

Méthode d'appariement de l'EPAS et de l'ESPS

Mise à jour

Anne Aligon
Nathalie Grandfils
Sophie Lebreton

Rapport n° 451 (biblio n° 1157bis)

Janvier 1997

I.S.B.N. : 87812-195-3

- Remerciements -

Les auteurs remercient :

- Mme GRANGE du CNUSC (centre national universitaire sud de calcul), et Catherine ORDONNEAU du CREDES pour leur aide dans l'exploitation informatique des données,*
- M. FERRAGU, directeur du département statistique de la CNAMTS, Me MERLIERE et M. HOLLA, également de la CNAMTS, pour leur relecture attentive de ce travail et leurs remarques.*
- Catherine SERMET, Andrée et Arié MIZRAHI, du CREDES, pour leurs commentaires et critiques sur ce travail,*
- Patricia DARFEUILLE pour la mise en page de ce rapport.*

- Avertissement -

Le CREDES mène depuis 1988, en collaboration avec le Département Statistique de la CNAMTS¹, une enquête sur la santé et la protection sociale (ESPS), auprès d'assurés sociaux et de leurs familles ; depuis 1994, avec l'aide du service statistique de la CANAM², l'enquête a été étendue aux professions indépendantes ; en 1996, avec l'aide de la MSA³, l'enquête s'est élargie aux assurés des régimes agricoles.

Ont participé au terrain de l'enquête en 1992 :

- l'institut de sondage ISL ainsi que l'IFOP pour les contacts avec les enquêtés
- l'Institut Panel Etudes pour le chiffrage médical.

Au CREDES, l'organisation du chiffrage et plus généralement de l'enquête, a été réalisée par Nathalie Grandfils et Philippe Le Fur. Catherine Ordonneau et Nathalie Grandfils ont constitué les bases de données relatives à l'enquête SPS. Le travail informatique a été réalisé par Sophie Lebreton, aidée pour certaines parties d'Anne Aligon.

Le travail présenté dans ce rapport met en évidence les difficultés rencontrées lors de la fusion de deux sources de données de natures différentes mais relatives aux mêmes individus : l'une administrative, l'autre issue d'une enquête en population générale.

Ces deux types de données comportent des incertitudes d'origines diverses :

- * l'EPAS (l'Echantillon Permanent d'Assurés Sociaux), sous-échantillon des fichiers de remboursement de l'Assurance Maladie, comporte un certain nombre de données utiles au remboursement mais non conformes à ce que l'on peut observer dans l'enquête : il s'agit par exemple de déménagements, de décès ou de changements de régimes d'Assurance maladie. Ce fichier, comme tout fichier administratif, n'est pas conçu à l'origine pour être utilisé dans le cadre de travaux de recherche.
- les fichiers de l'ESPS contiennent des données recueillies auprès des ménages résidant en France. Comme dans toute enquête, les résultats obtenus sont entachés de différents biais ; ces biais peuvent être dus à des omissions, volontaires ou non, de la part des enquêtés. Signalons également qu'un certain nombre de ménages refusent d'emblée de participer à l'enquête, introduisant alors un biais dans la population répondant à l'enquête. Enfin, la réponse à certaines questions, et particulièrement à celles touchant à la réglementation de l'Assurance Maladie sont à prendre avec précaution car il s'agit de domaines mal connus du grand public ; il s'agit par exemple de la situation administrative au regard de l'Assurance Maladie ("êtes-vous assuré ou ayant droit ? »)

La fusion de ces deux types de données, du fait de leurs incertitudes respectives mais aussi de leur complémentarité, constitue une richesse importante pour la recherche. Les difficultés rencontrées dans ce travail et détaillées dans le rapport qui suit devraient permettre d'une part d'apprécier la qualité des données recueillies dans l'enquête, d'autre part de mieux comprendre le contenu des fichiers administratifs.

¹ Caisse Nationale d'Assurance Maladie des Travailleurs Salariés

² Caisse Nationale d'assurance Maladie et Maternité des Professions Indépendantes

³ Mutualité Sociale Agricole

SOMMAIRE

SOMMAIRE	7
INTRODUCTION	1
PRÉSENTATION DES DONNÉES ET DE LA PROCÉDURE D'APPARIEMENT	5
1.1 L'ECHANTILLON PERMANENT D'ASSURÉS SOCIAUX DE LA CAISSE NATIONALE D'ASSURANCE MALADIE DES TRAVAILLEURS SALARIÉS (EPAS)	7
1.2 L'ENQUÊTE SANTÉ ET PROTECTION SOCIALE (ESPS)	8
1.3 LA PROCÉDURE D'ANONYMAT	9
LA MÉTHODE D'APPARIEMENT	11
INTRODUCTION	13
2.1 APPARIEMENT DES PERSONNES PROTÉGÉES	13
<i>1 - Les deux fichiers comportent des individus communs, mais aussi des individus propres à chacun des fichiers :</i>	<i>13</i>
2.2 APPARIEMENT DES PRESTATIONS	20
<i>2.2.1 Fusion des fichiers de prestations avec la totalité du fichier des personnes protégées de l'EPAS</i>	<i>21</i>
<i>2.2.2 Appariement des fichiers de prestations avec les 4 906 personnes communes à l'EPAS et à l'ESPS</i>	<i>22</i>
<i>2.2.3 Répartition des prestataires selon le type de prestations</i>	<i>26</i>
3.1 MÉTHODE DE REDRESSEMENT	31
3.2 POPULATION DE RÉFÉRENCE	32
3.3 CHOIX DES VARIABLES DE RÉFÉRENCE POUR LE REDRESSEMENT	33
3.4 CHOIX DE LA FONCTION DE DISTANCE	34
ANNEXE 1 RECHERCHE DE L'IDENTIFIANT À UTILISER POUR RÉALISER L'APPARIEMENT DES PERSONNES PROTÉGÉES DE L'EPAS AVEC LES ENQUÊTÉS DE L'ESPS (DÉMARCHE CHRONOLOGIQUE)	36
I DESCRIPTION DES FICHIERS INITIAUX	38
<i>I.1 Le fichier des personnes enquêtées dans l'ESPS</i>	<i>38</i>
<i>I.2 Le fichier ménages enquêtés dans l'ESPS</i>	<i>38</i>
<i>I.3 Le fichier des personnes protégées de l'EPAS</i>	<i>38</i>
<i>I.4 Le fichier des adresses à enquêter de l'ESPS</i>	<i>39</i>
II IDENTIFICATION DES ASSURÉS PRINCIPAUX (AP)	39
III IDENTIFICATION DES AYANTS DROIT	40
IV - EST-IL POSSIBLE D'AVOIR LE MEME IDENTIFIANT (N1 + ANNÉE DE NAISSANCE + MOIS DE NAISSANCE) POUR DES INDIVIDUS DIFFÉRENTS ?	41
<p>On dispose de deux dates de naissance, la seconde étant, selon des informations recueillies au département statistique de la CNAMTS, celle à considérer. Celle-ci présente toutefois de nombreuses valeurs manquantes ; afin d'améliorer les résultats, ces valeurs manquantes ont été remplacées par la première date de naissance. ..</p>	
V - FUSION DU FICHIER DES PERSONNES PROTÉGÉES (EPAS) AVEC LE FICHIER PERSONNE (ESPS) EN UTILISANT L'IDENTIFIANT PRÉCÉDEMMENT DÉFINI	42
ANNEXE 2 REDRESSEMENT DES DONNÉES	47

GLOSSAIRE.....	59
LISTE DES TABLEAUX ET DES SCHÉMAS	63

Introduction

Chaque année, le CREDES (Centre de Recherche, d'Etudes et de Documentation en Economie de la Santé) réalise l'Enquête Santé et Protection Sociale (ESPS) dont l'un des buts principaux est de compléter les données de remboursement issues de l'Echantillon Permanent d'Assurés Sociaux (EPAS) de la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAMTS).

L'EPAS a été élaboré il y a une quinzaine d'années et résulte d'une collaboration entre la CNAMTS et la Division d'Economie Médicale du CREDOC, aujourd'hui devenue CREDES. La base de sondage de cet échantillon est constituée par l'ensemble des assurés sociaux et de leurs ayants droit affiliés au Régime Général. Les assurés faisant partie de l'échantillon satisfont aux trois critères suivants : leur année de naissance est paire, leur mois de naissance est octobre et les deux derniers chiffres du numéro de leur commune de naissance sont identiques aux deux derniers chiffres de leur numéro d'ordre et ce, dans le même ordre ou dans l'ordre inverse. Tous les ayants droit des assurés sélectionnés sont inclus dans l'échantillon. Cette méthode de tirage assure une représentativité au 1/1200ème environ et permet une mise à jour permanente de l'échantillon par le biais des entrants et des sortants. Cet échantillon contient des renseignements administratifs ainsi que toutes les données de prestations relatives aux personnes le constituant.

L'enquête SPS a été mise en place pour compléter les informations de l'EPAS par des données individuelles, telles que des renseignements sociodémographiques, des données de morbidité ou encore des données de protection sociale (notamment le niveau de couverture complémentaire). Le CREDES enquête chaque année un quart de l'EPAS réparti en deux vagues, une vague au printemps, l'autre à l'automne.

L'objet du présent travail est de présenter la méthodologie utilisée pour réaliser l'appariement de l'EPAS avec les données recueillies dans l'enquête SPS. La première partie présente les différents fichiers utilisés dans ce travail. La deuxième partie présente les différentes méthodes successivement testées pour réaliser l'appariement des données et le bilan final ; enfin, la troisième partie détaille le redressement mis en place pour l'exploitation des données. Les chiffres qui figurent dans ce rapport sont ceux observés sur l'année 1992.

Pour les définitions de certains termes, voir le glossaire page 59

Présentation des données et de la procédure d'appariement

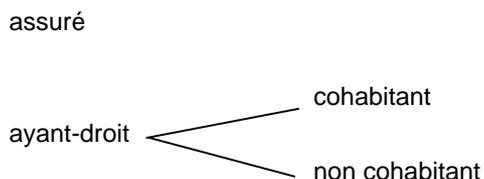
1.1 L'ECHANTILLON PERMANENT D'ASSURES SOCIAUX DE LA CAISSE NATIONALE D'ASSURANCE MALADIE DES TRAVAILLEURS SALARIES (EPAS)

Depuis une quinzaine d'années, la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés dispose d'un instrument de suivi des dépenses de l'Assurance Maladie grâce à l'Echantillon Permanent d'Assurés Sociaux (EPAS), échantillon alimenté par le CENTI (Centre National Du Traitement de l'Information). Cet échantillon est représentatif au 1/1200^eme des assurés au Régime Général et de leurs ayants droit⁴. Il est constitué de deux types de données :

- le fichier des personnes protégées
- les fichiers de prestations

Le **fichier des personnes protégées** contient des informations se rapportant aux personnes protégées par l'Assurance Maladie. Les personnes protégées comprennent les assurés et leurs ayants droit ; on parle de "grappes assurés" (cf. Schéma n° 1). On dispose dans ce fichier de renseignements sociodémographiques tels que le sexe, la date de naissance, et de données relatives à la situation des individus à l'égard de la réglementation de la protection sociale : régime de sécurité sociale, motif d'exonération du ticket modérateur, caisse primaire gestionnaire, etc. Il ne contient toutefois aucun renseignement sur la profession, la couverture complémentaire ou encore la morbidité des personnes protégées.

Schéma n° 1 La grappe assuré



Les **fichiers de prestations** recensent toutes les opérations liées au versement de prestations pour les personnes protégées présentes dans l'EPAS. Ces fichiers contiennent des informations relatives au consommateur (âge, sexe), au producteur de soins (spécialité du prescripteur, de l'exécutant, lieu des actes), à la nature des soins (cotation des actes effectués, date des soins) et aux flux monétaires engendrés (débours du consommateur, montant remboursé au consommateur ou directement au producteur dans le cas de tiers payant, taux de remboursement...).

Nous disposons pour ce rapport du fichier des personnes protégées en 1992 et du fichier de toutes les prestations relatives à des soins effectués au cours de l'année 1992, prestations remboursées en 1992 ou en 1993.

⁴ Cf. An. et Ar. Mizrahi - *Méthode de sondage - Enquête permanente dans les dossiers de Sécurité Sociale (Panel)* - CREDOC, biblio n° 395, 1978.

1.2 L'ENQUÊTE SANTÉ ET PROTECTION SOCIALE (ESPS)⁵

Chaque année, le CREDES cherche à enquêter un quart des personnes de l'EPAS ; ainsi, tous les quatre ans, toutes les personnes présentes dans l'EPAS auront été contactées par le CREDES. L'échantillon annuel résulte d'un tirage au sort dans l'EPAS. Ce tirage est basé sur la clé de contrôle du numéro d'immatriculation à la Sécurité sociale, clé qui varie de façon aléatoire entre 1 et 97. Ainsi, pour 1992, les assurés sélectionnés ont une clé de contrôle inférieure ou égale à 23.

