

DÉPENSES DE SANTÉ ET RESTES À CHARGE DES MÉNAGES : LE MODÈLE DE MICROSIMULATION OMAR

Rémi Lardellier, Renaud Legal, Denis Raynaud, Guillaume Vidal ()¹*

()Drees, bureau Dépenses de santé et Relations avec l'Assurance Maladie*

Introduction

La connaissance de la structure de financement des dépenses de santé est un enjeu d'importance. Le reste à charge permet de mesurer l'accessibilité financière des soins ou d'apprécier la solidarité entre malades et bien portants induite par le système d'assurance obligatoire.

Grâce aux Comptes de la santé, la structure de financement et la dynamique des dépenses de santé sont bien connues au niveau macroéconomique. Ainsi, en 2009, la Sécurité sociale a financé 75.5 % des dépenses de santé, les organismes complémentaires 13.8 %, les ménages 9.4 % et l'État 1.3 % au travers essentiellement de la CMU-C (source : Comptes de la santé).

En revanche, il n'existe pas à ce jour de sources d'information permettant de décliner l'exercice au niveau microéconomique. En effet, les données de l'assurance maladie renseignent uniquement sur la dépense individuelle présentée au remboursement et la part remboursée par l'assurance maladie obligatoire. Ces données ne permettent pas d'estimer la part de cette dépense restant à la charge de l'individu et de son ménage après intervention des organismes complémentaires, qui couvrent 93.9 % de la population. Les données des organismes complémentaires sont quant à elles souvent incomplètes et ne permettent pas d'avoir une vue d'ensemble, représentative pour l'ensemble des ménages, des remboursements qu'elles prennent en charge. Quant aux cas-types parfois mis en avant pour analyser la structure du financement des dépenses de santé et le reste à charge pour différents types de ménages, ils ont le défaut majeur de ne pas tenir compte de la fréquence réelle des cas décrits dans la population.

Pourtant, les structures de financement de la dépense de santé observées au niveau agrégé n'ont aucune raison d'être les mêmes pour tous les ménages, en raison notamment de l'hétérogénéité de la population face au risque maladie d'une part, et face à sa prise en charge par la Sécurité sociale et les organismes complémentaires d'autre part. Aussi, le reste à charge des ménages est-il probablement distribué de façon très hétérogène au sein de la population.

Afin de pallier l'absence de données disponibles, la Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (Drees) du Ministère de la Santé et des Sports, en collaboration avec la Direction Générale du Trésor du ministère de l'Économie, de l'Industrie et de l'Emploi a construit un outil de microsimulation (Omar, Outil de Microsimulation pour l'Analyse des Restes à charge) permettant d'apprécier la structure de financement des dépenses de santé au niveau microéconomique.

Pour ce faire, la Drees s'est appuyé sur l'échantillon issu de l'édition 2006 de l'enquête Santé et Protection Sociale conduite par l'Irdes, apparié avec les données de remboursement des caisses d'assurance maladie. Toutefois, pour être en mesure d'effectuer le partage entre assurance maladie obligatoire et assurance maladie complémentaire au niveau individuel, il a fallu compléter ces données avec des informations de garanties de couverture complémentaire. Pour cela, les

¹ Ce travail a bénéficié de présentations au séminaire microsimula de Paris School of Economics, au séminaire 3S de la Drees... Ces travaux font l'objet d'une collaboration avec l'Irdes et la DG-Trésor. Les auteurs remercient particulièrement Laurent Davezies, Valérie Albouy, Bérengère Mesqui, Claire Montialoux, Grégoire de Lagasnerie, Pierre-Yves Geoffard et Thierry Debrand pour leurs conseils et remarques méthodologiques, Anne-Marie Brocas et Lucile Olier pour leurs relectures attentives de l'article. Les auteurs restent évidemment seuls responsables des limites ou inexactitudes de leur travail.

informations tirées de l'enquête sur les contrats d'assurance complémentaire santé modaux de 2008 ont également été mobilisées. En fin de compte, l'application de ces garanties aux dépenses de santé constatées (ou imputées) a permis de simuler le cas échéant le montant remboursé par l'assurance complémentaire au niveau individuel et donc d'avoir une vision complète de la structure de financement au niveau microéconomique.

En raison d'un nombre important de valeurs ou de données manquantes, la construction d'Omar a nécessité de nombreuses imputations : imputation des revenus par équations d'imputations et résidus simulés, imputation de dépenses de santé et imputation de contrats d'assurance complémentaire par *hot deck* stratifié. Plus généralement, de nombreuses options méthodologiques ont été prises lors de la construction de cet outil. Le présent article est un article méthodologique qui les expose dans le détail.

L'article se compose de cinq parties. La première d'entre elles revient sur l'enjeu de la démarche. La deuxième partie décrit les données et leurs limites. La troisième présente l'intérêt et le déroulement de la méthode d'imputation des dépenses de santé. Une quatrième partie vient détailler les différentes étapes de l'imputation de contrats complémentaires qui permettent, *in fine*, de microsimuler les remboursements complémentaires. Enfin, la dernière partie, à travers la présentation d'un résultat - le montant des remboursements complémentaires par décile de niveau de vie - est l'occasion de tester la sensibilité des résultats à la méthode d'imputation des contrats et de montrer à travers le calcul de la structure de financement des dépenses de santé par décile de niveau de vie un exemple d'utilisation de ces travaux de microsimulation.

1. Enjeu de la démarche

1.1. Ce que nous dit la distribution des restes à charge sur notre système de santé

En France, la participation à l'assurance maladie publique est obligatoire, d'où l'appellation « régime obligatoire ». La philosophie d'ensemble de ce régime est souvent résumée par l'expression : « contribution de chacun selon ses moyens, couverture de chacun selon ses besoins ». Les cotisations maladie sont, en effet, assises sur les revenus, sans lien *a priori* avec le niveau de risque individuel et le niveau des remboursements. Les niveaux de remboursement des soins ne dépendent pas, en revanche, de la situation financière du patient, mais du degré de morbidité supposé et de la lourdeur conséquente de la prise en charge afférente (taux de remboursement de 80 % des frais hospitaliers, de 70 % des honoraires de médecins, *etc.*), du service médical rendu (médicaments à 35 % ou 65 %, 100%, *etc.*) et - le cas échéant - de la situation particulière du patient au regard des mécanismes d'exonération au moment des soins (ALD, hospitalisation de plus de 30 jours, maternité...).

Plus concrètement, on peut dire que le système d'assurance obligatoire français poursuit par essence deux objectifs distincts : un objectif de solidarité entre bien portants et malades (centré sur le coût de la maladie indépendamment des revenus) et un objectif plus englobant d'accessibilité financière des soins. En outre, dans une perspective dynamique, chaque individu pouvant au cours de sa vie connaître des problèmes de santé, le système d'assurance publique et obligatoire a aussi pour objectif de procurer une assurance de long terme contre le risque maladie ([21]).

Le reste à charge après remboursement par les seuls régimes de base, est pertinent pour apprécier la solidarité entre bien portants et malades induite par le système d'Assurance Maladie Obligatoire (AMO). Dans un système visant un niveau élevé de solidarité, les personnes les plus malades ne devraient pas avoir de dépenses sensiblement plus élevées que les autres personnes.

Le reste à charge qui intègre non seulement les remboursements de l'AMO mais aussi les remboursements des organismes complémentaires (Assurance Maladie Complémentaire, AMC) permet d'apprécier le second objectif, celui d'accessibilité financière des soins. On peut souhaiter que ce reste à charge soit faible pour les individus aux faibles revenus, ou que corolairement le taux d'effort correspondant soit raisonnable.

Les données de l'assurance maladie permettent d'apprécier le reste à charge après AMO seulement. Par contre, elles n'offrent que peu d'informations sur les individus, hormis leur âge, leur sexe, le fait qu'ils bénéficient ou non de la CMU-C ou du dispositif ALD. De ce point de vue, l'appariement Epas-

SPS (présenté ci-dessous) permet une analyse plus riche, en mobilisant les informations sociodémographiques (CSP, niveau d'études, état de santé déclaré...) de l'enquête SPS. Par contre, cet appariement ne permet pas d'opérer le partage du RAC après AMO entre remboursements OC et RAC après AMO et AMC : la vision du RAC après AMO et AMC n'existe donc pas à ce jour. Pourtant, de par son poids croissant aujourd'hui, l'intervention des organismes complémentaires ne peut rester absente des études portant sur notre système de santé.

Pour apporter un éclairage aux questions d'accessibilité financière des soins et illustrer à travers un exemple les utilisations potentielles des travaux de microsimulation détaillés dans cet article, nous présenterons en conclusion la structure de financement des dépenses de santé (AMO, AMC, reste à charge des ménages) par décile de niveau de vie.

1.2. La nécessaire analyse du rôle des couvertures maladie complémentaires

1.2.1. La couverture complémentaire santé est un déterminant important de l'accès aux soins

Les études empiriques, notamment les études américaines appuyées sur l'expérience menée par la Rand Corporation dans les années 70 ([30]) ou les études sur l'extension du programme Medicaid ([14]) ont, de longue date, démontré le rôle déterminant de l'assurance santé dans l'accès aux soins, à état de santé donné. En allégeant la contrainte financière que la consommation de soins exerce sur le budget des ménages, l'assurance santé permet de réduire les disparités de consommation de soins entre les ménages modestes et les autres et d'éviter notamment une sous-consommation de soins au regard de l'état de santé.

En France, de nombreux travaux empiriques ont mis en évidence un effet positif de la couverture complémentaire sur le recours aux soins ([10], [9], [16], [32], [33], [5]). Les dépenses de spécialistes, comme de soins ambulatoires, des personnes disposant d'une assurance complémentaire santé sont plus élevées toutes choses égales par ailleurs, et notamment à état de santé identique. La couverture complémentaire joue un rôle important dans l'accès des ménages modestes à des prestations peu remboursées par la Sécurité sociale, comme les dépenses d'optique et en soins dentaires ([26]).

C'est pourquoi les pouvoirs publics se sont efforcés de favoriser l'accès à la couverture universelle, notamment des plus modestes. Les études existantes confirment en effet que l'absence de couverture complémentaire dépend essentiellement du niveau des revenus et du milieu social ([25]). Globalement, le taux de couverture par une complémentaire santé augmente avec le niveau de vie (*cf.* Graphique 1).

La mise en place de la Couverture Maladie Universelle Complémentaire (CMU-C), sous condition de ressources, instaurée par la loi du 27 juillet 1999, a représenté à cet égard un progrès décisif (encadré n°1). Plus récemment, l'Aide à l'acquisition d'une Complémentaire Santé mise en place le 1^{er} janvier 2005 vise à faciliter l'accès des ménages dont les revenus dépassent légèrement le seuil d'éligibilité à la CMU-C à une complémentaire santé, en prenant en charge une partie des primes d'assurance.

Si la CMU-C et l'ACS permettent d'atténuer les inégalités en offrant une couverture complémentaire aux plus démunis, la proportion de personnes sans couverture complémentaire demeure toutefois bien plus élevée dans les premiers déciles de niveau de vie (20 % dans le premier décile contre 3 % dans le dernier).

Encadré 1 - La CMU-C et l'Aide à l'acquisition d'une Complémentaire Santé

La Couverture Maladie Universelle Complémentaire (CMU-C), instaurée par la loi du 27 juillet 1999, permet d'accorder une couverture complémentaire gratuite aux Français et aux étrangers en situation régulière. Ce dispositif sous conditions de ressources permet la prise en charge avec dispense d'avance de frais du ticket modérateur, du forfait journalier et des frais supplémentaires concernant les prothèses dentaires, l'orthopédie dento-faciale et certains dispositifs médicaux à usage individuel (lunettes, audio-prothèses...). Par ailleurs, pour les patients concernés par ce dispositif, les professionnels de santé ont l'obligation de respecter les tarifs reconnus par la Sécurité sociale. Financé intégralement par les organismes complémentaires depuis 2009, le dispositif comptait 4,1 millions de bénéficiaires au 31 décembre 2009 pour les 3 grands régimes d'assurance maladie.

L'enquête Santé et Protection Sociale permet de connaître les caractéristiques sociodémographiques des bénéficiaires de la CMU-C. Il s'agit d'une population jeune, dont un tiers appartient à une famille monoparentale, issue d'un milieu social modeste. Les bénéficiaires de la CMU-C déclarent un moins bon état de santé que le reste de la population du même âge. Le taux de personnes souffrant d'une affection de longue durée (ALD) est d'ailleurs plus élevé parmi les bénéficiaires de la CMU-C que parmi les personnes ayant une couverture privée, à âge équivalent ([5], [6], [7]).

Par rapport à l'absence de couverture, la CMU-C permet de diminuer le renoncement aux soins pour raisons financières ([8]). « Toutes choses égales par ailleurs », le bénéfice de la CMU-C permet de diviser par trois le risque de renoncement aux soins par rapport à une personne sans assurance complémentaire ([33]).

