}$	$\Pi_{i,j}^{(2)}$	$M_{i,j}^{(3)}$	$\Pi_{i,j}^{(3)}$	$R_{i,j}$	$\Pi_{i,j}^{(T)}$

Ainsi l’indice i identifie l’assuré numéro i , et l’indice j , la période numéro j de l’histoire des remboursements de l’assuré numéro i .

4.2 Quelques repères concernant la psychiatrie en cabinet privé

On utilise un jeu de données de 5’773 assurés choisis “arbitrairement.”²²

La répartition du nombre de périodes par trajectoire de remboursements (donc par assuré) est la suivante:²³

²²Ce sont en fait les 5’773 premiers assurés d’un fichier contenant tous les assurés ayant été au bénéfice de remboursements pour des soins reçus en psychiatrie.

²³Il y a donc, par exemple, 608 trajectoires dans lesquelles on ne décèle qu’une période. Il y a relativement peu de trajectoires avec deux périodes, et un quart des trajectoires comportent trois périodes.

Nb périodes	Fréq.	%
1	608	10.53
2	163	2.82
3	1431	24.79
4	544	9.42
5	924	16.01
6	610	10.57
7	559	9.68
8	400	6.93
9	273	4.73
10	163	2.82
11	52	0.90
12	25	0.43
13	12	0.21
14	7	0.12
15	1	0.02
16	1	0.02
Total	5773	≈ 100

Ces 5'773 trajectoires produisent en tout 28'040 périodes parmi lesquelles il y en a 2875 sans remboursements. La notion de période n'est utile que si l'on peut lui associer, de manière "étroite," un libellé médical. Une mesure possible de la force de ce lien est le pourcentage des remboursements attribuable à chacune des trois catégories médicales associées aux trois pourcentages de remboursements les plus élevés, durant la période, soit la valeur des variables $\Pi_{i,j}^{(1)}$, $\Pi_{i,j}^{(2)}$, et $\Pi_{i,j}^{(3)}$. Numériquement, on a les chiffres suivants:²⁴

	Moy	Méd	Min	Max	Q_1	Q_3
$\Pi_{\bullet,\bullet}^{(1)}$	55.992	55.611	0.000	100.000	39.930	76.089
$\Pi_{\bullet,\bullet}^{(2)}$	19.067	20.404	0.000	50.000	7.365	29.954
$\Pi_{\bullet,\bullet}^{(3)}$	8.082	7.133	0.000	32.309	0.000	14.202
$\Pi_{\bullet,\bullet}^{(T)}$	83.141	95.388	0.000	100.000	82.820	100.000

On constate donc que les trois premiers domaines médicaux, en termes d'importance des remboursements, "expliquent" plus de 80% du montant des

²⁴ Q_1 est le premier quartile et Q_3 le troisième quartile: ils "marquent" respectivement les 25 et 75 pourcents des observations. Le tableau se lit comme suit:

1. 25% des pourcentages de coûts se trouvent entre Min et Q_1 ,
2. 25% des pourcentages de coûts se trouvent entre Q_1 , et $Méd$,
3. 25% des pourcentages de coûts se trouvent entre $Méd$ et Q_3 ,
4. 25% des pourcentages de coûts se trouvent entre Q_3 , et Max .

$\Pi_{\bullet,\bullet}^{(k)}$ indique que l'on examine la répartition des valeurs $\Pi_{i,j}^{(k)}$, k étant fixe, mais i et j prenant toutes les valeurs possibles. Ainsi, si on ne s'intéresse qu'au domaine médical qui est responsable de la plus grande partie des remboursements, exprimés en pourcents, dans une période donnée, on constate que pour 75% des périodes le pourcentage de remboursements dépasse 40%, et, pour 50% des périodes, il dépasse 55%.

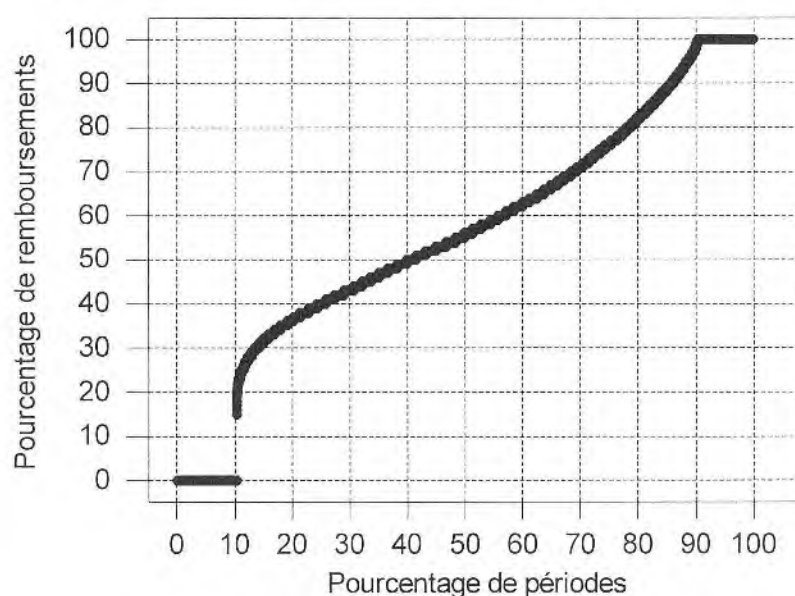
remboursements dans 75% des cas. Si l'on omet les périodes sans remboursements, ce qui est plus représentatif de l'étroitesse du lien entre remboursement et période, on obtient les chiffres suivants:

	Moy	Méd	Min	Max	Q_1	Q_3
$\Pi_{\bullet\bullet\bullet}^{(1)}$	62.389	59.024	14.537	100.000	44.893	78.089
$\Pi_{\bullet\bullet\bullet}^{(2)}$	21.245	22.199	0.000	50.000	12.962	29.966
$\Pi_{\bullet\bullet\bullet}^{(3)}$	9.006	8.620	0.000	32.309	0.000	14.944
$\Pi_{\bullet\bullet\bullet}^{(T)}$	92.639	97.401	39.127	100.000	87.811	100.000

On a donc ainsi "saisi" environ 90% des coûts pour 75% des assurés. Une question pendante est celle de savoir s'il faut plus ou moins de couverture.²⁵ On peut illustrer ce fait graphiquement.

Lien entre période et domaine médical:

domaine ayant contribué le plus aux remboursements

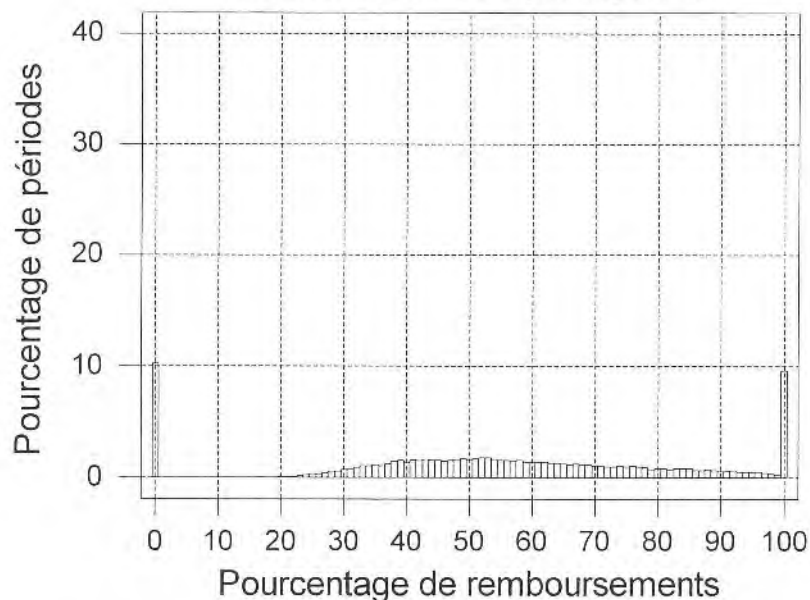


²⁵Ceci n'est que la première d'une longue série d'interrogations auxquelles il faut répondre avant que l'on se soucie d'écrire des programmes permettant de traiter efficacement toutes les données de la BDL.

On voit donc que pour 60% des périodes, le domaine médical associé aux remboursements les plus importants “épuise” au moins 50% des remboursements. On voit également qu’il y a environ 10% des périodes pour lesquelles le domaine médical associé aux remboursements les plus importants “épuise” l’intégralité des remboursements, et enfin qu’il y a environ 10% de périodes pour lesquelles il n’y a pas de remboursements. On retrouve ces caractéristiques dans l’histogramme qui suit:

Répartition des pourcentages de remboursements

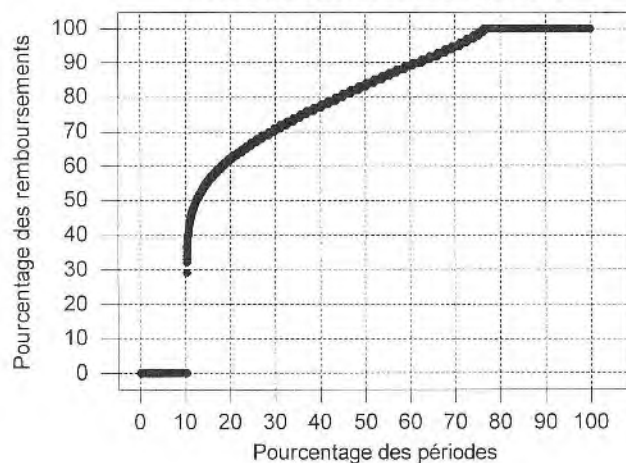
attribuables au domaine médical occasionnant
les remboursements les plus importants



On peut recommencer l’exercice en cumulant les remboursements issus des deux ou trois domaines médicaux ayant contribué le plus aux remboursements. On constate que parallèlement les effets cumulés “absorbent” une part de plus en plus importante des remboursements et que chaque “ajout” y contribue de moins en moins. Voici les graphiques:

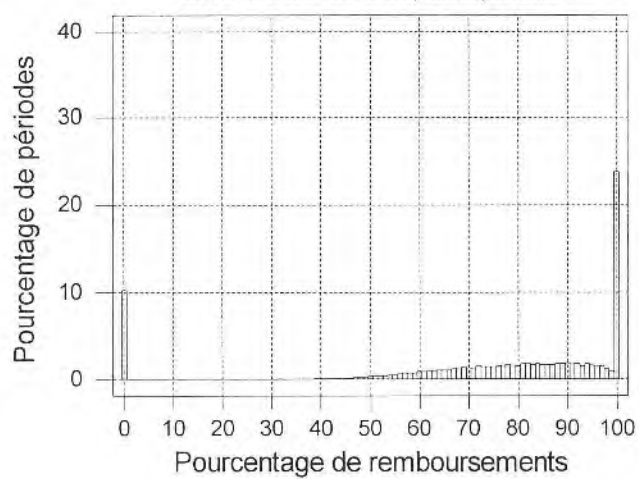
Lien entre période et domaine médical:

les deux domaines ayant occasionné
les remboursements les plus importants



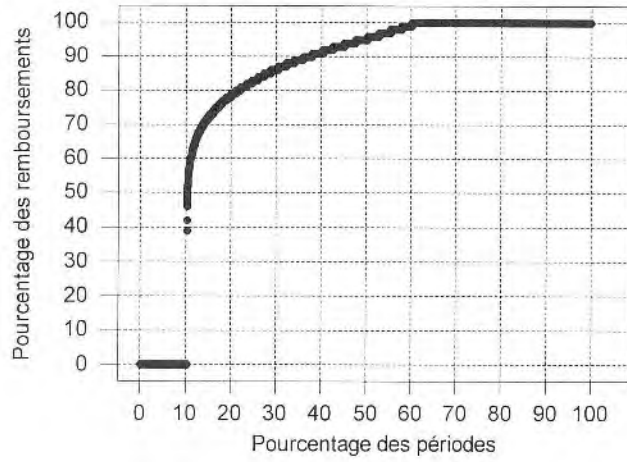
Répartition des pourcentages de remboursements

attribuables aux deux domaines médicaux occasionnant
les remboursements les plus importants



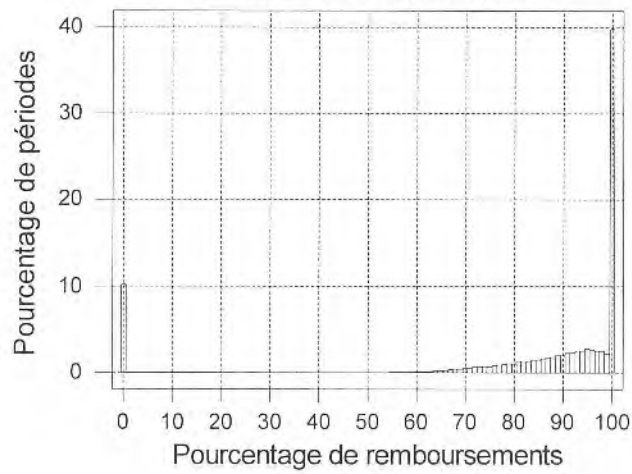
Lien entre période et domaine médical:

les trois domaines ayant occasionné
les remboursements les plus importants



Répartition des pourcentages de remboursements

attribuables aux trois domaines médicaux occasionnant
les remboursements les plus importants



4.3 La psychiatrie en cabinet privé sporadique et la psychiatrie en cabinet privé régulière

Ce qui suit veut illustrer une utilisation possible de l'instrument de lecture qui a été développé, celle de la définition de "catégories pertinentes." Comme les données sont celles de la psychiatrie, on peut vouloir distinguer, à des fins d'étude, une psychiatrie "occasionnelle" ou "accidentelle" d'une psychiatrie "régulière," pour, par exemple, en faire émerger similitudes et dissimilarités. La psychiatrie "régulière" pourrait être "définie" comme celle qui concerne des personnes souffrant de troubles mentaux, nécessitant un suivi régulier et prolongé, alors que la psychiatrie "occasionnelle" serait "définie" comme celle qui concerne des personnes ayant besoin d'un soutien "moral" ponctuel. En pratique ce seront des psychiatres qui devront énoncer les définitions opérationnelles. Ici on en produit une *ad hoc*, pour les besoins de l'illustration: sa pertinence reste à établir, si elle peut l'être.

On abrégera "psychiatrie en cabinet privé sporadique" par PCPS et "psychiatrie en cabinet privé régulière" par PCPR.²⁶ Une manière de définir un cas de PCPR est de poser que la période j pour l'assuré numéro i relève de la PCPR si au moins l'une des variables $M_{i,j}^{(1)}$, $M_{i,j}^{(2)}$, $M_{i,j}^{(3)}$ prend la valeur *PSYCH*. Tout assuré qui ne relève pas de la PCPR relève par suite de la PCPS (pour autant qu'il y ait un remboursement attribuable à la psychiatrie en cabinet privé). En pratique cela signifie que la psychiatrie en cabinet privé doit représenter une part "non négligeable" des coûts (remboursements) au cours d'au moins une période. On peut ensuite restreindre encore le champ de la définition en exigeant, par exemple, que les remboursements correspondants ou la durée de la période correspondante dépassent un certain seuil, ou encore qu'il y ait plusieurs périodes, dans la trajectoire de l'assuré, pour lesquelles la psychiatrie contribue de manière importante aux remboursements. On commencera par ne considérer que le cas le plus simple, le premier cité. On voit donc que les diverses définitions que l'on peut donner sont purement empiriques: cela est inévitable puisque l'on ne décèle la "maladie" qu'au travers du filtre des remboursements et des nombres de prestations. C'est donc à l'usage que l'on peut juger de la pertinence d'une telle définition, d'où la

²⁶Ces termes sont moins "évocatifs" que ceux qui ont été utilisés au paragraphe précédent, mais ils ont l'avantage d'être neutres et plus précis!

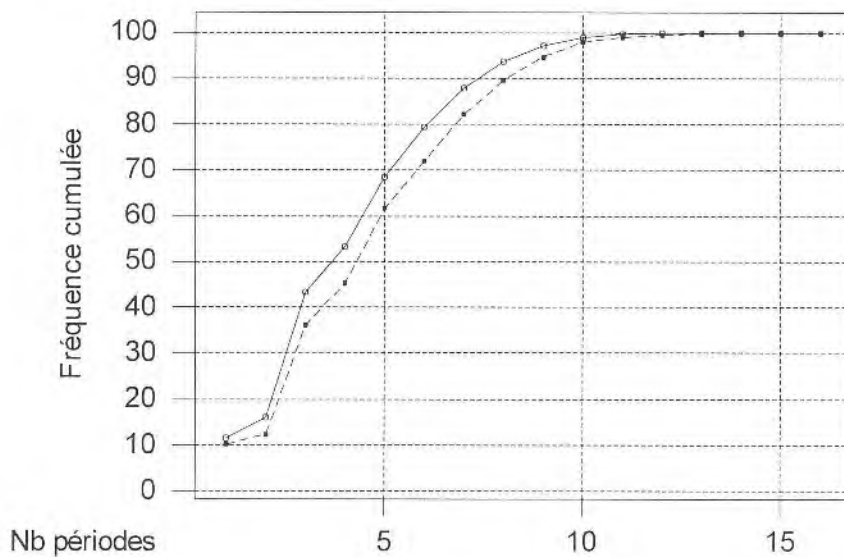
nécessité de l'Etape 2 (page 9) décrite dans les prémisses.

On va commencer par examiner le sens qu'a le "concept" de PCPR, défini comme il l'a été ci-dessus.

4.3.1 Effectifs de la PCPR et de la PCPS dans le corpus utilisé

L'effectif de la PCPR est de 4'157 assurés et celui de la PCPS, de 1'616 assurés. Il y a donc environ 29% des cas de psychiatrie en cabinet médical qui sont considérés, selon le critère retenu, comme des cas "sporadiques."²⁷ Voici les répartitions respectives des nombres de périodes par assuré pour les deux catégories:

Répartition du nombre de périodes selon les catégories PCPR et PCPS



Disque plein: PCPR, Disque vide: PCPS

²⁷Ces chiffres n'ont, *stricto sensu*, pas de signification concrète puisque le groupe d'assurés retenu ne représente pas une "population autonome" ou "circonscrite" (il s'agit des premiers assurés d'un fichier: on peut cependant espérer être en présence d'un échantillon "raisonnablement aléatoire").