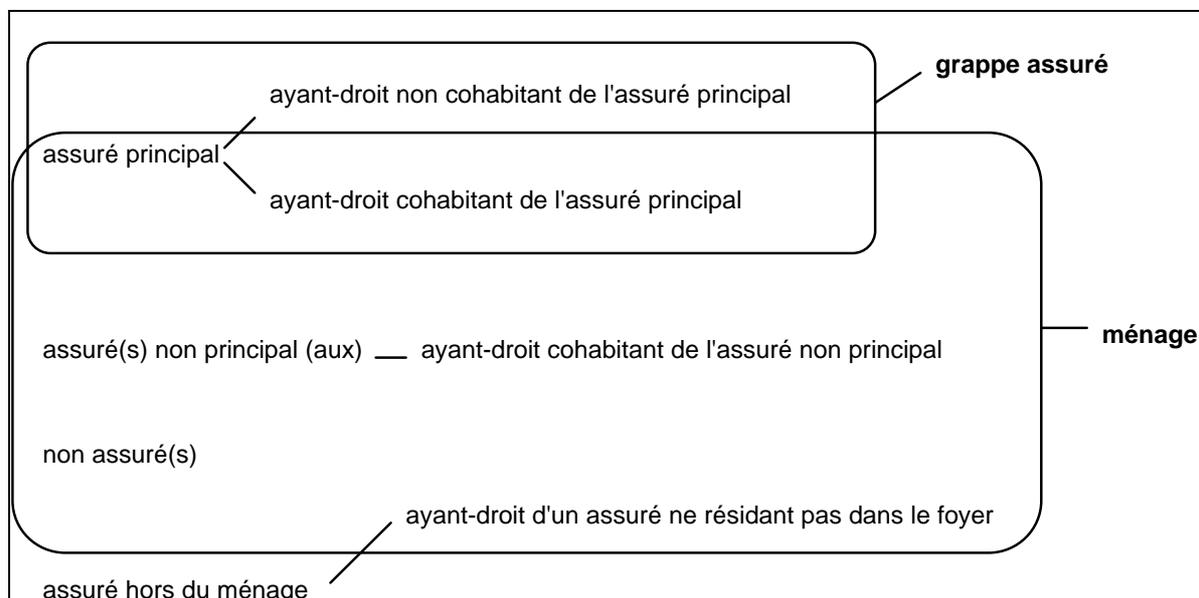
L'enquête se déroule en deux vagues, une vague au printemps et une vague en automne pour tenir compte des variations saisonnières de la consommation de soins et de la morbidité.

Elle combine deux modes de recueil des données :

- Le téléphone pour les assurés sociaux qui le possèdent et dont le numéro d'appel a pu être trouvé,
- Le face à face, c'est-à-dire un déplacement d'enquêteur pour ceux ne disposant pas du téléphone ou ceux dont le numéro n'a pu être trouvé (listes rouge, orange...).

L'enquête SPS permet de recueillir des renseignements sur la grappe assuré, c'est-à-dire sur l'assuré tiré au sort (appelé par la suite assuré principal) et ses ayants droit. Elle recueille également des données concernant les autres membres du ménage de l'assuré principal, qu'ils soient assurés (ils sont alors appelés assurés non principaux) ou ayants droit (ils sont alors ayants droit d'un assuré non principal). Le nombre de personnes enquêtées dans l'ESPS est donc plus important que dans l'EPAS puisque tout le ménage de l'assuré principal est enquêté.

Schéma n° 2
Les personnes des ménages enquêtés dans l'ESPS



⁵ Cf. A. Bocognano - *Méthode et déroulement de l'Enquête sur la Santé et la protection sociale*, CREDES 1992, Biblio n° 935.

A quelques exceptions près, la grappe assuré est incluse dans le ménage car la plupart des ayants droit de l'assuré principal vivent sous le même toit que l'assuré principal.

Pour tous les membres du ménage, on recueille :

- **des renseignements sociodémographiques** : âge, sexe, profession, occupation principale, département de résidence, revenus, niveau d'étude, nationalité...
- **des renseignements sur la protection sociale** : régime de Sécurité sociale, motif éventuel d'exonération du ticket modérateur, protection par une couverture complémentaire maladie...
- **des données de morbidité** : maladies dont sont atteints les enquêtés au jour de l'enquête (morbidité prévalente), facteurs de risque tels que tabac et alcool, degré d'invalidité...
- **des données sur leur consommation de soins** au cours d'un mois : toutes leurs consommations de soins (séances de médecins, d'auxiliaires, de dentistes, analyses médicales, acquisitions pharmaceutiques, hospitalisations..) sont consignées par les enquêtés sur un carnet de soins auto-administré.

1.3 LA PROCÉDURE D'ANONYMAT

La réalisation de l'enquête SPS implique un échange d'information entre la CNAMTS et le CREDES. En effet, seule la CNAMTS dispose de l'EPAS de façon nominative. Le tirage au sort des assurés à enquêter est donc effectué par le département statistique de la CNAMTS qui attribue aux personnes ainsi tirées au sort un identifiant n0 qui vient remplacer le Numéro National d'Identité (NNI). Les coordonnées nominatives de ces personnes sont directement adressées aux sociétés d'enquête par la CNAMTS, sociétés qui, outre la CNAMTS sont seules à détenir la liaison entre le n0 et les coordonnées de la personne et ce, uniquement pour la durée de l'enquête. Une fois l'enquête réalisée, les sociétés d'enquête transmettent au CREDES les données recueillies rendues anonymes par l'intermédiaire de l'identifiant n0.

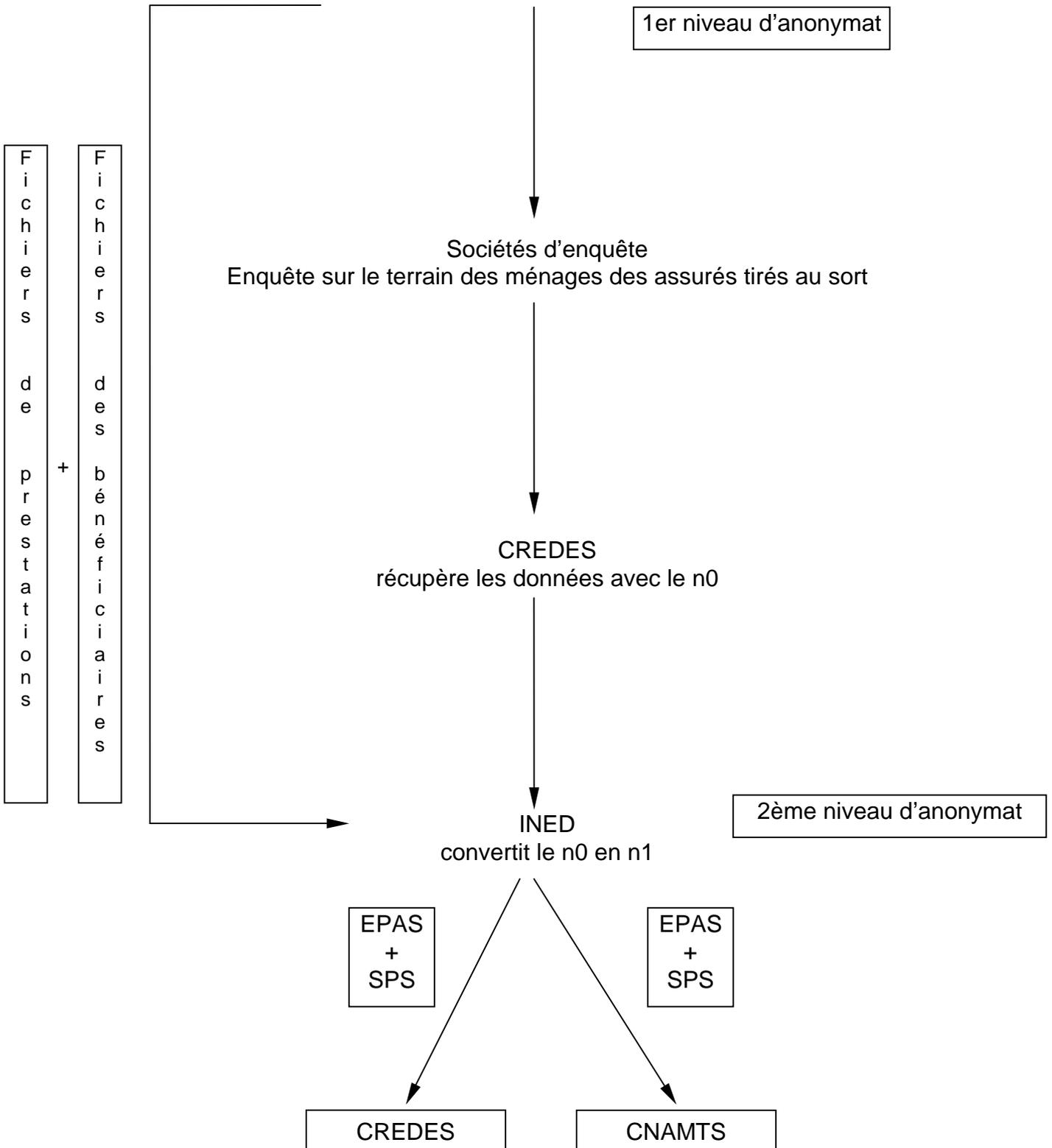
Un second niveau d'anonymat est obtenu par la conversion de l'identifiant n0 en un nouvel identifiant n1. C'est un organisme tiers, l'INED⁶, qui est chargé de cette conversion et lui seul détient les règles de passage de n0 à n1. Pour ce faire, le CREDES adresse les données issues de l'enquête de l'année n à l'INED ; de son côté, la CNAMTS adresse à l'INED le fichier des bénéficiaires de l'EPAS à jour au 31 Décembre de l'année n-1 ainsi que le fichier des prestations versées au cours de l'année n pour l'ensemble des bénéficiaires de l'EPAS. Les n0 sont alors transformés en n1 dans tous les fichiers ; le CREDES et la CNAMTS récupèrent alors l'ensemble de ces fichiers (EPAS et ESPS) avec les nouveaux identifiants n1 (cf. schéma récapitulatif).

Ainsi, ni le CREDES, ni la CNAMTS ne disposent d'informations nominatives concernant les personnes enquêtées.

La CNAMTS et le CREDES ont obtenu le consentement de la CNIL (Commission Nationale Informatique et Liberté) pour mener à bien cette opération.

⁶ INED (Institut National d'Etudes Démographiques)

CENTI - CNAMTS
Tirage au sort du 1/4 des assurés
du fichier des personnes protégées de l'EPAS
Affectation d'un n0 à la place du NIR



La méthode d'appariement

INTRODUCTION

Plusieurs niveaux d'appariement sont successivement présentés dans cette partie :

1er niveau : appariement de la population (cf 2.1)

* appariement de la « grappe assuré » de l'EPAS avec la « grappe assuré » reconstituée dans l'enquête SPS

* appariement des individus de l'EPAS et des individus SPS, sans prendre en compte leur qualité de bénéficiaires

2ème niveau : appariement de la population protégée et des prestations (2.2)

* appariement des prestations de l'EPAS et des bénéficiaires de l'EPAS (2.2.1)

* appariement des prestations de l'EPAS et du fichier des personnes appariées (2.2.2)

2.1 APPARIEMENT DES PERSONNES PROTÉGÉES

Cette partie du travail a pour but de fusionner les informations recueillies dans l'ESPS auprès des ménages enquêtés et celles contenues dans le fichier des personnes protégées de l'EPAS sur les assurés principaux et leurs ayants droit. Ceci correspond à l'appariement de la "grappe assuré" présente dans l'EPAS avec la "grappe assuré" correspondante dans les ménages enquêtés.

Les grandes lignes de cette démarche sont résumées ci-dessous. Le détail du cheminement est présenté en annexe.

Ce travail présente deux difficultés majeures :

1 - Les deux fichiers comportent des individus communs, mais aussi des individus propres à chacun des fichiers :

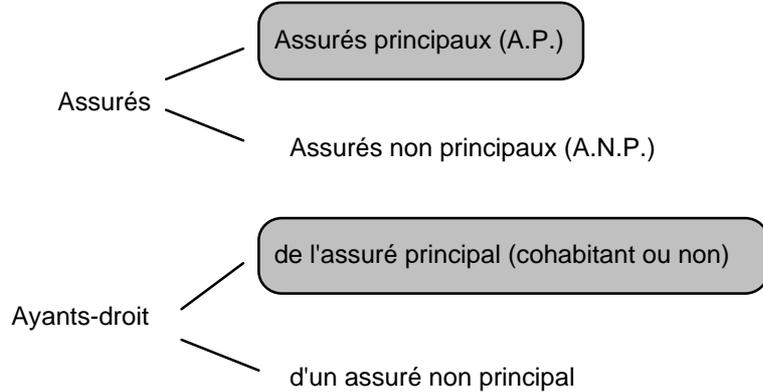
- les individus communs correspondent à l'intersection de l'ensemble des "grappes assuré" (c'est-à-dire l'assuré principal et ses ayants droit) et de l'ensemble des ménages ayant accepté de participer à l'enquête.
- les individus propres à chacun des fichiers : dans l'ESPS, il s'agit de tous les autres membres des ménages ayant participé à l'enquête, et dans l'EPAS, de toutes les "grappes assurés" issues de ménages qui n'ont pas participé à l'enquête pour diverses raisons (refus, déménagement...) (cf. Schéma n° 3).

2 - Les ménages dans l'ESPS et la "grappe assuré" dans l'EPAS sont identifiés de façon identique par le n1 mais les membres du ménage et les membres de la grappe assuré ne sont pas identifiés de la même façon dans les deux fichiers :

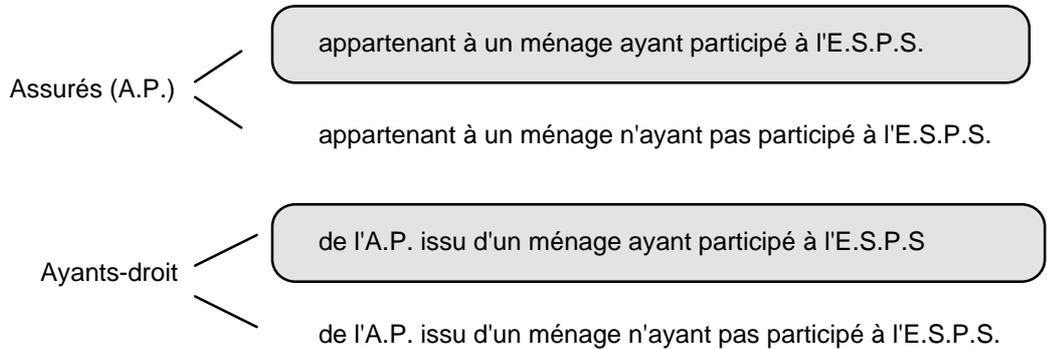
- dans l'ESPS, les membres d'un ménage sont identifiés par un numéro d'individu appelé "nind". L'assuré principal possède toujours le numéro d'individu 1. Les autres membres du ménage sont théoriquement numérotés par l'enquêteur par âge décroissant. Une personne est ainsi repérée par son n1 et son nind.
- dans l'EPAS, ce sont les ayants droit de l'assuré qui sont numérotés et cette numérotation prend en compte la qualité de l'ayant droit (conjoint, enfant...).