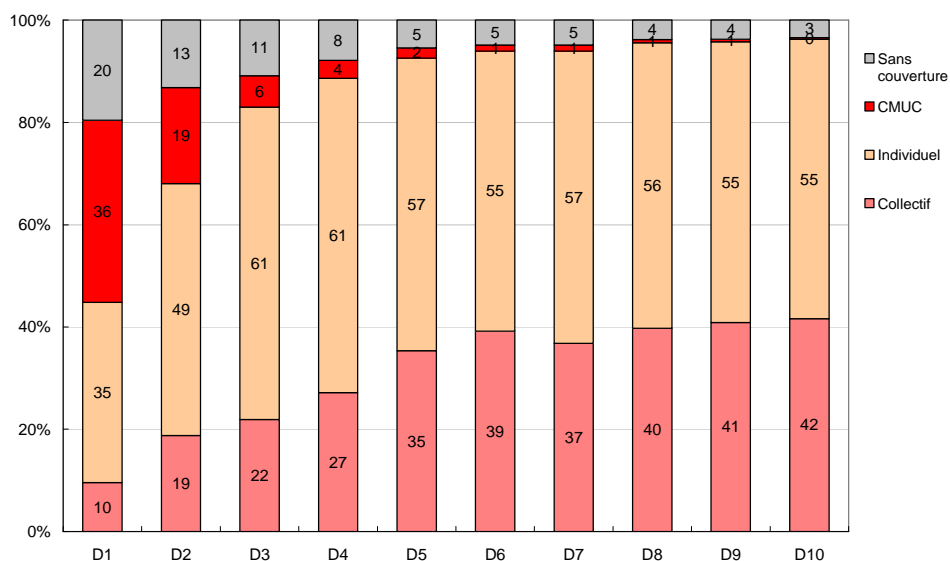
1.2.2. Le niveau de couverture offert par les complémentaires est très hétérogène

Le marché de la complémentaire santé se répartit entre trois types d'organismes : les mutuelles, les sociétés d'assurances et les institutions de prévoyance. Les bénéficiaires d'une couverture complémentaire santé sont assurés dans 58 % des cas par un contrat individuel et dans 42 % des cas par un contrat collectif, acquis par l'intermédiaire de l'entreprise (cf. encadré 2).

Le niveau de garanties des personnes assurées par les contrats modaux des organismes complémentaires santé varie fortement selon le type de contrat adopté : les contrats collectifs sont en moyenne de meilleure qualité que les contrats individuels ([12], [17]). Le niveau varie également selon le type d'organisme et l'âge du souscripteur.

En définitive, le niveau des remboursements au titre de l'assurance santé complémentaire dont dispose un individu dépend assez peu de son état de santé. En revanche, il varie fortement selon son niveau social. Les études empiriques existantes montrent sans surprise que les revenus et le statut par rapport à l'emploi conditionnent largement l'accès à un niveau donné de remboursements des soins ([4]). Avec une proportion croissante de contrats collectifs, et symétriquement une proportion décroissante d'individus non couverts, au fur et à mesure que le niveau de vie s'élève, les résultats de l'appariement Epas-SPS illustrent ce résultat (cf. Graphique 1).

Graphique 1 – Couverture maladie complémentaire et niveau de vie



Source : Appariement Epas-SPS 2006, traitement de la Drees

Champ : Ménage ordinaire, France métropolitaine

Lecture : 10% des individus appartenant à un ménage du premier décile de niveau de vie sont couverts par un contrat collectif contre 42% des individus appartenant à un ménage du dernier décile.

Ainsi, l'hétérogénéité des couvertures complémentaires rend indispensable une observation microéconomique des garanties offertes par ces contrats. Il serait en effet illusoire de vouloir approcher les remboursements complémentaires par une simple imputation globale de remboursements observés à l'échelle macroéconomique. Or, nous l'avons évoqué précédemment, les systèmes d'informations disponibles à ce jour ne permettent pas de connaître les garanties et les remboursements de l'AMC à l'échelle individuelle. En attendant la mise en place de nouveaux systèmes d'information (cf. projet Monaco), la microsimulation de ces données absentes constitue l'une des solutions.

Encadré 2 - La protection sociale complémentaire d'entreprise

Conduite pour la première fois en 2003 par l'Irdes auprès des établissements et reconduite en 2009 grâce à un cofinancement de la Cnamts et de la Drees, l'enquête PSCE (Protection Sociale Complémentaire d'Entreprise) a permis de mettre en évidence une grande hétérogénéité de l'offre de contrats collectifs selon la taille des entreprises et les secteurs d'activité. Plus l'entreprise comporte de salariés et plus elle est susceptible de proposer un contrat collectif à ses employés. Les salariés des petites entreprises du secteur des services et encore plus de l'industrie restent majoritairement exclus de la protection maladie collective. Enfin, ce sont les cadres et les professions intermédiaires qui ont les probabilités les plus élevées d'avoir une complémentaire maladie d'entreprise ([12], [17]).

Les résultats de l'édition 2009 de l'enquête PSCE montrent que tous les établissements de plus de 500 salariés proposent une couverture complémentaire d'entreprise contre seulement 34 % des très petites entreprises (TPE de moins de 10 salariés), confirmant ainsi le lien entre taille de l'entreprise et offre de complémentaire. En 2009, les cadres ont toujours plus de chance d'avoir une couverture d'entreprise. En 2009, le principal changement par rapport à 2003 tient à la recomposition récente de l'offre de complémentaire santé des entreprises aux salariés sous l'effet de la loi Fillon du 21 août 2003 ([23]).

Avant celle-ci, les employeurs et les salariés bénéficiaient d'exonérations fiscales et sociales sur leurs cotisations respectives quel que soit le type de contrat de complémentaire santé collective souscrit. La loi Fillon du 21 août 2003 a modifié les conditions d'exonérations fiscale et sociale. Au-delà d'une période transitoire prenant fin le 31 décembre 2008, les contrats doivent désormais respecter un certain nombre de contraintes pour pouvoir bénéficier de ces exonérations. Le régime doit notamment avoir un caractère collectif. Les contrats doivent être « responsables ». L'adhésion doit avoir un caractère obligatoire.

Afin de conserver les exonérations fiscales et sociales, bon nombre d'entreprises ont refondu leur offre de contrats, si bien qu'en 2009, le tiers des contrats datent de moins de deux ans. Sous l'effet de la loi Fillon, en 2009 le ou les contrats sont exclusivement obligatoires dans plus de trois établissements sur quatre.

2. Sources

2.1. Appariement de l'enquête SPS avec les données de l'Assurance maladie

A ce jour en France, seules deux sources offrent à la fois des informations fines sur les caractéristiques sociodémographiques des individus et des données précises sur leur consommation médicale : l'enquête Insee Handicap-Santé appariée aux données de l'assurance maladie (2003 et 2008) et l'appariement Epas-ESPS (bisannuel). L'élaboration d'Omar se fonde sur l'appariement Epas-ESPS 2006.

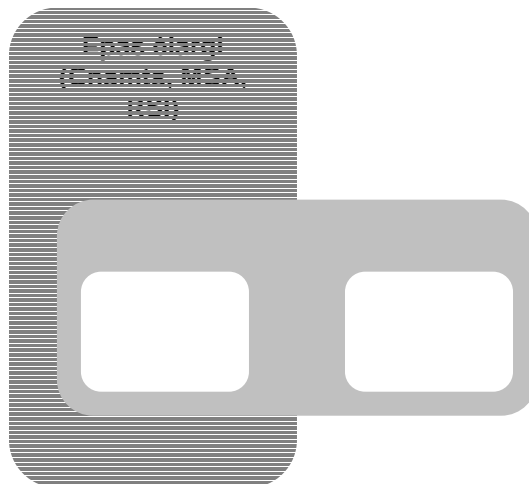
L'enquête Santé et Protection Sociale est une enquête bisannuelle réalisée par l'Irdes qui interroge tous les deux ans 8 000 ménages ordinaires environ, soit 22 000 individus résidant en France métropolitaine. Elle permet de recueillir des données sur les caractéristiques sociodémographiques (âge, sexe, revenus, catégorie socioprofessionnelle, composition du ménage...), sur les couvertures complémentaires ainsi que sur l'état de santé. La base de sondage servant à définir les individus interrogés est par construction incluse dans l'Échantillon Permanent des Assurés Sociaux (Epas), de manière à appairer les données de l'enquête avec les données remboursements de l'assurance maladie.

L'Epas est un échantillon au 1/600ième d'assurés sociaux du régime général. En tant que fichier administratif de liquidation, l'Epas permet de connaître les dépenses médicales présentées au remboursement. Ce champ ne recouvre donc pas l'ensemble des dépenses, contrairement au champ des comptes de la santé. Il s'agit d'une particularité importante de l'étude présentée ici. Un traitement des données brutes a permis de reconstituer par type de soins et pour l'année 2006 les agrégats suivants : le montant total des dépenses de santé présentées au remboursement, des remboursements par la Sécurité sociale, des tickets modérateurs, des dépassements et des participations forfaitaires ventilés par grands postes de soins (omnipraticien, spécialiste, pharmacie, biologie, infirmière, masseur-kinésithérapeute, optique, dentiste et hôpital). Enfin, la base de sondage de l'enquête SPS est en réalité un Epas élargi puisqu'à l'Epas de la Cnamts présenté ci-dessus s'ajoutent deux bases de données similaires relevant de la Mutualité Sociale Agricole (MSA) et du Régime Social des Indépendants (RSI). Ainsi, l'appariement Epas-ESPS permet de couvrir les trois grands régimes de l'assurance maladie.

En fin de compte, l'appariement de l'échantillon SPS avec les données de l'Epas permet de compléter les informations précédentes avec des données de consommation de soins. Toutefois, la vision « assurés et ayants-droit » offerte par l'Epas ne coïncidant pas avec la vision « ménages » de l'enquête SPS, l'appariement permet de récupérer les dépenses de santé pour seulement 10 000 individus environ (cf. Figure 1).

En 2006, 53 % des individus interrogés dans l'enquête SPS n'étaient pas présents dans l'Epas. De fait, les dépenses de ces individus ne sont pas connues. Nous avons donc choisi de les imputer.

Figure 1 - Articulation de l'enquête SPS de l'Irdes et des données de l'assurance maladie



Source : appariement Epas-ESPS 2006

2.2. L'enquête de la Drees sur les contrats complémentaires les plus souscrits

L'enjeu de la démarche, tel qu'il a été exposé précédemment, étant de microsimuler les remboursements de l'AMC, l'appariement Epas-ESPS ne peut suffire à mener ces travaux. En matière de couverture maladie complémentaire, l'enquête sur les contrats les plus souscrits (ECPS), menée par la Drees, constitue la source d'informations utilisée dans Omar.

2.2.1. Une source d'information précieuse

Chaque année, la Drees mène une enquête auprès des trois types d'OC (sociétés d'assurances, instituts de prévoyance et mutuelles), dite « enquête sur les contrats modaux ». Les questions portent sur les contrats couvrant le plus grand nombre de personnes, qu'ils fassent encore partie de l'offre commerciale ou non. Ils sont appelés contrats modaux.

Pour les contrats individuels, les questions portent sur les trois contrats couvrant le plus grand nombre de personnes au sein de l'ensemble des contrats individuels. Ils sont dénommés 1^{er} contrat modal, 2^{ème} contrat modal et 3^{ème} contrat modal. Pour les contrats collectifs, les questions portent sur les deux contrats couvrant le plus grand nombre de personnes au sein de l'ensemble des contrats collectifs, que l'adhésion soit obligatoire ou facultative, que le contrat soit "sur-mesure" ou standard. Ils sont dénommés 1^{er} et 2^{ème} contrat modal.

Les contrats couvrant les bénéficiaires de la CMU complémentaire sont hors du champ de l'enquête. Les organismes concernés relèvent de la France métropolitaine et des Dom. L'enquête permet d'une part de rassembler quelques informations sur l'activité globale de l'organisme et d'autre part de décrire de façon détaillée les caractéristiques et les garanties proposées par les cinq contrats modaux.

2.2.2. Quatre classes de contrats complémentaires

Pour réduire l'hétérogénéité en termes de couverture et de garanties des quelques 900 contrats collectés par la Drees, une typologie a été construite au moyen d'une Classification Ascendante Hiérarchique (CAH) réalisée à la suite d'une Analyse en Composante Principale (ACP) ([2]). Cette typologie aboutit à un regroupement des contrats en quatre classes : les contrats A qui offrent les meilleures garanties jusqu'aux contrats D offrant les moins bonnes.

La classe d'appartenance – le niveau de couverture – d'un contrat est très dépendante du type de contrat (collectif/individuel) et du type d'organisme complémentaire qui le propose (mutuelle, IP, assureur privé). Par ailleurs, la structure par âge des assurés varie fortement d'une classe à une autre, signe d'un lien fort entre âge et niveau de couverture (cf. Tableau 1). Pourtant la structure par âge au sein des contrats n'est demandée dans l'enquête que depuis 2008. Pour tenir compte de cette

information fortement reliée à la qualité de la couverture, nous avons choisi de considérer les données 2008 de l'enquête Drees sur les contrats modaux. Ce décalage temporel n'est toutefois pas gênant, la répartition des contrats (A, B, C, D) en fonction du type d'organisme et du type de contrat ayant très peu évoluée entre 2006 et 2008.

Tableau 1 - Répartition des contrats les plus souscrits A, B, C et D en fonction du type d'organisme, du type de contrat et de l'âge de l'assuré

		Type d'organisme		Institut de prévoyance		Société d'assurance		
		Mutuelle						
		Coll.	Ind.	Coll.	Ind.	Coll.	Ind.	
Classe de qualité du contrat	Tranche d'âge		Assurés de moins de 25 ans					
	A	14%	2%	19%	9%	42%	3%	
	B	10%	36%	42%	47%	37%	22%	
	C	14%	29%	38%	43%	20%	54%	
	D	62%	33%	0%	0%	0%	21%	
	Tranche d'âge		Assurés de 25 à 59 ans					
	A	39%	2%	25%	11%	48%	5%	
	B	31%	57%	42%	44%	47%	22%	
	C	26%	33%	33%	44%	5%	57%	
	D	4%	8%	0%	1%	0%	16%	
	Tranche d'âge		Assurés de 60 ans et plus					
	A	7%	3%	25%	12%	28%	5%	
	B	11%	53%	43%	29%	69%	23%	
C	80%	37%	33%	58%	2%	52%		
D	2%	6%	0%	1%	0%	20%		

Source : Enquête Drees sur les contrats les plus souscrits, 2008

Champ : Contrats modaux

Note de lecture : 14% des assurés de moins de 25 ans couverts par un contrat collectif de mutuelle ont un contrat de qualité A (qualité supérieure).