Voici les données à l'origine du graphique précédent:

Nb Pér	PCPS			PCPR		
	Fréq	%	% Cumulé	Fréq	%	% Cumulé
1	186	11,5099	11,510	422	10,1516	10,152
2	72	4,4554	15,965	91	2,1891	12,341
3	443	27,4134	43,379	988	23,7671	36,108
4	161	9,9629	53,342	383	9,2134	45,321
5	245	15,1609	68,502	679	16,3339	61,655
6	176	10,8911	79,394	434	10,4402	72,095
7	137	8,4777	87,871	422	10,1516	82,247
8	94	5,8168	93,688	306	7,3611	89,608
9	58	3,5891	97,277	215	5,1720	94,780
10	28	1,7327	99,010	135	3,2475	98,027
11	11	0,6807	99,691	41	0,9863	99,014
12	3	0,1856	99,876	22	0,5292	99,543
13	2	0,1238	100,000	10	0,2406	99,783
14	0	0,0000	100,000	7	0,1684	99,952
15	0	0,0000	100,000	1	0,0241	99,976
16	0	0,0000	100,000	1	0,0241	100,000

On voit immédiatement que les deux catégories (PCPR et PCPS) se distinguent par le nombre de périodes que l'on trouve dans les trajectoires des remboursements des assurés qui en relèvent: il y a plus de périodes dans les trajectoires de la catégorie PCPR que dans celles de la catégorie PCPS (la fonction de répartition de la catégorie PCPS croît plus rapidement que celle de la PCPR). On peut également remarquer que les nombres de périodes 1, 2 et 4 sont relativement moins nombreux, pour les deux catégories. On a donc ici une première confirmation qu'il est judicieux de définir les catégories ("naturelles") PCPR et PCPS. Il faut cependant se garder d'extrapoler ces faits et interprétations à toute la population! Ils ne sont en effet valables, pour l'instant, comme on l'a déjà relevé, que pour les assurés dont on étudie les données. Mais on voit bien sur cet exemple l'intérêt qu'il y a à étudier ces données sous l'angle qui est ici choisi.

4.3.2 Domaines médicaux présents dans les remboursements relatifs à la PCPR et à la PCPS

On peut se demander si la "fréquentation médicale" des assurés PCPR se distingue de celle des assurés PCPS. Une première réponse peut être obtenue en examinant les fréquences (calculées à partir des périodes individuelles, et non pas par assuré) avec lesquelles les divers domaines médicaux sont mis à contribution:

No de	Domaine	PCPR			PCPS		
Domaine	Médical	%1	%2	%3	%1	%2	%3
1	AGN-M	0.3902	0.1204	0.0482	0.5495	0.1923	0.1099
2	AUN+M	0.1541	0.0096	0.0289	0.5632	0.0962	0.0824
3	AUN-M	0.0482	0.0193	0.0048	0.0962	0.1374	0.0412
4	CARDL	0.6407	0.9441	0.7803	0.9615	1.2225	1.2225
5	CHIRG	0.8863	1.5848	1.3825	1.1538	2.9121	2.0604
6	DERMA	0.5202	0.7322	1.0983	0.4670	1.2500	1.2912
7	ENDOC	0.1445	0.2649	0.2264	0.2473	0.4121	0.3571
8	GASTR	0.7563	1.0934	0.8333	0.9341	1.3874	1.4560
9	GENER	19.7543	19.1715	11.8690	22.7610	22.3214	13.6951
10	GYN-M	0.6021	1.4403	2.3410	0.5220	1.4286	2.1566
11	HEMAT	0.1975	0.1830	0.1927	0.2747	0.3846	0.3709
12	JAMBS	0.1156	0.1397	0.1590	0.1648	0.2335	0.1511
13	LABOR	1.1127	3.1455	6.0405	1.2088	3.5989	7.0879
14	NEURL	0.1782	0.3757	0.5058	0.1923	0.5357	0.8104
15	NO	3.8632	3.9451	4.6098	5.3022	5.0549	5.9066
16	NOT-ILL	10.3227	20.3420	35.1686	10.0549	18.1456	30.1236
17	OPHTL	0.9489	1.6378	3.0684	1.2637	2.8159	4.0934
18	OTORL	0.3565	0.6696	1.1898	0.4670	0.8929	1.4698
19	PHARM	23.0829	17.8902	10.3468	22.5549	18.7775	11.4011
20	PNEUM	0.3131	0.4094	0.2938	0.3846	0.5769	0.4670
21	PSYCH	16.9798	13.7958	9.8025	0.0000	0.0000	0.0000
22	RADLG	0.9200	1.2187	1.3873	1.0989	1.6346	2.4588
23	RHUMT	0.9875	1.1175	0.9586	1.3599	1.5522	1.6896
24	SGN-M	10.6021	4.1426	3.5645	17.0467	5.6456	4.5604
25	SOINS	3.3622	3.9162	2.7890	4.8489	5.6319	4.9176
26	SREAD	1.2717	0.5443	0.2071	2.2253	1.0989	0.5357
27	SUV-M	1.1272	0.6936	0.6214	2.9258	1.3049	0.8929
28	TROPL	0.0771	0.0771	0.0771	0.0549	0.0962	0.0687
29	UROLG	0.2842	0.3757	0.4046	0.3159	0.6593	0.5220

Ce tableau²⁸ s'interprète comme suit. Si l'on prend en compte l'ensemble des périodes constituant les trajectoires des remboursements dont bénéficient les assurés PCPR, il y a 0.3902% de ces périodes pour lesquelles la catégorie médicale ayant contribué le plus aux remboursements de la période est AGN-M, soit les soins ambulatoires en établissement hospitalier de soins généraux, sans maternité. On retrouve tout d'abord, dans ce tableau, la définition donnée des catégories PCPR et PCPS: les pourcentages de la psychiatrie, pour la catégorie PCPS, sont nuls. Ces mêmes pourcentages valent respectivement, pour la catégorie PCPR, de 17, 14, et 10.²⁹ Ces pourcentages sont plus faibles que ceux qui correspondent aux libellés "généraliste" et "pharmacie." On peut remarquer également que pour les autres catégories médicales, les pourcentages sont en général plus élevés pour les assurés PCPS que pour

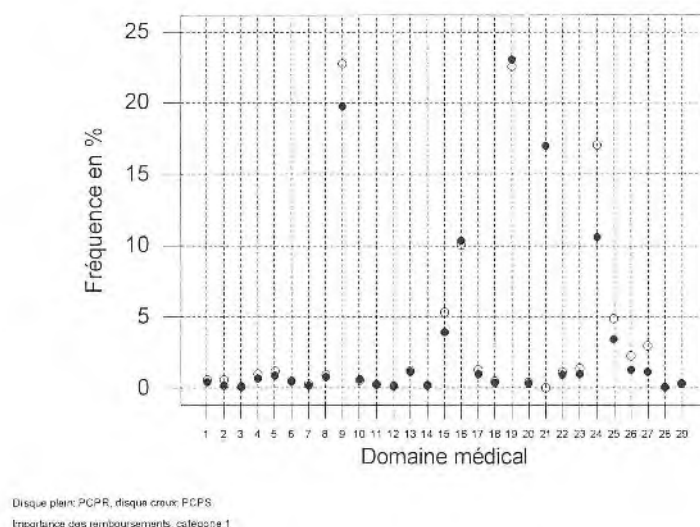
²⁸L'ordre des domaines médicaux est l'ordre lexicographique du tableau (1 correspond à AGN-M et 29 correspond à UROLG).

²⁹On a arrondi ces pourcentages qui valent en fait 16.9798, 13.7958 et 9.8025.

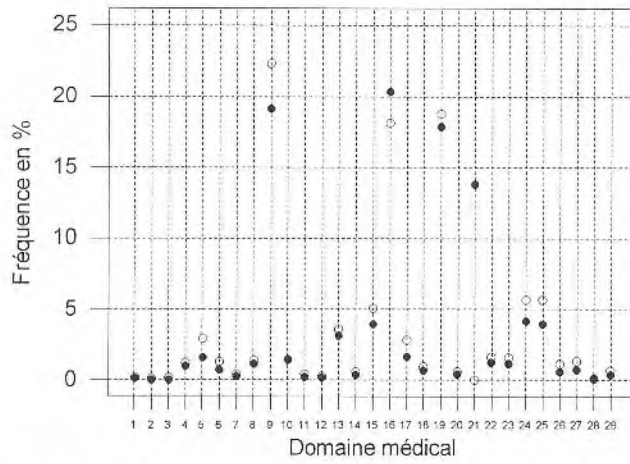
les assurés PCPR. Une première raison de ce fait est que les pourcentages PCPS sont ventilés sur un nombre plus petit (de 1) de catégories médicales. Une seconde pourrait être, ce qui confirmerait les remarques de la section 4.3.1, que la catégorie PCPS est une catégorie d'assurés chez lesquels un mal-être "diffus" pousse l'assuré à une "fréquentation médicale" peu différenciée. On peut remarquer finalement qu'il y a peu de catégories médicales pour lesquelles les pourcentages sont importants.

Si l'on représente graphiquement (voir ci-dessous) les données de ce tableau, ressemblances et dissemblances sont encore plus visibles et la structure des pourcentages est immédiatement perceptible. Là où les distinctions, sur le graphique, sont possibles, la différence entre les fréquences relatives aux divers domaines est toujours "en faveur" de la PCPS, sauf dans les domaines 16 et 21. Or 21 représente la psychiatrie et 16 représente les périodes pendant lesquelles il n'y a pas de remboursements. Qualitativement, les répartitions sont très semblables, et les trois quarts des domaines médicaux (29 parmi 41) retenus sont concernés (pour identifier le domaine médical utiliser le tableau de la page 27).

Fréquence de mise à contribution des divers secteurs médicaux

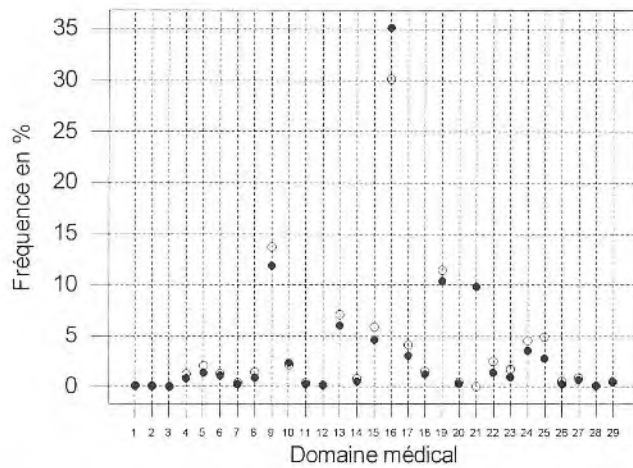


Fréquence de mise à contribution des divers secteurs médicaux



Disque plein: PCPR, disque creux: PCPS
 Importance des remboursements: catégorie 2

Fréquence de mise à contribution des divers secteurs médicaux



Disque plein: PCPR, disque creux: PCPS
 Importance des remboursements: catégorie 3

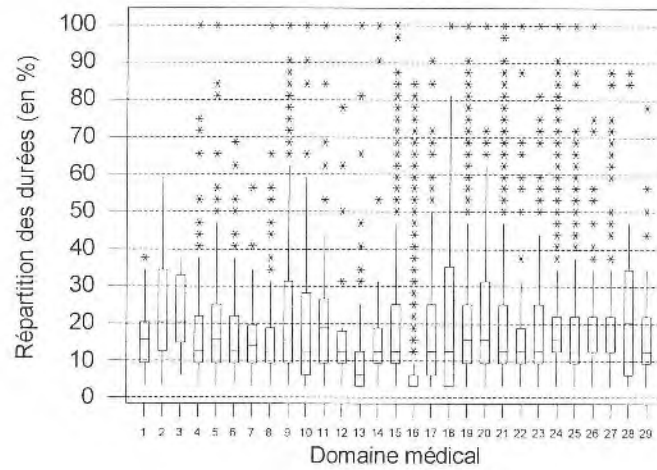
4.3.3 Comparaison des durées des périodes selon la catégorie (régulière ou sporadique) et le domaine médical

Le tableau qui suit contient, pour chaque catégorie médicale, et les deux catégories d'assurés (PCPR et PCPS) les descripteurs numériques habituels de la répartition des durées des périodes relevant de ces catégories. Les durées sont en catégories de pourcents correspondant aux 32 divisions du temps retenues pour l'approximation (ce qui conduit à des anomalies numériques mineures: 3.119 et 3.120 par exemple).

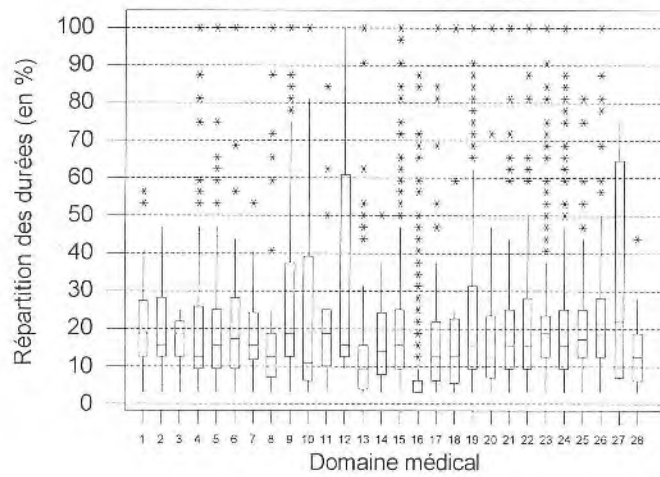
No de	Dom	PCPR						PCPS					
Dom	Méd	Eff	Méd	Min	Max	Q1	Q3	Eff	Méd	Min	Max	Q1	Q3
1	AGN-M	81	15.619	3.119	37.494	9.369	20.306	40	18.740	3.120	56.240	12.490	27.340
2	ALN-IM	32	20.310	3.120	59.370	12.490	34.370	41	15.620	3.120	46.870	12.490	28.120
3	AUN-M	10	20.310	6.240	37.490	14.840	32.810	7	18.740	3.120	24.990	12.490	21.870
4	CARDL	133	12.490	3.120	99.990	9.370	21.870	70	12.490	3.120	99.990	9.370	25.780
5	CHIRG	184	15.620	3.120	99.990	9.370	24.990	84	15.620	3.120	99.990	9.370	24.990
6	DERMA	108	12.490	3.120	68.740	9.370	21.870	34	17.180	3.120	99.990	9.370	28.120
7	ENDOC	30	14.060	3.120	56.240	9.370	19.530	18	15.620	3.120	53.120	11.710	24.210
8	GASPR	157	12.490	3.120	99.990	9.370	18.740	68	12.490	3.120	99.990	7.030	18.740
9	GENER	4101	15.619	3.119	99.994	9.369	31.244	1657	18.744	3.119	99.994	12.494	37.494
10	CYN-M	125	12.490	3.120	99.990	6.240	28.120	38	10.930	3.120	99.990	6.240	39.060
11	HEMAT	41	18.740	3.120	99.990	9.370	26.560	20	18.740	3.120	84.370	10.150	24.990
12	JAMBS	24	12.490	3.120	78.120	9.370	17.960	12	15.620	9.370	99.990	12.490	60.930
13	LABOR	231	6.244	3.119	99.994	3.119	12.494	88	9.370	3.120	99.990	3.900	15.620
14	NEURL	37	12.490	3.120	99.990	9.370	18.740	14	14.060	3.120	40.990	7.810	24.210
15	NO	802	12.494	3.119	99.994	9.369	24.994	386	15.619	3.119	99.994	9.369	24.994
16	NOT-ILL	2143	3.119	3.119	84.369	3.119	6.244	732	3.119	3.119	87.494	3.119	6.244
17	OPHTL	197	12.490	3.120	90.620	6.240	24.990	92	12.490	3.120	99.990	6.240	21.870
18	OTORL	74	12.490	3.120	99.990	3.120	35.150	34	12.490	3.120	99.990	5.460	22.650
19	PHARM	4792	15.619	3.119	99.994	9.369	24.994	1642	15.619	3.119	99.994	9.369	31.244
20	PNEUM	65	15.620	3.120	99.990	9.370	31.240	28	12.490	3.120	99.990	7.030	23.430
21	PSYCH	3525	12.494	3.119	99.994	9.369	24.994	0	0.000	0.000	0.000	0.000	0.000
22	RADLG	191	12.490	3.120	99.990	9.370	18.740	80	15.620	3.120	99.990	9.370	24.990
23	RHUMT	205	12.490	3.120	99.990	9.370	24.990	99	15.620	3.120	99.990	9.370	28.120
24	SGN-M	2201	15.619	3.119	99.994	12.494	21.869	1241	18.744	3.119	99.994	12.494	23.431
25	SOINS	698	12.494	3.119	99.994	9.369	21.869	353	15.620	3.120	99.990	9.370	24.990
26	SREAD	264	15.619	3.119	99.994	12.494	21.869	162	17.180	3.120	81.240	12.490	24.990
27	SUV-M	234	15.619	3.119	87.494	12.494	21.869	213	18.740	3.120	99.990	12.490	28.120
28	TROPL	16	20.310	3.120	87.490	6.240	34.370	4	21.900	6.200	75.000	7.000	64.800
29	UROLG	59	12.490	3.120	78.120	9.370	21.870	23	12.490	3.120	43.740	6.240	18.740

On remarque que les effectifs sont très inégaux, les catégories dominantes étant celles des personnes qui n'ont pas bénéficié de remboursements, celles de la médecine de généraliste et de son pendant, la pharmacie, celle de la psychiatrie, et enfin celle des séjours en établissement hospitalier de soins généraux, sans maternité. Si l'on tient compte des effectifs globaux, les répartitions des effectifs dans les catégories PCPR et PCPS sont semblables. Les répartitions sont fortement asymétriques, les moyennes étant toujours bien plus élevées que les médianes. Comme le montrent les graphiques qui suivent, cette asymétrie est due en particulier à la présence de nombreuses valeurs excentriques. Les descripteurs numériques appropriés sont donc la médiane et les quartiles.

Répartition des durées des périodes PCPR,
 par domaine médical à l'origine
 des remboursements les plus importants

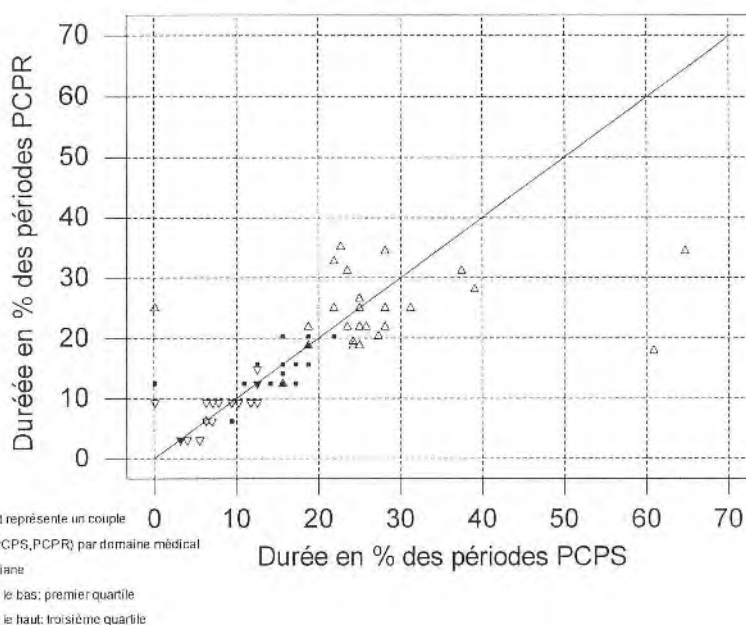


Répartition des durées des périodes PCPS,
 par domaine médical à l'origine
 des remboursements les plus importants



Pour les classes de faibles effectifs, ni les descripteurs numériques ni les graphiques n'ont vraiment de sens: ils sont inclus pour faciliter la lecture. Une comparaison globale de ces répartitions s'obtient en situant, par rapport à l'identité, les mêmes indicateurs des catégories PCPR et PCPS, par spécialité médicale. Il vient le graphique suivant:

Comparaison des durées des périodes PCPR et PCPS relatives aux secteurs médicaux principaux



On constate que les deux tiers des points sont en dessous de la diagonale, ce qui signifie que les valeurs pour la catégorie PCPR sont plus petites que les valeurs correspondantes de la catégorie PCPS, et c'est surtout vrai pour les médianes et les troisièmes quartiles. Les répartitions relatives à la PCPR sont donc plus concentrées et de valeurs moins élevées que celles de la classe PCPS. On retrouve ainsi l'aspect "diffus" de cette dernière. Si l'on ne retient

que les différences de plus de 3% de durée, ce qui correspond à environ 45 jours, on a les différences “significatives” suivantes:

Domaine	PCPR < PCPS	PCPR > PCPS
AGN-M	⊗	
AUN+M		⊗
DERMA	⊗	
GENER	⊗	
JAMBS	⊗	
LABOR	⊗	
NO	⊗	
PNEUM		⊗
RADLG	⊗	
RHUMT	⊗	
SCN-M	⊗	
SOINS		
SUV-M	⊗	

On observe de nouveau la tendance des deux groupes à se distinguer, bien que les différences soient plus difficiles à “lire.”