Schéma n° 3
Personnes théoriquement présentes dans l'EPAS et l'ESPS (1992)

E.S.P.S.
9 563 personnes
dont 9 554 ayant un mois et une année de naissance renseignés
dont 9 092 appartenant à un ménage où l'assuré principal déclare être au Régime général



E.P.A.S.
11 645 lignes ayant un mois et une année de naissance renseignés
correspondant à 11 303 personnes ayant un mois et une année de naissance renseignés
et à 10 176 personnes n'appartenant pas à un régime particulier



○ personnes présentes dans l'ESPS et dans l'EPAS

Plusieurs méthodes ont été testées pour réaliser l'appariement :

- **La première méthode** consiste à appairer séparément les assurés principaux et leurs ayants droit.

Les assurés principaux sont repérés simplement par un numéro d'individu codé 1 dans l'ESPS et par leur qualité de bénéficiaire dans l'EPAS. L'extraction de ces personnes des fichiers initiaux et l'appariement par le n1 ne posent pas de difficultés.

Quant aux ayants droit, la seule façon de les identifier de manière unique consiste à prendre leur n1 et leur date de naissance.

Cette méthode n'a pas été retenue car elle présente l'inconvénient de partir de l'hypothèse que la qualité du bénéficiaire (assuré/ayant droit) déclaré par l'enquêté et celui relevé dans l'EPAS sont identiques ; elle exclut donc la possibilité d'appairer une personne se déclarant par exemple assurée dans l'ESPS alors qu'elle est en fait ayant droit au regard de la CNAMTS.

- **La seconde méthode** consiste à créer dans les deux fichiers un identifiant individuel commun et d'appairer les deux fichiers sans prendre en compte la situation administrative des personnes (assurés/ayants droit). Plusieurs identifiants ont été progressivement testés (Cf. annexe n° 1). L'identifiant finalement retenu pour une personne est constitué de :
 - *n1* : numéro de ménage
 - *mois et année de naissance* (le jour de naissance n'a pas été pris en compte car très souvent manquant ou forcé à 1 dans l'EPAS)
 - un *rang* dans le ménage, créé de façon identique dans les deux fichiers et essentiellement destiné à distinguer les jumeaux et les conjoints nés le même mois de la même année (30 cas de jumeaux, 1 cas de triplés, 1 cas de quadruplés et 15 cas de ménages où deux personnes sont nées le même mois de la même année ; cf. annexe n° 1)
 - le *sexe*, ajouté en dernier lieu pour éviter d'avoir des personnes appariées, mais ayant un sexe différent dans les deux fichiers.

Cette dernière méthode conduit à retenir 5 373 individus communs aux deux fichiers⁷. 3 719 individus sont présents dans l'ESPS et non trouvés dans l'EPAS et 5 925 individus sont présents dans l'EPAS et non enquêtés dans l'ESPS (voir détails en annexe).

Parmi ces 5 373 individus, 467⁸ sont bénéficiaires d'un régime particulier. Les caisses primaires d'assurance maladie ne gèrent pas la totalité de ces régimes, les prestations concernant les personnes affiliées à ces régimes sont partielles dans les fichiers dont nous disposons ; ces 467 personnes sont par conséquent exclues de la population appariée.

Il reste donc 4906 personnes communes aux deux fichiers.

⁷ Pour plus de détails, cf annexe

⁸ Les régimes particuliers sont les suivants : grands invalides de guerre (codes 130 et 540 - 1 individu concerné), fonctionnaires et ouvriers de l'état (code 230 - 304 individus), agents statutaires de l'EDF-GDF (code 243 - 30 individus concernés), étudiants cotisants (code 250 - 8 individus concernés), retraités : Algérie O-M, Fonctionnaires, ouvriers d'état (code 260 - 83 individus concernés), agents EDF-GDF en retraite (code 270 - 8 individus concernés), agents des collectivités locales (codes 290, 300, 310, 320 - 28 individus concernés), étudiants exonérés de cotisations (code 350 - 2 individus concernés) ; un code régime (code 180 des adultes handicapés) a été exclu par erreur et concernait 3 individus.

Bilan de l'appariement des personnes :

4 906 individus communs (au sens de l'identifiant retenu)
3 719 individus non trouvés dans l'EPAS mais présents dans l'ESPS
5 270 individus non trouvés dans l'ESPS mais présents dans l'EPAS

a) Parmi les 4 906 individus communs :

- 2 741 personnes sont assurés principaux des deux côtés (soit 55,9 % de l'effectif),
- 1 437 sont ayants droit de l'assuré principal (29,3 % de l'effectif) ; au total ce sont donc 85,2⁹ % d'individus qui ont la même "qualité de bénéficiaire" des deux côtés.

Les 14,8 % restants se décomposent de la façon suivante :

- 429 (8,7 %) personnes qui se déclarent assurées dans l'ESPS et qui sont ayants droit pour l'EPAS (le cas inverse n'existe pas c'est-à-dire qu'aucune personne n'est assurée pour l'EPAS et ayant droit pour l'ESPS)
- 294 (6 %) qui se déclarent dans l'ESPS ayant droit d'un assuré autre que l'assuré principal
- 2¹⁰ personnes sont assurés principaux pour l'EPAS et assurés non principaux pour l'ESPS
- 1 est ayant droit pour l'EPAS et non protégé pour l'ESPS
- 2 personnes sont ayant droit d'un assuré principal côté ESPS et ont une qualité de bénéficiaire inconnue dans l'EPAS

Tableau n° 1
Qualité de bénéficiaire observée des personnes appariées dans l'ESPS et dans l'EPAS

EPAS \ ESPS	Assuré	Aydt conjoint-concubin	Aydt enfant	Modalité inconnue	TOTAL
Assuré principal	2 741	0	0	0	2 741
Assuré non principal	2	419	10	0	431
Aydt conjoint de L'AP	0	319	1	0	320
Aydt enfant de l'AP	0	0	1 115	0	1 115
Aydt autre de l'AP	0	0	2	2	4
Aydt d'un ANP	0	0	294	0	294
Non assuré	0	1	0	0	1
TOTAL	2 743	739	1 422	2	4 906

⁹ Ces 4 178 personnes correspondent aux chiffres grisés dans le tableau 1 ci-dessus

¹⁰ Ces 2 individus (assurés non principaux dans l'ESPS et assurés dans l'EPAS) correspondent à une situation théoriquement impossible due probablement à une erreur au moment du remplissage des questionnaires

b) Parmi les 3 719 individus qu'on ne trouve pas dans l'EPAS :

- 2 531 assurés non principaux (68,1 %),
- 825 ayants droit d'un assuré autre que l'assuré principal (22,2 %),
- 3 non assurés (0,1 %).

Il est normal de ne pas trouver ces personnes dans l'EPAS puisqu'ils ne sont ni assurés principaux, ni ayants droit de l'assuré principal.

- 69 assurés principaux (1,9 %),
- 291 ayant droit de l'assuré principal (7,8 %).

Soit 360 individus qu'il est anormal de ne pas trouver dans l'EPAS et pour lesquels l'identifiant "num" n'existe pas dans l'EPAS ; le num étant composé du n1, de l'année de naissance, du mois de naissance, du rang et du sexe, la discordance peut venir de l'un ou de l'autre de ses composants :

- pour ces 360 individus, le n1 composant leur num existe bien dans l'EPAS.
- le rang a été attribué de la même façon dans les deux fichiers.

Il s'agit donc de discordance sur les autres composantes : année de naissance, mois de naissance, sexe.

c) Parmi les 5 270 individus qu'on ne trouve pas dans l'ESPS (chutes, adresses non exploitées) :

- 3 090 assurés principaux soit 58,6 %
- 2 180 ayants droit de l'assuré principal soit 41,4 %

Soit 3 090 grappes assurés concernées.

- 2 881 grappes parmi les 3 090 ci-dessus sont recensées dans le fichier des chutes de l'enquête SPS ; on dispose donc d'un motif de non réponse.
- 154 des 209 autres grappes ne sont pas des chutes mais correspondent à des ménages d'assurés principaux étant assurés à un régime de Sécurité Sociale non rattaché au Régime Général (militaires, agricoles, SNCF ...), ménages non retenus pour l'appariement.
- Les 55 grappes restantes correspondent probablement à des assurés rentrés dans l'EPAS entre 1992 et 1994¹¹, après la réalisation de l'ESPS.

Les motifs de chutes des 2 881 grappes recensées dans l'ESPS peuvent être regroupés en deux catégories : les « hors-champs » et les « refus ». Les hors-champ comprennent les adresses pour lesquelles il n'y a pu avoir aucun contact avec le ménage, qu'il s'agisse de coordonnées inexactes dans le fichier, d'un changement de coordonnées entre le moment du tirage et le moment de l'enquête ou encore de personnes décédées, de personnes en institution ou de personnes résidant trop loin. Les « refus » regroupent les adresses pour lesquelles un contact a été établi mais où le ménage n'a pas accepté de participer.

¹¹ On rappelle (cf annexe) que le fichier des personnes protégées est un fichier à jour début 1994 et qu'il y a eu des entrants et des sortants entre 1992 et 1994.

Tableau n° 2
Motifs de non réponse

Motif de non réponse	Effectif	% du total de chutes
HORS CHAMP LIES A L'EPAS	917	31,8%
Décès ¹²	34	
Pb d'adresses ou de n° de téléphone, dont :		
<i>Inconnu à cette adresse</i>	669	
<i>Inconnu à ce numéro</i>	78	
<i>Adresses incomplètes ou inexistantes</i>	46	
<i>Numéro non attribué</i>	90	
HORS CHAMP LIES A L'ENQUETE	382	13,3%
Pb d'adresses ou de n° de téléphone, dont :	366	
<i>Etranger, Corse, Outremer</i> ¹³	24	
<i>Changement de domicile</i>	342	
Personnes âgées en institution	16	
REFUS	1582	54,9% ¹⁴
Raison de santé	215	
Problèmes d'alphabétisation	17	
Personnes indisponibles	310	
Refus (pas le temps, pas intéressé...)	1 040	
Ensemble	2881	100,0%

¹² Il est logique de retrouver des personnes décédées puisque ces personnes restent dans le fichier durant les 3 ans qui suivent leur décès.

¹³ Ces personnes ainsi que les personnes âgées en institution sont exclues du champ de l'enquête.

¹⁴ Rappelons ici que le taux de participation global de l'enquête (nombre de ménages ayant accepté rapporté au nombre de personnes dans le champ de l'enquête) est de 74% (cf Santé, Soins et protection Sociale en 1992 - rapport CREDES n°962)

Tableau n° 3 : bilan
Ventilation des personnes présentes dans les fichiers initiaux à l'issue de l'appariement

EPAS ESPS	Assuré	Ayant droit	Modalité inconnue	TOTAL des individus communs	Individus communs appartenant à un régime particulier	Individus non trouvés dans l'EPAS mais présents dans l'ESPS	Ensemble SPS
Assuré principal	2 741	0	0	2 741		69	
Assuré non principal	2	429	0	431		2531	
Ayant droit de L'AP	0	1437	2	1439		291	
Ayant droit d'un ANP	0	294	0	294		825	
Non assuré	0	1	0	1		3	
TOTAL des individus communs	2 743	2 161	2	4 906	467	3 719	9092

Individus non trouvés dans ESPS et présents dans l'EPAS	3090	2180	0	5 270
Ensemble EPAS (hors Régimes particuliers)				10 176

CONCLUSION :

Si l'on considère que seuls, les 360 individus précédemment cités, pour lesquels les dates de naissance ou le sexe sont discordants entre les deux fichiers, auraient pu être appariés en plus, on peut conclure qu'au total, 93,2 % des personnes potentiellement appariables sont appariées¹⁵.

Ce pourcentage ne prend toutefois pas en compte les personnes qui ont été supprimées dès le départ des deux sources de données, en particulier celles dont les dates de naissance n'étaient pas renseignées.

¹⁵ Ce pourcentage est calculé de la façon suivante : 4 906 (individus appariés) / (4906+360)

2.2 APPARIEMENT DES PRESTATIONS

Pour chacune des personnes protégées de l'EPAS, nous disposons des prestations relatives aux liquidations effectuées en 1992 et en 1993.

Les données de prestations sont organisées en 7 fichiers correspondant chacun à une catégorie différente de prestations ; 3 fichiers concernent les prestations en nature :

- les **actes professionnels** comportant tous les actes effectués par les praticiens (médecins, dentistes, auxiliaires médicaux ou biologistes),
- les **prestations sanitaires diverses** regroupant les prestations de pharmacie, d'appareillages ou de prothèses, les frais de transport (ambulance, taxis...) et les frais de cures thermales,
- les **hospitalisations-séjours** regroupant toutes les prestations relatives à des frais spécifiquement hospitaliers ; 4 autres fichiers concernent les prestations en espèces :
 - les **rentes accident du travail**,
 - les **pensions d'invalidité**,
 - les **indemnités journalières**,
 - les **prestations capital-décès**.

Tous ces fichiers possèdent une partie commune donnant des renseignements administratifs et identifiant le personne protégée. Chaque type de prestations doit être ensuite étudié séparément.

Le travail d'appariement s'effectue en deux étapes :

- dans la première, on apparie les fichiers de prestations avec la totalité du fichier des personnes protégées de l'EPAS,
- puis on ne retient que les personnes communes à l'EPAS et l'ESPS.

Cette méthode permet de contrôler le plus exactement possible les prestations qui ne s'apparient pas.

2.2.1 Fusion des fichiers de prestations avec la totalité du fichier des personnes protégées de l'EPAS

Dans un premier temps, nous avons cherché à évaluer le taux de prestataires au cours de l'année 1992 c'est-à-dire le nombre de personnes protégées qui ont perçu au moins une prestation¹⁶ en nature relative à des soins effectués au cours de l'année 1992 et remboursés en 1992 ou 1993.

Pour ce faire, l'identifiant individuel défini dans la partie précédente (n1 + mois et année de naissance + rang + sexe) a été reconstitué dans tous les fichiers de prestations grâce aux renseignements contenus dans la partie commune de ces fichiers.