2.3. Représentativité des sources

Les dépenses enregistrées dans l'Epas sont uniquement celles présentées au remboursement. Ainsi par exemple, les dépenses d'automédication n'apparaissent pas dans Omar, alors qu'elles sont comptabilisées dans les Comptes de la santé. D'une manière générale, l'écart entre les résultats présentés dans les Comptes de la santé et ceux disponibles dans Omar, doit être apprécié suivant la différence de champ existant entre ces deux bases de données : les Comptes de la santé couvrent les départements d'outre-mer alors qu'Omar est restreint à la France métropolitaine.

Par ailleurs, l'enquête SPS n'interroge pas les ménages en institutions. Par conséquent, ces ménages aux dépenses élevées ne rentrent pas dans le champ de l'outil de microsimulation. Enfin, la taille de l'échantillon (environ 10 000 dépenses observées seulement) est faible pour apprécier correctement les dépenses hospitalières, alors même que les dépenses élevées sont très concentrées.

L'enquête Drees sur les contrats ne concerne que les contrats modaux, c'est-à-dire les plus souscrits. Par conséquent, les contrats de niches (offrant une très forte couverture ou à l'inverse proposant une couverture très faible), peu souscrits, sont notamment sous représentés, voire absents de l'enquête, et donc absents également du modèle de microsimulation.

Plus généralement, les contrats recueillis par l'enquête Drees couvrent 63,2% des personnes couvertes en individuel et 22,2% des personnes couvertes en collectif ([20]). La faible représentativité en termes de personnes couvertes des contrats collectifs de l'enquête peut conduire à s'interroger sur la représentativité de ces contrats en termes de niveau de couverture. D'après une comparaison des résultats de la dernière enquête Drees sur les contrats modaux avec les résultats de la dernière édition de l'enquête Protection Sociale Complémentaire d'Entreprise (PSCE 2009) réalisée par l'Irdes,

il semblerait que la qualité des contrats collectifs de l'enquête Drees sur les contrats modaux, et donc d'Omar, soit légèrement sous-estimée ([20]).

Finalement, ces problèmes de représentativité des contrats modaux sont susceptibles de biaiser nos résultats de restes à charge après Assurance maladie et couverture maladie complémentaire. Toutefois, l'imputation des contrats issus de l'enquête Drees conduit à retrouver au niveau agrégé des chiffres très proches de ceux des Comptes de la santé (cf. Annexe 4). Certes cela ne garantit pas la fiabilité du partage Assurance Maladie Obligatoire / Assurance Maladie Complémentaire offert par l'outil de microsimulation au niveau microéconomique. Mais, a contrario, des chiffres agrégés très éloignés des Comptes auraient invalidé l'outil. Ce n'est pas le cas.

3. L'imputation des dépenses de santé au niveau individuel

La première étape dans l'élaboration d'Omar va consister, à partir des deux premières sources d'information présentées ci-dessous (ESPS et appariement Epas-ESPS), à imputer aux individus présents dans ESPS mais non-appariés à l'Epas, des dépenses de santé (cf. Figure 1).

3.1. Quelques spécificités des dépenses de santé

Les dépenses de santé offrent deux types de spécificités dont les imputations devront tenir compte. Certaines sont purement arithmétiques : par exemple la dépense totale portée au remboursement doit être égale à la somme du remboursement de la Sécurité sociale, du ticket modérateur, des dépassements et des participations forfaitaires. D'autres sont d'ordre statistique : par exemple, les assurés ayant des dépenses importantes en omnipraticien auront généralement des dépenses élevées en pharmacie (cf. Tableau 2), si bien qu'une forte corrélation doit exister entre ces deux postes au niveau des dépenses observées, comme des dépenses imputées.

Tableau 2 - Matrice de corrélation entre les dépenses totales des différents segments de soins

Dépenses	Biologie	Dentistes	Hospit	Infirmières	Kiné	Omni	Optique	Pharmacie	Spé
Biologie	1,00	0,06	0,23	0,17	0,15	0,39	0,08	0,41	0,36
Dentistes	0,06	1,00	0,00	0,01	0,02	0,03	0,05	0,05	0,02
Hospit	0,23	0,00	1,00	0,12	0,11	0,20	0,01	0,24	0,07
Infirmiers	0,17	0,01	0,12	1,00	0,17	0,23	0,00	0,22	0,13
Kiné	0,15	0,02	0,11	0,17	1,00	0,28	0,05	0,16	0,06
Omni	0,39	0,03	0,20	0,23	0,28	1,00	0,07	0,38	0,15
Optique	0,08	0,05	0,01	0,00	0,05	0,07	1,00	0,05	0,03
Pharmacie	0,41	0,05	0,24	0,22	0,16	0,38	0,05	1,00	0,24
Spé	0,36	0,02	0,07	0,13	0,06	0,15	0,03	0,24	1,00

Source : Appariement Epas-SPS 2006, Irdes, Cnamts, RSI, MSA, - traitement Drees

Champ : Ménage ordinaire, France métropolitaine, dépenses observées

Note de lecture : 1^{ère} ligne deuxième colonne, le coefficient de corrélation linéaire entre les dépenses en omnipraticien et la biologie est de 0,39.

3.2. Choix d'une méthode d'imputation

Deux méthodes peuvent, *a priori*, être envisagées pour imputer les dépenses de santé des individus de la base ESPS n'ayant pu être appariés avec l'Epas. La première est une méthode d'imputation par équation et résidus simulés, la seconde est une méthode de *hot deck*.

Avant de s'intéresser aux avantages et aux limites de chacune des méthodes, il convient de rappeler les objectifs de cette imputation. Ainsi, le point de départ de ce travail est de pouvoir reconstituer, pour chacun des ménages présents dans l'enquête SPS, l'ensemble de ses dépenses de santé, de ses remboursements AMO et de ses remboursements AMC. Le point essentiel à retenir ici est la reconstitution des remboursements AMC. En effet, afin de pouvoir simuler ces remboursements il faut

pouvoir appliquer les garanties d'un contrat complémentaire aux dépenses de santé des individus. Or, les niveaux de remboursements ne sont pas équivalents pour tous les postes et tous ne sont pas spécifiés sous la même forme par les organismes d'AMC (taux ou forfait). Si l'on veut simuler ces garanties, il faut donc que les dépenses de santé des individus soient présentées à un niveau désagrégé.

Si l'on décide d'imputer les dépenses de santé à partir d'équations et de résidus simulés, plusieurs étapes apparaissent. Dans un premier temps, il faut que la modélisation mise en œuvre détermine à la fois la consommation ou la non-consommation mais aussi, le cas échéant, le montant de cette consommation engagée (*Two parts model* par exemple). Surtout, il faut que cette modélisation soit effectuée pour chacun des postes de soins pour lesquels les organismes d'AMC spécifient un niveau de remboursement. Dès lors ce n'est donc plus une équation qu'il faut envisager mais plus d'une dizaine d'équations. Enfin, le Tableau 2 ci-dessus a montré que les postes de soins étaient, hors optique et dentaire, fortement corrélés entre eux : l'imputation n'est plus une succession d'équations mais bien un système d'équations toutes corrélées les unes avec les autres. Si ce système a une solution, il faut encore songer à ajouter un résidu simulé à chaque valeur imputée.

Parallèlement, les techniques de *hot deck* permettent aussi de traiter le même problème. Ces techniques consistent à répliquer des données existantes plutôt qu'à en inventer de nouvelles comme le ferait une imputation par équations. Pour imputer des données, il faut que la population soit segmentée entre répondants et non-répondants : la méthode consiste alors à associer à chaque non-répondant un répondant, et à dupliquer la donnée existante pour remplacer la non-réponse. En fin de compte, le *hot deck* est le pendant non-paramétrique de l'imputation par équations et résidus simulés. Ici, le résidu est empirique.

L'avantage des techniques *hot deck* est majeur : là où l'imputation par équations impose un lourd système d'équations pour déterminer les dépenses de santé individuelle, le *hot deck* permet, par l'association d'un donneur et d'un receveur de dupliquer immédiatement chacun des postes de soins. Ces valeurs imputées étant des dépenses constatées, les deux contraintes évoquées plus haut sont immédiatement satisfaites (1^o) dépenses = remboursements AMO + TM ; 2^o) postes corrélés).

Parmi les méthodes de *hot deck* existantes, celle que nous avons retenue est le *hot deck* aléatoire stratifié. Aléatoire, car le donneur est sélectionné par un tirage aléatoire et deux jeux d'imputation n'associeront pas le même donneur à un receveur précis (contrairement au *hot deck* métrique par exemple). Stratifié, car le tirage aléatoire du donneur se fait parmi une population de donneurs ayant les mêmes caractéristiques que le receveur.

Il est alors possible de résumer les étapes à mettre en œuvre pour imputer les dépenses de santé des individus non appariés à l'Epas :

- 1) cerner les facteurs explicatifs de la consommation de soins ;
- 2) choisir parmi ces facteurs des variables, de manière à définir des strates d'imputation ;
- 3) ventiler les individus au sein de ces strates ;
- 4) pour un receveur appartenant à une classe donnée, tirer au sort (avec remise) un donneur parmi les donneurs appartenant à la même strate ;
- 5) appliquer au receveur le profil de dépense du donneur tiré au sort.

3.3. Mise en œuvre du *hot deck* stratifié

3.3.1. Choix des strates d'imputation

La mise en œuvre du *hot deck* suppose que les donneurs aient des caractéristiques proches des receveurs, autrement dit de raisonner à l'échelle de strates d'imputations. Ces strates sont construites au moyen de variables auxiliaires qui expliquent bien la dépense.

L'analyse précédente (Tableau 2) montre que les dépenses d'optique et de dentaire ne sont pas corrélées avec les autres dépenses de soins. Elles doivent donc dépendre de variables auxiliaires différentes. C'est pourquoi ces deux types de dépenses ont été imputés séparément. De manière à bâtir les strates d'imputation pour la mise en œuvre du *hot deck* stratifié, trois modèles de consommation de soins ont donc été estimés au préalable. L'un pour modéliser la dépense en optique, l'autre la dépense en dentaire, le dernier pour modéliser le reste de la dépense de santé.

Du fait de l'existence d'un nombre importants d'individus n'ayant aucune dépense de santé dans l'année, la distribution des dépenses de santé présente un point d'accumulation en 0. D'un point de vue économétrique, cette particularité oblige à modéliser la dépense en deux temps. D'abord on modélise le fait d'avoir ou non une dépense de santé, à l'aide d'un modèle probit: c'est l'équation de participation. Ensuite, on modélise le montant de la dépense (ou son logarithme, compte tenu de la distribution log-normale des dépenses de santé), conditionnellement au fait d'avoir recouru : c'est l'équation de dépense.

$$\begin{cases} P_i^* = X'_{2i} \cdot \beta_2 + u_{i2} & \text{(équation de participation)} \\ \log(dep_i) = (X'_{1i} \cdot \beta_1 + u_{i1}) P_i & \text{(équation de dépense)} \end{cases}$$

Deux types de méthode économétriques peuvent être utilisées dans ce cas de figure : le modèle à deux parties (*Two Parts model*), utilisé notamment aux États-Unis dans les travaux de la RAND ([30]) ou le modèle de sélection *Sample selection model* ([24]). Ces deux méthodes ne sont pas équivalentes. La première s'intéresse à l'espérance de la dépense conditionnellement à une consommation positive, et les deux équations de participation et de dépense peuvent alors être estimées de manière indépendante. La deuxième s'intéresse à l'espérance non conditionnelle de la dépense, si bien que dès lors que la décision de consommer et le montant de dépense ne sont pas indépendants, il est nécessaire dans l'estimation de tenir compte de la corrélation des résidus des deux équations, en introduisant dans la deuxième équation l'inverse du ratio de Mills estimé à partir de la première équation.

Les travaux cherchant à mesurer les déterminants de la dépense de santé privilégient généralement l'estimation par un modèle de sélection car l'intérêt se porte sur l'espérance non conditionnelle de dépense ([31]). En revanche, les travaux de microsimulation privilégient plutôt l'estimation par un *Two part model* ([15]) car l'objectif principal n'est pas d'interpréter les coefficients estimés de la deuxième équation. De nombreux travaux théoriques et empiriques débattent du meilleur choix possible ([27], [3]). En particulier, des travaux de simulation par la méthode Monte-Carlo ont montré que l'estimation par *Two part model* pouvaient obtenir de meilleurs résultats que les modèles de sélection même quand le vrai modèle générateur des données est un modèle de sélection ([29]). Leung et Yu ont montré que ce cas de figure pouvait arriver s'il y avait trop de collinéarité entre l'inverse du ratio de Mills et les variables explicatives de la deuxième équation.