4.4 Un examen plus détaillé des données de la classe PCPR

Du point de vue de la “consommation médicale” c’est la classe PCPR qui est en principe la classe “intéressante.”³⁰ On va donc procéder à l’examen un peu plus détaillé de cette classe.

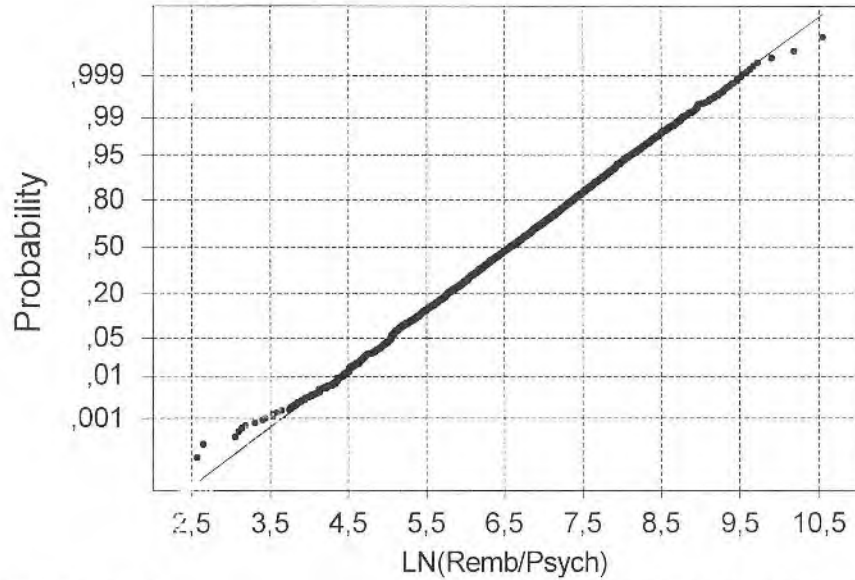
4.4.1 La répartition des remboursements pour les soins en psychiatrie

Du point de vue de l’étude des coûts, ce sont les répartitions des remboursements par domaine médical qui produisent l’une des informations les plus utiles. Or la répartition des logarithmes naturels des montants remboursés pour soins psychiatriques en cabinet médical, “sur” toutes les périodes durant lesquelles de tels soins sont prodigués, semble être normale, comme le montre le graphique suivant:³¹

³⁰Bien que la “consommation médicale superflue,” et donc la classe PCPS, soit aussi une “dimension” qui alimente fréquemment les discussions relatives à la LAMal.

³¹LN signifie “logarithme naturel”.

Loi des remboursements (unité statistique: la période)



Average: 6,56787
 StDev: 0,943392
 N: 8424

Anderson-Darling Normality Test
 A-Squared: 0,421
 P-Value: 0,324

4.4.2 Comparaison des répartitions des remboursements, dans la classe PCPR, pour les divers domaines médicaux

La répartition des remboursements par période, domaine médical et importance relative, dans une période, des remboursements, peut être tabulée. Voici tout d'abord les chiffres pour les domaines médicaux à l'origine des remboursements, au cours d'une période, relativement les plus importants:³²

³²On retrouve, dans les répartitions, les asymétries caractéristiques de ce type de données, ainsi que les très nombreuses valeurs excentriques: pour l'exploitation de ces données, il faut donc toujours soit les transformer, soit travailler avec les médianes. Les remboursements médians les plus élevés se trouvent, par ordre décroissant, dans les domaines AUN+M, AUN-M, SGN-M, SREAD, SUV-M, AGN-M. Par contre, pour les montants remboursés les plus élevés, on trouve, toujours par ordre décroissant, SGN-M, NO, SUV-M, SREAD, AUN+M, GENER, PSYCH, et les domaines GENER et PSYCH sont

Domaine Médical	II ₁						
	Effectif	Moyenne	Médiane	Minimum	Maximum	Quartile 1	Quartile 3
AGN-M	81	4377.0	3380.0	94	14925	2294.0	6151.0
AUN+M	32	10035.0	5488.0	345	41695	3545.0	17768.0
AUN-M	10	4879.0	4733.0	280	9758	1055.0	7842.0
CARDL	133	1005.5	820.0	48	6574	601.0	1124.0
CHIRG	184	1382.0	827.0	49	12079	439.0	1355.0
DERMA	108	742.0	617.0	33	3155	299.2	1036.2
ENDOC	30	796.9	674.0	158	2682	510.8	946.0
GASTR	157	869.7	781.0	56	4137	584.5	961.5
GENER	4101	1232.7	907.0	18	38970	563.0	1517.0
GYN-M	125	489.2	282.0	43	6712	136.0	607.0
HEMAT	41	1229.0	929.0	80	3935	635.0	1630.0
JAMBS	24	842.0	700.0	183	2067	433.0	1202.0
LABOR	231	291.9	162.0	21	2664	70.0	377.0
NEURL	37	862.1	709.0	31	2607	517.5	1250.5
NO	802	4580.0	1255.0	6	130362	556.0	3169.0
NOT-ILL	2143	0.0	0.0	0	0	0.0	0.0
OPHTL	197	1017.0	333.0	24	9096	135.0	1083.0
OTORL	74	435.4	275.0	54	2078	145.8	625.0
PHARM	4792	1364.6	1017.5	3	16882	621.0	1667.8
PNEUM	65	946.6	842.0	22	3104	579.5	1174.5
PSYCH	3525	1653.8	1180.0	38	38072	711.5	1991.0
RADLG	191	823.0	771.0	136	3309	503.0	1133.0
RHUMT	205	1158.7	940.0	57	7017	616.0	1320.5
SGN-M	2201	7044.0	4410.0	21	171600	1960.0	8478.0
SOINS	698	1070.0	788.0	36	14656	619.0	1177.0
SREAD	264	6267.0	4408.0	53	59101	2469.0	7765.0
SUV-M	234	6884.0	3525.0	101	72530	1030.0	8018.0
TROPL	16	890.0	805.0	59	2289	526.0	1129.0
UROLG	59	857.0	642.0	105	6232	426.0	1005.0

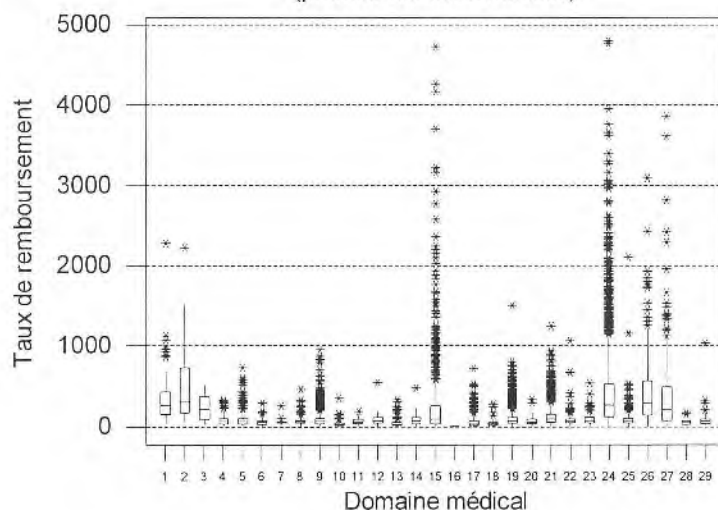
Voici une représentation graphique de ces données (les montants y sont divisés par la durée, et on ne retient que les valeurs inférieures à 5'000, pour ne pas écraser les graphiques):

essentiellement comparables. Les établissements de soins posent donc un problème pratique d'analyse difficile puisqu'ils sont associés aux montants les plus élevés, et que l'on ne sait leur associer de domaine médical spécifique. C'est dans ce cadre que l'on utilisera la classification par cartes de Kohonen des trajectoires de remboursements: cette technique permet de regrouper les trajectoires simultanément par forme et montant des remboursements. On a que *the Kohonen self-organizing map (SOM) network performs a mapping from a continuous input space to a discrete output space, preserving the topological properties of the input. This means that points close to each other in the input space are mapped to the same or neighbouring PE's [processing elements] in the output space. [...] This property is very interesting for some applications that require the preservation of an input-space metric. [...] The SOM preserves the data densities of the input space reasonably well; that is, regions with more data points are mapped to larger regions in the output space.* (Voir: J.C. Principe, N.R. Euliano, W.C. Lefebvre, Neural and Adaptive Systems, Wiley, New York (2000), pages 348, 350 et 353.)

Répartition des remboursements

sous forme de taux par période

(premier domaine médical)



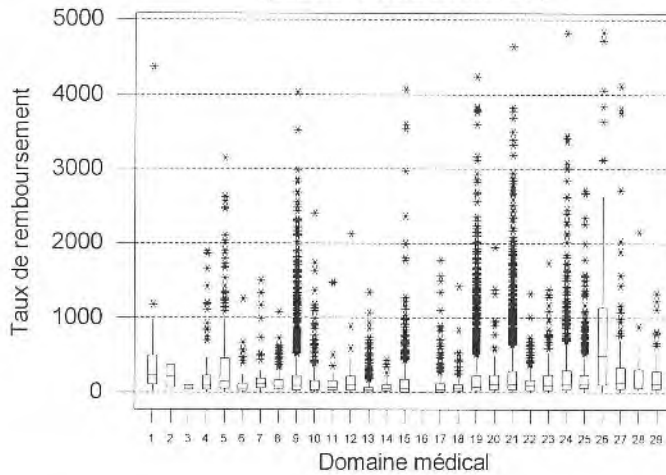
Taux retenus: moins de 5000

Pour le deuxième domaine médical on a de même:

Domaine Médical	Π_2						
	Effectif	Moyenne	Médiane	Minimum	Maximum	Quartile 1	Quartile 3
AGN-M	25	3563.0	1718.0	278	32000	632.0	3700.0
AUN+M	2	2224.0	2224.0	588	3860	*	*
AUN-M	4	334.5	389.0	140	420	198.5	416.0
CARDL	196	863.7	538.5	52	11356	351.2	880.3
CHIRG	329	1815.0	729.0	27	19156	320.0	2540.0
DERMA	152	475.7	349.5	41	7865	197.5	523.5
ENDOC	55	891.3	666.0	154	3432	454.0	1041.0
GASTR	227	707.7	586.0	43	2985	436.0	835.0
GENER	3980	756.7	539.5	14	8479	305.0	950.0
GYN-M	299	649.9	252.0	33	7234	136.0	510.0
HEMAT	38	646.0	405.0	57	3382	302.0	759.0
JAMBS	29	639.0	497.0	41	2962	357.0	618.0
LABOR	653	289.5	194.0	13	1840	90.0	401.0
NEURL	78	457.8	408.0	75	2113	277.0	557.7
NO	819	722.0	322.0	8	19175	105.0	794.0
NOT-ILL	4223	0.0	0.0	0	0	0.0	0.0
OPHTL	340	447.6	228.0	51	6150	120.3	513.2
OTORL	139	393.1	263.0	42	3776	153.0	464.0
PHARM	3714	803.8	537.0	3	10131	277.8	993.2
PNEUM	85	738.8	632.0	30	4423	363.0	944.0
PSYCH	2864	835.3	557.5	13	11042	327.3	990.0
RADLG	253	665.2	541.0	90	3527	288.0	872.0
RHUMT	232	786.1	606.0	71	4102	318.2	992.0
SGN-M	860	1289.3	600.0	15	40746	303.0	1228.5
SOINS	813	880.0	653.0	15	13969	479.5	962.0
SPREAD	113	4024.0	2068.0	25	38678	633.0	5337.0
SUV-M	144	1324.0	699.0	22	17418	306.0	1214.0
TROPL	16	801.0	629.0	116	2871	305.0	918.0
UROLG	78	1206.0	717.0	83	5309	324.0	1482.0

Répartition des remboursements

sous forme de taux par période
(deuxième domaine médical)



Taux retenus: moins de 5000

Remarques:

1. Comme le tableau reflète les remboursements dont l'importance est moindre que celle, correspondante, des remboursements à l'origine du tableau précédent, on s'attend à des valeurs généralement plus petites. En fait, on a la situation suivante, pour les effectifs.³³

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
81	32	10	133	184	108	30	157	4101	125	41	24	231	37	802
25	2	4	196	329	152	55	227	3980	299	38	29	653	78	819
10	6	1	162	287	228	47	173	2464	486	40	33	1254	105	957
16	17	18	19	20	21	22	23	24	25	26	27	28	29	
2143	197	74	4792	65	3525	191	205	2201	698	264	234	16	59	
4223	340	139	3714	85	2864	253	232	860	813	113	144	16	78	
7301	637	247	2148	61	2035	288	199	740	579	43	129	16	84	

On voit donc que se maintiennent, tout en perdant en importance, les domaines GENE, PHARM, PSYCH.

2. Les domaines relativement "peu sollicités" du tableau précédent se maintiennent ou augmentent en importance relative.

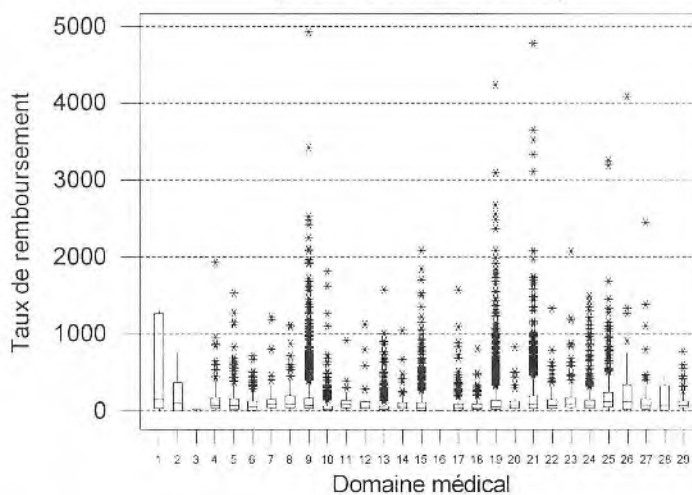
³³La première ligne représente les effectifs du tableau Π_1 , la deuxième ceux du tableau Π_2 , et la troisième ceux du tableau Π_3 . Les numéros des colonnes font référence aux domaines médicaux: 1 := AGN-M, 2 := AUN+M, ...

Finalement, pour le troisième domaine médical, il vient:

Domaine Médical	H ₃						
	Effectif	Moyenne	Médiane	Minimum	Maximum	Quartile 1	Quartile 3
AGN-M	10	1792.0	884.0	40	6710	246.0	3264.0
AUN+M	6	991.0	525.0	107	3474	184.0	1697.0
AUN-M	1	189.0	189.0	189	189	*	*
CARDL	162	549.3	405.5	49	4387	238.5	698.0
CHIRC	287	546.1	348.0	21	6123	179.0	614.0
DERMA	228	290.1	229.0	30	2137	127.3	373.0
ENDOC	47	572.8	405.0	53	2312	274.0	696.0
GASTR	173	566.4	488.0	40	2339	295.5	708.0
GENER	2464	563.4	390.1	3	11284	200.0	697.0
GYN-M	486	238.4	154.0	16	4671	107.0	262.0
HEMAT	40	531.0	440.5	62	2678	217.2	623.0
JAMBS	33	395.3	276.0	88	1881	202.0	411.5
LABOR	1254	213.7	142.0	18	2433	70.0	263.0
NEURL	105	420.7	357.0	120	1553	261.0	524.5
NO	957	403.6	186.0	6	3689	97.0	479.0
NOT-ILL	7301	0.0	0.0	0	0	0.0	0.0
OPHTL	637	258.3	165.0	38	2643	110.5	309.0
OTORL	247	252.3	192.0	35	1302	118.0	303.0
PHARM	2148	538.9	323.0	4	9958	158.0	639.7
PNEUM	61	514.6	391.0	18	3795	246.5	633.0
PSYCH	2035	565.5	393.0	23	9010	231.0	658.0
RADLG	288	477.2	364.5	72	2850	231.5	629.7
RHUMT	199	615.1	455.0	31	4740	261.0	758.0
SGN-M	740	468.7	307.0	16	8003	184.3	552.5
SOINS	579	722.5	582.0	18	8353	376.0	819.0
SREAD	43	1198.0	572.0	39	9339	134.0	1540.0
SUV-M	129	486.5	289.0	22	6536	140.0	547.5
TROPL	16	495.0	286.0	38	1545	114.0	844.0
UROLG	84	481.6	373.0	135	2097	247.8	578.8

Répartition des remboursements

sous forme de taux par période
(troisième domaine médical)



Taux retenus: moins de 5000

Remarque: Les évolutions relatives aux deux tableaux précédents se maintiennent et s'accroissent.