La lecture cumulée de tous les fichiers de prestations nous a permis de comptabiliser 38 050 identifiants différents. Le fichier des personnes protégées contient 43 032 identifiants (24 581 assurés et 18 451 ayants droit) ayant un mois et une année de naissance renseignés (non compris les bénéficiaires de régimes particuliers). Seuls, 30 025 de ces identifiants (19 041 assurés et 10 984 ayants droit) s'apparient, avec le fichier des personnes protégées, ce qui signifie que 8 025 personnes percevant des prestations ne sont pas retrouvées selon les critères retenus dans le fichier des personnes protégées.

L'étude de ces 8 025 personnes (4 468 assurés et 3 557 ayants droit) montre que dans la grande majorité des cas, l'année de naissance n'est pas renseignée dans les fichiers de prestations. Un autre identifiant a alors été mis en place pour ces personnes, identifiant qui prend en compte le numéro de la grappe assuré (n1), la qualité de la personne protégée (assuré/ayant droit) et le sexe.

1 099 personnes (1098 assurés et 1 personne de situation administrative inconnue) parmi les 8 025 s'apparient alors de manière unique avec le fichier des personnes protégées selon cet identifiant.

Parmi les 6 926 individus restant, 4 411 ont un numéro de grappe assuré (n1) n'existant pas dans le fichier des personnes protégées, soit 2 515 individus pouvant encore être appariés. 668 ayants droit parmi les 2 515 individus précédents sont issus d'une grappe assuré n'ayant qu'un seul ayant droit dans le fichier des personnes protégées. On pose donc l'hypothèse qu'il s'agit du même ayant droit dans les deux fichiers. 1 ayant droit parmi les 2515 individus précédents est issu d'une grappe assuré n'ayant qu'un ayant droit enfant et qu'un ayant droit conjoint.

Au total, 1 768 individus parmi les 8 025 sont identifiés de manière unique dans le fichier des personnes protégées. Pour ces 1 768 personnes, les dates de naissance et le sexe figurant dans le fichier des personnes protégées sont alors affectés dans les fichiers de prestation.

On obtient donc ainsi 31 793 prestataires, parmi lesquels 20 139 assurés, 11 653 ayants droit et 1 personne de situation administrative inconnue, soit un taux de prestataires pour les soins effectués au cours de l'année 1992 de 73,9 %¹⁷ sur l'ensemble de l'EPAS, de 81,9% pour les assurés et de 63,2% pour les ayants droit.

¹⁶ On s'est limité aux actes professionnels, aux prestations sanitaires diverses et aux prestations hospitalières

¹⁷ $(30025+1768)/43032=73,9 \%$

2.2.2 Appariement des fichiers de prestations avec les 4 906 personnes communes à l'EPAS et à l'ESPS

Après avoir corrigé les identifiants dans le fichier des prestations selon la précédente méthode, on dénombre 3 858 prestataires d'au moins une prestation relative à un soin effectué au cours de l'année 1992 parmi les 4 906 personnes protégées s'appariant entre l'EPAS et l'ESPS, ce qui donne un taux de prestataires de 78,6 %. Ce pourcentage, plus élevé que le précédent, s'explique par le fait que le fichier des 4 906 personnes protégées est "nettoyé" par l'enquête SPS puisqu'il ne comprend que les personnes enquêtées, qui sont effectivement affiliées au Régime général.

Sur les 3 858 prestataires de l'EPAS, 2 319 (60,1 %) sont assurés dans l'ESPS et 1 539 sont ayants droit. Le taux de consommateurs d'au moins une prestation relative à des soins au cours de l'année 1992 est donc de 84,6 % ($2\,319/2\,741$) pour les assurés.

Il reste 1 048 personnes sur 4 906, soit 21,4 % qui n'ont perçu aucune prestation relative à des soins effectués en 1992.

Etude des 1 048 personnes non prestataires :

Parmi les 1 048 personnes n'ayant perçu aucune prestation au cours de l'année 1992, 561 ont déclaré au moins une consommation au cours des trois semaines d'enquête susceptible de donner lieu à un remboursement.

Dans l'ESPS, en 4 semaines, parmi ces 561 personnes :

254 ont eu au moins une fois recours à un médecin ou à un dentiste.

237 ont consommé au moins une fois de la pharmacie prescrite.

34 ont eu recours au moins une fois à un auxiliaire.

34 ont eu au moins une analyse.

6 ont eu recours au moins une fois à l'hospitalisation.

Ces 561 personnes sont, au sens de l'EPAS à 22,3 % des assurés et à 78,7 % des ayants droit. Au sens de l'ESPS elles sont à 63,8% des assurés, 36,0% des ayants droit et 0,2% non protégées.

Enfin, on a cherché à voir si la vague d'enquête avait une influence sur le fait d'être ou non répertorié dans le fichier des prestations. Parmi ces 561 personnes, 44 % ont été enquêtés lors de la première vague de l'enquête et 56 % lors de la deuxième vague.

Parmi les 247 enquêtés lors de la 1ère vague, 101 se déclarent assurés non principaux dans l'ESPS alors qu'ils apparaissent comme ayants droit dans l'EPAS, 56 se déclarent ayants droit d'un assuré autre que l'assuré principal et 1 se déclare non protégé soit 158 personnes pour lesquelles on peut faire la même hypothèse que précédemment et pour lesquelles il est donc normal de ne pas trouver leurs prestations dans l'EPAS. Il reste donc 89 personnes enquêtées lors de la 1ère vague et ayant déclaré des consommations susceptibles d'être présentées au remboursement et n'ayant visiblement pas donné lieu à des prestations au cours de l'année 1992 ou 1993.

Parmi les 314 enquêtés lors de la 2ème vague, 132 se déclarent assurés alors qu'ils apparaissent comme ayants droit dans l'EPAS et 71 se déclarent ayants droit d'un assuré autre que principal soit 203 personnes pour lesquelles, selon l'hypothèse précédente, il est donc normal de ne pas trouver leurs prestations dans l'EPAS. Il reste donc 111 personnes enquêtées lors de la 2ème vague (entre octobre et décembre 1992), ayant déclaré des consommations pendant l'enquête mais n'ayant pas donné lieu à des prestations en 1992 ou 1993.

Parmi les 4 906 personnes appariées, 728 n'ont pas la même situation administrative dans l'EPAS et dans l'ESPS (Cf. 2.1.) à savoir que 431 se déclarent assurés non principaux dans l'ESPS alors qu'elles sont 429 ayants droit et 2 assurées dans l'EPAS, 294 se déclarent ayants droit d'un autre assuré que l'assuré principal dans l'ESPS alors qu'elles apparaissent ayants droit de l'assuré principal dans l'EPAS, 2 se déclarent ayants droit dans l'ESPS et ont une situation administrative inconnue dans l'EPAS, et 1 se déclare non protégé dans l'ESPS. Si l'on fait l'hypothèse que la qualité de bénéficiaire déclarée par les enquêtés est en dehors de (outre ceux qui déclarent être non assurés), les prestations de ces 727 personnes ne devraient pas apparaître dans les fichiers de prestations de l'EPAS.

Parmi ces 727 personnes, seules 182 ont perçu au moins une prestation au cours de l'année 1992. 162 ont perçu au moins une prestation en actes professionnels et 163 au moins une prestation sanitaire diverse.

Ces 182 personnes sont à 99,5 % des ayants droit (1 seul assuré : 0,5 %). 49,2 % sont des ayants droit enfants et 49,2 % des ayants droit conjoints (1,1 % sont des ayants droit autres).

Pour ces 182 personnes, on peut donc penser qu'il s'agit d'une erreur de déclaration de leur situation administrative pendant l'enquête ou que celle-ci a changé entre le tirage au sort et l'enquête sur le terrain.

Pour les 545 autres personnes, on peut faire l'hypothèse que leurs déclarations dans l'ESPS sont exactes (à savoir qu'ils sont assurés non principaux ou ayant droit d'un autre assuré que l'assuré principal) et que, par conséquent, il est normal de ne pas trouver leurs prestations dans l'EPAS puisqu'ils ne font pas partie du champ de l'EPAS.

La population de référence pour l'étude des prestations devient alors 4 361 personnes (4 906 -545) au lieu de 4 906.

CONCLUSION :

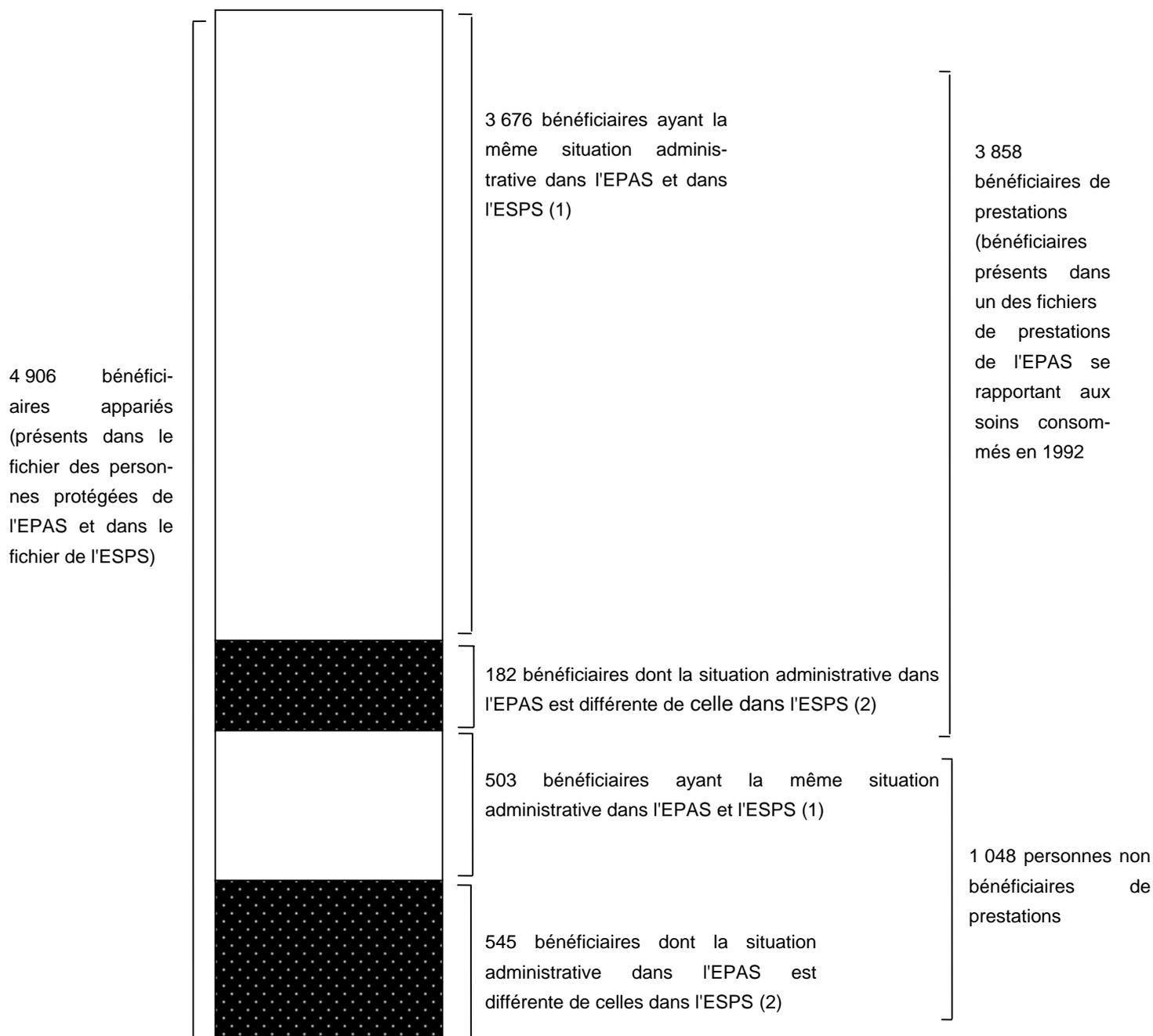
Le taux de prestataires d'au moins une prestation pour les assurés est de : 88,2 %. Le taux de prestataires pour les ayants droit est de 89,0 %¹⁸.

Le schéma n°4 (page suivante) résume la démarche pour l'appariement des prestations. Le tableau n° 4 donne la ventilation du taux de prestataires selon l'âge et le sexe.

¹⁸ Nombre d'ayants-droit ayant perçu au moins une prestation en 1992 : 1441

Nombre d'ayants-droit protégés parmi la population de référence précédemment définie : 1619

Schéma n° 4 Récapitulatif de l'appariement des prestations



4 906 - 545 = 4 361 bénéficiaires retenus pour l'étude des prestations

(1) Dans l'EPAS et dans l'ESPS, les bénéficiaires (assurés et ayants droit) sont rattachés à un N1 (ou à un NIR) correspondant aux critères de tirage de l'EPAS.

(2) Dans l'EPAS, les bénéficiaires (assurés et ayants-droit) sont rattachés à un N1 (ou à un NIR) ne correspondant pas aux critères de tirage de l'EPAS. En revanche, dans l'ESPS, les bénéficiaires sont rattachés à un N1 (ou à un NIR) ne correspondant pas aux critères de tirage de l'EPAS.