Dans la présente étude, compte tenu de notre objectif de microsimuler des dépenses de soins plutôt que de mesurer les facteurs explicatifs de la dépense pour pouvoir les interpréter, l'estimation de ces deux équations a donc été faite dans le cadre d'un *Two Parts Model*, qui présente de meilleures propriétés empiriques. Les résultats des estimations sont présentés dans les tableaux 8 à 13 en annexes.

Dans le cadre d'une imputation par *hot deck* stratifié, il n'est pas possible de retenir la totalité des facteurs explicatifs de la dépense de santé dans la construction des classes d'imputation. En effet, un nombre trop important de variables conduirait à des sous-populations de taille très faibles avec très peu de donneurs. C'est pourquoi, dans une logique de parcimonie, nous avons sélectionné les déterminants les plus pertinents, quitte à ne pas retenir toutes les variables significatives au seuil de 5%.

3.3.2. Le problème des donneurs en nombre insuffisant ou des donneurs inexistant

Malgré le principe de parcimonie précédent, il se peut qu'au sein de certaines strates le nombre de donneurs soit encore trop faible (voire nul). C'est pourquoi il est nécessaire de s'imposer deux contraintes techniques supplémentaires, dans la mise en œuvre du *hot deck* :

- un nombre minimal de donneurs par classe (C1) ;
- le fait que le nombre de donneurs doit être supérieur au nombre de receveurs (C2)

Ces deux contraintes permettent notamment d'éviter qu'un individu atypique soit utilisé un trop grand nombre de fois.

Soumis à ces contraintes et à l'absence de donneurs dans certaines classes, le *hot deck* s'effectue de manière itérative. Dans un premier temps, les classes d'imputation sont formées à partir de toutes les variables auxiliaires. L'imputation par donneurs est réalisée pour tous les receveurs des classes pour lesquelles les contraintes C1 et C2 sont respectées. Pour les classes pour lesquelles l'une de ces contraintes ne serait pas vérifiée, on supprime alors une des variables auxiliaires ayant servi à la définition des strates et on reforme des strates sur la base des variables auxiliaires restantes. Le choix de retirer une variable se fonde sur son niveau de significativité aussi bien dans l'équation de participation que dans l'équation de montant. L'imputation par donneurs est alors à nouveau réalisée pour tous les receveurs des classes pour lesquelles les contraintes C1 et C2 sont respectées. Ce processus itératif se poursuit jusqu'à ce que tous les individus soient imputés. Au fil des différentes étapes, les variables auxiliaires sont retirées les unes après les autres.

Après un premier essai d'imputation *hot deck*, l'ordre des variables peut éventuellement être modifié à la marge afin d'imputer un plus grand nombre d'individus au cours des premières étapes.

Le Tableau 3 ci-dessous fait apparaître le déroulement du *hot deck* pour les dépenses hors optique et dentaire.

Tableau 3 - Déroulement de l'imputation par *hot deck* des dépenses hors optique et dentaire

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour	9 ^{ème} tour
Nb d'individus auxquels sont affectés des dépenses de santé	3227	1717	1679	2242	1755	571	208	299	4
	27,58%	14,67%	14,35%	19,16%	15,00%	4,88%	1,78%	2,56%	0,03%
Variables de stratification	Exonération	x	x	x	x	x	x	x	x
	Consultation d'un omnipraticien	x	x	x	x	x	x	x	x
	Régime	x	x	x	x	x	x	x	
	Sexe et âge	x	x	x	x	x	x	x	
	État de santé perçu	x	x	x	x	x	x		
	Nature de la couverture complémentaire	x	x	x	x	x			
	Nb de personnes dans le ménage	x	x	x	x				
	Niveau de revenus	x	x	x					
	Consultation d'un spécialiste	x	x						
	Consommation de médicaments	x							
R ² du modèle associé (équation de montant)	0,4074	0,3997	0,3765	0,3764	0,3744	0,3687	0,3588	0,288	0,2824

Source : Omar 2006.

Champ : individus non appariés avec l'Epas.

Note de lecture : avec la totalité des variables de stratification (dix) et sous la condition que le ratio receveurs/donneurs n'excède pas 1,4, il est possible d'imputer 3227 individus au premier tour de la procédure *hot deck*.

Il convient ainsi de noter que plus de la moitié de la population est appariée avant le quatrième tour et que le R² des modèles associés à ces stratifications est généralement compris entre 0,41 et 0,36. Moins de 3% de la population est traitée lors des deux dernières étapes où le R² des modélisations associées est de 0,28.

3.4. Calage des dépenses sur les données hospitalières de l'Epas

Du fait des sources utilisées, l'outil de microsimulation sous-estime les dépenses hospitalières. En effet, le taux de réponse à l'enquête SPS est plus faible chez les personnes en mauvaise santé. Par ailleurs, seules les personnes en «ménage ordinaire» sont interrogées dans l'enquête SPS. Les personnes vivant en institution, qui ont potentiellement des dépenses hospitalières importantes, ont

été exclues du champ de l'enquête. Les données issues de SPS ont tendance à sous-estimer les dépenses hospitalières par rapport à l'Epas. Ce biais peut être redressé en calant les données SPS sur la « queue de distribution » des dépenses hospitalières de l'Epas, c'est-à-dire obtenir une proportion de personnes ayant des dépenses hospitalières importantes équivalente à celle de l'Epas. Ce calage doit être fait à la fois au niveau des individus et des ménages.

Les variables de calage retenues sont celles utilisées dans SPS pour préserver la représentativité initiale de l'échantillon : âge, sexe, régime d'affiliation et taille du ménage ([1]) auxquelles sont ajoutées une variable de dépenses hospitalières et la proportion d'ALD par niveau de dépenses à l'hôpital (cf. Annexe 3). Le calage est réalisé à partir de la macro Calmar mise au point par l'Insee ([34]).

3.5. Simulation de non-réponse et test de la méthode

Afin de s'assurer que la méthode mise en place ne conduit pas à imputer des valeurs aberrantes, un essai d'imputation a été effectué sur des individus pour lesquels les dépenses de santé sont connues (i.e. présentes dans l'Epas). Le Tableau 4 compare les moyennes obtenues après *hot deck* avec les « vraies » moyennes pour les dépenses en omnipraticien.

Pour produire ces résultats, un échantillon représentant un tiers de la population appariée entre l'Epas et l'ESPS a été sélectionné aléatoirement. Pour ces 3135 individus, les dépenses de santé sont donc observées et n'ont pas besoin, dans Omar, d'être imputées par *hot deck*. Pour les besoins de l'exercice, il a été considéré ici que ces dépenses de santé n'étaient pas connues : un tiers des personnes appariées est donc considéré comme non-répondant.

Ensuite, la procédure d'imputation détaillée précédemment a été mise en place. Le tiers de non-répondants a été imputé par *hot deck* stratifié sur la base des valeurs observées pour les deux tiers restants de l'appariement. La non-réponse étant fictive, il est alors possible de confronter valeurs observées et valeurs imputées.

Tableau 4 - Comparaison des montants moyens de l'Epas et du *hot deck* pour le poste "Omnipraticiens" sur non-réponse fictive

Poste de soins "Omnipraticiens"	Observé dans l'Epas	Imputé par <i>hot deck</i> stratifié
Taux de consommateurs	83,4%	82,2%
<i>Pour les consommateurs, montant moyen des...</i>		
... dépenses	145,3 €	144,8 €
... remboursements AMO	103,8 €	103,3 €
... tickets modérateurs	31,9 €	31,5 €
... dépassements	6,0 €	6,4 €
... participations forfaitaires	3,6 €	3,6 €

Source : Appariement Epas-SPS 2006, traitements de la Drees

Champ : Échantillon de 3135 individus sélectionné par tirage aléatoire dans l'appariement Epas-ESPS.

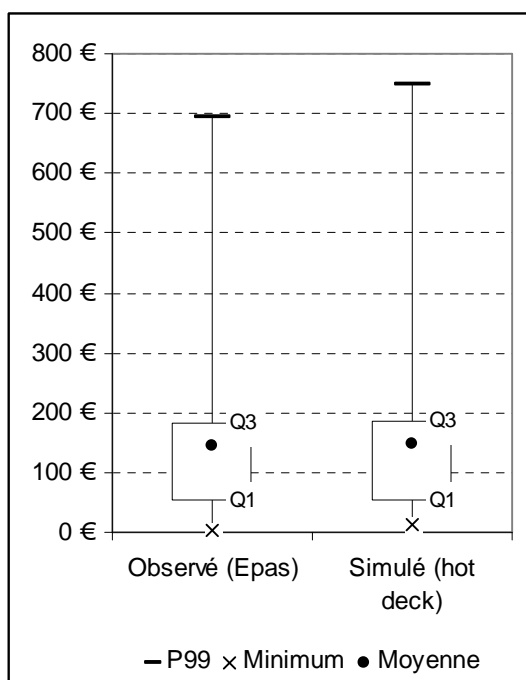
Note de lecture : pour un tiers des individus présents dans l'appariement Epas-ESPS, le taux de consommateurs en Omnipraticiens est 83,4% et les dépenses moyennes de ces consommateurs s'élèvent à 145€. Lorsque ces individus sont traités par *hot deck* le taux est plus faible, à 82,2% et les dépenses moyennes sont quasiment identiques.

Ces résultats sont tout à fait satisfaisants. Le taux de consommateurs ne varie que de 1,2 points entre les deux sources. Concernant les montants, les variations apparaissant sont très faibles: dépenses, remboursements AMO, tickets modérateurs, dépassements et participations forfaitaires ne diffèrent jamais plus d'un euro, en moyenne, suivant la source.

Néanmoins, les résultats présentés ci-dessus sont des résultats moyens. Dans un second temps, il est indispensable de comparer les deux sources à un niveau plus fin. Ainsi, le Graphique 3 ci-dessous présente les diagrammes en boîte des deux séries. Si les valeurs centrales (moyennes, 1^{er} quartile et 3^{ème} quartile) ne font pas apparaître de divergences majeures entre l'Epas et les données simulées, il convient toutefois de noter les différences observables aux extrémités de la distribution. Le 99^{ème} percentile est ainsi supérieur de 60€ pour les données simulées par rapport aux données observées.

Surtout, c'est l'étude de la répartition des écarts au niveau individuel entre dépenses observées et imputées qui fait apparaître une limite essentielle de la méthode d'imputation. Certes, en moyenne et même en distribution, les résultats du *hot deck* sont tout à fait acceptables. Pourtant, il ne faudrait pas oublier qu'au niveau individuel, la valeur imputée a très peu de chances d'être équivalente à celle observée : c'est ce que le Graphique 2 fait apparaître.

Graphique 1 - Distribution des dépenses de santé en omnipraticiens, observées dans l'Epas et simulées par hot deck

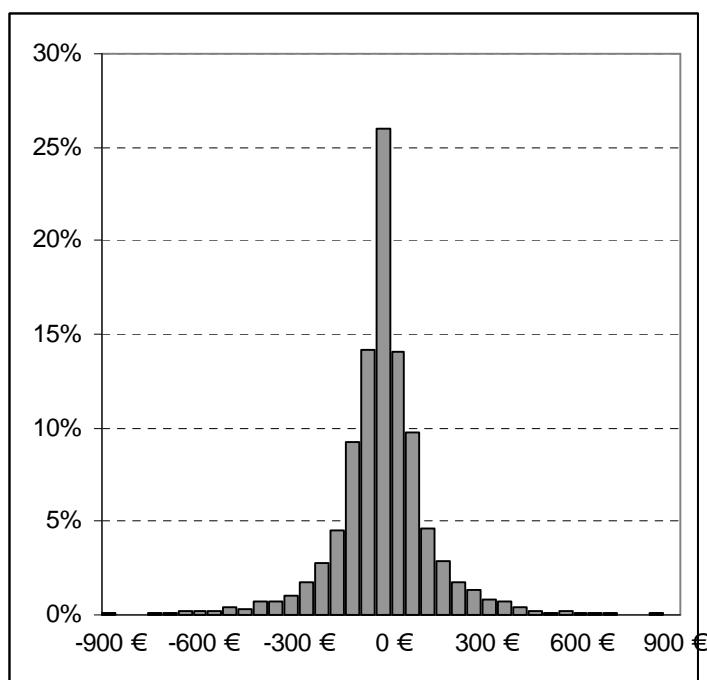


Source : appariement Epas-SPS 2006, traitements de la Drees.

Champ : Échantillon de 3135 individus sélectionné par tirage aléatoire dans l'appariement Epas-ESPS. Individus ayant au moins une dépense en omnipraticiens.

Note de lecture : dans l'Epas, 99% des dépenses observées sont inférieures à 690€. Lorsque ces dépenses sont simulées par hot deck, 99% d'entre elles sont inférieures à 750€.

Graphique 2 - Répartition des écarts entre dépenses observées et dépenses simulées (poste omnipraticiens)



Source : appariement Epas-SPS 2006, traitements de la Drees.

Champ : Échantillon de 3135 individus sélectionné par tirage aléatoire dans l'appariement Epas-ESPS ; cinq écarts supérieurs en valeur absolue à 1000€ ne sont pas représentés ici.

Note de lecture : pour 26% des individus, l'écart entre dépenses observées et dépenses simulées est compris dans l'intervalle]-25€ ; +25€].

74% des écarts entre dépenses observées en omnipraticiens et dépenses simulées sont supérieurs, en valeur absolue, à 25€. Cette variabilité est en fait à rapprocher du R² des modélisations *Two Parts Model* présentées précédemment. Ainsi, lorsqu'avec l'ensemble des variables de stratification, le modèle pour les dépenses hors optique et dentaire explique à peine 41% de la variance initiale, cela signifie bien que 59% ne peuvent être expliqués par le modèle.