4.4.3 Assurés de la PCPR pour lesquels il y a exactement une période de remboursements dans le domaine de la psychiatrie

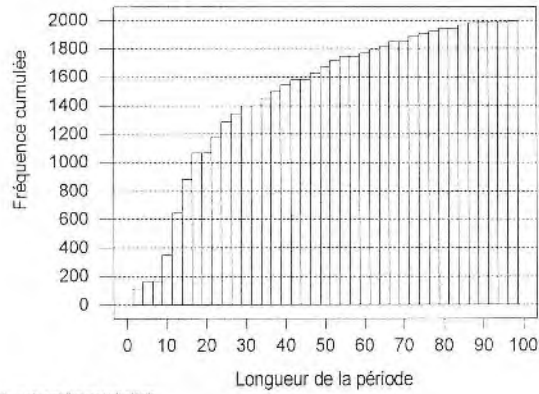
2'416 assurés relèvent d'une telle catégorie et pour 422 d'entre eux la période couvre les quatre ans d'observation.³⁴ Les fréquences des ces durées sont les suivantes:

Durée	Fréquence
3.1189	113
6.2439	53
9.3689	185
12.4939	294
15.6189	238
18.7439	184
21.8689	116
24.9939	102
28.1189	59
31.2439	57
34.3689	49
37.4939	49
40.6189	48
43.7439	37
46.8689	45
49.9939	44
53.1189	44
56.2439	28
59.3689	26
62.4939	25
65.6189	24
68.7439	30
71.8689	35
74.9939	21
78.1189	19
81.2439	19
84.3689	25
87.4939	14
90.6189	10
96.8689	1
99.9939	422

Il y a exactement 32 durées qui sont issues de l'approximation des trajectoires utilisée, et chaque durée représente une augmentation d'environ 45 jours par rapport à la durée qui la précède. Il y a donc, par exemple, 113 assurés pour lesquels la "période de psychiatrie" a duré environ 45 jours. Voici la répartition de la durée de ces périodes:

³⁴On pourrait interpréter ces assurés là comme des assurés suivant une psychothérapie: il faudrait encore vérifier que la trajectoire des remboursements qui leur est associée soit linéaire un signe de remboursements "réguliers et égaux."

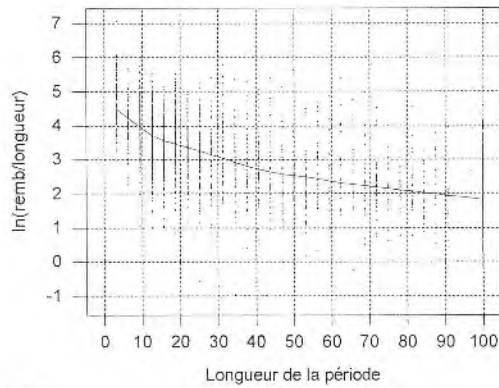
Longueur (en %, <100%) des périodes de remboursements avec remboursements pour soins en psychiatrie



Assurez-vous toujours d'y avoir une seule période avec des remboursements pour des soins en psychiatrie

La figure qui suit montre le lien qui semble exister entre le taux de remboursement mesuré par rapport à la longueur de la période correspondante et cette longueur: le taux est d'autant plus élevé que la période est courte.³⁵ La variation est très importante.

Lien entre taux de remboursement et longueur de la période



Message: loess

³⁵ \ln désigne le logarithme naturel.

4.4.4 Assurés de la PCPR pour lesquels il y a exactement deux périodes de remboursements dans le domaine de la psychiatrie

Les faits intéressants sont ici les liens qui peuvent exister entre périodes. On peut les représenter dans un tableau comme suit.³⁶

	[0,10]	[10,20]	[20,30]	[30,40]	[40,50]	[50,60]	[60,70]	[70,80]	[80,90]	[90,100]	Tot Périj
[0,10]	24	46	20	5	9	0	5	1	0	1	87
[10,20]	71	55	28	14	16	5	14	13	12	0	102
[20,30]	33	44	19	4	11	10	7	13	0	0	45
[30,40]	22	21	10	3	8	3	2	0	0	0	13
[40,50]	13	17	4	11	1	1	0	0	0	0	1
[50,60]	5	15	5	1	2	0	0	0	0	0	0
[60,70]	4	5	7	1	0	0	0	0	0	0	0
[70,80]	2	8	0	0	0	0	0	0	0	0	0
[80,90]	11	13	0	0	0	0	0	0	0	0	0
[90,100]	6	0	0	0	0	0	0	0	0	0	0
Tot Am	167	123	26	13	2	0	0	0	0	0	

Le nombre de cas pour lesquels la seconde période est ainsi plus courte que la première est de 331, ceux pour lesquels il n'y a pas de changement de 102, et ceux pour lesquels la seconde est plus longue que la première de 248.

Une interprétation de ces chiffres, qu'il faudrait bien sûr confirmer, car elle est quelque peu "rapide," est que

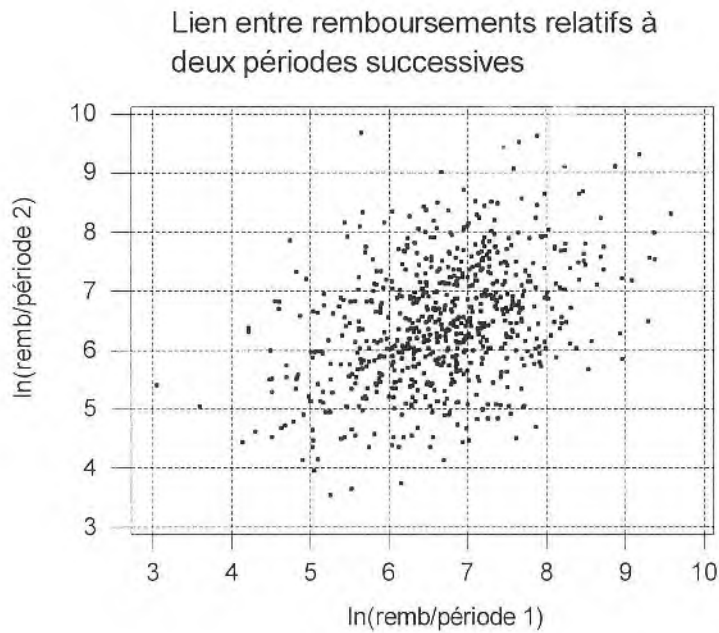
- le pourcentage de cas pour lesquels il y a amélioration est de $100 \times \frac{331}{681} \approx 46.5\%$,
- le pourcentage de cas pour lesquels il y a péjoration est de $100 \times \frac{248}{681} \approx 36.5\%$,
- le pourcentage de cas pour lesquels il n'y a pas de changement est de $100 \times \frac{102}{681} \approx 15\%$.

Si on fait la même opération sur les (logarithmes naturels des) remboursements, on constate que les remboursements ont tendance à augmenter:

³⁶En regroupant les durées pas déciles. Ce regroupement est fait pour faciliter la lecture et prendre en compte la variation. Néanmoins la longueur des périodes n'est pas strictement liée à la psychiatrie et le maintien des 32 catégories pourrait changer l'image. Cette tentative est faite pour explorer les limites de la méthode.

]2,3]]3,4]]4,5]]5,6]]6,7]]7,8]]8,9]	Tot Aug
]2,3]	0	0	3	1	0	0	0	4
]3,4]	0	8	13	17	9	0	0	39
]4,5]	2	11	48	80	32	3	0	115
]5,6]	0	6	39	108	81	15	1	97
]6,7]	0	3	13	55	63	21	4	25
]7,8]	0	0	4	14	13	4	1	1
]8,9]	0	0	1	1	4	2	1	0
Tot Dim	2	20	57	70	17	2	0	681

La figure suivante exhibe aussi le lien qui existe entre remboursements pour les deux périodes:

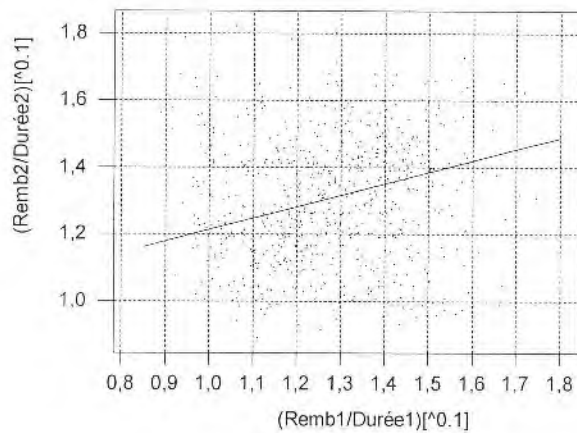


Dans ce qui précède, on a utilisé les remboursements “bruts.” Or on a déjà vu l’intérêt qu’il peut y avoir à pondérer ces remboursements par la durée de la période correspondante. Si l’on transforme de plus les données en utilisant la transformation définie par la méthode de Box et Cox³⁷ et que l’on

³⁷Qui suggère $x \mapsto \frac{x^\alpha - 1}{\alpha}$, avec $\alpha = 0.1$.

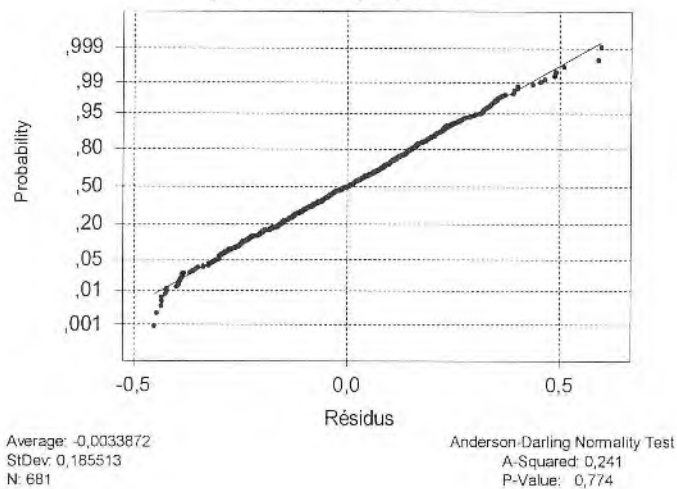
ajuste une droite résistante au résultat, on obtient la relation d'ajustement suivante ([^0.1] donne la valeur de α de la transformation de Box et Cox):

Lien entre remboursements relatifs à deux périodes successives



Les résidus en sont remarquablement normaux, comme l'indique le diagramme qui suit.

Répartition des résidus pour l'ajustement qui précède



4.4.5 Assurés de la PCPR pour lesquels, au cours d'au moins une période, la psychiatrie est à l'origine des remboursements les plus importants

Sur les 4'157 assurés de la PCPR, il y en a 1'869, soit environ 45%, qui ont, dans leur trajectoire de remboursements, une période durant laquelle la psychiatrie est la source des remboursements les plus importants. La répartition, "sur" ces 1'869 assurés, du nombre de périodes qui ont, "en première position," des remboursements pour de la psychiatrie, est la suivante:

Nb Pér	Fréq	%
1	1152	61,6372
2	325	17,3890
3	164	8,7747
4	90	4,8154
5	58	3,1033
6	34	1,8192
7	19	1,0166
8	11	0,5886
9	10	0,5350
10	3	0,1605
11	1	0,0535
12	1	0,0535
13	1	0,0535

On remarque que ce sont avant tout des trajectoires à période unique, soit celles que l'on pourrait attribuer aux "psychothérapies."

4.5 Les profils psychiatriques

On peut se poser la question de savoir si certains "profils" de domaines médicaux, que l'on définit comme des suites, de longueur variable, de domaines médicaux que l'on trouve dans les trajectoires de domaines médicaux associées à celles des remboursements, se retrouvent plus fréquemment, dans ces trajectoires, que d'autres. Les données disponibles (quelques milliers d'assurés) ainsi que la relative brièveté de la période considérée, couplée au choix de périodes relativement longues (l'unité de temps est de 45 jours) font qu'il y a peu d'espoir de voir des résultats spectaculaires. Voici tout de même les séquences les plus fréquentes:

Fréq	Support	Confiance	Séquence	
429	10,319942270	26,87969925	161616	⇒ 161616
165	3,969208564	36,42384106	91616	⇒ 161616
134	3,223478470	39,18128655	191616	⇒ 161616
132	3,175366851	40,36697248	211616	⇒ 161616
126	3,031031994	42,14046823	190916	⇒ 161616
119	2,862641328	30,43478261	190921	⇒ 161616
103	2,477748376	29,59770115	192109	⇒ 161616
100	2,405580948	6,26566416	161616	⇒ 91616
92	2,213134472	29,48717949	211909	⇒ 161616
88	2,116911234	19,42604857	91616	⇒ 91616
86	2,068799615	5,38847118	161616	⇒ 190916
86	2,068799615	43,87755102	192116	⇒ 161616
83	1,996632187	5,20050125	161616	⇒ 211616
93	2,237190281	21,67832168	161616	⇒ 161616 ⇒ 161616

Explications:

Code	Interprétation
161616	:= NotIII-NotIII-Not-III
91616	:= Gener-NotIII-NotIII
191616	:= Pharm-NotIII-NotIII
211616	:= Psych-NotIII-NotIII
190916	:= Pharm-Gener-NotIII
190921	:= Pharm-Gener-Psych
192109	:= Pharm-Psych-Gener
211909	:= Psych-Pharm-Gener

On recense, dans les données, 29 domaines médicaux, et chaque période retient les trois domaines médicaux ayant produit les montants de remboursements les plus élevés: cela fait $29 \times 28 \times 27 = 21'924$ profils possibles! C'est à peu près le nombre de séquences disponibles (20'760). Il n'est donc pas surprenant que l'on ne décèle que peu de séquences de profils. Un support de 10% pour la séquence de profils 161616 \Rightarrow 161616 signifie qu'il y a 10% d'assurés pour lesquels on trouve deux séquences avec le profil 161616, c'est-à-dire deux séquences pour lesquelles il n'y a pas de remboursements. Une confiance de 27% signifie que parmi les couples de profils dont un des profils est 161616 il y en a 27% qui contiennent, comme deuxième profil, encore une fois le profil 161616.

5 Evaluation de la méthode selon les normes du mandat

Ce cadre est formulé dans un document dont on va reprendre exactement les termes que l'on va commenter en faisant référence à ce qui précède.

5.1 Ausgangslage

5.1.1 *Die Daten mit Hilfe von Kohorten- und Längsschnittansätzen aufzubereiten und zu analysieren:*

Dans le cadre de ce mandat, on a fait un certain nombre d'expériences d'échantillonnage qui ont clairement montré que toutes les catégories habituelles telles que l'âge ou le sexe sont impuissantes à saisir la réalité des remboursements et que l'unité statistique avec laquelle il faut travailler est l'historique des remboursements et des prestations de chaque assuré. Les cohortes³⁸ dans leur évolution temporelle ainsi que les analyses longitudinales sont alors automatiquement prises en compte. L'étude des cohortes figées à un instant de temps est facilitée par le fait que l'approximation des trajectoires retenue se fait selon un découpage régulier du temps (contrairement aux observations elles-mêmes qui sont produites "aléatoirement et irrégulièrement") et calcule des moyennes sur ces découpages.

5.1.2 *Die verwendeten Methoden sowie die verwendeten Kategorisierungs- und Klassifikationssysteme allgemein verständlich darzustellen und überprüfen:*

Le détail des méthodes utilisées relève des mathématiques et de la programmation. Une partie de ce détail se trouve en annexe. Cependant les idées sont assez simples et les algorithmes de base bien connus: ce qui reste à faire relève de l'organisation des calculs et de la comparaison des méthodes. Cette organisation doit émerger de l'Etape 2 (page 9).

Si l'on considère les remboursements seulement, les données sont sous la forme suivante: l'intervalle d'observation est $[0, T]$ où 0 représente le premier janvier 1991 et T le 31 décembre 1994. Les factures concernant l'assuré

³⁸Il serait utile de fixer le langage et le langage le plus spécifique est l'anglais. Ainsi A. Fink, dans *How to Design Surveys* (SAGE Publications, Thousands Oaks, CA (1995)), définit: *Cross-sectional designs result in a portrait of one or many groups at one point in time. A cohort is a group of people that have something in common and who remain part of the group over an extended period of time. Cohorts come in two varieties. Type A focuses on the same population each time survey data are collected, although the samples may be different. Type B, sometimes called a panel study, focuses on the same sample.* On voit qu'ici, les unités statistiques étant les trajectoires, le seul problème se trouve du côté des décès.

numéro i (il y en a n_i) “arrivent” à des dates

$$0 \leq t_1^{(i)} < t_2^{(i)} < t_3^{(i)} < \dots < t_{n_i}^{(i)} \leq T$$

et sont à l’origine des remboursements respectifs

$$r_1^{(i)}, r_2^{(i)}, r_3^{(i)}, \dots, r_{n_i}^{(i)}.$$

On construit alors, pour des raisons techniques, la suite des remboursements cumulés:

$$R_1^{(i)}, R_2^{(i)}, R_3^{(i)}, \dots, R_{n_i}^{(i)},$$

où

$$\begin{aligned} R_1^{(i)} &= r_1^{(i)} \\ R_2^{(i)} &= r_1^{(i)} + r_2^{(i)} \\ R_3^{(i)} &= r_1^{(i)} + r_2^{(i)} + r_3^{(i)} \\ &\vdots \qquad \qquad \qquad \vdots \end{aligned}$$

Si l’on veut comparer deux suites de remboursements (cumulés), relatifs aux assurés i_1 et i_2 , on a la difficulté que les deux suites de “dates”

$$0 \leq t_1^{(i_1)} < t_2^{(i_1)} < t_3^{(i_1)} < \dots < t_{n_{i_1}}^{(i_1)} \leq T$$

$$0 \leq t_1^{(i_2)} < t_2^{(i_2)} < t_3^{(i_2)} < \dots < t_{n_{i_2}}^{(i_2)} \leq T$$

ne coïncident pas. C’est pourquoi l’on introduit une suite de “dates virtuelles” suffisamment “fine,”³⁹ soit

$$0 < t_1(n) < t_2(n) < t_3(n) < \dots < t_{2^n-1}(n) < T,$$

pour que les dates des deux suites précédentes y soient contenues. Il faut alors étendre la suite des valeurs des remboursements de manière cohérente pour qu’il y ait “remboursement” aux dates virtuelles que l’on vient d’introduire.⁴⁰ On obtient alors des *trajectoires* de remboursements que l’on peut comparer.

³⁹On remarquera que la suite ainsi construite est *indépendante* de l’assuré. De plus le facteur n définit l’écart entre deux “dates” successives.

⁴⁰Cette opération introduit une redondance importante qui est temporaire mais qui a une incidence sur l’importance des calculs qu’il faut faire.

On dira ainsi que deux trajectoires sont proches et donc semblables si l'écart des remboursements à chacune des "dates" retenues ne dépasse pas un certain seuil $\epsilon > 0$, fixé *a priori*:

$$\max_j | R_{t_j(n)}^{(i_1)} - R_{t_j(n)}^{(i_2)} | \leq \epsilon.$$

Du point de vue des volumes informatiques à traiter, ces trajectoires à l'information redondante posent problème. C'est pourquoi on les "résume" à l'aide d'une des "technologies" de "compactification" des données parmi les plus efficaces, celle des développements en série dans des bases d'ondelettes. *In fine* une trajectoire aura la forme suivante (pour l'assuré numéro i):⁴¹

$$R_1^{(i)}(n), R_2^{(i)}(n), R_3^{(i)}(n), \dots, R_{32}^{(i)}(n).$$

Il y a par suite trois voies qui permettent de créer des catégories et des classifications. Il y a tout d'abord les moyens habituels, sexe, âge, ... , dont on a vu l'insuffisance. Il y a ensuite les catégories "construites" comme l'assuré qui relève de la PCPR (voir la section 4.3, page 24) , et il y a finalement celles qui s'obtiennent en regroupant les trajectoires "semblables." La finalité de ces dernières est la maîtrise des volumes et le recours à des techniques statistiques qui dépassent la description pure (à des fins de prévision, ou d' "explication," par exemple). La finalité des premières est évidemment l'interprétation. Dans tous les cas la compréhension des concepts utilisés ne devrait pas être insurmontable.