Tableau n° 4
Taux de prestataires selon l'âge, le sexe et la qualité de bénéficiaire

Tx de prestataires (%)	Assurés	Ayants droit conjoints	Ayants droit enfants	Total ¹⁹ ayants droit	TOTAL
Moins de 2 ans	0,0	0,0	93,6	93,6	93,6
de 2 à 9 ans	0,0	0,0	88,0	88,0	88,0
de 10 à 19 ans	47,6	100,0	86,4	86,2	84,8
de 20 à 29 ans	83,1	95,1	88,9	94,0	84,1
de 30 à 39 ans	87,1	90,7	0,0	90,7	87,6
de 40 à 49 ans	87,5	94,9	0,0	94,9	82,6
de 50 à 59 ans	88,9	90,7	0,0	90,8	89,3
de 60 à 69 ans	93,7	94,3	0,0	94,3	93,7
de 70 à 79 ans	96,7	89,5	0,0	89,5	96,0
80 ans et plus	98,6	100,0	0,0	100	98,6
Total	88,2	92,7	87,8	89,0	88,5
Homme	88,2	80,0	87,1	87,0	87,8
Femme	88,1	93,2	88,4	90,3	89,0

Tableau n° 4 bis
Effectif des personnes prestataires selon l'âge, le sexe et la qualité de bénéficiaire

Effectifs des prestataires	Assurés	Ayants droit conjoints	Ayants droit enfants	Total ayants droit	TOTAL
Moins de 2 ans	0	0	102	102	102
de 2 à 9 ans	0	0	453	453	454
de 10 à 19 ans	10	2	497	498	508
de 20 à 29 ans	439	39	8	47	486
de 30 à 39 ans	545	88	0	88	633
de 40 à 49 ans	503	93	0	93	596
de 50 à 59 ans	336	88	0	89	425
de 60 à 69 ans	339	50	0	50	389
de 70 à 79 ans	176	17	0	17	193
80 ans et plus	69	2	0	2	71
Total	2 417	379	1 060	1 441	3 858
Homme	1 201	12	535	547	1 748
Femme	1 216	367	525	894	2 110

¹⁹ Le total des ayants-droit inclut les ayants-droit conjoints, les ayants-droit enfants et les ayants-droit autres

Tableau n° 4 ter
Répartition des personnes protégées selon l'âge, le sexe et la qualité de bénéficiaire

Effectifs des personnes protégées	Assurés	Ayants droit conjoints	Ayants droit enfants	Total ayants droit	TOTAL
Moins de 2 ans	0	0	109	109	109
de 2 à 9 ans	0	0	515	515	515
de 10 à 19 ans	21	2	575	578	599
de 20 à 29 ans	528	41	9	50	578
de 30 à 39 ans	626	97	0	97	723
de 40 à 49 ans	575	98	0	98	673
de 50 à 59 ans	378	97	0	98	476
de 60 à 69 ans	362	53	0	53	415
de 70 à 79 ans	182	19	0	19	201
80 ans et plus	70	2	0	2	72
Total	2 742	409	1 208	1 619	4 361
Homme	1 361	15	614	629	1 990
Femme	1 381	394	594	990	2 371

Tableau n° 5
Situation administrative estimée des personnes appariées dans l'ESPS et dans l'EPAS

EPAS \ ESPS	Assuré	Aydt conjoint-concubin	Aydt enfant	Modalité inconnue	TOTAL
Assuré principal	2 741	0	0	0	2 741
Assuré non principal	1	89	7	0	97
Aydt conjoint de l'AP	0	319	1	0	320
Aydt enfant de l'AP	0	0	1 115	0	1 115
Aydt autre de l'AP	0	0	2	2	4
Aydt d'un ANP	0	0	83	0	83
Non protégé	0	1	0	0	1
TOTAL	2 742	409	1 208	2	4 361

(En grisé, qualité de bénéficiaire concordante EPAS/ESPS)

2.2.3 Répartition des prestataires selon le type de prestations

Parmi les 4 361 personnes protégées appariées avec les personnes enquêtées dans l'ESPS :

- ◇ 3 858, soit 88,5 %, ont donc perçu au moins une prestation au cours de l'année 1992,
- ◇ 3 782, soit 86,7 %, ont perçu au moins une prestation d'actes professionnels en un an,
- ◇ 3 661, soit 84,0 %, ont perçu au moins une prestation "sanitaire diverse" en un an,

La structure démographique des personnes protégées et des prestataires est la suivante :

Tableau n° 6
Structure démographique des personnes protégées et prestataires
(en % par rapport à l'effectif renseigné pour une variable donnée)

Critères socio-démographiques (source CREDES)	Personnes protégées	Prestataires d'au moins un type de prestations	Prestataires d'au moins une prestation en actes prof.	Prestataires d'au moins une prestation sanitaire diverse
Effectifs	4 361	3 858	3 782	3 661
Age				
Moins de 2 ans	2,5	2,6	2,5	2,8
2 - 9 ans	11,8	11,7	11,5	11,8
10-19 ans	13,7	13,2	13,3	12,8
20-29 ans	13,3	12,6	12,5	12,2
30-39 ans	16,6	16,4	16,4	16,1
40-49 ans	15,4	15,4	15,5	15,5
50-59 ans	10,9	11,0	11,1	11,3
60-69 ans	9,5	10,1	10,2	10,5
70-79 ans	4,6	5,0	5,1	5,2
80 ans et plus	1,7	1,8	1,9	1,9
Sexe				
Homme	45,6	45,3	45,1	44,9
Femme	54,4	54,7	54,9	55,1
Occupation principale				
Actif	41,5	42,8	42,8	42,7
Retraité	13,3	14,3	14,5	14,8
Chômeur	5,3	4,6	4,6	4,5
Autre inactif	10,4	9,7	9,8	9,9
Enfant-militaire	29,5	28,7	28,3	28,2
PCS individuelle Nb = 1766				
Artisans	2,3	2,2	2,2	2,1
Cadres supérieurs	7,8	7,8	7,7	7,6
Prof. Intermédiaires	23,8	24,5	24,5	24,7
Employés	33,4	33,6	33,8	33,7
Ouvriers qualifiés	23,7	23,4	23,4	23,2
Ouvriers non qualifiés	8,8	8,4	8,4	8,5
Exonération du ticket modérateur				
Oui	9,1	9,8	10,0	10,2
Non	90,5	89,7	89,6	89,4
Ne sait pas	0,4	0,4	0,4	0,4
Couverture complémentaire				
Oui	86,8	88,2	88,6	88,7
Non	13,0	11,6	11,3	11,2
Ne sait pas	0,1	0,1	0,1	0,1
Hospitalisation < 3 mois				
Oui	3,9	4,2	4,2	4,2
Non	96,1	95,7	95,7	95,8

Redressement des données appariées

3.1 METHODE DE REDRESSEMENT

La population appariée diffère sensiblement de l'échantillon permanent d'assurés sociaux du Régime général (EPAS) puisque cet échantillon n'a pas été apparié dans sa totalité.

Le redressement de la population appariée va consister à modifier les poids de sondage des individus de façon à « caler » l'échantillon, pour un certain nombre de variables jugées importantes, sur des structures reconnues fiables.

Le but de la procédure est de déterminer une nouvelle variable de poids (notée W_k) pour un individu k , qui permet de bien estimer le vrai total connu X (population de référence) et qui soit construite de telle sorte qu'elle s'éloigne le moins possible de la pondération d'origine (d_k =poids de sondage).

Cette méthode de redressement est appelée « calage sur marges ».

Soient s la population appariée

X_k les variables sur lesquelles se fait le redressement. Il ne doit pas y avoir de colinéarité entre les variables.

Au départ, on a : $\sum_{(k \in s)} d_k \cdot X_k \neq X$

et on veut que pour notre échantillon s : $\sum_{(k \in s)} W_k \cdot X_k = X$

On arrive ainsi au problème suivant d'inconnues W_k , $k \in s$

Minimiser : $\sum_{(k \in s)} D(W_k, d_k)$

sous contrainte : $\sum_{(k \in s)} W_k \cdot X_k = X$

Il faut choisir une fonction de distance D pour mesurer les proximités entre les poids cherchés W_k et les poids de sondage d_k .

Pour résoudre le problème on se donne une fonction F au choix et on détermine λ par la méthode de Newton (par itérations successives) vérifiant :

$$\sum_{(k \in s)} d_k \cdot F(X_k \cdot \lambda) \cdot X_k = X \quad \text{et} \quad W_k = d_k \cdot F(X_k \cdot \lambda)$$

L'expression $F(X_k \cdot \lambda)$ apparaît comme un coefficient multiplicateur de redressement du poids d'origine d_k .

Les différentes fonctions F généralement utilisées sont de 4 types :

- **méthode linéaire** : $F(x) = 1 + x$

- **méthode raking ratio** : $F(x) = \exp(x)$

Cette méthode classique de redressement est aussi connue sous le nom de méthode RAS ou IPF (Iterative Proportional Fitting).

$$\text{- méthode logit : } F(x) = \frac{L(U-1) + U(1-L) \exp(Ax)}{U-1 + (1-L) \exp(Ax)} \quad \text{avec } A = \frac{U-L}{(1-L)(U-1)}$$

La forme logistique de la fonction F donne son nom à cette méthode que l'on peut aussi caractériser comme étant une méthode raking ratio tronquée aux deux extrémités, de façon à ce que les rapports W_k/d_k soient bornés inférieurement par L et supérieurement par U.

$$\text{- méthode linéaire tronquée : } F(x) = \begin{cases} 1 + x & \text{si } L-1 \leq x \leq U-1 \\ L & \text{si } x < L-1 \\ U & \text{si } x > U-1 \end{cases}$$

Il s'agit de la méthode linéaire tronquée aux deux extrémités.

Dans la pratique, la détermination des valeurs L_{max} et U_{min} ($L < 1$ et $U > 1$) se fait par approximations successives.

Les critères de choix de la fonction F parmi les quatre formes citées sont :

- la plus faible dispersion,
- la plus faible étendue,
- l'allure générale de la distribution.

Le redressement pas calage sur marges est réalisé avec le logiciel appelé CALMAR de l'INSEE mis au point en SAS par Olivier Sautory. Il consiste à prendre comme paramètres les effectifs marginaux des variables qualitatives sur lesquelles on veut se caler, ainsi que l'expression de la fonction F (ou celle de la distance D) qui reste un libre choix de l'utilisateur.

3.2 POPULATION DE REFERENCE

La population de référence théorique devrait être constituée des personnes vivantes au 1/01/1992 (correspondant au quart annuel de 1992), couvertes par le Régime Général ou l'un des régimes rattachés et étant assuré dans l'EPAS. Cette population n'est pas déterminable rigoureusement du fait de certaines adresses de l'EPAS classées « hors-champ » à l'issue de l'enquête. Elle est donc approximée comme il est expliqué ci-dessous.

Dans l'EPAS de 1992, on dénombre 6 511 personnes assurées. Sont éliminées les personnes :

- demeurant dans les DOM-TOM (17).
- ayant comme régime de Sécurité Sociale un régime particulier (666).
- « hors champ » pour des raisons liées à l'EPAS : adresse incomplète ou numéro de téléphone non attribué, inconnu à cette adresse ou à ce numéro de téléphone (868)²⁰.

²⁰ A noter que parmi ces 868 assurés exclus de la population de référence, 566 ont perçu au moins une prestation en 92 ou en 93 et relative à un soin effectué en 92. Pour ceux-ci, on peut imaginer qu'il s'agit, non d'un problème de coordonnées dans l'EPAS de base, mais plutôt d'un problème de remontée de l'adresse jusqu'à la société d'enquête. Il est donc envisageable, pour l'appariement d'une année ultérieure, d'étudier ces personnes et d'éventuellement les prendre en compte dans la population de référence.

- décédées avant 1992 (la date de décès n'étant pas renseignée dans l'EPAS, nous avons pris en compte les assurés principaux ayant comme motif de chute le décès. Nous avons fait l'hypothèse que ceux qui n'ont pas de prestations en 1992 sont décédés avant 1992) (20 personnes).

La population de référence compte donc 4 940 personnes assurées.

La répartition en terme de participation à l'enquête²¹ est la suivante :

Hors champ	395	(8%)
Impossibilité	558	(11.3%)
Refus	1035	(21%)
Interview	2952	(59.8%)

Les « hors champ » comprennent les motifs suivants : personne non rencontrée ou non jointe, changement de domicile, absent de longue durée, décédé.

L'impossibilité résulte d'une raison de santé, d'une indisponibilité, de l'analphabétisation, de l'inaccessibilité du logement ou de l'institutionnalisation de la personne.

Le coefficient de redressement sera identique pour tous les individus de la grappe assuré.

3.3 CHOIX DES VARIABLES DE REFERENCE POUR LE REDRESSEMENT

Les variables retenues sont au nombre de trois : l'âge, le sexe, et la Zone d'Etude et d'Aménagement du Territoire (ZEAT). Ce sont les seules variables sociodémographiques dont on dispose dans l'EPAS.

Distribution en pourcentages des variables de référence dans la population de référence (4 940 assurés) et dans l'échantillon (2 850 assurés) :

	Population	Echantillon
AGE		
Moins de 26 ans	15,5	15,4
de 26 à 34 ans	18,2	18,9
de 35 à 42 ans	16,3	18,3
de 43 à 52 ans	16,2	17,5
de 53 à 66 ans	18,4	18,2
Plus de 66 ans	15,4	11,6
SEXE		
Masculin	50,2	48,2
Féminin	49,8	51,8
ZEAT		
Région parisienne	22,7	18,5
Bassin parisien nord	11,2	11,7
Bassin parisien sud	6,7	7,2
Nord	6,2	6,4
Est	8,4	9,5
Ouest	12,3	13,8
Sud-Ouest	9,2	9,8
Centre Est	12,1	12,6
Méditerranée	10,7	10,3
Corse	0,5	0,2

²¹ Cette répartition diffère légèrement de celle obtenue page 18, probablement en raison de corrections intervenues entre les diverses étapes de ce rapport.

Les poids de sondage dk sont tous égaux au coefficient d'extrapolation (4940/2850).
Le redressement a été réalisé pour les quatre méthodes : linéaire, raking ratio, logit et linéaire tronquée.

3.4 CHOIX DE LA FONCTION DE DISTANCE

Le critère qui a été privilégié est celui de la plus faible étendue. Nous avons préféré ne pas avoir de poids trop petits ou trop importants. Les résultats sont présentés en Annexe 2.

Les marges obtenues avec la pondération finale de l'échantillon sont identiques aux marges de la population, et ce, pour les quatre fonctions de distance.

Méthode linéaire :

La méthode « linéaire » est la méthode qui donne le plus petit écart-type (0,186) : cela résulte de sa définition même. Son étendue est de 1,597.

Méthode raking ratio :

La méthode « raking ratio » donne la plus forte étendue : 2,829 en raison de ses forts poids (3,97) qui sont plus élevés qu'avec les autres méthodes. En revanche, les valeurs minimales sont les plus élevées. L'écart-type est de 0,28.