Observer l'âge, l'état de santé perçu et diverses variables socio-démographiques ne permet évidemment pas de déduire précisément la dépense de santé d'un individu. Seules des données très

fin de diagnostic pourraient améliorer significativement la variance expliquée. Dès lors, les strates créées pour l'imputation, sur la base de ces variables, restent hétérogènes. Ces écarts observés dans le Graphique 2 sont donc la conséquence de la variabilité intra-strates du *hot deck*. Au total l'objectif de la microsimulation n'est pas de reconstruire l'existant individu par individu, mais de construire une base qui reproduit bien les distributions observées dans la réalité comme l'illustre le Graphique 3.

4. Simulation des remboursements complémentaires au niveau individuel

Les données de l'Epas appariées aux données ESPS ne renseignent pas sur le partage des dépenses non remboursées par la Sécurité sociale entre ce qui est pris en charge par les organismes complémentaires et ce qui reste à la charge des ménages.

Le calcul des dépenses remboursées au titre de la couverture complémentaire est immédiat pour les personnes qui en sont dépourvues et pour les bénéficiaires de la CMUC. En revanche, il est plus complexe pour les personnes ayant une couverture complémentaire privée puisque les niveaux de garantie changent selon les contrats. Les données de l'enquête SPS ne permettent cependant pas de définir précisément ces niveaux de remboursement.

Il a donc fallu simuler pour chaque individu les remboursements de son assurance complémentaire en appliquant les garanties du contrat aux dépenses constatées lesquelles, compte tenu des sources utilisées, correspondent aux seules dépenses présentées au remboursement. Pour ce faire, il faut toutefois connaître les garanties précises du contrat d'assurance complémentaire détenu par l'individu. Or les données de l'enquête SPS ne renseignent pas sur les niveaux exacts de remboursement, mais seulement sur le type de contrat (individuel ou collectif) et la nature de l'organisme assureur (assurance, mutuelle, institut de prévoyance). Il a donc été nécessaire de mobiliser une troisième source d'informations, la base des quelque 900 contrats issus de l'enquête Drees 2006 sur les contrats modaux, pour lesquels les garanties exactes sont connues. Pour rapprocher ces sources, il a fallu attribuer à chaque individu de SPS ayant une couverture privée un contrat issu de l'enquête sur les contrats modaux.

4.1. Rapprochement des deux sources d'informations

Les données de l'enquête SPS renseignent également sur le type de contrat (individuel ou collectif), le type d'organisme (IP, mutuelle, assureur) et l'âge de l'individu. Le rapprochement des deux sources s'effectue alors sur la base de ces informations de l'enquête SPS et des informations du Tableau 1 précédent.

L'individu statistique de ces opérations de rapprochement est le ou les souscripteurs du ou des contrats du ménage. Il est en effet important que les individus ayant déclaré dans ESPS être rattachés à un même contrat soient, effectivement, couverts par les mêmes garanties. Omar procède donc en deux temps, traitant d'abord le cas des ouvrants-droit, et en déduisant ensuite les caractéristiques de l'AMC des ayants-droit.

L'association d'un contrat modal à chaque ouvrant-droit présent dans Omar se fait en deux étapes. La première consiste à déterminer la classe de qualité du contrat complémentaire et se fonde sur les caractéristiques individuelles déclarées dans ESPS : $X_i = (\text{âge}, \text{type de contrat}, \text{type d'organisme})$. L'ouvrant-droit se voit alors attribuer la distribution marginale du Tableau 1 qui correspond à ses caractéristiques X_i , c'est-à-dire un jeu de quatre probabilités ($PA(X_i)$, $PB(X_i)$, $PC(X_i)$, $PD(X_i)$) correspondant aux quatre classes de couvertures identifiées (A, B, C, ou D). Par ce tirage aléatoire du niveau de garantie, la distribution globale de la qualité des contrats est respectée dans Omar.

Une fois cette première étape de détermination de la qualité du contrat réalisée, l'attribution d'un contrat modal à proprement parler peut être effectuée. Cette allocation se fait par tirage aléatoire d'un contrat parmi ceux de l'enquête sur les contrats les plus souscrits. Le choix se fait suivant les caractéristiques associées à l'ouvrant-droit : contrats de même type (collectif/individuel), proposés par le même type d'organisme et appartenant au même niveau de garantie. Ce tirage n'est pas à probabilités égales. En effet, les contrats de l'enquête Drees couvrent un nombre variable d'individus. Afin d'en tenir compte, la probabilité pour un contrat d'être tiré est proportionnelle au nombre de

personnes ayant souscrit ce contrat. Ainsi, plus un contrat couvre un nombre important d'ouvrants-droit dans l'enquête Drees, plus il a de chances d'être sélectionné.

A la fin de cette étape, un contrat de l'enquête Drees a été attribué à chaque ouvrant-droit déclaré dans l'enquête SPS et par voie de conséquences, à ses ayants droit. Pour chacun de ces individus, on connaît donc les garanties pour un certain nombre de soins (spécialiste, optique, dentaire *etc.*).

4.2. La simulation des remboursements de l'organisme complémentaire

La dernière étape de ce travail de microsimulation consistait alors à appliquer ces garanties aux dépenses de l'individu, dépenses constatées ou imputées, pour déterminer le montant simulé des remboursements au titre de la couverture complémentaire (ROC).

4.2.1. Remboursement des actes médicaux

Pour les actes médicaux, les garanties des assurances complémentaires sont exprimées en pourcentage de la base de remboursement, si bien que le montant maximal remboursé par l'organisme complémentaire est obtenu par la formule suivante :

$$ROC_{max} = (RSS + TM) \times \text{garantieOC} - RSS$$

où RSS désigne le montant remboursé par la Sécurité sociale, TM le montant de ticket modérateur et garantieOC le taux de remboursement de l'organisme complémentaire (exprimé en pourcentage de la base de remboursement).

Cependant, il est possible que les dépenses de l'assuré soient inférieures à ce remboursement maximum (par exemple, en l'absence de dépassements d'honoraires). De ce fait, le montant effectivement remboursé par l'organisme complémentaire est en réalité donné par la formule suivante :

$$ROC = \min(ROC_{max}, \text{dépenses} - RSS)$$

4.2.2. Remboursement des soins et prothèses dentaires

En matière de soins dentaires, les données de l'Epas ne permettent pas de distinguer les dépenses imputables aux prothèses de celles dues aux soins conservateurs. Ces dépenses sont en effet regroupées dans une variable unique « dépense en dentaire ». C'est pourquoi le premier travail est de distinguer les individus qui ont eu une pose de prothèse dentaire des autres. Pour ce faire, nous considérons qu'un individu dont la base de remboursement annuelle est inférieure à celle d'une prothèse (107,5€) a eu uniquement des soins conservateurs.

Les individus dont la base de remboursement est supérieure à 107,5€ se sont fait poser une prothèse. Pour ces individus, il faut alors distinguer dans la dépense agrégée en dentaire la part liée aux prothèses de celle liée aux soins conservateurs. Pour ce faire, nous faisons l'hypothèse qu'un individu ne peut s'être fait poser plus de trois prothèses dans l'année, les autres dépenses étant de fait liées à des soins conservateurs. Formellement, le partage de la dépense agrégée est fait comme suit :

$$\text{dépenses agrégées en dentaire} = \text{dépenses en prothèse} + \text{dépenses en soins conservateurs}$$

avec :

$$\text{dépenses en prothèse} = 107,5 \times \min\left(\text{Ent}\left(\frac{\text{dépenses agrégées}}{107,5}\right), 3\right)$$

$$\text{dépenses en soins conservateurs} = \text{dépenses agrégées} - \text{dépenses en prothèse}$$

4.2.3. Remboursement des dépenses d'optique

Les données de l'Epas dont nous disposons agrègent les dépenses en optique au sein d'une seule variable « dépense en optique ». De son côté, l'enquête Drees distingue les garanties couvrant les verres simples et les garanties s'appliquant aux verres complexes. Considérant qu'il est rare d'acheter plusieurs paires dans l'année, nous avons supposé que le montant de la dépense observé relevait d'un seul achat de lunettes. Ensuite, la confrontation de la base de remboursement observée avec celle de verres complexes standards nous a permis d'inférer si l'individu avait acheté des verres simples ou des verres complexes. La dernière opération a consisté à appliquer la garantie adéquate du contrat attribué à l'individu lors de la précédente étape.

4.2.4. Remboursement des dépenses en pharmacie ou en hospitalisation

Nous disposons uniquement de données agrégées par poste. De fait, cela ne nous permet pas de ventiler la dépense de pharmacie entre ce qui relève de médicaments pris en charge à 65% ou à 35%. Concernant les dépenses hospitalières, nous ne sommes pas capables de distinguer ce qui relève du ticket modérateur à proprement parler de ce qui relève du forfait journalier dans le montant non remboursé par la Sécurité sociale : dans nos données, la variable « Ticket modérateur » englobe les deux montants.

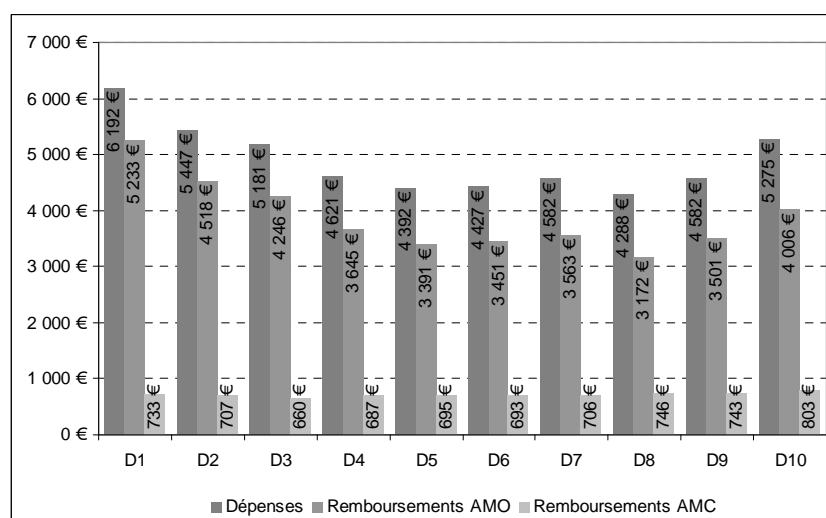
Cependant, les contrats complémentaires ne remboursant pas les médicaments pris en charge à 35% ou le ticket modérateur à l'hôpital sont minoritaires. Aussi, sur ces deux postes de dépenses, faute d'avoir des informations détaillées, nous avons considéré que les contrats de couverture complémentaire remboursaient systématiquement le montant du ticket modérateur.

5. Sensibilité des résultats à l'attribution aléatoire d'un contrat : le cas des remboursements complémentaires

L'objet de cette partie est de tester, à travers la présentation de premiers résultats, la sensibilité des valeurs simulées au caractère aléatoire de la méthode d'imputation des contrats de couverture maladie complémentaire. La variable d'intérêt choisie est le montant des remboursements versés par l'AMC. Celui-ci est en effet directement dépendant des spécificités du contrat imputé.

Dans un premier temps, le Graphique 4 ci-dessous fait apparaître les dépenses de santé, les remboursements de l'AMO et ceux de l'AMC tels qu'ils sont simulés dans Omar.

Graphique 4 - Dépenses de santé, remboursements AMO et remboursements AMC des ménages selon le décile de niveau de vie



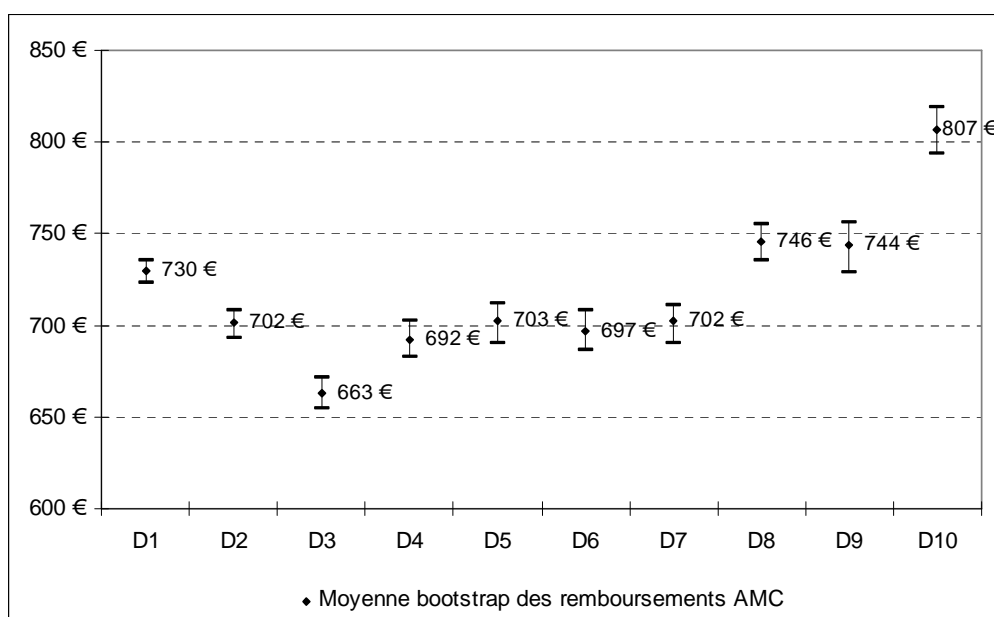
Champ : Ménage ordinaire, France métropolitaine ; dépenses présentées au remboursement.
Source : Omar 2006, Drees.