⁴¹Le choix de 32 termes est temporaire et a été choisi parce que cela paraissait satisfaisant lors d'une première mise en oeuvre: le passage à 16 ou 64 termes n'est "qu'un" problème de paramétrage et de puissance de calcul. La présence du n est là pour rappeler que l'on a fait un choix particulier d'unité de temps. Les valeurs 16, 32 et 64 sont également liées à une durée d'observation qui est ici de 4 ans. Si cette durée doublait, les chiffres seraient en principe et respectivement de 32, 64 et 128. Il n'y a pas cependant de raison impérative, en l'état, de garder, quelle que soit la durée d'observation, des multiples de ces valeurs.

5.1.3 *Die Programme so zu verallgemeinern, dass sie auf unterschiedliche Fragestellungen und Datenbestände angewendet und durch beliebige - fachkundige - Benutzer verwendet werden können. Voraussetzung dafür ist eine entsprechende Anleitung und eine detaillierte Dokumentation:*

Ce que l'on propose tout d'abord c'est une méthode générale de lecture de la BDL à partir de considérations statistiques qui s'efforcent d'en exploiter la dimension temporelle, ce qui semble indispensable. La généralisation de son utilisation ne devrait donc présenter "que" des problèmes d'adaptation. En fait, si l'on accepte le langage et les concepts du *Data Warehouse* et des *Data Marts*, l'exploitation de la BDL devra probablement se faire, pour surmonter le problème des volumes, au travers d'une suite de *Data Marts*. Si de plus, on souhaite minimiser, pour l'exploitation, la charge informatique, il faudra probablement créer les interfaces appropriées.

Le résultat de l'application de la méthode est fait de deux fichiers, l'un d'approximation des trajectoires, l'autre des segments constitutifs de ces trajectoires. Ce dernier fichier peut être exploité directement en recourant à SQL. Le premier a une vocation plus technique puisqu'il sert à opérer sur les entités statistiques de base.

5.2 *Auftrag*

5.2.1 *... erarbeitet einen Bericht über die Verwendungsmöglichkeiten neuester statistischen Methoden und Informationslösungen zur Analyse von Längsschnitt und Kohortendaten. Der Bericht hat insbesondere aufzuzeigen wie sich im Rahmen der Administrativdatenstatistik stehenden Probleme in Zusammenhang mit der Analyse grosser Datenmengen und zur Fallklassifikation nutzen lassen.*

Les techniques d'analyse longitudinale sont non seulement des techniques récentes et en pleine évolution, mais elles sont les seules qui permettent d'exploiter l'information probablement la plus utile des données de la BDL, à savoir la structure de corrélation intrinsèque à chaque histoire: c'est la seule information qui permette l'étude sérieuse de la dynamique du système. Or le choix des histoires individuelles comme entités statistiques premières est indispensable à ces études longitudinales. La segmentation facilite l'étude

descriptive de cette dimension.

La prise en compte de volumes s'exprime de plusieurs manières.

1. Tout d'abord on fait appel à la technologie de compression la plus efficace que l'on connaisse à l'heure actuelle, ce qui permet de remplacer des séries *irrégulières* de temps de réalisation d'événements par des séries *régulières souvent sensiblement plus courtes*. Ainsi, non seulement on *réduit le volume*, mais encore on le *structure*. Cette technologie permet la concaténation des fichiers et donc le traitement des données par lots. Cela permet en particulier de pallier à un désavantage non négligeable d'une telle approche qui est qu'elle exige des calculs intermédiaires importants.
2. Le recours aux trajectoires comme entités de base permet la segmentation par catégories de coûts et de dynamique: cette approche a le mérite de réduire la variation et faciliter par suite l'émergence de l'information utile.
3. Il faut aussi tenir compte du fait que les volumes à traiter ne sont pas ceux de la BDL tout entière. Ainsi, la psychiatrie peut "naturellement" se décomposer en psychiatrie hospitalière, ambulatoire et de cabinet. Quand on voudra comparer ces domaines, on travaillera sur des fichiers de sortie plus restreints.

5.2.2 *Parallel zum Bericht sind die praktischen Arbeiten anhand der vorhandenen Daten zu illustrieren und die notwendigen Programme aufzubauen, damit diese für die zukünftige Administrativdatenstatistik in den produktiven Betrieb übernommen werden können:*

Un certain nombre d'illustrations ont été présentées dans ce document. La partie la moins développée est celle des programmes et il y a à cela diverses raisons.

1. Si l'on veut gérer les volumes, il faut en général exploiter toutes les particularités de la situation que l'on connaît. Or l'usage de la méthode que l'on a décrite est encore en gestation et cela ne semble pas avoir beaucoup de sens que de faire de très grands efforts en matière d'informatique

quand on n'est pas à même de spécifier ce que ce système informatique doit être.

2. La mise en oeuvre de la méthode que l'on a décrite peut se faire de multiples manières et il n'y a pas eu d'étude comparative: on s'est contenté de faire des choix apparaissant raisonnables au moment de ce choix, mais rien n'exclut qu'il y ait d'autres venues plus utiles ou plus efficaces. Pour en décider, il faut recourir à une expérience plus grande dans l'analyse des données disponibles que celle dont on dispose à ce jour.
3. La première programmation de la méthode a été faite avec SAS-IML. La raison en fut qu'on voulait minimiser l'effort de programmation en recourant à un système préprogrammé qui de plus "accueillerait directement" les fichiers produits pour en permettre l'analyse statistique. Ces programmes initiaux devaient également aider à rechercher la manière la plus efficace d'organiser les calculs.

La programmation qui a été produite fut irréfléchie et ne correspondait pas aux normes de qualité minimales auxquelles on peut s'attendre dans ce domaine. De plus, les programmes produits ont achoppé sur le nombre maximal de prestations par facture fixé à 16 à la suite semblait-il d'un malentendu dont il est difficile de retracer l'origine.

Ces programmes ont néanmoins produit les fichiers sur lesquels les analyses de ce document sont basées, et en cela l'opération de programmation n'a été qu'un échec partiel. Elle a permis cependant de vérifier un fait qui devrait être évident, mais qui ne l'est pas assez souvent, à savoir que lors de démarches complexes il est risqué de faire de l'*outsourcing* en matière de programmation.

5.3 *Teilschritte*

5.3.1 *Aufzeigen der zusätzlichen Analysemöglichkeiten/Notwendigkeit von Kohorten-, Längsschnitt- und Pfadanalysen für die Bearbeitung von Problemstellungen, wie sie im Rahmen des Administrativdatenstatistik vorkommen:*

On a donné ailleurs dans ce document des réponses à cette interrogation générale. On ne va donc qu'aborder les points unter Berücksichtigung

1. der vorhandenen Daten und deren Qualität:

Il est difficile de dire quoi que ce soit sur la qualité des données en amont de l'usage qui en est fait ici. Il semble cependant au vu des premières analyses qu'il est possible d'utiliser les données disponibles à bon escient. Par ailleurs il serait possible d'y apporter à peu de frais une amélioration de première importance. En effet, dans la BDL, quand un assuré est soigné en hôpital pour des raisons autres que psychiatriques, on ne peut associer aux prestations dont il bénéficie un domaine médical. Or il semble que la statistique des hôpitaux (Office fédéral de la statistique) contienne une telle information. Tous les efforts devraient être faits pour compléter la BDL avec cette information. Enfin une forme de contrôle de qualité peut être réalisée.

2. der fehlenden Diagnosen:

On a montré comment on peut, au travers des périodes de remboursement et des domaines médicaux qu'on leur associe, et donc de la méthode proposée, faire émerger des informations que l'on peut utiliser en lieu et place des diagnostics. L'efficacité de ce substitut ne peut être démontrée qu'à l'usage. De plus, comme on l'a vu, il reste encore à trouver le lien le plus étroit possible entre période et domaine médical.

3. der Notwendigkeit einer Fallklassifikation, die in Rahmen der Datenanalyse 'empirisch' aufzustellen und periodisch zu überprüfen/anzupassen ist:

La méthode suivie a pour résultat des fichiers d'exploitation neutres qui ne présument en rien les conceptualisations qui peuvent en être tirées, mais qui, par contre, permettent de les produire puis de les évaluer.

4. der Datenmenge (inkl. Erörterung, ob mit Stichproben, aggregierten Angaben etc. ähnliche Resultate erreichbar sind/die Durchführung vereinfacht werden kann):

Presque toutes les répartitions que l'on rencontre dans la BDL semblent très fortement asymétriques et présentent un très grand nombre de valeurs excentriques: toutes les agrégations doivent par suite faire la preuve qu'il n'y a pas surestimation des grandeurs considérées. Les indicateurs qui ont été développés à ce jour sont très faiblement liés aux données. Les échantillons qui ont été formés (typiquement sur le principe des quotas) n'ont pas permis d'accéder au "réel," et d'appréhender les faits pertinents. La raison en est que la dimension temporelle n'a pas été prise en compte. C'est pourquoi on a choisi de travailler avec l'essence des données, soit cette dimension temporelle, et l'on a vu dans les illustrations qu'une foule de faits potentiellement intéressants peuvent être mis en évidence par ce moyen.

5. des gesetzlichen Auftrags:

Il semble que seules des approches du type de celle qui a été développée permettent en fait de satisfaire les injonctions de la loi.

5.3.2 *Erörterung der Vor- und Nachteile des Vorgehens verglichen mit anderen Analysemethoden:*

On a fait, dans ce qui précède, de nombreuses remarques sur les avantages et les inconvénients de la méthode proposée. On ne fera ici que les rappeler et résumer. Du côté des *avantages*, on a:

1. On attaque les problèmes de fond, de front, et on exploite la nature essentielle des données.
2. On produit les moyens de conduire des analyses statistiques complexes qui seules permettent peut-être de répondre aux questions auxquelles on aimerait donner réponse.
3. On se donne une chance de comprendre le fonctionnement de l'assurance maladie.

Rien de ce qui a été fait à ce jour et que l'on a pu examiner ne peut remplacer ce qui est proposé.

Du côté des *désavantages*, on a:

1. La méthode conduit à des exploitations plus complexes que celles mises en oeuvre à ce jour, et par suite plus difficiles à diffuser.
2. La méthode exige un engagement supérieur à celui qui a existé.
3. La méthode présuppose une volonté de *comprendre* ce qui se passe dans l'assurance maladie plutôt que le simple désir de calculer l'évolution de quelques grands agrégats.

5.3.3 *Allgemein verständliche Erläuterung der in der HMO/-Bonus-Evaluation erarbeiteten konkreten Lösungen sowie Illustration anhand konkreter Beispiele:*

Rien n'a encore été fait dans ce domaine qui présuppose que l'on dispose d'un instrument complet d'analyse, ce qui est loin d'être le cas.⁴²

5.3.4 *Programme:*

Comme on l'a indiqué, le seul effort de programmation qui a été fait a produit le résultat minimal que l'on dispose du fichier qui permet d'attaquer l'Etape 2 (page 9) du processus qui est engagé, mais que l'on n'en a tiré aucune autre information utile. Trop de préalables manquent encore pour que l'on puisse se prononcer sérieusement sur les questions d'informatique. Par contre, étant donné que la méthode conduit à la production de fichiers plats, la compatibilité avec les systèmes informatiques traditionnels ne pose que des problèmes de volume et de temps de calcul. Finalement, la méthode est applicable à toutes les données de structure équivalente.

⁴²Il y aurait cependant des travaux internes à l'OFAS qui n'ont pas été publiés.

5.3.5 *Die theoretischen Überlegungen und die darauf aufbauende Programmierung ist anhand von Datenanalysen mit den Daten des Rechnungsjahre 1991-1994 sowie 1997 und 1998 auszutesten und zu illustrieren:*

La partie (section) 4 de ce document contient un assortiment d'analyses qui a pour but d'illustrer ce que l'on peut attendre de la méthode proposée aussi bien que d'exposer sa nature. On n'a travaillé que sur les données 1991-1994 car en principe cette méthode doit s'appliquer en continu plutôt que sporadiquement.

Mit Hilfe von Längsschnitt- und Querschnittanalysen ist insbesondere aufzuzeigen⁴³

1. *die Verteilung der Kosten und deren Entwicklung nach*
 - (a) Versichertenmerkmalen (Alter, Geschlecht, Wohnort, etc.)
 - (b) Leistungserbringer (kategorien) und Leistungen
 - (c) geeigneter Fall- und Leistungsklassifikation
 - (d) Versicherungsform
 - (e) Versicherer

Les illustrations relatives aux groupes PCPS et PCPR sont typiques de ce que l'on ferait pour produire les répartitions mentionnées ci-dessus. Mais des analyses systématiques de ce type demandent une codification et une systématisation des procédures d'analyse qui n'a pas encore été fait. C'est le rôle de l'Étape 2 (page 9) du processus.

2. *die gesundheits- und sozialpolitischen Wirkungen des Gesetzes (z.B. Risikoausgleich, Analyse von Tarifwirkungen, Möglichkeiten der Rückversicherung etc.)*

On pourra essayer de répondre à de telles questions au terme de l'Étape 2 (page 9): le corpus de données utilisées pour l'illustration de la méthode est à cette fin incomplet et de plus il reste à préciser la manière dont ces phénomènes peuvent être reflétés par les données.

⁴³Cet aspect est consubstantiel aux choix fait des unités statistiques, à savoir les trajectoires.

6 Description de quelques algorithmes

6.1 Organisation générale des algorithmes

Les algorithmes dont la description suit sont des *fonctions*, au sens de la programmation en C, par exemple. Plus le nombre d'assurés est grand, plus l'organisation de l'évaluation de ces fonctions devient importante pour le déroulement efficace des calculs. À ce jour, cette question n'a pas été étudiée, et ne doit l'être que lorsque la partie conceptuelle du travail sera stabilisée.

Voici la liste des fonctions et leur rôle:

1. Transformation: **trajectoire brute** \longrightarrow **trajectoire dyadique**

On suppose que l'on a observé, pour l'assuré numéro i , aux instants (qui lui sont propres)

$$0 \leq t_1^{(i)} < t_2^{(i)} < \dots < t_{n_i}^{(i)} \leq T,$$

des valeurs d'une fonction f_i , dont le graphe est celui d'un escalier "à marches irrégulières," soit

$$f_i(t_1^{(i)}) < f_i(t_2^{(i)}) < \dots < f_i(t_{n_i}^{(i)}).$$

On a que $f_i(t)$ est une constante pour $t \in [t_j^{(i)}, t_{j+1}^{(i)}[$. $[0, T]$ désigne la période durant laquelle la fonction f_i a été observée (par exemple, 0 correspond au premier janvier 1991, et T au 31 décembre 1994). La fonction f_i peut représenter les remboursements cumulés, les nombres cumulés de prestations ou de factures, ou tout autre fonction numérique du même genre, toujours pour l'assuré numéro i .

Les couples $\{(t_j^{(i)}, f_i(t_j^{(i)}))\}$, $1 \leq j \leq n_i$ constituent l'*input* de la transformation. L'*output* doit avoir la forme suivante:

$$\{(k, f_i(k)), 1 \leq k \leq 2^n\}.$$

Dans le cas des remboursements, le seul qui a été pris en compte, n représente à la fois la durée $[0, T]$ et le nombre maximum de prestations sur une facture.⁴⁴

2. Transformation: $\{(k, f_i(k)), 1 \leq k \leq 2^n\} \longrightarrow$ ondelette de Haar

Il s'agit d'obtenir la représentation

$$f_i(t) = \sum_{l,m} \gamma_{l,m}^{(i)} h_{l,m}(t),$$

où les nombres $\gamma_{l,m}^{(i)}$ sont spécifiques à f_i , et les fonctions $h_{l,m}$ particulièrement simples, et indépendantes de f_i . Ces dernières n'ont, dans la suite, qu'un rôle latent.

3. Calcul d'une approximation de f_i

C'est l'opération de décimation: un certain nombre des coefficients $\gamma_{l,m}^{(i)}$ sont remplacés par 0. Le choix des coefficients ainsi "annulés" est basé sur le "niveau" de ces coefficients (c'est pour cela qu'il y a deux indices): tous les coefficients dont le niveau est supérieur à un seuil donné sont ainsi "mis à zéro." Ce seuil détermine la précision de l'approximation et doit donc aussi être "paramétré."

Si l'on désigne par $\hat{\gamma}_{l,m}^{(i)}$ les coefficients qui résultent de la décimation, on obtient une fonction

$$h_i(t) = \sum_{l,m} \hat{\gamma}_{l,m}^{(i)} h_{l,m}(t),$$

et c'est à cette fonction qu'on applique la transformée de Haar inverse. Le résultat est l'approximation \hat{f}_i de f_i recherchée. On dispose de 2^n valeurs de cette fonction, et cela est trop. C'est pourquoi on écrit $n = p + q$ et que l'on ne retient que 2^p valeurs. Ce sont ces 2^p valeurs qui servent d'approximation à f_i . Jusqu'à présent on a choisi $p = 5$, mais il faut pouvoir expérimenter avec cette valeur qui doit donc aussi entrer dans les programmes comme paramètre.

⁴⁴Pour les premiers essais, ce nombre avait été estimé à 16, ce qui s'est révélé insuffisant. C'est donc un nombre qui reste à déterminer, mais qui doit être une puissance de 2. Pour une raison d'efficacité des calculs, ce nombre maximum doit être le plus petit possible. Ce doit être un paramètre de la fonction.

4. Recherche des périodes

Les périodes sont définies à partir de la “dérivée” de l’approximation (on calcule en fait les différences successives). La trajectoire de cette “dérivée” ressemble à un relief montagneux fait d’une succession de pics plus ou moins importants. L’algorithme recherche le pic le plus élevé, identifie sa base, puis met les données du pic à 0, sur cette base. Il recherche ensuite le pic le plus élevé et recommence l’opération. L’algorithme s’arrête quand les pics restants ont une hauteur “insignifiante,” qui est définie par des seuils qui sont passés à l’algorithme comme paramètres. Chaque base de pic définit une période et ces périodes sont “re-numérotées” pour obtenir l’ordre temporel de leur déroulement.

5. Identification des domaines médicaux prépondérants

Pour chaque période est calculée la somme des montants remboursés ainsi que les pourcentages de cette somme représentés par les divers domaines médicaux retenus (il y en a 41 en ce moment). Pour la présentation des résultats, ont été retenus les trois pourcentages les plus élevés, mais 2 ou 4 auraient tout aussi bien pu être retenus. Ce nombre doit aussi figurer comme paramètre dans les algorithmes.

À la fin de l’étape numéro 3, on doit pouvoir disposer d’un fichier d’approximations de trajectoires, et à la fin de l’étape numéro 5, d’un fichier de périodes, de pourcentages et de domaines médicaux.