Méthode logit :

La méthode « logit » provoque une concentration des poids près des deux bornes. Le premier quartile vaut 0,85 et le quantile d'ordre 0,90 vaut 1,27. De plus, l'étendue est minimale (1,20). L'écart-type est de 0,20.

Méthode linéaire tronquée :

La méthode « linéaire tronquée » donne une distribution voisine de celle obtenue avec la méthode « logit » avec cependant une concentration encore plus forte autour de la borne inférieure (0,8), due au fait que la méthode « linéaire tronquée » permet d'atteindre effectivement ces bornes, alors que la méthode « logit » ne permet que de s'en approcher asymptotiquement. L'étendue est de 1,5 et l'écart-type est de 0,19.

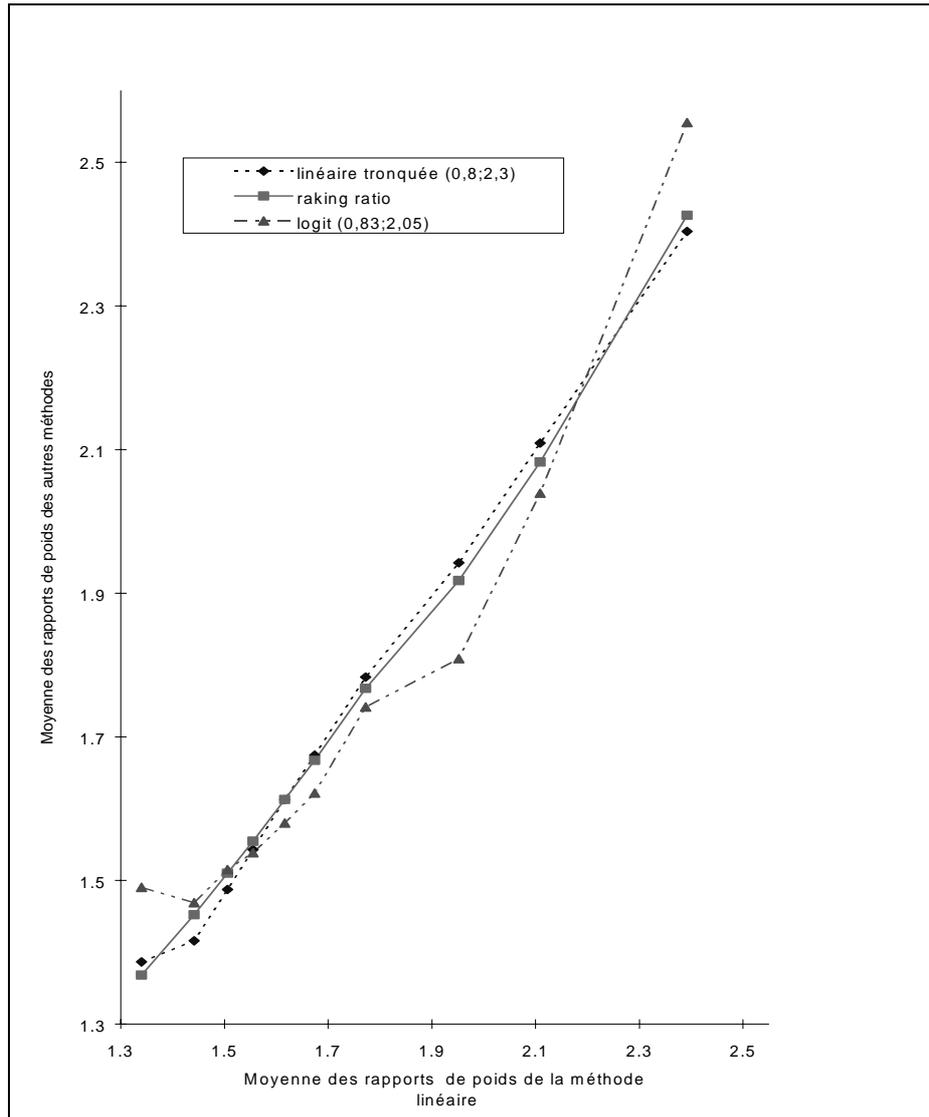
Les ménages ont été classés selon les valeurs croissantes des rapports de poids de la méthode linéaire. A partir de ce classement 10 groupes d'effectifs égaux par découpage en tranches des rapports de poids de cette méthode ont été définis. Pour chacun de ces groupes la moyenne, l'écart-type, les valeurs minimales et maximales des autres méthodes de pondération ont été calculés et sont présentés en Annexe 2.

Le graphique suivant représente les courbes : moyenne dans un groupe des rapports de poids des méthodes linéaire tronquée, logit et raking ratio en fonction de la moyenne dans un groupe de rapports de poids de la méthode linéaire.

Ce graphique permet de visualiser les commentaires faits précédemment.

La méthode retenue est la méthode logit (0.83,2.05).

Graphique n° 1
Moyenne dans un groupe de rapports de poids des méthodes raking ratio, logit, et linéaire tronquée en fonction de la moyenne dans un groupe de rapports de poids de la méthode linéaire



Annexe 1

**Recherche de l'identifiant à utiliser pour réaliser
l'appariement des personnes protégées de l'EPAS
avec les enquêtés de l'ESPS
(démarche chronologique)**

Méthode d'appariement des personnes protégées de l'EPAS avec les enquêtés de l'ESPS 1992

VARIABLES - CLÉS :

- n1** : numéro de ménage attribué par l'INED (présent dans l'EPAS et dans l'ESPS)
nind : numéro d'individu à l'intérieur d'un ménage (propre à l'ESPS)
ss : donne la qualité de bénéficiaire de la personne (assuré ou ayant droit) (propre à l'ESPS)
aydt : numéro de l'individu dont est ayant droit la personne (propre à l'ESPS)
benran : qualité de personne protégée (assuré / ayant droit) (propre à l'EPAS)
benrng : rang Laser de la personne protégée (propre à l'EPAS)

I DESCRIPTION DES FICHIERS INITIAUX

I.1 Le fichier des personnes enquêtées dans l'ESPS

Il contient 9 563 lignes dont 4 correspondent à des ayants droit du ménage non cohabitants.
9 554 personnes parmi ces 9 563 ont un mois et une année de naissance renseignés.
161 ménages, correspondant à 462 personnes ont été retirés parce que l'assuré principal était affilié à un régime non rattaché au régime général.

Soit 9 092 lignes dont 3 088 correspondant à un assuré principal.

I.2 Le fichier ménages enquêtés dans l'ESPS

Il contient 3 266 lignes dont 3 261 N1 différents.
Les 5 N1 en double (4 107, 4 151, 20 995, 44 728, 45 728) sont éliminés pour l'appariement.

==> La fusion du fichier des personnes et du fichier des ménages de l'enquête SPS donne
9 092 lignes correspondant à des personnes dont le mois et l'année de naissance sont
renseignés ; 3 088 sont des assurés principaux.

44 individus ont le mois de naissance non renseigné. 2 d'entre eux existent dans l'EPAS et ont
l'année de naissance identique; on leur attribue le mois de naissance de l'EPAS qui est renseigné.

I.3 Le fichier des personnes protégées de l'EPAS

Il contient 49 427 lignes dont :
- 49 415 lignes pour lesquelles ou le mois ou l'année de naissance sont renseignés
- 12 lignes avec le mois et l'année de naissance non renseignés

Ce fichier contient l'ensemble des personnes protégées de l'EPAS en 1994 ; seulement un quart de
ces personnes ont été tirées au sort pour l'enquête de 1992.

I.4 Le fichier des adresses à enquêter de l'ESPS

Il contient 6 649 N1 dont 6 645 N1 différents. Parmi ces 6 645 ménages différents, 6 511 sont communs avec l'EPAS.

Les 134 N1 présents dans l'ESPS et qui n'existent pas dans l'EPAS correspondent probablement à des grappes sorties de l'EPAS entre 1992 et 1994 :

- 114 sont des chutes d'ESPS, dont 45 pour cause de décès
- 20 sont enquêtés dont 4 ont un régime non rattaché au régime général, 15 ont un régime rattaché au régime général et 1 est recensé dans le fichier ménage d'ESPS et n'existe pas dans le fichier personne d'ESPS

La fusion de ce fichier avec le fichier des personnes protégées de l'EPAS fournit 11 645 lignes correspondant aux adresses tirées au sort pour l'enquête de 1992 avec le mois et l'année de naissance renseignés.

Parmi ces 11 645 lignes, 6 792 correspondent à un assuré principal et 4 853 à un ayant droit.

II IDENTIFICATION DES ASSURÉS PRINCIPAUX (AP)

Le numéro du ménage (n1) suffit normalement à identifier les Assurés principaux puisqu'il n'y a qu'un Assuré principal par ménage. Ceci suppose toutefois de filtrer les deux fichiers sur la donnée "AP oui/non" (EPAS : benran=1 ; ESPS : nind=1) ; ceci exclut donc la possibilité de retrouver un assuré principal n'ayant par exemple pas été placé en 1ère ligne dans nos questionnaires principaux.

Par cette méthode, on obtient les résultats suivants :

Fichier personnes protégées (EPAS) :

- 11 645 lignes ayant un mois et une année de naissance renseignés
- 6 796 correspondant à un Assuré principal (benran = 1),
- 6 511 Assurés principaux (on ne retient qu'une ligne par individu - ligne correspondant à la date de dernière utilisation la plus récente).

Fichier personne (ESPS) :

- 9 554 lignes avec un mois et une année de naissance renseignés dont 9 092 lignes correspondant à des personnes protégées par le Régime Général ou l'un des régimes rattachés
- 3 088 assurés principaux (nind = 1).

La fusion de ces deux fichiers d'assurés principaux selon le n1 donne 3 073 individus, 15 n1 n'ont pas été appariés. Ces derniers existent dans ESPS mais n'existent pas dans l'EPAS. Parmi ces 3 073 individus, 52 ont soit le sexe, soit l'année de naissance différents dans les deux fichiers. Ceci nous donnerait donc 3 021 assurés principaux appariés.

Remarques :

Si l'on compare aux résultats de participation²² de l'ESPS en 1992 :

- 6 517 adresses ont été exploitées par l'ESPS contre 6 511 assurés principaux ayant une année de naissance renseignée dans l'EPAS
- 3 266 acceptations contre 3 088 assurés principaux assurés au Régime Général ou à l'un des régimes rattachés et ayant une année de naissance renseignée dans l'ESPS

III IDENTIFICATION DES AYANTS DROIT

La seule façon d'identifier un ayant droit est de considérer son n1 et sa date de naissance. Deux méthodes ont été utilisées :

1ère méthode :

On ne retient que les ayants droit dans les deux fichiers c'est-à-dire :

- benran \neq 1 pour l'EPAS
- nind \neq 1 et ss \neq 1 pour l'ESPS

Ce qui donne pour les effectifs :

- EPAS :
 - 11 645 lignes ayant 1 mois et une année de naissance renseignés
 - 4 854 lignes (en enlevant benran = 1 et en ne conservant que benrng = 1)
 - 4 792 ayants droit (en ne conservant qu'une ligne par individu - ligne correspondant à la date d'utilisation la plus récente)
- ESPS :
 - 9 554 lignes dont 9 092 lignes correspondant à des assurés au Régime Général (ou rattachés) ayant un mois et une année de naissance renseignés
 - 3 005 ayants droit (en enlevant nind = 1 et ss = 1)

La fusion de ces deux fichiers donne :

- a) avec la date de naissance entière (jour + mois + année) : 1 823 ayants droit
- b) avec le mois et l'année de naissance : 1 887 ayants droit
- c) avec uniquement l'année de naissance : 1 936 ayants droit

2ème méthode :

On exclut seulement les assurés principaux de l'ESPS c'est-à-dire qu'on garde en plus de la méthode précédente les assurés autres que principaux.

Les effectifs de départ deviennent :

- EPAS : 4 792 ayants droit (inchangé)
- ESPS : 6 004 assurés non principaux ou ayants droit (nind \neq 1)

²² Cf. santé, soins et protection sociale en 92, biblio n° 982, CREDES, 1993.

La fusion de ces deux fichiers donne :

- a) avec la date de naissance complète : 2 260 individus
- b) avec le mois et l'année de naissance : 2 356 individus
- c) avec l'année de naissance seule : 2 425 individus

CONCLUSIONS DES DEUX MÉTHODES :

- 1- Il semble que le mois et l'année de naissance permettent d'identifier assez bien un individu; il faut noter au passage que le jour de naissance semble très mal renseigné, surtout dans l'EPAS où il est très souvent forcé à 1 ou 0.
- 2- Si l'on prend comme identifiant d'un individu le n1, le mois et l'année de naissance, la comparaison des deux méthodes permet de mettre en évidence que 469 personnes²³ sont ayants droit dans l'EPAS et assurés non principaux dans l'ESPS (il semble s'agir de conjoints ou d'enfants passés au régime étudiant).

Cette méthode qui consiste à appairer séparément les assurés principaux de leurs ayants droit en partant de l'hypothèse que la qualité de bénéficiaire (assuré/ayant droit) déclaré par l'enquêté et celui relevé dans l'EPAS sont identiques exclut la possibilité d'appairer une personne se déclarant par exemple assurée dans l'ESPS alors qu'elle est en fait ayant droit au regard de la CNAMTS.

IV - EST-IL POSSIBLE D'AVOIR LE MÊME IDENTIFIANT (N1 + ANNÉE DE NAISSANCE + MOIS DE NAISSANCE) POUR DES INDIVIDUS DIFFÉRENTS ?

Il existe effectivement un certain nombre d'individus ayant un même identifiant ; les n1 étant uniques, il s'agit de mois et d'années de naissances identiques.

a) Cas des jumeaux, triplés et quadruplés

Dans ce cas, le jour de naissance est lui aussi identique.

Dans l'EPAS, il existe une variable "benrng" qui attribue un rang aux jumeaux ; dans l'ESPS, rien ne permet par contre de les identifier. Dans l'ESPS, on relève :

- 30 cas de jumeaux,
- 1 cas de triplés,
- 1 cas de quadruplés.