Les résultats du Graphique 4 sont issus d'une simulation unique à partir d'Omar. Les trois indicateurs dépendent du montant des dépenses simulées, mais surtout, les remboursements complémentaires sont directement tributaires de l'attribution de tel ou tel contrat modal aux individus de l'échantillon.

Afin de tester la qualité de la méthodologie employée, des tests de sensibilité ont été effectués. Ainsi, une procédure bootstrap a été mise en place permettant de produire 150 jeux de données microsimulées. Entre chaque itération, les dépenses de santé associées aux individus restent inchangées : seules les étapes d'attribution d'un niveau de garantie et, par la suite, d'un contrat modal sont renouvelées. Dès lors, la seule source de variation des résultats relève de l'imputation de la couverture maladie complémentaire.

Avec les 150 versions d'Omar ainsi produites, il est possible de calculer des moyennes bootstrap (et leur intervalle de confiance à 95%). Ce sont ces résultats qui apparaissent dans le Graphique 5 ci-dessous.

Graphique 5 - Moyenne bootstrap des remboursements AMC (niveau ménage) selon le décile de niveau de vie et intervalle de confiance à 95%



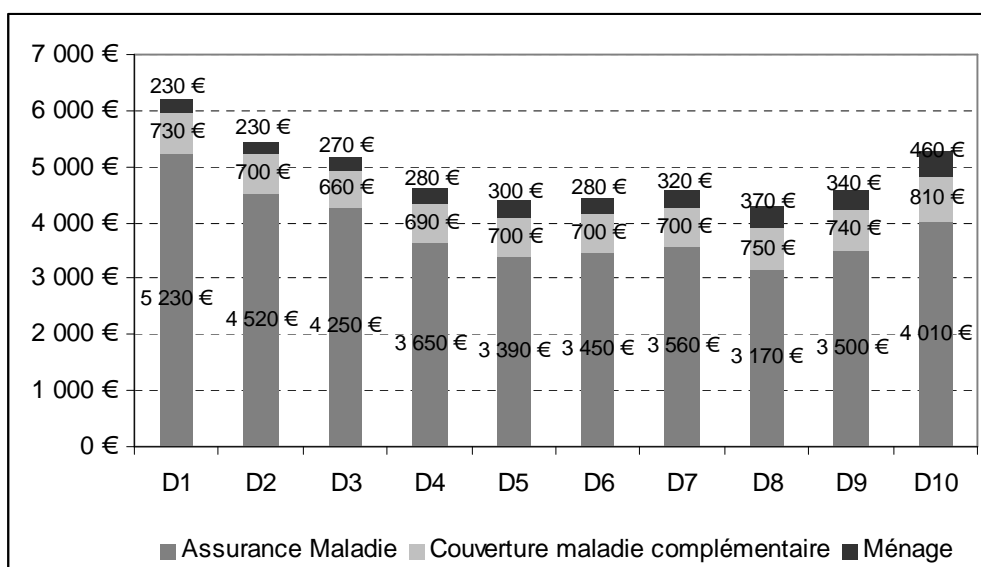
Champ : Ménage ordinaire, France métropolitaine ; dépenses présentées au remboursement.
Source : Omar 2006, Drees.

Ces résultats sont tout à fait rassurants quant à la faible sensibilité des remboursements d'AMC simulés à l'aléa d'imputation. Ainsi, l'erreur c'est-à-dire l'écart entre la borne supérieure et la borne inférieure de l'intervalle de confiance est, pour l'ensemble des déciles de niveau de vie, comprise entre 12€ (1^{er} décile de niveau de vie) à 27€ (9^{ème} décile de niveau de vie). Rapportée au remboursement moyen, cette incertitude s'élève au maximum à 3,6% de la moyenne bootstrap estimée.

Dès lors, même si les montants calculés via Omar ne doivent pas être étudiés à la dizaine près, les intervalles de confiance présentés ci-dessus permettent d'apprécier certains phénomènes tels que les remboursements d'assurance maladie complémentaire plus élevés des ménages appartenant aux trois déciles de niveaux de vie les plus favorisés.

Avec cette confirmation de la faible volatilité des montants simulés, c'est finalement l'objectif de départ qui est atteint. Il est désormais possible de simuler la structure de financement des dépenses de santé à un niveau plus fin que celui disponible dans les Comptes de la santé. Le graphique ci-dessous fait ainsi apparaître cette structure suivant le décile de niveau de vie des ménages.

Graphique 6 - Structure de financement des dépenses de santé des ménages selon le décile de niveau de vie



Source : Omar 2006

Champ : ménages ordinaires ; dépenses présentées au remboursement de l'Assurance Maladie.

Note de lecture : les dépenses de santé d'un ménage du 1er décile de niveau de vie s'élèvent en moyenne à 6190€ ; sur ce montant, 5230€ sont pris en charge, en moyenne, par l'Assurance Maladie, 730€ par la couverture maladie complémentaire et 230€ restent à la charge du ménage.

La simulation de la structure de financement selon le décile de niveau de vie pourra se prolonger vers des analyses microéconomiques sur l'accessibilité financière aux soins ou sur les enjeux redistributifs des assurances maladie.

6. Perspectives ultérieures

Les quelques résultats présentés dans cette étude montrent les potentialités offertes par Omar pour mieux évaluer le poids des dépenses de santé dans le budget des ménages.

A l'avenir, la Drees prévoit d'actualiser et d'améliorer cet outil tous les deux ans, suivant la mise à disposition des données de l'appariement Epas-ESPS. De plus, une réflexion est en cours afin d'améliorer l'imputation des contrats d'assurance complémentaire santé : il s'agirait de tenir compte de l'appréciation subjective de l'assuré sur les niveaux de garanties offerts par le contrat qu'il détient. Par ailleurs, conjointement au renouvellement de l'enquête sur les contrats les plus souscrits, il est prévu d'étudier les changements résultant d'une modification de la classification des contrats suivant leur niveau de garantie. En effet, des travaux méthodologiques sur la typologie des contrats d'assurance maladie complémentaire vont être reconduits par la Drees à partir des données 2010 de l'enquête sur les contrats. Ces travaux permettront de tester la sensibilité des résultats de l'actualisation du modèle Omar aux différentes typologies essayées.

Enfin, la Drees a intégré ce modèle de microsimulation dans le modèle Ines. Cette intégration permet de conduire des analyses redistributives globales puisqu'Ines contient des informations sur les cotisations individuelles versées à la Sécurité sociale. Ainsi, l'intégration de ces deux modèles permet de mettre en regard, au niveau individuel, le volet prestations et le volet cotisations et primes à la fois pour les assurances publique et privée. Dès lors, la réunion de ces deux outils permet de compléter et d'affiner les travaux macroéconomiques et microéconomiques existant sur l'analyse de la redistribution opérée par l'assurance maladie ([11]) et d'estimer l'impact de mesures portant sur les règles de remboursement par les différents organismes de couverture obligatoire ou facultative. Une première étude, réalisée grâce à l'outil Ines-Omar et portant sur les transferts entre tranches d'âge a été publiée dans l'ouvrage des Comptes de la santé 2010.

Finalement, le fait qu'Omar soit un outil encore naissant et amené à évoluer dans les prochaines années empêche pour le moment d'en faire un outil pour mener des analyses longitudinales. L'intérêt

d'Omar réside d'une part, dans sa capacité à fournir des premiers chiffrages de l'intervention des organismes d'AMC et, d'autre part, à rendre possible des comparaisons transversales, entre déciles de niveau de vie pour une année donnée par exemple. Ces travaux de microsimulation ne sont qu'un substitut imparfait à l'absence d'observation de données réelles qui seraient directement issues d'un appariement des données de l'assurance maladie obligatoire et des assurances complémentaires. Des expérimentations en ce sens sont en cours, comme le projet Monaco mis en œuvre par l'Irdes avec la collaboration de la Cnamts et des organismes complémentaires volontaires sous l'égide de l'Institut des données de santé (IDS). Il est souhaitable que ce type de rapprochement de données se développe, comme le recommande vivement le Conseil National de l'Information Statistique (Cnis).

Bibliographie

- [1] Allonier C., Dourgnon P., Rochereau T., « Enquête sur la Santé et la Protection Sociale », *Rapport Irdes*, n°1800, juin 2010.
- [2] Arnould M-L, Vidal G., « Typologie des contrats les plus souscrits auprès des complémentaires santé en 2006 », *Études et résultats*, n°663, octobre 2008.
- [3] Beeuwkes Buntin M., Zaslavsky A.M., « Too much about two-part models and transformation? Comparing methods of modelling Medicare expenditures », *Journal of Health Economics*, vol. 23 n°3, pp 525-542, mai 2004.
- [4] Bocognano A., Couffinhal A., Dusmenil S., Grignon M., « La complémentaire maladie en France : qui bénéficie de quels remboursements ? », *Questions d'économie de la Santé*, n°32, octobre 2000.
- [5] Boisguérin B., « État de santé et recours aux soins des bénéficiaires de la CMU : un impact qui se consolide entre 2000 et 2003 », *Études et résultats*, n°294, mars 2004.
- [6] Boisguérin B., « Les allocataires des minima sociaux : CMU, état de santé et recours aux soins », *Études et résultats*, n°603, octobre 2007.
- [7] Boisguérin B., « Quelles caractéristiques et quel recours aux soins pour les bénéficiaires de la CMUC ? », *Études et résultats*, n°675, janvier 2009.
- [8] Boisguérin B., Després C., Dourgnon P., Fantin R., Legal R., « Étudier l'accès aux soins des assurés CMU-C, une approche par le renoncement aux soins », in Enquête sur la santé et la protection sociale 2010, *Rapport Irdes*, n°1800, juin 2010.
- [9] Breuil-Genier P., Grandfils N., Raynaud D., « Revenus, assurance et santé : le problème de l'accès aux soins des plus démunis », *Les cahiers du Gratice*, n°15, pp 243-76, 1998.
- [10] Caussat L., Glaude M., « Dépenses médicales et couverture sociale », *Économie et Statistique*, n°265, pp 31-43, 1993.
- [11] Caussat L., Le Minez S., Raynaud D., « L'assurance-maladie contribue-t-elle à redistribuer les revenus ? », *Dossiers Solidarité et Santé*, n°1, 2005.
- [12] Couffinhal A., Perronnin M., « Accès à la couverture complémentaire maladie en France : une comparaison des niveaux de remboursement », *Questions d'économie de la Santé*, n°47, février 2004.
- [13] Couffinhal, Grandfils N., Grignon M., Rochereau T., « La complémentaire maladie d'entreprise. Premiers résultats nationaux d'une enquête menée fin 2003 auprès de 1700 établissements », *Questions d'économie de la Santé*, n°83, juin 2004.
- [14] Currie J., « Child Health in Developed Countries », *Handbook of Health Economics*, vol. 1B, pp. 1054-1090, 2000.
- [15] Dormont B., Hubert H., Grignon M., « Health expenditures growth: reassessing the threat of ageing » *Health economics*, vol.15 n°9, septembre 2006.
- [16] Dourgnon P., Grignon M., Jusot F., « L'assurance maladie réduit-elle les inégalités sociales de santé ? Une revue de littérature », *Questions d'économie de la Santé*, n°43, décembre 2007.
- [17] Francesconi C., Perronnin M., Rochereau T., « La complémentaire maladie d'entreprise : niveaux de garanties des contrats selon les salariés et le secteur d'activité », *Questions d'économie de la Santé*, n°112, Septembre 2006.

- [18] Franc. C., Perronnin M., « Aide à l'acquisition d'une assurance maladie complémentaire : une première évaluation du dispositif ACS », *Questions d'économie de la santé*, n°121, avril 2007.
- [19] Francesconi C., Perronnin M., Rochereau T., « Complémentaire maladie d'entreprise : contrats obligatoires ou facultatifs, lutte contre l'antisélection et conséquences pour les salariés », *Questions d'économie de la Santé*, n°115, novembre 2006.
- [20] Garnero M., Rattier MO, « Les contrats les plus souscrits auprès des complémentaires santé en 2008 », *Études et Résultats*, n°752, février 2011.
- [21] Geoffard P.Y., « Assurance maladie : la gestion du risque long » *Revue d'économie politique*, vol. 110 n°4, pp. 457 -482, décembre 2000.
- [22] Grignon M., Kambia-Chopin B. (2010) : « Quelle subvention optimale pour l'achat d'une complémentaire santé ? », *Questions d'économie de la Santé*, n°153, avril 2010.
- [23] Guillaume S., Rochereau T., « La complémentaire santé collective : une offre aux salariés très inégale selon les entreprises », *Questions d'économie de la Santé*, n°154, mai 2010.
- [24] Heckman J.J., « Sample selection bias as a specification error » *Econometrica*, n°47pp 153-161, janvier 1979.
- [25] Kambia-Chopin B., Perronnin M., Pierre A., Rochereau T., « La complémentaire santé en France en 2006 : un accès qui reste inégalitaire », *Questions d'économie de la Santé*, n°132, mai 2008.
- [26] Lengagne P., Perronin M., « Impact des niveaux de garanties des complémentaires santé sur les consommations de soins peu remboursées par l'Assurance maladie : le cas des lunettes et des prothèses dentaires », *Questions d'économie de la Santé*, n°100, novembre 2004.
- [27] Leung S.F, Yu S., « One the choice between sample selection and two-part models », *Journal of Econometrics*, vol. 72 n°1-2, pp 197-229, mai 1996.
- [28] Lollivier S., « Modèles univariés et modèles de durée sur données individuelles », *Document de travail, Méthodologie Statistique, Insee*, n°9702, 1997.
- [29] Manning W., Duan N., Rogers W., « Monte-Carlo evidence on the choice between sample selection and two-part models » *Journal of econometrics*, n°35, pp 59-82, 1987.
- [30] Newhouse J.-P., «Free for All? Lessons from the RAND experiment», Harvard University Press, 1993.
- [31] Pichetti S., Raynaud D., Vidal G., « Les déterminants individuels des dépenses de santé », *Traité d'économie et de gestion de la santé*, juillet 2009.
- [32] Raynaud D., « Les déterminants individuels des dépenses de santé », *Études et résultats*, n°182, juillet 2002.
- [33] Raynaud D., « Les déterminants individuels des dépenses de santé : l'influence de la catégorie sociale et de l'assurance maladie complémentaire », *Études et résultats*, n°378, février 2005.
- [34] Sautory O., « La macro calmar : redressement d'un échantillon par calage sur marges », *Document de travail*, n°F9310, novembre 1993.