6.2 Préparation des données

Il s’agit d’écrire les données d’entrée sous forme dyadique. Supposons en effet qu’il y ait, dans la période considérée, M jours, et que les factures de l’assuré numéro i (il y a I assurés) arrivent aux jours

$$m_1^{(i)} < \dots < m_{p_i}^{(i)}.$$

Chaque facture recense en plus un certain nombre de prestations qui, dans un premier temps, sont gérées individuellement: c’est pourquoi chaque jour $m_j^{(i)}$ est décomposé en un certain nombre de tranches

$$m_j^{(i)} = m_{j,1}^{(i)} < \dots < m_{j,q_j}^{(i)} < m_{j+1}^{(i)}.$$

L'approche utilisée présuppose que l'on aie⁴⁵

$$m_{j,k}^{(i)} = k \in [1 : 2^n].$$

Ce type d'identité n'ayant lieu qu'exceptionnellement, l'idée est de choisir n suffisamment grand pour que, si l'on écrit les entiers de $[1 : 2^n]$ sous la forme suivante:

$$1 = n_1 < \dots < n_l < \dots < n_{2^n} = 2^n,$$

l'attribution

$$m_{j,k}^{(i)} \longrightarrow n_l$$

soit univoque, c'est-à-dire qu'un seul entier n_l corresponde à chacune des valeurs $m_{j,k}^{(i)}$. On introduit ainsi une modification des données qui, si l'on prend en compte les autres facteurs d'incertitude, ne doit pas porter à conséquence. D'autres approches sont possibles et elles devraient être testées.

On parvient ainsi à un premier programme qui ne prend en compte, pour des raisons de lisibilité, qu'une variable, les remboursements par exemple. L'extension à plusieurs variables est "évidente." Les valeurs en **entrée** en sont:

nbDays := nombre de jours dans la période de facturation prise en compte (**param**)
nbDivs := nombre de périodes dans une journée (**param**: puissance de 2)
indaynb := numéros des jours pour lesquels il y a une facture (**data**)
indata := donnée de la facture dont on veut construire la trajectoire dyadique (**data**)

Les valeurs en **sortie** en sont:

outDivnb := séquence 1, 2, 3, ..., 2^n
outData := valeur de la donnée en *outDivnb*
dyaDay := forme dyadique des numéros des jours pour lesquels il y a une facture

Remarques:

⁴⁵La notation $[m : n]$, $m < n$, signifie tous les entiers $m, m+1, m+2, \dots, n-2, n-1, n$.

1. Ces programmes calculent la “trajectoire” d’un assuré comme s’il était le seul, d’où le “calcul” de la variable *outDivnb*. En pratique, et pour éviter tout calcul redondant, cette variable ne sera “calculée” qu’une fois.
2. La variable *dyaDay* contient les instants de *outDivnb* auxquels *outData* change de valeur.

Etape 1: Recherche de la puissance dyadique appropriée (la valeur de n dans 2^n)

Supposons que $nbDays = 365$: on cherche alors n tel que

$$2^{n-1} < 365 \leq 2^n,$$

ce qui donne $n = 9$, puisque $2^8 = 256$ et $2^9 = 512$. De plus, si l’on pense qu’il y a au plus 16 prestations par facture, on ajoutera à $n = 9$, 4 unités, puisque $2^4 = 16$, et la valeur de n utilisée dans le programme sera $n = 13$.

Le **code** pour obtenir n (la variable *power* ci-dessous) est le suivant:

```

I      ← 0
J      ← 1
left   ← 2I
right  ← 2J
WHILE [(nbDays < left) ∨ (nbDays > right)]
    I ← J
    J ← I + 1
    left ← 2I
    right ← 2J
END    (while)
power ← J + ln2(nbDays)

```

Etape 2: Valeurs dyadiques des jours de facturation

Il s’agit de transformer des valeurs contenues dans l’intervalle $[1 : 365]$ (celles de *indaynb*) en valeurs correspondantes de l’intervalle $[1 : 2^n]$. On écrit à cette fin

$$y = \alpha x, \quad 2^n = \alpha 365, \quad n = power,$$

ce qui donne

$$\alpha = \frac{2^n}{365},$$

puis

$$dyaDay [l] = \begin{cases} [\alpha \times indaynb [l]] & \text{si } \alpha \times indaynb [l] - [\alpha \times indaynb [l]] < \frac{1}{2} \\ [\alpha \times indaynb [l]] + 1 & \text{si } \alpha \times indaynb [l] - [\alpha \times indaynb [l]] \geq \frac{1}{2} \end{cases}$$

Le **code** pour obtenir *dyaDay* est le suivant:

$$\begin{aligned} factor &\leftarrow 2^{power} / nbDays \\ dyaDay &\leftarrow round(factor \times indaynb) \end{aligned}$$

Etape 3: Certaines factures reçues par les caisses dans l'année courante sont datées de l'année précédente: elles ont un numéro négatif. Les assurances ne s'intéressant qu'aux flux financiers de l'année comptable, on commence l'année avec le cumul des "impayés" de l'année précédente (dans les cas pour lesquels la variable d'intérêt est celle des remboursements). S'il n'y a pas d'"impayés," on commence à 0. Sinon il faut éliminer les numéros négatifs et obtenir les bonnes valeurs issues de l'année précédente, jusqu'à l'arrivée de la première facture de l'année courante.

Le **code** pour obtenir ces "valeurs initiales" est à la page suivante.

Résultat: On obtient en sortie 2^n valeurs de remboursements cumulés, soit la trajectoire temporelle de ces remboursements.

<pre> I IF ← 1 (indaynb[I] < 0) WHILE (indaynb[I] < 0) I ← I + 1 (while) END J ← I - 1 K ← 1 WHILE (outDivnb[K] < dyaDay[I]) outData[K] ← indata[I] K ← K + 1 (while) END ELSE K ← 1 WHILE (outDivnb[K] < dyaDay[I]) outData[K] ← 0 K ← K + 1 (while) END (if) left ← indaynb - I + 1 (left > 0) curr ← dyaDay[I] J ← I + 1 next ← dyaDay[J] IF (curr < next) WHILE (outDivnb(K) < next) outData[K] ← indaynb[I] K ← K + 1 (while) END ELSE outData[K] ← indaynb[I] K ← K + 1 (if) END left ← left - 1 curr ← next I ← I + 1 next ← dyaDay[I] (while) END I ← I - 1 WHILE (outDivnb(K) ≤ 2^{power} - 1) outData[K] ← indata[I] K ← K + 1 (while) END </pre>	<pre> Elimination des numéros i 0 Remplissage jusqu'à l'arrivée de la première facture de l'année S'il n'y a pas de numéros i 0, on insère des valeurs nulles Nombre de jours avec facture au cours de la période considérée (indaynb := nombre de jours avec facture) Pour chaque jour avec facture, on regarde s'il y a plusieurs prestations: - si ce n'est pas le cas, on complète avec les valeurs courantes l'intervalle allant à la prochaine facture -si c'est le cas, on passe au jour de facturation suivant Actualisation des indices Prolongement de la trajectoire avec la dernière valeur enregistrée </pre>
--	--

6.3 Transformation de Haar, “décimation” et transformation de Haar inverse

Etant donnés 2^n nombres f_1, \dots, f_{2^n} , représentant les valeurs d’une fonction f aux “instants” $1, \dots, 2^n$, on recherche une approximation de f “économique.” Voici une manière de procéder illustrée à l’aide du cas $n = 3$.

6.3.1 Transformation de Haar

Les données en entrée sont

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \end{bmatrix}, \text{ par exemple, } \underline{f} = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 4 \\ 8 \\ 6 \\ 9 \\ 9 \end{bmatrix}.$$

L’esquisse de l’algorithme est la suivante:

$$\begin{aligned} & i = 1, j = 2, n = 2^p, \tilde{f} \leftarrow \underline{f} \\ & \text{pour } k = 1 \text{ jusqu'à } p \\ & \quad n \leftarrow \frac{n}{2} \\ & \text{pour } l = 0 \text{ jusqu'à } n - 1 \\ & \quad \zeta \leftarrow j \times l \\ & \quad \eta \leftarrow \zeta + i \\ & \quad \hat{f}_\zeta \leftarrow \frac{\tilde{f}_\zeta + \tilde{f}_\eta}{2} \\ & \quad \hat{f}_\eta \leftarrow \frac{\tilde{f}_\zeta - \tilde{f}_\eta}{2} \\ & \text{fin de boucle sur } l \\ & \quad \tilde{f} \leftarrow \hat{f} \end{aligned}$$

$$i \longleftarrow j$$

$$j \longleftarrow 2j$$

fin de boucle sur k

Avec les données de l'exemple, on a:

1. Début de l'algorithme

$$i = 1, \quad j = 2$$

2. $k = 1 : n = 4$

(a) $l = 0 : i = 1, j = 2 \longrightarrow \zeta = 0, \eta = 1$

$$\hat{f}_0 = \frac{\tilde{f}_0 + \tilde{f}_1}{2} = \frac{3 + 1}{2} = 2$$

$$\hat{f}_1 = \frac{\tilde{f}_0 - \tilde{f}_1}{2} = \frac{3 - 1}{2} = 1$$

(b) $l = 1 : i = 1, j = 2 \longrightarrow \zeta = 2, \eta = 3$

$$\hat{f}_2 = \frac{\tilde{f}_2 + \tilde{f}_3}{2} = \frac{0 + 4}{2} = 2$$

$$\hat{f}_3 = \frac{\tilde{f}_2 - \tilde{f}_3}{2} = \frac{0 - 4}{2} = -2$$

$l = 2 : i = 1, j = 2 \longrightarrow \zeta = 4, \eta = 5$

$$\hat{f}_4 = \frac{\tilde{f}_4 + \tilde{f}_5}{2} = \frac{8 + 6}{2} = 7$$

$$\hat{f}_5 = \frac{\tilde{f}_4 - \tilde{f}_5}{2} = \frac{8 - 6}{2} = 1$$

(c) $l = 3 : i = 1, j = 2 \longrightarrow \zeta = 6, \eta = 7$

$$\hat{f}_6 = \frac{\tilde{f}_6 + \tilde{f}_7}{2} = \frac{9 + 9}{2} = 9$$

$$\hat{f}_7 = \frac{\tilde{f}_6 - \tilde{f}_7}{2} = \frac{9 - 9}{2} = 0$$

fin de boucle k = 1 \longrightarrow

$$\underline{\tilde{f}} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ -2 \\ 7 \\ 1 \\ 9 \\ 0 \end{bmatrix}$$

$i = 2, j = 4$

3. $k = 2 : n = 2$

(a) $l = 0 : i = 2, j = 4 \longrightarrow \zeta = 0, \eta = 2$

$$\hat{f}_0 = \frac{\tilde{f}_0 + \tilde{f}_2}{2} = \frac{2 + 2}{2} = 2$$

$$\hat{f}_2 = \frac{\tilde{f}_0 - \tilde{f}_2}{2} = \frac{2 - 2}{2} = 0$$

(b) $l = 1 : i = 2, j = 4 \longrightarrow \zeta = 4, \eta = 6$

$$\hat{f}_2 = \frac{\tilde{f}_2 + \tilde{f}_3}{2} = \frac{7 + 9}{2} = 8$$

$$\hat{f}_3 = \frac{\tilde{f}_2 - \tilde{f}_3}{2} = \frac{7 - 9}{2} = -1$$

fin de boucle k = 2 \longrightarrow

$$\underline{\tilde{f}} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -2 \\ 8 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

$i = 4, j = 8$

4. $k = 3 : n = 1$

(a) $l = 0 : i = 4, j = 8 \longrightarrow \zeta = 0, \eta = 4$

$$\hat{f}_0 = \frac{\tilde{f}_0 + \tilde{f}_4}{2} = \frac{2 + 8}{2} = 5$$

$$\hat{f}_4 = \frac{\tilde{f}_0 - \tilde{f}_4}{2} = \frac{2 - 8}{2} = -3$$

fin de boucle $k = 3 \longrightarrow$

$$\underline{\tilde{f}} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ -2 \\ -3 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

\equiv la représentation de la trajectoire!

5. fin de l'algorithme

Remarque: f_0 est la moyenne des valeurs de la fonction et jouit donc d'un statut "privilegié:" c'est pourquoi on recourt à l'indice 0.

6.3.2 Niveau des coefficients

Pour approcher, on décime: certains des coefficients sont mis à zéro. Une manière de repérer les coefficients que l'on va annuler consiste à leur attribuer une profondeur. Cette profondeur est calculée à l'aide de l'algorithme suivant qui ne doit être exécuté qu'une fois (pour autant qu'on ne change pas la division dyadique).

Soit κ_i la valeur de l'indice en position i

$$\kappa_1 \longleftarrow 0$$

pour $i = 0$ *jusqu'à* $p - 1$
 $j = 2^{p-i-1} - 1$
pour $k = 0$ *jusqu'à* j
 $l \leftarrow (2^i + 1) + k2^{i+1}$
 $\kappa_l \leftarrow p - i$
fin de la boucle sur k
fin de la boucle sur i

Dans le cas de l'exemple considéré, on obtient le vecteur

$$\begin{bmatrix} 0 \\ 3 \\ 2 \\ 3 \\ 1 \\ 3 \\ 2 \\ 3 \end{bmatrix}$$

6.3.3 “Décimation”

Selon la précision souhaitée, on décide que tous les coefficients plus profonds qu'un seuil donné, qui définit la précision résultante, sont mis à zéro. Ainsi, dans l'exemple, si l'on met tous les coefficients à zéro, sauf le premier, on approche la fonction par sa moyenne. L'approximation typique remplacerait les coefficients de profondeur 3 par 0, ce qui donne, pour la transformée de Haar de l'approximation,

$$\underline{f}_{approx}^H = \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ -3 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

6.3.4 Reconstruction de (l'approximation de) la trajectoire

A partir des coefficients obtenus en 2.3, on peut récupérer (l'approximation de) la trajectoire à l'aide de l'algorithme suivant:

$$i = 2^{p-1}, j = 2i, m = 1$$

$$\tilde{f}_{approx}^H \leftarrow \text{vecteur des coefficients}$$

pour l = 1 jusqu'à p

pour k = 0 jusqu'à m - 1

$$\zeta \leftarrow jk$$

$$\eta \leftarrow \zeta + i$$

$$\tilde{f}_{\zeta}^H \leftarrow \tilde{f}_{\zeta}^H + \tilde{f}_{\eta}^H$$

$$\tilde{f}_{\eta}^H \leftarrow \tilde{f}_{\zeta}^H - \tilde{f}_{\eta}^H$$

fin de la boucle sur k

$$j \leftarrow i, i \leftarrow \frac{i}{2}, m \leftarrow 2m$$

fin de la boucle sur l

6.3.5 Réduction du nombre de valeurs de la fonction

Pour des raisons d'économie de calcul on ne retient de l'approximation qu'un nombre réduit de valeurs, les valeurs de l'approximation étant très redondantes. Supposons donc que la profondeur de "décimation" soit $p + 1$. L'approximation sera constante sur des intervalles dyadiques de longueur 2^p . On ne retiendra de l'approximation que les valeurs aux positions

$$2^{p-1} + 0 \times 2^p, 2^{p-1} + 1 \times 2^p, 2^{p-1} + 2 \times 2^p, \dots, 2^{p-1} + (2^p - 1) \times 2^p,$$

qui sont les points milieu de segments successifs de "longueur" 2^p .

6.4 Segmentation des trajectoires

6.4.1 Segmentation proprement dite

A partir d'une trajectoire (réduite selon 6.2), de remboursements par exemple, on produit une décomposition de celle-ci en périodes successives. A cette fin, on utilise la "dérivée" de la trajectoire: les périodes correspondent alors soit à des "pics" soit à des "paliers." L'algorithme commence par rechercher le "pic" le plus "élevé," puis évalue l'étendue de sa base et en fait une période. Le "pic" est alors "éliminé" et l'algorithme recherche alors le "pic" le plus "élevé" parmi ceux qui restent: l'opération précédente recommence. Quand il n'y a plus de "pics," la recherche s'arrête: la décomposition est terminée. En pratique, il faut décider ce qu'est un "pic:" on utilise à cette fin des seuils.