- cas de deux jumeaux de sexes différents : on va affecter les mêmes rangs que la variable "benrng" dans l'EPAS,
- cas de jumeaux de même sexe, de triplés ou de quadruplés : on va affecter de façon aléatoire les rangs de la variable "benrng" (on n'est donc pas sûr d'avoir attribué le rang 1 dans l'ESPS au jumeau qui a le rang 1 dans l'EPAS puisque rien ne permet de les différencier).

²³ 469 = 2356 - 1887 (cf page précédente)

b) Cas de deux personnes (souvent deux conjoints) nées le même mois de la même année

On note 15 cas de ce type. Les jours de naissance les différencient, on va attribuer aussi bien dans l'EPAS que dans l'ESPS, une valeur du rang (choisie à 8) pour l'une de ces deux personnes.

c) Pour tous les autres individus, la valeur du rang est mise à 1, aussi bien dans le fichier des personnes protégées de l'EPAS que dans le fichier personne de l'ESPS (le tri à plat de la variable "benrng" ayant révélé des modalités étranges et inexpliquées du rang - 0, 9 - en quantité non négligeable, nous avons jugé préférable de les mettre à 1).

CONCLUSION :

- **La méthode retenue** pour l'appariement consiste à créer dans les deux fichiers un identifiant individuel commun et à apparier les deux fichiers sans prendre en compte la qualité de bénéficiaire des personnes (assurés/ayants droit). Plusieurs identifiants ont été progressivement testés lors de l'appariement réalisé sur les données de l'année 1988. L'identifiant retenu pour une personne est constitué de :
 - *n1* : numéro de ménage
 - *mois et année de naissance* (le jour de naissance n'a pas été pris en compte car très souvent manquant ou forcé à 1 dans l'EPAS)
 - un *rang* dans le ménage, créé de façon identique dans les deux fichiers et essentiellement destiné à distinguer les jumeaux et les conjoints nés le même mois de la même année (30 cas de jumeaux, 1 cas de triplés, 1 cas de quadruplés et 15 cas de ménages où deux personnes sont nées le même mois de la même année)
 - le *sexe*, ajouté en dernier lieu pour éviter d'avoir des personnes appariées, mais ayant un sexe différent dans les deux fichiers.

L'identifiant retenu est :

$\text{num} = n1 + \text{année de naissance} + \text{mois de naissance} + \text{rang} + \text{sexe}$
--

On dispose de deux dates de naissance, la seconde étant, selon des informations recueillies au département statistique de la CNAMTS, celle à considérer. Celle-ci présente toutefois de nombreuses valeurs manquantes ; afin d'améliorer les résultats, ces valeurs manquantes ont été remplacées par la première date de naissance.

V - FUSION DU FICHIER DES PERSONNES PROTÉGÉES (EPAS) AVEC LE FICHIER PERSONNE (ESPS) EN UTILISANT L'IDENTIFIANT PRÉCÉDEMMENT DÉFINI

- Fichier personnes protégées : 11 645 lignes

Une fois l'identifiant constitué, 675 lignes avec le même identifiant ont été relevées, dont 324 doublons et 9 triplets.

Il existe dans le fichier une date de dernière utilisation correspondant à la dernière fois où une modification a été faite dans l'enregistrement d'un individu. On observe deux cas de figure :

- * ou bien la date de dernière utilisation la plus récente est antérieure à 1993 : on retient alors la plus récente
- * ou bien la date de dernière utilisation la plus récente se situe en 1993 ou 1994 :
s'il s'agit de modification concernant le jour de naissance, i.e. si parmi les lignes doubles dans l'EPAS, on constate une modification sur le jour de naissance, on retient la ligne sur laquelle le jour de naissance coïncide avec celui relevé dans l'ESPS ;
s'il ne s'agit pas d'une modification sur le jour de naissance, on retient arbitrairement une seule de ces lignes.

Soit 342 lignes ainsi éliminées.

On obtient alors 11 303 lignes.

96 parmi les 11 303 individus sont étrangers et sont repérés par leur mois de naissance ayant la valeur 50. 10 parmi ces 96 individus ont l'année et le jour de naissance identiques dans l'ESPS, c'est pourquoi on leur attribue le mois de naissance d'ESPS.

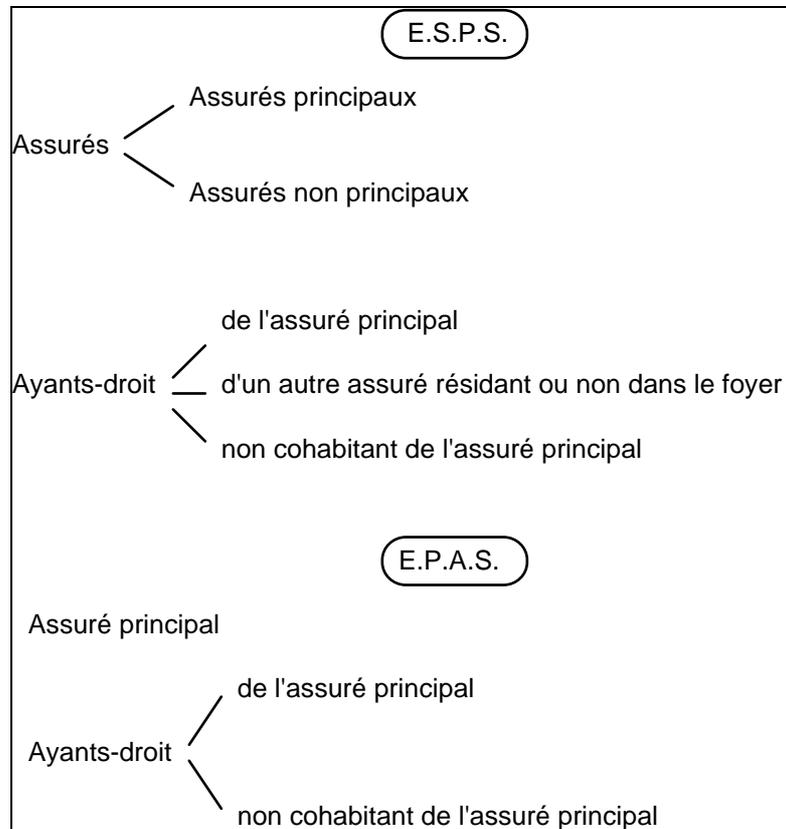
En enlevant les dates de naissances inconnues et en ne conservant qu'une ligne par individu (ligne correspondant toujours à la date la plus récente de dernière utilisation), on obtient 11 303 individus.

- Fichier (personne + ayant droit non cohabitants) : 9 563 lignes
dont 9 554 ayant un mois et une année de naissance renseignés
dont 9 092 appartenant à un ménage où l'assuré principal déclare être au Régime Général ou à l'un des régimes rattachés.

La fusion nous donne :

5 334 individus (au sens de num) communs 3 758 individus non trouvés dans l'EPAS 5 969 individus non trouvés dans l'ESPS
--

Quels sont les différents cas de figure ?



Parmi les 3 758 individus non trouvés dans l'EPAS

2 531 (soit 68,1%) sont assurés non principaux.

825 (soit 22,2%) sont ayants droit d'un assuré non principal.

3 (soit 0,1%) sont non assurés.

Soit 90,4% des individus qu'il est normal de ne pas trouver dans l'EPAS puisqu'ils ne sont ni assurés principaux, ni ayants droit de l'assuré principal.

92 (soit 1,9%) assurés principaux.

307 (soit 8,2%) ayants droit de l'Assuré principal

Soit 399 individus qu'il est anormal de ne pas trouver dans l'EPAS et pour lesquels l'identifiant Num n'existe pas dans l'EPAS. Pour ces 399 individus, l'appariement par rapport au N1 et l'année de naissance a été tenté.

Parmi les 399, 320 ne s'apparient pas du fait d'une incohérence de l'année de naissance avec l'ESPS ou d'une absence d'individu.

Les 79 restants s'apparient par rapport au n1 et l'année de naissance ; parmi ces 79 grappes :

* **19** assurés (étrangers) ont un mois de naissance à 50 dans l'EPAS. On leur affecte le mois de naissance recueilli dans l'ESPS

* 42 assurés ont une date de naissance différente entre les deux sources de données :

- **18** cas ont une année et un jour de naissance identiques mais un mois de naissance différent de 1 ; on pose l'hypothèse qu'il s'agit des mêmes individus et on leur affecte le mois de naissance recueilli dans l'enquête SPS

- 9 cas ont une année et un jour de naissance identiques mais un mois de naissance différent de 2

- 5 cas ont une année et un jour de naissance identiques mais un mois différent de plus

de

2

- **2** cas ont une année de naissance identique mais une inversion du jour et du mois de naissance. On pose l'hypothèse qu'il s'agit des mêmes individus et on leur affecte les renseignements recueillis dans l'enquête SPS

- 3 cas ont une année de naissance identique mais un mois et un jour de naissance différents.

- 5 autres cas.

* 17 cas ont un sexe différent entre EPAS et ESPS (dont 1 assuré principal).

* 1 cas où le rang diffère entre EPAS et ESPS (en fait il s'agit de jumeaux dans l'enquête SPS et d'un seul individu ayant ce N1 et cette date de naissance dans l'EPAS).

==> Soit 39 individus (correspondant aux cas en gras) ayant des renseignements proches entre les deux sources de données et qu'on apparie en leur affectant les renseignements recueillis dans l'enquête SPS.

Parmi les 5 969 individus non trouvés dans ESPS

3 513 sont assurés principaux (58,8%).

2 456 sont ayants droit (41,2%).

5 320 individus correspondent à des ménages n'ayant pas participé à l'enquête (3 270 grappes assurés concernées).

Compte tenu des choix précédents, l'appariement donne :

5 373 individus (au sens de num) communs

3 719 individus non trouvés dans l'EPAS

5 930 individus non trouvés dans l'ESPS

Parmi ces 5 373 individus, certains ont comme régime de Sécurité Sociale un régime particulier (fonctionnaires et ouvriers de l'Etat, agents EDF-GDF...). Etant donné que les caisses primaires d'assurance maladie ne gèrent pas la totalité de ces mutuelles décompteuses, celles ci sont exclues de notre population appariée.

Bilan de l'appariement des personnes :

4 906 individus communs (au sens de l'identifiant retenu)

3 719 individus non trouvés dans l'EPAS mais présents dans l'ESPS

5 270 individus non trouvés dans l'ESPS mais présents dans l'EPAS

Annexe 2

Redressement des données

Méthode : linéaire
Statistiques sur les rapports de poids (= pondérations finales / pondérations initiales)
et sur les pondérations finales

Univariate Procedure

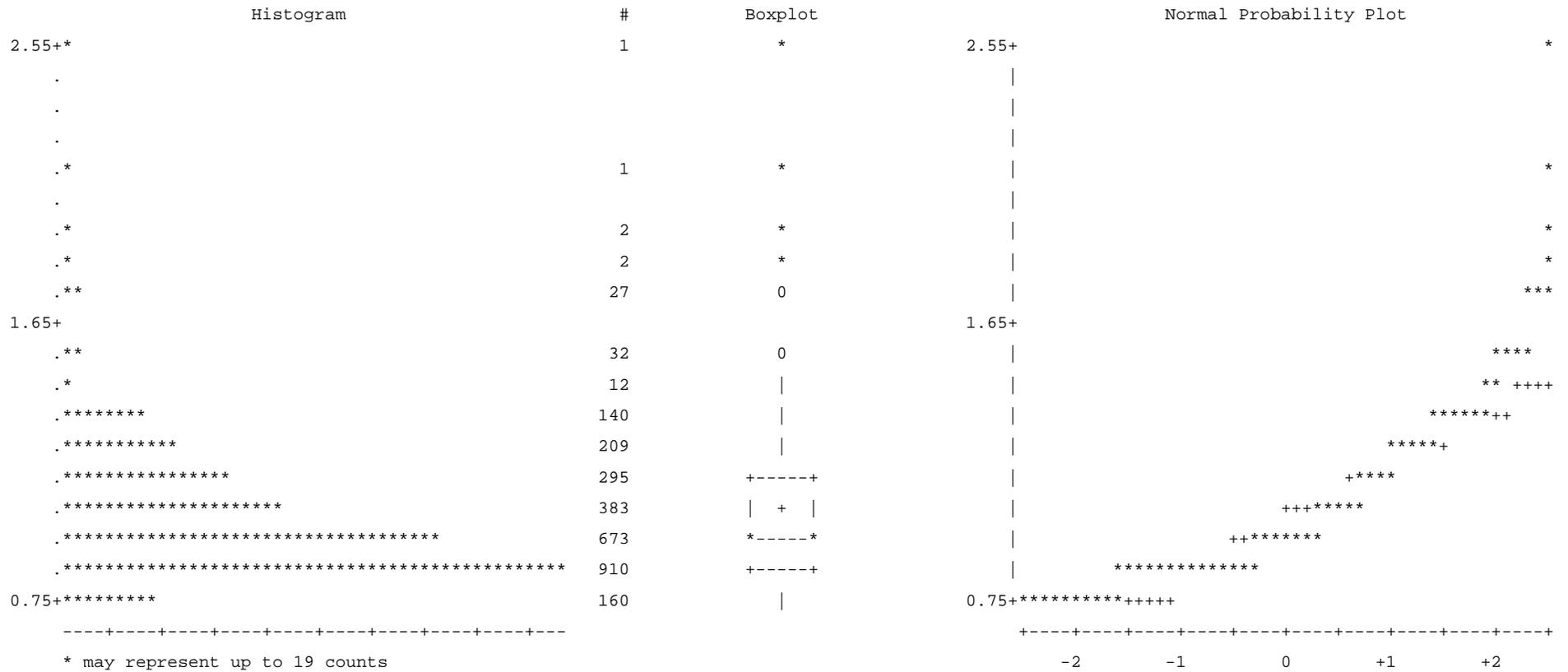
Variable=_F_ Rapport de poids

Moments				Quantiles(Def=5)				Extremes			
N	2847	Sum Wgts	2847	100% Max	2.32031	99%	1.61261	Lowest	ID	Highest	ID
Mean	1.001054	Sum	2850	75% Q3	1.108774	95%	1.302142	0.722714(46171541)		1.971448(02544561)	
Std Dev	0.186103	Variance	0.034635	50% Med	0.949142	90%	1.261179	0.722714(45886521)		2.008273(41371681)	
Skewness	1.248135	Kurtosis	2.556876	25% Q1	0.867956	10%	0.808355	0.722714(40927521)		2.008273(44091661)	
USS	2951.573	CSS	98.56979	0% Min	0.722714	5%	0.773724	0.722714(04362561)		2.089814(03511301)	
CV	18.59076	Std Mean	0.003488			1%	0.742982	0.722714(03963541)		2.32031(00213201)	
T:Mean=0	287.0098	Pr> T	0.0001	Range	1.597596						
Num ^= 0	2847	Num > 0	2847	Q3-Q1	0.240817						
M(Sign)	1423.5	Pr>= M	0.0001	Mode	1.141994						
Sgn Rank	2027064	Pr>= S	0.0001								
D:Normal	0.148052	Pr>D	<.01								