Annexes

Annexe 1 : L'imputation des revenus

Tableau 5 - Proportion et nombre de ménages ayant répondu en montant ou en tranche ou n'ayant pas répondu à la question sur les revenus

Type de réponse	Données avec pondération		Données sans pondération	
	Proportion (%)	Nombre de ménages	Proportion (%)	Nombre de ménages
Revenu en montant	67	5 430	66	5 335
Revenu par tranche	9	731	10	802
Revenu inconnu	24	1 968	24	1 992

Source : Appariement Epas-SPS 2006, Irdes, Cnamts, RSI, MSA, - traitement Drees
Champ : Ménage ordinaire, France métropolitaine

Des sources de revenus très hétérogènes

Suspectant des sources de revenus assez hétérogènes selon le statut vis-à-vis de l'emploi des membres du ménage, nous avons distingués *a priori* quatre types de ménages : les ménages dont les adultes sont uniquement des actifs occupés, ceux dont les adultes sont uniquement des inactifs ou des chômeurs, ceux dont les adultes sont uniquement des retraités et les autres (*cf.* tableau 6). Sont considérés comme *actifs occupés* les individus qui occupent un travail au moment de l'enquête, qui sont en congé maternité ou qui sont en congé longue maladie. Sont considérées comme *retraités* les personnes qui se déclarent à la retraite ou retiré des affaires ou celles qui ont plus de soixante-cinq ans et qui n'exercent plus d'activité professionnelle. Ce dernier critère permet de classer dans cette catégorie les personnes qui ont déclaré une PCS mais qui n'exercent plus leur métier et les veuves et les veufs de plus de 65 ans qui perçoivent une pension de réversion. Sont rassemblés dans la catégorie *inactifs ou chômeurs* les adultes qui ne sont ni classés comme actifs occupés ni comme retraités. Cette classe comprend donc les individus scolarisés de plus de seize ans, les enfants, les demandeurs d'emploi, les hommes et les femmes au foyer de moins de 65 ans, les inactifs avec pension d'invalidité etc.

Tableau 6 - Part des différentes sources de revenus selon le type de ménage

	Actifs occupés	Retraités	Inactifs ou chômeurs	Autres
% dans l'ensemble des ménages	45.8%	19.8%	8.5%	25.9%
Revenus d'activité	91.6	1.3	5.7	67.1
Pensions de retraite	0.4	90.5	4.0	14
Revenus de transfert	6.1	4.3	84.3	16.2
Autres	1.9	3.9	6.0	2.7

Source : Appariement Epas-SPS 2006, Irdes, Cnamts, RSI, MSA, - traitement Drees
Champ : Ménage ordinaire, France métropolitaine

L'estimation de quatre équations de revenu

Compte tenu de la distribution log-normale des revenus, c'est le logarithme du revenu qui est modélisé, et non le revenu lui-même. Par ailleurs, il est possible que l'équation de revenu souffre d'un biais de sélection. C'est le cas si certaines variables inobservables jouent à la fois sur la décision de répondre à la question du revenu et sur le montant du revenu déclaré. Pour tester l'existence d'un tel phénomène et le corriger le cas échéant, il convient donc d'estimer l'équation d'imputation de revenu au moyen d'un modèle *Tobit* généralisé en faisant l'hypothèse d'une distribution jointe des résidus (en l'occurrence, *via* une loi normale bivariée) :

$$\begin{cases} P_i^* = X_{2i}' \cdot \beta_2 + u_{i2} & \text{(équation de réponse)} \\ \log(R_i) = (X_{1i}' \cdot \beta_1 + u_{i1}) P_i & \text{(équation de revenu)} \end{cases}$$

$$\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

L'existence d'un biais de sélection se matérialise par un coefficient de corrélation des résidus ρ significativement non nul.

L'estimation des quatre modèles *Tobits* généralisés¹ est obtenue par la *proc qlim* de SAS (méthode du maximum de vraisemblance). Les résultats montrent qu'il n'existe pas de biais de sélection (ρ non significativement différent de 0) (tableau 4).

Tableau 7 - Coefficients de corrélation des résidus des 4 modèles Tobit généralisés de revenu

Catégorie de ménages	ρ estimé	p-value
actifs	-0,093	0,4066
retraités	-0,0303	0,9467
inactifs	0,0065	0,98
mixtes	-0,0798	0,7356

Quel que soit le modèle ce coefficient n'est jamais significativement différent de zéro (cf. tableau 7). Au final, il est donc possible d'estimer directement l'équation de revenu pour obtenir une estimation non biaisée de β_1 .

Environ 10% des individus ont déclaré leur revenu en tranches. En notant E_1 (resp. E_2) le sous-échantillon des individus qui ont fourni leur revenu en euros (resp. en tranches), et R_1 et R_2 les bornes respectivement inférieure et supérieure des tranches déclarées, nous estimons β_1 et de σ_1 par maximisation de la fonction de vraisemblance suivante au moyen de la *proc lifereg* sous SAS :

$$L = \prod_{i \in E_1} \phi\left(\frac{\log(R_i) - X'_{1i}\beta_1}{\sigma_1}\right) \prod_{i \in E_2} \left(\Phi\left(\frac{\log(R_{2i}) - X'_{1i}\beta_1}{\sigma_1}\right) - \Phi\left(\frac{\log(R_{1i}) - X'_{1i}\beta_1}{\sigma_1}\right) \right)$$

De cette manière, nous prenons en compte l'information des individus qui ont déclaré leurs revenus en montant mais aussi en tranche ([28]).

Imputation des revenus manquants par la méthode des résidus simulés

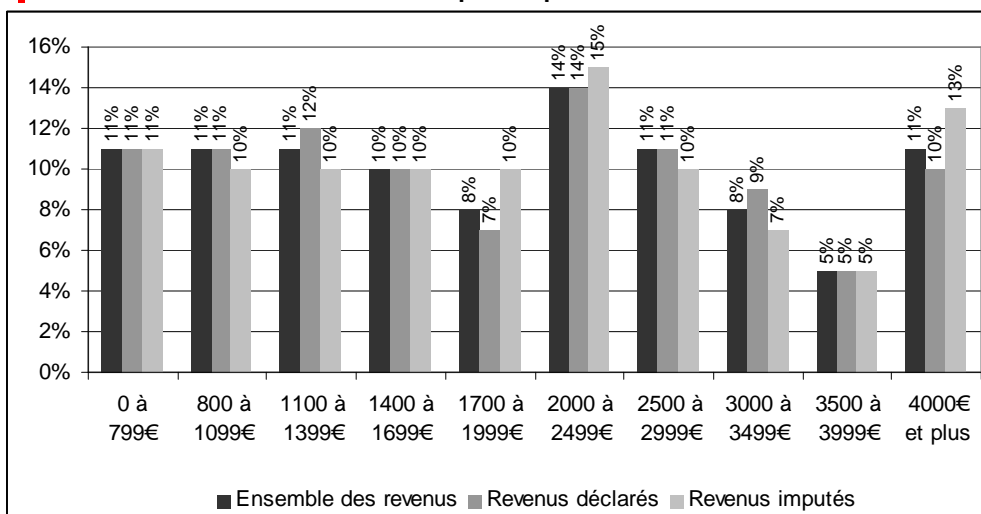
Les revenus des non-répondants (non-réponse totale et non-réponse partielle, *i.e.* en tranche) sont alors imputés par la méthode des résidus simulés. Pour cela, on simule le logarithme du revenu d'un individu de caractéristiques observables X_i qui n'aurait pas déclaré son revenu à l'aide de l'espérance conditionnelle du revenu à laquelle on ajoute une perturbation :

$$(\log(R_i))^{sim} = E(\log(R_i)|X_i) + u_i^{sim} = X'_{1i}\hat{\beta}_1 + u_i^{sim}$$

Si bien que $R_i^{sim} = \exp(X'_{1i}\hat{\beta}_1 + u_i^{sim})$.

Le résidu simulé u_i^{sim} est tiré dans une loi normale de variance $\hat{\sigma}_1^2$. Pour les individus qui ont déclaré leur revenu en tranches, tant que le revenu simulé n'est pas dans la tranche, un nouveau résidu est tiré au sort. Pour les autres, on contraint le revenu simulé à être supérieur au minimum des revenus déclarés et inférieur au maximum des revenus déclarés (pour la catégorie de ménages correspondant). Ce dernier choix suppose implicitement que les revenus minimum et maximum par catégorie de ménages sont bien observés. Cette hypothèse peut paraître forte, mais elle permet d'éviter d'introduire aléatoirement des revenus qui seraient atypiques, et par ailleurs, elle n'a pas d'impact sur les résultats car les revenus sont ensuite traités en tranche (déciles).

Graphique 7 - Distribution des revenus avant et après imputation



Source : Omar 2006.

Champ : ensemble des ménages

Annexe 2 : Résultats des estimations des 3 modèles Tobit généralisés pour modéliser la dépense

Tableau 8 - Modélisation de la présence ou non d'une consommation hors optique et dentaire

Variable	Modalité	Coefficient estimé	Écart-type
Constante		1,73 ***	0,11
Nature de la couverture complémentaire	Aucune	-0,48 ***	0,12
	Privée	réf.	réf.
	CMU	0,25 *	0,13
Régime	MSA	0,33 **	0,16
	RSI	-0,19	0,14
	Cnamts et autres	réf.	réf.
Exonération	Non et NSP	réf.	réf.
	Oui	0,99 ***	0,17
Consultation d'un omnipraticien	0	-1,09 ***	0,1
	1 ou 2	réf.	réf.
	3 ou 4	0,37 ***	0,13
	> 5	0,54 ***	0,15
	NSP	-0,64 ***	0,15
Consultation d'un spécialiste	0	réf.	réf.
	> 1	0,64 ***	0,1
	NSP	0,31 **	0,12
Consommation de médicaments	Oui	0,29 **	0,11
	Non	réf.	réf.
	Inconnu	0,14	0,14
État de santé perçu	Dégradé	0,08	0,14
	Bon à très bon	réf.	réf.
	Inconnu	0,02	0,18
Sexe et âge	0 à 19 ans	0,90 ***	0,11
	40 à 59 ans	0,33 ***	0,1
	60 à 69 ans	0,48 ***	0,17
	> 70 ans	0,79 ***	0,2
	Homme de 20 à 39 ans	réf.	réf.
	Femme de 20 à 39 ans	0,55 ***	0,13
Niveau de revenus	Élevé	-0,01 *	0,09
	Moyen	réf.	réf.
	Faible	-0,28 ***	0,09
Nb de personnes dans le ménage	1 à 3	réf.	réf.
	> 4	-0,13 *	0,08

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le pourcentage de paires concordantes est de 73,3

Tableau 9 - Modélisation du montant de la consommation hors optique et dentaire

<i>Variable</i>	<i>Modalité</i>	<i>Coefficient estimé</i>	<i>Écart-type</i>
Constante		7,15***	0,1
Nature de la couverture complémentaire	Aucune	-0,57***	0,06
	Privée	-0,28***	0,04
	CMU	réf.	réf.
Régime	MSA	-0,57***	0,07
	RSI	réf.	réf.
	Cnamts et autres	-0,40***	0,06
Exonération	Non et NSP	-0,91***	0,04
	Oui	réf.	réf.
Consultation d'un omnipraticien	0	-0,45***	0,07
	1 ou 2	-0,22***	0,06
	3 ou 4	0,22***	0,06
	> 5	0,56***	0,06
	NSP	réf.	réf.
Consultation d'un spécialiste	0	-0,44***	0,04
	> 1	0,09**	0,04
	NSP	réf.	réf.
Consommation de médicaments	Oui	0,17***	0,05
	Non	-0,22***	0,05
	Inconnu	réf.	réf.
État de santé perçu	Dégradé	0,35***	0,07
	Bon à très bon	0,01	0,07
	Inconnu	réf.	réf.
Sexe et âge	0 à 19 ans	0,11**	0,05
	40 à 59 ans	0,54***	0,05
	60 à 69 ans	0,86***	0,06
	> 70 ans	1,10***	0,06
	Homme de 20 à 39 ans	réf.	réf.
	Femme de 20 à 39 ans	0,55***	0,05
Niveau de revenus	Élevé	réf.	réf.
	Moyen	-0,01	0,03
	Faible	-0,01	0,04
Nb de personnes dans le ménage	1 à 3	0,15***	0,03
	> 4	réf.	réf.