Les variables utilisées sont les suivantes:⁴⁶

⁴⁶Une trajectoire complète contient 2^n valeurs, la trajectoire approchée 2^p . Δ est l'opérateur qui prend les différences successives:

$$\text{si } \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{m-1} \\ x_m \end{bmatrix}, \text{ alors } \Delta[\underline{x}] = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_m - x_{m-1} \end{bmatrix}.$$

\widehat{R}	:=	vecteur en entrée de l'algorithme "contenant" 2^P valeurs de remboursements
π_C	:=	pourcentage courant
π_m	:=	pourcentage minimum
π_M	:=	pourcentage maximum
π_S	:=	pourcentage suivant
ε_C	:=	différence courante
ε_m	:=	différence minimum
ε_M	:=	différence maximum
ε_S	:=	différence suivante
I_D	:=	localisation du début de période
I_C	:=	position courante
I_E	:=	localisation de fin de période
I_M	:=	position du pourcentage maximum
I_S	:=	position suivante
n_P	:=	nombre maximum de périodes (devrait être donné comme paramètre; dépend de la durée d'observation)
\underline{P}	:=	vecteur contenant les étiquettes identifiant les périodes
\underline{N}	:=	vecteur contenant les étiquettes identifiant les observations non utilisés
C	:=	indicateur qui signifie "continuer"
σ_π	:=	seuil qui rend les pourcentages "significatifs"
σ_δ	:=	seuil qui rend les différences "significatives"
λ_π	:=	pourcentage minimum à gauche
ρ_π	:=	pourcentage minimum à droite
λ_δ	:=	différence minimum à gauche
ρ_δ	:=	différence minimum à droite


```

 $\widehat{R}_\Delta \leftarrow \Delta \left[ \widehat{R} \right]$ 
 $\pi_{\widehat{R}_\Delta} \leftarrow 100 \times \frac{\widehat{R}_\Delta}{\max \left[ \widehat{R}_\Delta \right]}$ 
 $\pi_M \leftarrow 0$ 
 $I_M \leftarrow 1$ 
DO (I = 1 : 2p)
  IF (  $\pi_{\widehat{R}_\Delta} [I] > \pi_M$  )
     $\pi_M \leftarrow \pi_{\widehat{R}_\Delta} [I]$ 
     $I_M \leftarrow I$ 
  END (if)
END (do)
DO (I = 1 : 2p)
   $P [I] \leftarrow n_P$ 
   $N [I] \leftarrow 1$ 
END (do)

 $\pi_m \leftarrow 0$ 
 $\pi_M \leftarrow 100$ 
 $\delta_m \leftarrow 0$ 
 $\delta_M \leftarrow \widehat{R}_\Delta [I_M]$ 
 $I_B \leftarrow 1$ 
 $I_E \leftarrow 2^p$ 
C ← (  $\pi_M - \pi_m > \sigma_\pi$  ) ∧ (  $\delta_M - \delta_m > \sigma_\delta$  )
WHILE ( C = 1 )
   $I_C \leftarrow I_M$ 
   $\pi_C \leftarrow \pi_M$ 
   $\delta_C \leftarrow \delta_M$ 

   $I_S \leftarrow I_C - 1$ 
   $\pi_S \leftarrow \pi_{\widehat{R}_\Delta} [I_S]$ 
   $\delta_S \leftarrow \widehat{R}_\Delta [I_S]$ 
  WHILE [ (  $\pi_S < [1 + \epsilon] \times \pi_C$  ) ∧ (  $\pi_S > 0$  ) ]
     $I_C \leftarrow I_S$ 
     $\pi_C \leftarrow \pi_S$ 
     $\delta_C \leftarrow \delta_S$ 

     $I_S \leftarrow I_S - 1$ 
     $\pi_S \leftarrow \pi_{\widehat{R}_\Delta} [I_S]$ 
     $\delta_S \leftarrow \widehat{R}_\Delta [I_S]$ 
  END (while)

   $I_B \leftarrow I_C$ 
   $\lambda_\pi \leftarrow \pi_C$ 
   $\lambda_\delta \leftarrow \delta_C$ 

```

Les valeurs de \widehat{R} sont croissantes

Les valeurs de \widehat{R}_Δ sont positives

Localisation du pourcentage maximum

On commence avec une seule période

Le remboursement le plus important

On étend la période tant que les remboursements en valeur absolue et relative sont assez importants
 ♣ Recherche de période

On se positionne au maximum courant

On explore la période qui précède le maximum

On exige une différence d'au moins $\epsilon > 0$

continuation du WHILE \clubsuit

$I_C \leftarrow I_M$
 $\pi_C \leftarrow \pi_M$
 $\delta_C \leftarrow \delta_M$

$I_S \leftarrow I_C + 1$

$\pi_S \leftarrow \widehat{\pi_{R_\Delta}} [I_S]$

$\delta_S \leftarrow \widehat{\delta_{R_\Delta}} [I_S]$

WHILE $[(\pi_S < [1 \vee c] \times \pi_C) \wedge (\pi_S > 0)]$

$I_C \leftarrow I_S$
 $\pi_C \leftarrow \pi_S$
 $\delta_C \leftarrow \delta_S$

$I_S \leftarrow I_S + 1$

$\pi_S \leftarrow \widehat{\pi_{R_\Delta}} [I_S]$

$\delta_S \leftarrow \widehat{\delta_{R_\Delta}} [I_S]$

END (while)

$I_B \leftarrow I_C$
 $\rho_\pi \leftarrow \pi_C$
 $\rho_\delta \leftarrow \delta_C$

$\pi_m \leftarrow \lambda_\pi \wedge \rho_\pi$
 $\delta_m \leftarrow \lambda_\delta \wedge \rho_\delta$

IF $[(\pi_M - \pi_m > \sigma_\pi) \wedge (\delta_M - \delta_m > \sigma_\delta)]$

$n_P \leftarrow n_P - 1$

DO $(I = I_B : I_R)$

IF $(N[I] = 1)$

$\frac{P}{N}[I] \leftarrow \frac{n_P}{N}$

END (if)

END (do)

$\pi_M \leftarrow 0$
 $I_M \leftarrow 1$

DO $(i = 1 : 2^p)$

IF $\left[\left(\widehat{\pi_{R_\Delta}} [i] > \pi_M \right) \wedge (N[i] = 1) \right]$

$\pi_M \leftarrow \widehat{\pi_{R_\Delta}} [i]$

$I_M \leftarrow i$

END (if)

END (do)

$\delta_M \leftarrow \widehat{\delta_{R_\Delta}} [I_M]$

ELSE C $\leftarrow 0$

END (while)

On recommence l'opération à droite du maximum

On exige une différence d'au moins $\epsilon > 0$

Si la période qui vient d'être formée est assez marquée, on l'enregistre puis on actualise les valeurs

Recherche d'une nouvelle période à partir du "maximum suivant"

Localisation du pourcentage maximum

On n'examine plus les pourcentages relatifs aux périodes précédentes

Si non on arrête

6.4.2 Périodes

La démarche de segmentation ignore l'ordre chronologique: il est utile de numéroter les périodes dans cet ordre. Voici le code:

```
curr ← P[1]
next ← P[2]
PERD ← 1

P[1] ← PERD
IF (next = curr)
    P[2] ← PERD
ELSE
    PERD ← PERD + 1
    P[2] ← PERD
END (if)

NB ← 2p - 2
ind ← 3
WHILE (NB > 0)
    curr ← next
    next ← P[ind]
    IF (curr = next)
        P[ind] ← PERD
    ELSE
        PERD ← PERD + 1
        P[ind] ← PERD
    END (if)
    NB ← NB - 1
    ind ← ind + 1
END (while)
```

P Contient les indicateurs des périodes dans un ordre non chronologique

6.4.3 Spécialités médicales et pourcentages des remboursements associés aux périodes

Les périodes sont utiles si l'on sait leur associer "étroitement" des domaines médicaux et des montants de remboursements. On va donc calculer, par période, le pourcentage des remboursements attribuables à chaque domaine médical retenu (41 dans un premier temps). \underline{x} désigne le vecteur des valeurs

$$2^{p-1} + 0 \times 2^p, 2^{p-1} + 1 \times 2^p, 2^{p-1} + 2 \times 2^p, \dots, 2^{p-1} + (2^p - 1) \times 2^p.$$

Alors si I est un des indicateurs de période (apparaissant dans le vecteur \mathcal{P}), \underline{x}_I est le vecteur des éléments de \underline{x} qui ont la même position que les valeurs I de \mathcal{P} . Voici le code:

```

DO (I = 1 : perd)
    M [I] ← 2I - 1 | max (xJ)
END (do)
neg ← 0
J ← 1
WHILE (dayDay [J] < 0)
    nb ← nb + 1
    J ← J + 1
END (while)
low ← nb
DO (I = 1 : perd)
    right ← M [I]
    left ← low + 1
    WHILE ((left ≤ dayDay [J]) ∧ (dayDay [J] ≤ right))
        CASE {cdProv [J] = 0
            tot [1] ← amnt [J] + tot [1]
        CASE {cdProv [J] = 1
            tot [2] ← amnt [J] + tot [2]
        CASE ...
        CASE {cdProv [J] = 41
            tot [41] ← amnt [J] + tot [41]
        J ← J + 1
    END (while)
    IF (SUM [tot] > 0)
        π ← 100 ×  $\frac{tot}{SUM[tot]}$ 
    ELSE π = 0
    END (if)
    IF (I = 1)
        pcts ← [1 | π]
    ELSE
        tPcts ← [I | π]
        STACK pcts ONTO tPcts
    END (if)
    left → right
END (do)

```

"Date" du dernier point de la période

Elimination des jours de remboursements de la période qui précède celles que l'on vient de créer

cdProv est le code du domaine médical du prestataire
 J est une "date" de remboursement
 amnt est le montant remboursé

Il y a 41 domaines médicaux

I est un vecteur dont les composantes valent I: il y en a autant que d'indices I

«Aspects de la sécurité sociale»

Rapports de recherche et expertises selon le thème/programme

Fournisseur: Les rapports peuvent être obtenus en utilisant le numéro de commande (cf. N° de commande OFCL) à l'adresse suivante: Office fédéral des constructions et de la logistique OFCL, 3003 Berne ou par voie électronique en cliquant sur le numéro de commande.

Assurance-maladie / Analyse des effets de la LAMal

Nr. N° N°	Autor/inn/en, Titel auteur/s, titre Autore/i, titolo	Bestell-Nr. N° de commande N° di ordinazione
1/94	<i>Fischer, Wolfram (1994):</i> Möglichkeiten der Leistungsmessung in Krankenhäusern: Überlegungen zur Neugestaltung der schweizerischen Krankenhausstatistik.	318.010.1.94d
1/94	<i>Fischer, Wolfram (1994):</i> Possibilités de mesure des Prestations hospitalières: considérations sur une réorganisation de la statistique hospitalière.	318.010.1.94f
4/94	<i>Cranovsky, Richard (1994):</i> Machbarkeitsstudie des Technologiebewertungsregister.	318.010.4.94d
5/94	<i>BRAINS (1994):</i> Spitex-Inventar.	318.010.5.94d
5/94	<i>BRAINS (1994):</i> Inventaire du Spitex.	318.010.5.94f
1/97	<i>Fischer, Wolfram (1997):</i> Patientenklassifikationssysteme zur Bildung von Behandlungsfallgruppen im stationären Bereich.	318.010.1.97d
3/97	<i>Schmid, Heinz (1997):</i> Prämien genehmigung in der Krankenversicherung (Expertenbericht).	318.010.3.97d
3/97	<i>Schmid, Heinz (1997):</i> Procédure d'approbation des primes dans l'assurance-maladie (Expertise).	318.010.3.97f
6/97	<i>Latzel, Günther; Andermatt, Christoph; Walther, Rudolf (1997):</i> Sicherung und Finanzierung von Pflege- und Betreuungsleistungen bei Pflegebedürftigkeit. Band I und II.	318.010.6.97d
1/98	<i>Baur, Rita; Hunger, Wolfgang; Kämpf, Klaus; Stock, Johannes (1998):</i> Evaluation neuer Formen der Krankenversicherung. Synthesebericht.	318.010.1.98d
1/98	<i>Baur, Rita; Hunger, Wolfgang; Kämpf, Klaus; Stock, Johannes (1998):</i> Rapport de synthèse: Evaluation des nouveaux modèles d'assurance-maladie.	318.010.1.98f
2/98	<i>Baur, Rita; Eyett, Doris (1998):</i> Die Wahl der Versicherungsformen. Untersuchungsbericht 1.	318.010.2.98d
3/98	<i>Baur, Rita; Eyett, Doris (1998a):</i> Bewertung der ambulanten medizinischen Versorgung durch HMO-Versicherte und traditionell Versicherte. Untersuchungsbericht 2.	318.010.3.98d
4/98	<i>Baur, Rita; Eyett, Doris (1998b):</i> Selbstgetragene Gesundheitskosten. Untersuchungsbericht 3. (vergriffen)	318.010.4.98d
5/98	<i>Baur, Rita; Ming, Armin; Stock, Johannes; Lang, Peter (1998):</i> Struktur, Verfahren und Kosten der HMO-Praxen. Untersuchungsbericht 4.	318.010.5.98d
6/98	<i>Stock, Johannes; Baur, Rita; Lang, Peter; Conen, Dieter (1998):</i> Hypertonie-Management. Ein Praxisvergleich zwischen traditionellen Praxen und HMOs.	318.010.6.98d
7/98	<i>Schütz, Stefan et al. (1998):</i> Neue Formen der Krankenversicherung: Versicherte, Leistungen, Prämien und Kosten. Ergebnisse der Administrativdatenuntersuchung, 1. Teil.	318.010.7.98d
8/98	<i>Känzig, Herbert et al. (1998):</i> Neue Formen der Krankenversicherung: Alters- und Kostenverteilungen im Vergleich zu der traditionellen Versicherung. Ergebnisse der Administrativdatenuntersuchung, 2. Teil.	318.010.8.98d
9/98	<i>Sottas, Gabriel et al. (1998):</i> Données administratives de l'assurance-maladie: Analyse de qualité, statistique élémentaire et base pour les exploitations.	318.010.9.98f
15/98	<i>Greppi, Spartaco, Rossel, Raymond, Strüwe, Wolfram (1998):</i> Der Einfluss des neuen Krankenversicherungsgesetzes auf die Finanzierung des Gesundheitswesens.	318.010.15.98d
15/98	<i>Greppi, Spartaco; Rossel, Raymond; Strüwe, Wolfram (1998):</i> Les effets de la nouvelle loi sur l'assurance-maladie dans le financement du système de santé.	318.010.15.98f
21/98	<i>Balthasar, Andreas (1998):</i> Die sozialpolitische Wirksamkeit der Prämienverbilligung in den Kantonen.	318.010.21.98d
21/98	<i>Balthasar, Andreas (1998):</i> Efficacité sociopolitique de la réduction de primes dans les cantons.	318.010.21.98f
1/99	<i>Spycher, Stefan (1999):</i> Wirkungsanalyse des Risikoausgleichs in der Krankenversicherung.	318.010.1.99d
2/99	Kurzfassung von Nr. 1/99.	318.010.2.99d
2/99	Condensé du n° 1/99.	318.010.2.99f

3/99	<i>Institut de santé et d'économie ISE en collaboration avec l'Institut du Droit de la Santé IDS (1999):</i> Un carnet de santé en Suisse? Etude d'opportunité.	318.010.3.99f
4/99	<i>Faisst, Karin; Schilling, Julian (1999):</i> Inhaltsanalyse von Anfragen bei PatientInnen- und Versichertenorganisationen.	318.010.4.99d
10/99	<i>Faisst, Karin; Schilling, Julian (1999):</i> Qualitätssicherung – Bestandesaufnahme.	318.010.10.99d
3/00	<i>Spycher, Stefan (2000):</i> Reform des Risikoausgleichs in der Krankenversicherung? Studie 2: Empirische Prüfung von Vorschlägen zur Optimierung der heutigen Ausgestaltung.	318.010.3.00d
4/00	<i>Stürmer, Wilhelmine; Wendland, Daniela; Braun, Ulrike (2000):</i> Veränderungen im Bereich der Zusatzversicherung aufgrund des KVG.	318.010.4.00d
5/00	<i>Greppi, Spartaco; Ritzmann, Heiner; Rossel, Raymond; Siffert, Nicolas (2000):</i> Analyse der Auswirkungen des KVG auf die Finanzierung des Gesundheitswesens und anderer Systeme der sozialen Sicherheit.	318.010.5.00d
5/00	<i>Greppi, Spartaco; Ritzmann, Heiner; Rossel, Raymond; Siffert, Nicolas (2000):</i> Analyse des effets de la LAMal dans le financement du système de santé et d'autres régimes de protection sociale.	318.010.5.00f
6/00	<i>Bundesamt für Sozialversicherung (2000): Arbeitstagung des Eidg. Departement des Innern:</i> Massnahmen des KVG zur Kostendämpfung/ La LAMal, instrument de maîtrise des coûts/ Misure della LAMal per il contenimento dei costi.	318.010.6.00
7/00	<i>Hammer, Stephan (2000):</i> Auswirkungen des KVG im Tarifbereich.	318.010.7.00d
11/00	<i>Spycher, Stefan; Leu, Robert E. (2000):</i> Finanzierungsalternativen in der obligatorischen Krankenpflegeversicherung.	318.010.11.00d
12/00	<i>Polikowski, Marc; Lauffer, Régine; Renard, Delphine; Santos-Eggimann, Brigitte (2000):</i> Analyse des effets de la LAMal: Le «catalogue des prestations» est-il suffisant pour que tous accèdent à des soins de qualité?	318.010.12.00f
14/00	<i>Ayer, Ariane; Despland, Béatrice; Sprumont, Dominique (2000):</i> Analyse juridique des effets de la LAMal: Catalogue des prestations et procédures.	318.010.14.00f
15/00	<i>Baur, Rita; Braun, Ulrike (2000):</i> Bestandsaufnahme besonderer Versicherungsformen in der obligatorischen Krankenversicherung.	318.010.15.00d
2/01	<i>Balthasar, Andreas (2001):</i> Die Sozialpolitische Wirksamkeit der Prämienverbilligung in den Kantonen: Monitoring 2000.	318.010.2.01d
2/01	<i>Balthasar, Andreas (2001):</i> Efficacité sociopolitique de la réduction de primes dans les cantons.	318.010.2.01f
3/01	<i>Peters, Matthias; Müller, Verena; Luthiger, Philipp (2001):</i> Auswirkungen des Krankenversicherungsgesetzes auf die Versicherten.	318.010.3.01d
4/01	<i>Baur, Rita; Heimer, Andreas (2001):</i> Wirkungsanalyse KVG: Information der Versicherten.	318.010.4.01d
5/01	<i>Balthasar, Andreas; Bieri, Oliver; Furrer, Cornelia (2001):</i> Evaluation des Vollzugs der Prämienverbilligung.	318.010.5.01d
5/01	<i>Balthasar, Andreas; Bieri, Oliver; Furrer, Cornelia (2001):</i> Evaluation de l'application de la réduction de primes.	318.010.5.01f
6/01	<i>Hammer, Stephan; Pulli, Raffael; Iten, Rolf; Eggimann, Jean-Claude (2001):</i> Auswirkungen des KVG auf die Versicherer.	318.010.6.01d
7/01	<i>Bundesamt für Sozialversicherung (2001):</i> Persönlichkeitsschutz in der sozialen und privaten Kranken- und Unfallversicherung (Expertenbericht).	318.010.7.01d
7/01	<i>Office fédéral des assurances sociales (2001):</i> Protection de la personnalité dans l'assurance- maladie et accidents sociale et privée (Rapport d'experts).	318.010.7.01f
8/01	<i>Hammer, Stephan; Pulli, Raffael; Schmidt, Nicolas; Iten, Rolf; Eggimann, Jean-Claude (2001):</i> Auswirkungen des KVG auf die Leistungserbringer.	318.010.8.01d
9/01	<i>Battaglia, Markus; Junker, Christoph (2001):</i> Auswirkungen der Aufnahme von präventiv- medizinischen Leistungen in den Pflichtleistungskatalog, Teilbericht Impfungen im Schulalter.	318.010.9.01d
10/01	<i>Sager, Fritz; Rüeßli, Christian; Vatter, Adrian (2001):</i> Auswirkungen der Aufnahme von präventiv- medizinischen Leistungen in den Pflichtleistungskatalog. Politologische Analyse auf der Grundlage von drei Fallbeispielen.	318.010.10.01d
11/01	<i>Faisst, Karin; Fischer, Susanne; Schilling, Julian (2001):</i> Monitoring 2000 von Anfragen an PatientInnen- und Versichertenorganisationen.	318.010.11.01d
12/01	<i>Hornung, Daniel; Röthlisberger, Thomas; Stiefel, Adrian (2001):</i> Praxis der Versicherer bei der Vergütung von Leistungen nach KVG.	318.010.12.01d
13/01	<i>Haari, Roland; Schilling, Karl (2001):</i> Kosten neuer Leistungen im KVG. Folgerungen aus der Analyse der Anträge für neue Leistungen und Unterlagen des BSV aus den Jahren 1996-1998.	318.010.13.01d