Méthode : raking ratio
Statistiques sur les rapports de poids (= pondérations finales / pondérations initiales)
et sur les pondérations finales

Univariate Procedure

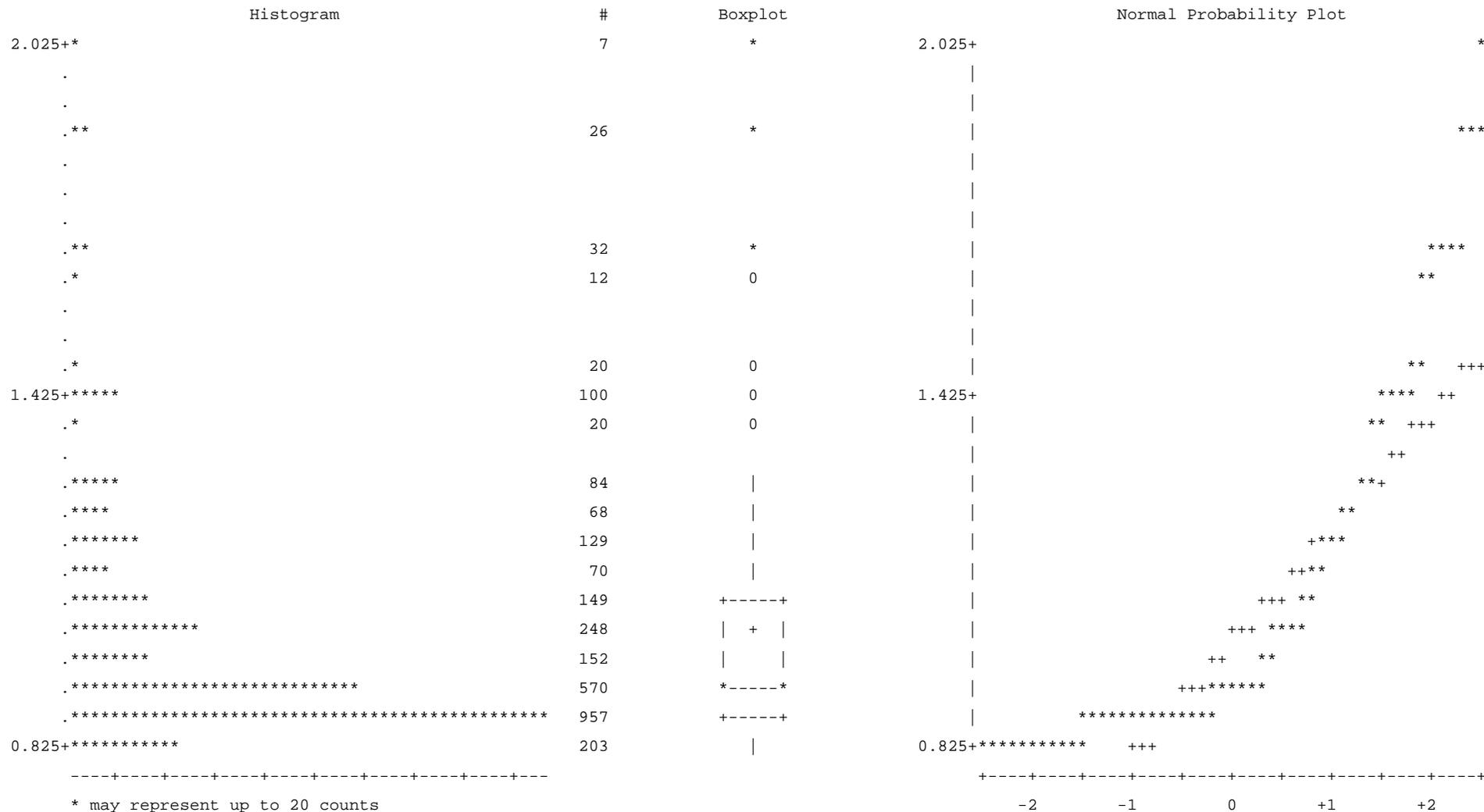
Variable=_F_		Rapport de poids						Extremes			
Moments				Quantiles(Def=5)							
N	2847	Sum Wgts	2847	100% Max	2.593879	99%	1.72536	Lowest	ID	Highest	ID
Mean	1.001054	Sum	2850	75% Q3	1.112878	95%	1.318456	0.746436(46171541)		1.89736(02544561)	
Std Dev	0.187014	Variance	0.034974	50% Med	0.948727	90%	1.241524	0.746436(45886521)		1.982145(41371681)	
Skewness	1.611777	Kurtosis	4.501907	25% Q1	0.874133	10%	0.819266	0.746436(40927521)		1.982145(44091661)	
USS	2952.539	CSS	99.5362	0% Min	0.746436	5%	0.789613	0.746436(04362561)		2.149771(03511301)	
CV	18.68167	Std Mean	0.003505			1%	0.76286	0.746436(03963541)		2.593879(00213201)	
T:Mean=0	285.613	Pr> T	0.0001	Range	1.847443						
Num ^= 0	2847	Num > 0	2847	Q3-Q1	0.238745						
M(Sign)	1423.5	Pr>= M	0.0001	Mode	1.124416						
Sgn Rank	2027064	Pr>= S	0.0001								
D:Normal	0.154358	Pr>D	<.01								



Méthode : logit, inf=0.83, sup=2.05
Statistiques sur les rapports de poids (= pondérations finales / pondérations initiales)
et sur les pondérations finales

Univariate Procedure

Variable=_F_		Rapport de poids		Moments		Quantiles(Def=5)			Extremes		
N	2847	Sum Wgts	2847	100% Max	2.047796	99%	1.850128	Lowest	ID	Highest	ID
Mean	1.001054	Sum	2850	75% Q3	1.060896	95%	1.418843	0.837933(46171541)		2.030347(02399621)	
Std Dev	0.196823	Variance	0.038739	50% Med	0.923484	90%	1.270043	0.837933(45886521)		2.038669(41371681)	
Skewness	2.165581	Kurtosis	5.345066	25% Q1	0.873528	10%	0.854937	0.837933(40927521)		2.038669(44091661)	
USS	2963.255	CSS	110.2515	0% Min	0.837933	5%	0.847318	0.837933(04362561)		2.044364(03511301)	
CV	19.66154	Std Mean	0.003689			1%	0.840601	0.837933(03963541)		2.047796(00213201)	
T:Mean=0	271.379	Pr> T	0.0001	Range	1.209863						
Num ^= 0	2847	Num > 0	2847	Q3-Q1	0.187369						
M(Sign)	1423.5	Pr>= M	0.0001	Mode	1.075201						
Sgn Rank	2027064	Pr>= S	0.0001								
D:Normal	0.211511	Pr>D	<.01								



Méthode : linéaire tronquée, inf=0.8, sup=2.3
Statistiques sur les rapports de poids (= pondérations finales / pondérations initiales) et sur les pondérations finales

Univariate Procedure

Variable=_F_ Rapport de poids

Moments				Quantiles(Def=5)				Extremes			
N	2847	Sum Wgts	2847	100% Max	2.3	99%	1.624285	Lowest	ID	Highest	ID
Mean	1.001054	Sum	2850	75% Q3	1.120284	95%	1.312053	0.8(45663401)		1.966754(02544561)	
Std Dev	0.186696	Variance	0.034855	50% Med	0.949113	90%	1.263655	0.8(45583401)		2.013862(41371681)	
Skewness	1.307991	Kurtosis	2.552945	25% Q1	0.860534	10%	0.8	0.8(45478421)		2.013862(44091661)	
USS	2952.202	CSS	99.19873	0% Min	0.8	5%	0.8	0.8(44876481)		2.103918(03511301)	
CV	18.64997	Std Mean	0.003499			1%	0.8	0.8(44573461)		2.3(00213201)	
T:Mean=0	286.0985	Pr> T	0.0001	Range	1.5						
Num ^= 0	2847	Num > 0	2847	Q3-Q1	0.259749						
M(Sign)	1423.5	Pr>= M	0.0001	Mode	0.8						
Sgn Rank	2027064	Pr>= S	0.0001								
D:Normal	0.140762	Pr>D	<.01								

Caractéristiques des 10 groupes de rapports de poids (pondérations finales/pondérations initiales) des quatre fonctions de distance

	Moyenne	Ecart-type	Minimum	Maximum
Méthode linéaire				
Groupe 1	1.3400709	0.0451757	1.2527034	1.4011488
Groupe 2	1.4417475	0.0216958	1.4011488	1.4724596
Groupe 3	1.5058432	0.0169122	1.4724596	1.5324869
Groupe 4	1.5549709	0.0184756	1.5324869	1.5869774
Groupe 5	1.6162579	0.0173072	1.5869774	1.6451796
Groupe 6	1.6739641	0.0190721	1.6451796	1.7094301
Groupe 7	1.7724433	0.0351917	1.7094301	1.866143
Groupe 8	1.9522326	0.0613756	1.866143	2.054706
Groupe 9	2.1087827	0.0390806	2.054706	2.186044
Groupe 10	2.3922013	0.2634436	2.186044	4.0218701
Méthode raking ratio				
Groupe 1	1.3682537	0.0387597	1.2938224	1.4200609
Groupe 2	1.4523627	0.0188882	1.4200609	1.4840217
Groupe 3	1.5101597	0.0159326	1.483976	1.5397527
Groupe 4	1.5545747	0.0156776	1.5318315	1.5855514
Groupe 5	1.6127059	0.0168475	1.5855514	1.6444593
Groupe 6	1.6671573	0.0183735	1.642686	1.7043755
Groupe 7	1.7675629	0.0338888	1.7043755	1.8202768
Groupe 8	1.9174992	0.0583381	1.8148009	1.9924868
Groupe 9	2.0825517	0.044317	1.9846928	2.1519754
Groupe 10	2.4260428	0.310596	2.1519754	4.4960563
Méthode logit (0.83,2.05)				
Groupe 1	1.4904939	0.0114149	1.4692593	1.5126602
Groupe 2	1.46914	0.0098991	1.4524164	1.4825362
Groupe 3	1.515239	0.0133306	1.495513	1.5457626
Groupe 4	1.5387138	0.0127094	1.5144852	1.5621019
Groupe 5	1.5799842	0.0217787	1.5433261	1.6188925
Groupe 6	1.6216316	0.0253608	1.585724	1.6750374
Groupe 7	1.7416252	0.0497709	1.6380095	1.819688
Groupe 8	1.8088944	0.1044892	1.6380095	1.9892997
Groupe 9	2.0392338	0.1067075	1.7787589	2.2014077
Groupe 10	2.5552751	0.3457816	2.0743932	3.5495131
Méthode linéaire tronquée (0.8,2.3)				
Groupe 1	1.3866667	0	1.3866667	1.3866667
Groupe 2	1.4159624	0.0234584	1.3866667	1.4607715
Groupe 3	1.4875244	0.0219167	1.4493114	1.532217
Groupe 4	1.5428163	0.021416	1.5006125	1.5791435
Groupe 5	1.61249	0.0207664	1.5736844	1.649133
Groupe 6	1.674461	0.0223096	1.6451299	1.7247804
Groupe 7	1.7833106	0.03863	1.7205638	1.8416964
Groupe 8	1.9424224	0.0629134	1.8330128	2.0327948
Groupe 9	2.1091423	0.0429636	2.0327948	2.190335
Groupe 10	2.4038408	0.2628316	2.190335	3.9866667

Glossaire

EPAS :

Echantillon Permanent d'Assurés Sociaux de la CNAMTS.

ESPS :

Enquête Santé et Protection Sociale du CREDES.

Assuré Principal (AP) :

C'est l'assuré tiré au sort dans l'EPAS pour participer à l'enquête SPS.

Assuré non Principal (A.N.P.) :

Tous les assurés autres que principaux, c'est-à-dire non tirés au sort, et faisant partie du ménage d'un assuré principal.

Grappe Assuré :

C'est l'assuré principal et ses ayants droit.

Liquidation :

Système de règlement et d'enregistrement des prestations

Prestataire :

Un prestataire est une personne qui a perçu au moins une prestation relative à des soins effectués en 92 et remboursés en 92 ou 93. Le fichier des prestataires est appelé fichier des "consommants" par la CNAMTS.

Liste des tableaux et des schémas

Tableau n° 1	
Qualité de bénéficiaire observée des personnes appariées dans l'ESPS et dans l'EPAS.....	16
Tableau n° 2	
Motifs de non réponse.....	18
Tableau n° 3	
Ventilation des personnes présentes dans les fichiers initiaux à l'issue de l'appariement.....	19
Tableau n° 4	
Taux de prestataires selon l'âge, le sexe et la qualité de bénéficiaire.....	25
Tableau n° 5	
Situation administrative estimée des personnes appariées dans l'ESPS et dans l'EPAS.....	26
Tableau n° 6	
Structure démographique des personnes protégées et prestataires.....	27

Schéma n° 1	
La grappe assuré	7
Schéma n° 2	
Les personnes des ménages enquêtés dans l'ESPS	8
Schéma n° 3	
Personnes théoriquement présentes dans l'EPAS et l'ESPS (1992)	14
Schéma n° 4	
Récapitulatif de l'appariement des prestations.....	24

Achevé d'imprimer le : décembre 1996

Nom et adresse de l'imprimeur

Dépot légal : décembre 1996