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le R^2 du modèle vaut 0,4074

Tableau 10 - Modélisation de la présence ou non d'une consommation en optique

<i>Variable</i>	<i>Modalité</i>	<i>Coefficient estimé</i>	<i>Écart-type</i>
Constante		-0,20***	0,13
Nature de la couverture complémentaire	Aucune	-0,54***	0,13
	Privée	réf.	réf.
	CMU	-0,14	0,11
Spécialiste	Ophtalmologue	1,12***	0,07
	Autres et non	réf.	réf.
	NSP	0,58***	0,11
Lunettes	Oui	réf.	réf.
	Non	-1,70***	0,12
	Inconnu	-1,58***	0,16
Indicateur synthétique de problèmes de vue	Loin et près	0,45***	0,1
	Loin ou près	0,36***	0,09
	Aucun	réf.	réf.
	Inconnu	0,96***	0,14
Maladie vue	Non	-0,32***	0,11
	Oui	réf.	réf.
	NSP	0,45**	0,18
Classes d'âge	0 à 15 ans	0,14	0,1
	16 à 44 ans	réf.	réf.
	Homme de 45 à 64 ans	0,21**	0,09
	Femme de 45 à 64 ans	0,29***	0,09
	> 65	-0,28***	0,1
Nb de personnes dans le ménage	1 à 3	réf.	réf.
	> 4	0,16**	0,07
Niveau de revenus	Élevé	0,03	0,07
	Moyen	réf.	réf.
	Faible	-0,18**	0,07
Régime	MSA	0,01	0,12
	Cnamts et autres	réf.	réf.
	RSI	-0,02	0,12
Niveau d'études	1	réf.	réf.
	2	-0,08	0,07

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le pourcentage de paires concordantes est de 76,1%.

Tableau 11 - Modélisation du montant de la consommation en optique

<i>Variable</i>	<i>Modalité</i>	<i>Coefficient estimé</i>	<i>Écart-type</i>
Constante		5,21***	0,1274943
Nature de la couverture complémentaire	Aucune	0,59***	0,10
	Privée	1,07***	0,07
	CMU	réf.	réf.
Spécialiste	Autres et non	-0,01	0,08
	NSP	réf.	réf.
	Ophthalmologue	0	0,07
Lunettes	Oui	0,25**	0,12
	Inconnu	réf.	réf.
	Non	-0,01	0,12
Indicateur synthétique de problèmes de vue	Loin et près	0,11	0,09
	Loin ou près	-0,06	0,09
	Inconnu	réf.	réf.
	Aucun	-0,05	0,09
Maladie vue	Oui	-0,33**	0,13
	Non	-0,24**	0,12
	NSP	réf.	réf.
Classes d'âge	0 à 15 ans	-0,36***	0,07
	16 à 44 ans	-0,25***	0,06
	Homme de 45 à 64 ans	0,05	0,06
	Femme de 45 à 64 ans	0,16**	0,06
	> 65	réf.	réf.
Niveau de revenus	Élevé	réf.	réf.
	Faible	-0,22***	0,05
	Moyen	-0,11**	0,04
Nb de personnes dans le ménage	1 à 3	-0,03	0,04
	> 4	réf.	réf.
Régime	Cnamts et autres	-0,34***	0,07
	MSA	-0,41***	0,10
	RSI	réf.	réf.
Niveau d'études	1	0,15***	0,04
	2	réf.	réf.

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le R^2 du modèle vaut 0,2984

Tableau 12 - Modélisation de la présence ou non d'une consommation en dentaire

<i>Variable</i>	<i>Modalité</i>	<i>Coefficient estimé</i>	<i>Écart-type</i>
Constante		-0,97 ***	0,09
Nature de la couverture complémentaire	Aucune	-0,41 ***	0,09
	Privée	réf.	
	CMU	0,05	0,08
Classes d'âge	0 à 5 ans	-1,75 ***	0,17
	6 à 15 ans	0,67 ***	0,10
	Femmes de 16 à 29 ans	0,49 ***	0,11
	Hommes de 16 à 29 ans	réf.	
	Femmes de 30 à 49 ans	0,50 ***	0,10
	Hommes de 30 à 49 ans	0,19 **	0,10
	Femmes de 50 à 64 ans	0,52 ***	0,11
	Hommes de 50 à 64 ans	0,13	0,11
	Femmes de plus de 65 ans	0,09	0,11
Niveau de revenus	Élevé	0,24 ***	0,05
	Moyen	réf.	
	Faible	-0,1576 ***	0,05
Régime	MSA	0,11	0,09
	Cnamts et autres	réf.	
	RSI	0,07	0,09
Nb de personnes dans le ménage	1 à 3	réf.	
	> 4	0,05	0,05
État de la dentition	Inconnu	0,28 ***	0,09
	Mauvais	-0,29 ***	0,07
	Bon	réf.	
	Très bon	-0,11	0,07
Problèmes de dents	Oui	0,74 ***	0,06
	Non	réf.	
	Inconnu	-0,16 **	0,08
Prothèse	Dentier	-0,12	0,09
	Fixe	0,57 ***	0,06
	Inconnu	0,01	0,08
	Non	réf.	
	Redressement	1,64 ***	0,21

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le pourcentage de paires concordantes est de 67,4%

Tableau 13 - Modélisation du montant de la consommation en dentaire

<i>Variable</i>	<i>Modalité</i>	<i>Coefficient estimé</i>	<i>Écart-type</i>
Constante		5,61***	0,16
Nature de la couverture complémentaire	Aucune	-0,21*	0,11
	Privée	-0,14*	0,08
	CMU	réf.	réf.
Classes d'âge	0 à 5 ans	-1,31***	0,21
	6 à 15 ans	-0,38***	0,10
	Femmes de 16 à 29 ans	-0,35***	0,10
	Hommes de 16 à 29 ans	-0,18	0,11
	Femmes de 30 à 49 ans	-0,10	0,08
	Hommes de 30 à 49 ans	-0,06	0,09
	Femmes de 50 à 64 ans	-0,19**	0,09
	Hommes de 50 à 64 ans	0,01	0,09
	Femmes de plus de 65 ans	réf.	réf.
Niveau de revenus	Élevé	réf.	réf.
	Moyen	0,04	0,05
	Faible	0,00	0,06
Régime	MSA	-0,42***	0,12
	Cnamts et autres	-0,45***	0,09
	RSI	réf.	réf.
Nb de personnes dans le ménage	1 à 3	-0,13**	0,05
	> 4	réf.	réf.
Etat de la dentition	Inconnu	réf.	réf.
	Mauvais	0,2*	0,12
	Bon	-0,11	0,10
	Très bon	-0,39***	0,11
Problèmes de dents	Oui	0,06	0,16
	Non	-0,14*	0,08
	Inconnu	réf.	réf.
Prothèse	Dentier	0,3**	0,12
	Fixe	0,41***	0,10
	Inconnu	réf.	réf.
	Non	-0,03	0,08
	Redressement	1,7***	0,15

Note de lecture : * : Significatif au seuil de 10% ; ** : Significatif au seuil de 5% ; *** : significatif au seuil de 1%
Le R² du modèle vaut 0,1006

Annexe 3 : Résumés des déroulements des *hot deck* stratifiés

Tableau 14 - Imputation des dépenses en optique

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour	9 ^{ème} tour
Nombre d'individus auxquels sont imputées des dépenses en optique	5843	2309	436	1140	613	974	296	87	4
	49,93%	19,73%	3,73%	9,74%	5,24%	8,32%	2,53%	0,74%	0,03%
Variables de stratification	Spécialiste	x	x	x	x	x	x	x	x
	Nature de la couverture complémentaire	x	x	x	x	x	x	x	x
	Classes d'âge	x	x	x	x	x	x	x	
	Régime	x	x	x	x	x	x	x	
	Indicateur synthétique de problèmes de vue	x	x	x	x	x	x		
	Niveau de revenus	x	x	x	x	x			
	Lunettes	x	x	x	x				
	Niveau d'études	x	x	x					
	Maladie vue	x	x						
	Nb de personnes dans le ménage	x							
R ² du modèle associé (équation de montant)	0,2984	0,2981	0,2957	0,2911	0,2874	0,2742	0,2663	0,2553	0,1973

Tableau 15 - Imputation des dépenses en dentaire

	1 ^{er} tour	2 ^{ème} tour	3 ^{ème} tour	4 ^{ème} tour	5 ^{ème} tour	6 ^{ème} tour	7 ^{ème} tour	8 ^{ème} tour
Nombre d'individus auxquels sont imputées des dépenses en dentaire	5228	2048	2354	871	399	138	81	583
	44,68%	17,50%	20,12%	7,44%	3,41%	1,18%	0,69%	4,98%
Variables de stratification	Prothèse	x	x	x	x	x	x	x
	Classes d'âge	x	x	x	x	x	x	
	Régime	x	x	x	x	x	x	
	Problèmes de dents	x	x	x	x	x		
	Nature de la couverture complémentaire	x	x	x	x			
	Etat de la dentition	x	x	x				
	Niveau de revenus	x	x					
	Nb de personnes dans le ménage	x						
R ² du modèle associé (équation de montant)	0,1006	0,0992	0,0991	0,0878	0,0865	0,0849	0,0795	0,0593

Annexe 4 : Marges utilisées pour le recalage des individus pour tenir compte de la sous-représentation des dépenses hospitalières importantes

Tableau 16 - Marges utilisées dans le recalage des individus

Variables	Modalités	Marges
Age	0-19 ans	24,6
	20-39 ans	26,8
	40-59 ans	27,9
	60-79 ans	16,7
	80 ans et plus	4
Sexe	Homme	48,3
	Femme	51,7
Régime d'affiliation	Régime Général	86,8
	RSI	4,5
	MSA	5,9
	Autres	2,8
Taille du ménage	1	12,9
	2	25,9
	3	20,2
	4	22,9
	5 et plus	18,1
Distribution dépenses hospitalières	<223	84,9
	223-<845	5,1
	845-<1653	3
	1653-<3027	2
	3027-<6682	2
	6682-<11202	1
	11202-<23210	1
	23210<=	1
% ALD par niveau de dépense hospitalière	<223	9,2
	223-<845	0,9
	845-<1653	0,7
	1653-<3027	0,6
	3027-<6682	0,8
	6682-<11202	0,5
	11202-<23210	0,6
	23210<=	0,7
	Non ALD	86

Annexe 5 : Quelques caractéristiques sociodémographiques de l'échantillon

Tableau 17 - Taille des ménages, structure par sexe et âge et état de santé selon le décile de niveau de vie

Décile de niveau de vie	Nombre moyen de personnes par ménage	% de femmes	% de moins de 20 ans	% de plus de 60 ans	% de personnes en ALD
D1	2,7	53	34	15	14
D2	2,3	55	27	25	18
D3	2,4	54	25	28	19
D4	2,5	54	27	26	17
D5	2,4	51	28	19	12
D6	2,4	50	26	18	12
D7	2,4	50	23	19	14
D8	2,5	51	21	18	13
D9	2,3	50	18	19	12
D10	2,2	49	16	22	11
Ensemble	2,4	52	25	21	14

Champ : Ménage ordinaire, France métropolitaine.

Source : Appariement Epas-SPS 2006, Cnamts, MSA, RSI, IRDES, traitement de la Drees.

Annexe 6 : Comparaison des résultats d'Omar avec les Comptes de la santé

Omar présente une structure de financement des dépenses de santé relativement proche de celle décrite par les Comptes de la santé. Les écarts qui peuvent apparaître tiennent surtout à des différences de champ.

Omar aboutit à un taux de remboursement de la Sécurité sociale un peu plus important que celui des Comptes de la santé (79 % contre 77 %) car elle ne porte que sur les dépenses présentées au remboursement. Cette différence de champ est surtout perceptible pour la pharmacie, puisque l'automédication est pris en considération dans les Comptes de la santé alors qu'elle est exclut du champ de l'outil de microsimulation.

Les dépenses hospitalières sont sous-estimées dans Omar. Cet écart tient d'abord au champ retenu par l'enquête SPS qui n'interroge que les ménages ordinaires. Or, les personnes en institution ont tendance à avoir des dépenses hospitalières plus importantes. De plus, les personnes les plus malades ont un taux de réponse plus faible à l'enquête SPS. Enfin, certaines dépenses hospitalières ne sont pas individualisables. Ce type de dépense est inclus dans les Comptes de la santé, mais pas dans Omar.

Enfin, une part des écarts s'explique par les aléas statistiques dus à la méthodologie qui repose sur des imputations successives.

Tableau 18 - Comparaison d'Omar avec les Comptes de la santé (2006)

	Omar				Comptes de la santé			
	AMO	CMU-C	AMC	Ménage	AMO	CMU-C	AMC	Ménage
Médecins	73	2	17	9	71	2	18	9
Dentistes	37	3	33	26	35	2	35	28
Actes de biologie	72	2	22	4	74	2	21	3
Auxiliaires médicaux	84	1	13	1	81	1	9	9
Transports	89	1	7	3	94	1	3	2
Pharmacie	77	2	21	1	68	1	19	12
Autres biens médicaux	38	2	27	15	43	1	30	26
Total ambulatoire	68	2	21	10	65	1	21	13
Hôpital	96	0	2	2	92	1	4	3
Total	79	1	13	6	77	1	13	9

Champ : Ménage ordinaire, France métropolitaine.

Source : Omar et Comptes de la santé.

Les similitudes entre Omar et les Comptes de la santé n'équivalent pas à une validation d'Omar. Mais *a contrario*, des chiffres agrégés très éloignés des comptes auraient invalidé l'outil. Ce n'est pas le cas.