14/01	<i>Rüefli, Christian; Vatter, Adrian (2001):</i> Kostendifferenzen im Gesundheitswesen zwischen den Kantonen. Statistische Analyse kantonaler Indikatoren.	318.010.14.01d
14/01	<i>Rüefli, Christian; Vatter, Adrian (2001):</i> Les différences intercantionales en matière de coûts de la santé. Analyse statistique d'indicateurs cantonaux.	318.010.14.01f
15/01	<i>Haari, Roland et al. (2001):</i> Kostendifferenzen zwischen den Kantonen. Sozialwissenschaftliche Analyse kantonaler Politiken.	318.010.15.01d
16/01	<i>Bundesamt für Sozialversicherung (2001):</i> Wirkungsanalyse KVG, Synthesebericht.	318.010.16.01d
16/01	<i>Office fédéral des assurances sociales (2001):</i> Analyse des effets de la LAMal, Rapport de synthèse.	318.010.16.01f
2/02	<i>Zellweger, Ueli; Faisst, Karin (2002):</i> Monitoring 2001 von Anfragen an PatientInnen- und Versichertenorganisationen.	318.010.2.02d
3/02	<i>Matenza, Guido et al. (2002):</i> Stationen im Prozess der Anerkennung der psychologischen Psychotherapie.	318.010.3.02d
14/03	<i>Schönenberger, Urs; Bestetti, Gilberto (2003):</i> Telemedizinische Verfahren: Auf dem Weg zum Standard.	318.010.14/03d
14/03	<i>Schönenberger, Urs; Bestetti, Gilberto (2003):</i> Les procédés de télémedecine: sur la voie de la standardisation.	318.010.14/03f
18/03	<i>Gualtierotti, Antoine (2003):</i> Exploitation de la base de données de la LAMal : rapport relatif à un mandat.	318.010.18/03f

Invalidité / handicap

Nr. N° N°	Autor/inn/en, Titel auteur/s, titre Autore/i, titolo	Bestell-Nr. N° de commande N° di ordinazione
6/99	<i>Bachmann, Ruth; Furrer, Cornelia (1999):</i> Die ärztliche Beurteilung und ihre Bedeutung im Entscheidungsverfahren über einen Rentenanspruch in der Eidg. Invalidenversicherung.	318.010.6.99d
7/99	<i>Prinz, Christopher (1999):</i> Invalidenversicherung: Europäische Entwicklungstendenzen zur Invalidität im Erwerbsalter. Band 1 (Vergleichende Synthese).	318.010.7.99d
8/99	<i>Prinz, Christopher (1999):</i> Invalidenversicherung: Europäische Entwicklungstendenzen zur Invalidität im Erwerbsalter. Band 2 (Länderprofile).	318.010.8.99d
10/00	<i>Aarts, Leo; de Jong, Philipp; Prinz, Christopher (2000):</i> Determinanten der Inanspruchnahme einer Invalidenrente – Eine Literaturstudie.	318.010.10.00d

Prévoyance vieillesse / prévoyance professionnelle

Nr. N° N°	Autor/inn/en, Titel auteur/s, titre Autore/i, titolo	Bestell-Nr. N° de commande N° di ordinazione
2/94	<i>Bender, André; Favarger, M. Philippe; Hoesli, Martin (1994):</i> Evaluation des biens immobiliers dans les institutions de prévoyance.	318.010.2.94f
3/94	<i>Wüest, Hannes; Hofer, Martin; Schweizer, Markus (1994):</i> Wohneigentumsförderung – Bericht über die Auswirkungen der Wohneigentumsförderung mit den Mitteln der beruflichen Vorsorge.	318.010.3.94d
1/95	<i>van Dam, Jacob; Schmid, Hans (1995):</i> Insolvenzversicherung in der beruflichen Vorsorge.	318.010.1.95d
3/96	<i>Bundesamt für Sozialversicherung (1996):</i> Berufliche Vorsorge: Neue Rechnungslegungs- und Anlagevorschriften. Regelung des Einsatzes der derivativen Finanzinstrumente.	318.010.3.96d
3/96	<i>Office fédéral des assurances sociales (1996):</i> Prévoyance professionnelle: Nouvelles prescriptions en matière d'établissement des comtes et de placements. Réglementation concernant l'utilisation des instruments financiers dérivés.	318.010.3.96f
3/96	<i>Ufficio federale delle assicurazioni sociali (1996):</i> Previdenza professionale : Nuove prescrizioni in materia di rendconto e di investimenti. Regolamentazione concernente l'impiego di strumenti finanziari derivati.	318.010.3.96i
4/96	<i>Wechsler, Martin; Savioz, Martin (1996):</i> Umverteilung zwischen den Generationen in der Sozialversicherung und im Gesundheitswesen.	318.010.4.96d
2/97	<i>Infras (1997):</i> Festsetzung der Renten beim Altersrücktritt und ihre Anpassung an die wirtschaftliche Entwicklung. Überblick über die Regelungen in der EU.	318.010.2.97d
12/98	<i>Spycher, Stefan (1998):</i> Auswirkungen von Leistungsveränderungen bei der Witwenrente. Im Auftrag der IDA FiSo 2.	318.010.12.98d

16/98	<i>Bundesamt für Sozialversicherung (1998):</i> Forum 1998 über das Rentenalter/ sur l'âge de la retraite (Tagungsband).	318.010.16.98
18/98	<i>Koller, Thomas (1998):</i> Begünstigtenordnung in der zweiten und dritten Säule (Gutachten).	318.010.18.98d
18/98	<i>Koller, Thomas (1998):</i> L'ordre des bénéficiaires des deuxième et troisième piliers (Expertise).	318.010.18.98f
19/98	<i>INFRAS (1998):</i> Mikroökonomische Effekte der 1. BVG-Revision. (vergriffen)	318.010.19.98d
19/98	<i>INFRAS (1998):</i> Effets microéconomiques de la 1 ^{re} révision de la LPP. Rapport final.	318.010.19.98f
20/98	<i>KOF/ETHZ (1998):</i> Makroökonomische Effekte der 1. BVG-Revision. Schlussbericht.	318.010.20.98d
20/98	<i>KOF/ETHZ (1998):</i> Effets macroéconomiques de la 1 ^{re} révision de la LPP. Rapport final.	318.010.20.98f
2/00	<i>PRASA (2000):</i> Freie Wahl der Pensionskasse: Teilbericht.	318.010.2.00d
9/00	<i>Schneider, Jacques-André (2000):</i> A-propos des normes comptables IAS 19 et FER/RPC 16 e de la prévoyance professionnelle en suisse (Expertise).	318.010.9.00f
1/01	<i>Gognalons-Nicolet, Maryvonne; Le Goff, Jean-Marie (2001):</i> Retraits anticipés du marché du travail avant l'âge AVS: un défi pour les politiques de retraite en Suisse.	318.010.1.01f
17/01	<i>Bundesamt für Sozialversicherung (2001):</i> Zwei Berichte zum Thema Minimalzinsvorschriften für Vorsorgeeinrichtungen. Hauptbericht: Über die Möglichkeit, bei den Minimalzinsvorschriften für Vorsorgeeinrichtungen auf Real- statt Nominalzinsen abzustellen. Ergänzender Bericht: Über den Aspekt der Lebensversicherer im Problemkreis Minimalzinsvorschriften gemäss BVG.	318.010.17.01d
17/01	<i>Office fédéral des assurances sociales (2001):</i> Deux rapports sur le thème prescriptions de taux minimaux pour les institutions de prévoyance. Rapport principal : sur la possibilité de se fonder sur les taux d'intérêts réels et non sur les taux nominaux pour fixer les prescriptions de taux minimaux pour les institutions de prévoyance. Rapport complémentaire : sur l'aspect de la problématique de la réglementation du taux d'intérêts minimal LPP du point de vue des assureurs-vie.	318.010.17.01f
1/03	<i>Antille Gaillard, Gabrielle ; Bilger, Marcel ; Candolfi, Pascal ; Chaze, Jean-Paul ; Flückiger, Yves (2003):</i> Analyse des déterminants individuels et institutionnels du départ anticipé à la retraite. (IDA ForAlt)	318.010.1/03f
2/03	<i>Balthasar, Andreas; Bieri, Olivier; Grau, Peter; Künzi, Kilian; Guggisberg Jürg (2003):</i> Der Übergang in den Ruhestand - Wege, Einflussfaktoren und Konsequenzen. (IDA ForAlt)	318.010.2/03d
2/03	<i>Balthasar, Andreas; Bieri, Olivier; Grau, Peter; Künzi, Kilian; Guggisberg Jürg (2003):</i> Le passage à la retraite: trajectoires, facteurs d'influence et conséquences. (IDA ForAlt)	318.010.2/03f
3/03	<i>Bonoli, Giuliano ; Gay-des-Combes, Benoît (2003):</i> L'évolution des prestations vieillesse dans le long terme : une simulation prospective de la couverture retraite à l'horizon 2040. (IDA ForAlt)	318.010.3/03f
4/03	<i>Jans, Armin; Hammer, Stefan; Graf, Silvio ; Iten Rolf ; Maag, Ueli ; Schmidt, Nicolas; Weiss Sampietro, Thea (2003):</i> Betriebliche Alterspolitik – Praxis in den Neunziger Jahren und Perspektiven. (IDA ForAlt)	318.010.4/03d
4/03.1	<i>Graf, Silvio; Jans, Armin; Weiss Sampietro, Thea (2003):</i> Betriebliche Alterspolitik – Unternehmens- und Personenbefragung. Beilageband I. (IDA ForAlt)	318.010.4/03.1d
4/03.2	<i>Hammer Stefan ; Maag, Ueli; Schmidt, Nicolas (2003):</i> Betriebliche Alterspolitik – Fallstudien. Beilageband II. (IDA ForAlt)	318.010.4/03.2d
5/03	<i>Fux, Beat (2003):</i> Entwicklung des Potentials erhöhter Arbeitsmarktpartizipation von Frauen nach Massgabe von Prognosen über die Haushalts- und Familienstrukturen. (IDA ForAlt)	318.010.5/03d
6/03	<i>Baumgartner, Doris A. (2003):</i> Frauen in mittleren Erwerbsalter. Eine Studie über das Potenzial erhöhter Arbeitsmarktpartizipation von Frauen zwischen 40 und 65. (IDA ForAlt)	318.010.6/03d
7/03	<i>Wanner, Philippe ; Gabadinho, Alexis ; Ferrari, Antonella (2003):</i> La participation des femmes au marché du travail. (IDA ForAlt)	318.010.7/03f
8/03	<i>Wanner, Philippe ; Stuckelberger, Astrid ; Gabadinho, Alexis (2003):</i> Facteurs individuels motivant le calendrier du départ à la retraite des hommes âgés de plus de 50 ans en Suisse. (IDA ForAlt)	318.010.8/03f
9/03	<i>Widmer, Rolf ; Mühleisen, Sybille; Falta, Roman, P.; Schmid, Hans (2003):</i> Bestandesaufnahme und Interaktionen institutioneller Regelungen beim Rentenantritt. (IDA ForAlt)	318.010.9/03d
10/03	<i>Schluemp, Kurt (2003):</i> Finanzierungsbedarf in der AHV (inkl. EL). (IDA ForAlt)	318.010.10/03d
10/03	<i>Schluemp, Kurt (2003):</i> Besoins de financement de l'AVS (PC comprises). (IDA ForAlt)	318.010.10/03f
11/03	<i>Müller, André; van Nieuwkoop, Renger; Lieb, Christoph (2003):</i> Analyse der Finanzierungsquellen für die AHV. SWISSOLG – ein Overlapping Generations Model für die Schweiz. (IDA ForAlt)	318.010.11/03d

12/03	<i>Abrahamsen, Yngve; Hartwig, Jochen (2003): Volkswirtschaftliche Auswirkungen verschiedener Demographieszennarien und Varianten zur langfristigen Finanzierung der Alterssicherung in der Schweiz. (IDA ForAlt)</i>	318.010.12/03d
13/03	<i>Interdepartementale Arbeitsgruppe (2003): Synthesebericht zum Forschungsprogramm zur längerfristigen Zukunft der Alterssicherung. (IDA ForAlt)</i>	318.010.13/03d
13/03	<i>Groupe de travail interdépartemental (2003) : Rapport de synthèse du Programme de recherche sur l'avenir à long terme de la prévoyance vieillesse. (IDA ForAlt)</i>	318.010.13/03f
13/03	<i>Gruppo di lavoro interdipartimentale (2003) : Rapporto di sintesi del Programma di ricerca sul futuro a lungo termine della previdenza per la vecchiaia.(IDA ForAlt)</i>	318.010.13/03i

Politique sociale, questions familiales et économie

Nr. N° N°	Autor/inn/en, Titel auteur/s, titre Autore/i, titolo	Bestell-Nr. N° de commande N° di ordinazione
2/95	<i>Bauer, Tobias (1995): Literaturrecherche: Modelle zu einem garantierten Mindesteinkommen.</i>	318.010.2.95d
3/95	<i>Farago, Peter (1995): Verhütung und Bekämpfung der Armut: Möglichkeiten und Grenzen staatlicher Massnahmen.</i>	318.010.3.95d
3/95	<i>Farago, Peter (1995) : Prévenir et combattre la pauvreté: forces et limites des mesures prises par l'Etat.</i>	318.010.3.95f
1/96	<i>Cardia-Vonèche, Laura et al. (1996): Familien mit alleinerziehenden Eltern.</i>	318.010.1.96d
1/96	<i>Cardia-Vonèche, Laura et al. (1996): Les familles monoparentales.</i>	318.010.1.96f
4/97	<i>IPSO und Infrac (1997): Perspektive der Erwerbs- und Lohnquote.</i>	318.010.4.97d
5/97	<i>Spycher, Stefan (1997): Auswirkungen von Regelungen des AHV-Rentenalters auf die Sozialversicherungen, den Staatshaushalt und die Wirtschaft.</i>	318.010.5.97d
10/98	<i>Bauer, Tobias (1998): Kinder, Zeit und Geld. Eine Analyse der durch Kinder bewirkten finanziellen und zeitlichen Belastungen von Familien und der staatlichen Unterstützungsleistungen in der Schweiz Mitte der Neunziger Jahre.</i>	318.010.10.98d
11/98	<i>Bauer, Tobias (1998a): Auswirkungen von Leistungsveränderungen bei der Arbeitslosenversicherung. Im Auftrag der IDA FiSo 2.</i>	318.010.11.98d
13/98	<i>Müller, André; Walter, Felix; van Nieuwkoop, Renger; Felder, Stefan (1998): Wirtschaftliche Auswirkungen von Reformen der Sozialversicherungen. DYNASWISS – Dynamisches allgemeines Gleichgewichtsmodell für die Schweiz. Im Auftrag der IDA FiSo 2.</i>	318.010.13.98d
14/98	<i>Mauch, S.P., Iten, R., Banfi, S., Bonato, D., von Stokar, T., Schips, B., Abrahamsen, Y. (1998): Wirtschaftliche Auswirkungen von Reformen der Sozialversicherungen. Schlussbericht der Arbeitsgemeinschaft INFRAS/KOF. Im Auftrag der IDA FiSo 2.</i>	318.010.14.98d
17/98	<i>Leu, Robert E.; Burri, Stefan; Aregger, Peter (1998): Armut und Lebensbedingungen im Alter.</i>	318.010.17.98d
5/99	<i>Bundesamt für Sozialversicherung (1999): Bedarfsleistungen an Eltern (Tagungsband).</i>	318.010.5.99d
9/99	<i>OECD (1999): Bekämpfung sozialer Ausgrenzung. Band 3. Sozialhilfe in Kanada und in der Schweiz.</i>	318.010.9.99d
1/00	<i>Ecoplan (2000): Neue Finanzordnung mit ökologischen Anreizen: Entlastung über Lohn- und MWST-Prozente?</i>	318.010.1.00d
8/00	<i>Sterchi, Beat; Egger, Marcel; Merckx, Véronique (2000): Faisabilité d'un «chèque-service».</i>	318.010.8.00f
13/00	<i>Wyss, Kurt (2000): Entwicklungstendenzen bei Integrationsmassnahmen der Sozialhilfe.</i>	318.010.13.00d
13/00	<i>Wyss, Kurt (2000): Évolution des mesures d'intégration de l'aide sociale.</i>	318.010.13.00f
1/02	<i>Schiffbänker, Annemarie; Thenner, Monika; Immervoll, Herwig (2001): Familienlastenausgleich im internationalen Vergleich. Eine Literaturstudie.</i>	318.010.1.02d
4/02	<i>Soland, Rita; Stern, Susanne; Steinemann, Myriam; Iten, Rolf (2002): Zertifizierung familienpolitischer Unternehmen in der Schweiz.</i>	318.010.4.02d

Perspectives et développement de la sécurité sociale

Nr. N° N°	Autor/inn/en, Titel auteur/s, titre Autore/i, titolo	Bestell-Nr. N° de commande N° di ordinazione
10/95	<i>Eidg. Departement des Innern (1995): Bericht des Eidgenössischen Departementes des Innern zur heutigen Ausgestaltung und Weiterentwicklung der schweizerischen 3-Säulen-Konzeption der Alters-, Hinterlassenen- und Invalidenvorsorge</i>	318.012.1.95d
10/95	<i>Département fédéral de l'intérieur (1995) : Rapport du Département fédéral de l'intérieur concernant la structure actuelle et le développement futur de la conception helvétique des trois piliers de la prévoyance vieillesse, survivants et invalidité.</i>	318.012.1.95f
10/95	<i>Dipartimento federale dell'interno (1995) : Rapporto del Dipartimento federale dell'interno concernente la struttura attuale e l'evoluzione futura della concezione svizzera delle tre pilastri de la previdenza per la vecchiaia, i superstiti e l'invalidità.</i>	318.012.1.95i
1/96	<i>Interdepartementale Arbeitsgruppe „Finanzierungsperspektiven der Sozialversicherungen“ (IDA FiSo 1) (1996): Bericht über die Finanzierungsperspektiven der Sozialversicherungen (unter besonderer Berücksichtigung der demographischen Entwicklung).</i>	318.012.1.96d
1/96	<i>Groupe de travail interdépartemental « Perspectives de financement des assurances sociales » (IDA FiSo 1) (1996) : Rapport sur les perspectives de financement des assurances sociale (en regard en particulier à l'évolution démographique).</i>	318.012.1.96f
1/97	<i>Interdepartementale Arbeitsgruppe „Finanzierungsperspektiven der Sozialversicherungen“ (IDA FiSo 2) (1997): Analyse der Leistungen der Sozialversicherungen; Konkretisierung möglicher Veränderungen für drei Finanzierungsszenarien.</i>	318.012.1.97d
1/97	<i>Groupe de travail interdépartemental « Perspectives de financement des assurances sociales » (IDA FiSo 1) (1997) : Analyse des prestations des assurances sociales ; Concrétisation de modifications possibles en fonction de trois scénarios financiers.</i>	318.012.1.97f
15/03	<i>Stutz, Heidi; Bauer, Tobias (2003): Modelle zu einem garantierten Mindesteinkommen. Sozialpolitische und ökonomische Auswirkungen.</i>	318.010.15/03d
15/03	<i>Stutz, Heidi; Bauer, Tobias (2003): Modèles de revenu minimum garanti. Effets sociopolitiques et économiques.</i>	318.010.15/03f
15/03	<i>Stutz, Heidi ; Bauer, Tobias (2003) : Modelli relativi a un reddito minimo garantito. Ripercussioni sociopolitiche ed economiche.</i>	318.010.15/03i