

points de repere



L'Échantillon généraliste de bénéficiaires (EGB) est un échantillon permanent représentatif de la population protégée par l'assurance maladie, qu'elle ait ou non perçu des remboursements de soins. Il regroupe actuellement près de 500 000 bénéficiaires du régime général et doit, à terme, couvrir l'ensemble des régimes d'assurance maladie sur une période de 20 ans. Il contient des informations anonymes sur les caractéristiques socio-démographiques et médicales des bénéficiaires et les prestations qu'ils ont perçues.

Cet échantillon permet de réaliser des études longitudinales et de reconstituer l'ensemble du parcours de soins des patients sur une longue période, que ce soit en ville ou à l'hôpital, notamment grâce aux données du Programme de médicalisation des systèmes d'information (PMSI). Il permet également d'estimer la population protégée, en particulier sa répartition par âge et sexe. Il constitue enfin un excellent outil pour étudier le recours aux soins et les dépenses de santé par personne protégée et par poste, ainsi que leurs évolutions au cours du temps.

L'Échantillon généraliste de bénéficiaires : représentativité, portée et limites

La constitution d'un échantillon permanent de bénéficiaires de soins par la Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS) répond au besoin de mieux connaître et comprendre le recours aux soins, les trajectoires de soins et les dépenses de santé des assurés sociaux sur une longue période¹.

L'arrêté du 20 juin 2005² relatif à la mise en œuvre du système d'informations inter-régimes de l'assurance maladie (SNIIRAM) prévoit dans son article 3 qu' « un échantillon généraliste [...] représentatif des personnes protégées des régimes est constitué, afin d'assurer le suivi de la consommation de soins et des taux de recours aux soins. Sa durée de conservation est de vingt ans au-delà de l'année en cours ». Ce panel, appelé Échantillon généraliste de bénéficiaires (EGB), est géré par la CNAMTS et a fait l'objet d'un accord de la Commission nationale de l'informatique et des libertés (CNIL)³.

● Laurence de Roquefeuil, Anna Studer, Anke Neumann, Yvon Merlière (CNAMTS)

¹ Polton D. Les bases de données de l'Assurance maladie, un potentiel pour la recherche en économie de la santé. La lettre du Collège n°1, Collège des économistes de la santé, mars 2009.

² Arrêté paru au Journal officiel du 19 août 2005.

³ Accords CNIL AT/CPZ/SVT/JB/DP/CR05222O du 14 juin 2005 et DP/CR071761 du 28 août 2007.

Il a pour principe d'associer aux informations socio-démographiques des bénéficiaires, contenues dans les bases de données administratives des différents régimes de sécurité sociale, les données de remboursement de soins de ces personnes, extraites du SNIIRAM. Toutes ces informations, bien qu'individualisées, sont rendues anonymes.

L'EGB résulte d'un sondage au 1/97^e sur le numéro de sécurité sociale⁴ (NIR) des bénéficiaires de l'assurance maladie, qu'ils soient ouvriers de droit ou ayants droit, et recense près de 500 000 bénéficiaires. S'il ne couvre actuellement que le régime général à l'exception des fonctionnaires et des étudiants, il a vocation à s'étendre à l'ensemble des régimes d'assurance maladie ; dès 2010, les ressortissants de la Mutualité sociale agricole (MSA) et du Régime social des indépendants⁵ (RSI) devraient y être intégrés.

A l'heure où l'EGB devient accessible à un nombre croissant d'utilisateurs internes et externes à l'Assurance maladie et où l'enquête Santé et protection sociale (ESPS), menée depuis 1988 par l'Institut de recherche et de documentation en économie de la santé (IRDES), se fera en 2010 à partir de cet échantillon, il convient d'apporter un éclairage sur sa représentativité, sa portée et ses limites.

● Du SNIIRAM à l'EGB

Le SNIIRAM est un système médicalisé d'informations mis en œuvre par la CNAMTS à compter du 1^{er} janvier 2003, en application de l'article L 161-28-1 du code de la Sécurité sociale. Cette base de données individualisées et

Encadré 1

Contenu du SNIIRAM

Outre des données concernant les dépenses de santé et le parcours de soins des patients de tous les régimes d'assurance maladie, le SNIIRAM contient :

- des informations de nature médicale : numéro de l'affection de longue durée⁶, code CIM 10 (classification internationale des maladies, 10^e version) de ses diagnostics et données médicalisées hospitalières du Programme de médicalisation des systèmes d'information (PMSI) ;
- des données socio-démographiques des patients (sexe, âge, caisse d'affiliation, département, commune de résidence, affiliation à la couverture maladie universelle) ;
- des informations sur les professionnels de santé et les établissements de soins ;
- des éléments médico-administratifs, notamment :
 - les codes CIP⁷ de l'ensemble des médicaments remboursés, à l'exception des médicaments rétrocédés à l'hôpital (médicaments délivrés par les pharmacies des hôpitaux) et des médicaments en sus du GHS⁸ (Groupe homogène de séjour, défini dans le cadre de la tarification à l'activité) ; le codage du médicament représente 82 % des dépenses de médicaments remboursés en 2007 pour le régime général en France métropolitaine ;
 - les codes détaillés des actes de biologie, des dispositifs médicaux (prothèses et appareillage principalement) et, depuis janvier 2007, de l'ensemble des actes techniques médicaux (codage CCAM⁹), à l'exception de ceux des anatomo-cytopathologistes.

anonymes est exhaustive : elle recense toutes les dépenses de santé des différents régimes d'assurance maladie (encadré 1).

Les données du SNIIRAM sont conservées pendant deux ans au-delà de l'année en cours. Cette durée peut être insuffisante pour assurer le suivi de la consommation de soins de certaines catégories de bénéficiaires ou connaître l'impact d'actions à long terme. En outre, le SNIIRAM ne fournit aucun renseignement sur les personnes qui ne perçoivent pas de remboursement au cours d'une période donnée, appelées « non consommateurs ». L'EGB a été mis en place pour remédier à ces deux inconvénients : c'est un échantillon de bénéficiaires consommateurs ou non, dont les données de remboursement

sont conservées vingt ans au-delà de l'année en cours. Les premières informations de l'EGB concernant la consommation de soins datent de janvier 2003.

L'intégration de cet échantillon comme un sous-produit du SNIIRAM oblige les différents régimes à garantir le strict anonymat des informations qu'il contient et à respecter les règles de sécurité du SNIIRAM (certaines données, dites sensibles, ne peuvent être croisées que par les utilisateurs qui en ont l'habilitation). L'arrêté ministériel du 20 juin 2005 relatif au SNIIRAM réglemente la mise en place de l'EGB et précise quels organismes extérieurs à l'Assurance maladie sont autorisés à l'exploiter (encadré 2).

⁴ Le numéro de sécurité sociale d'un individu est son numéro d'inscription au répertoire des personnes physiques (NIR). Un NIR est attribué par l'INSEE à chaque personne résidant sur le territoire français, ce numéro reste identique tout au long de sa vie.

⁵ Artisans, commerçants, professions libérales.

⁶ Au 31 décembre 2007, près de 8 millions de personnes relevaient de ce dispositif.

⁷ CIP : club inter pharmaceutique. Ce code est le numéro de l'autorisation de mise sur le marché attribué à chaque spécialité pharmaceutique.

⁸ L'intégration du codage de la rétrocession hospitalière est prévue pour janvier 2009 ; la date du codage des médicaments facturés en sus du forfait n'est pas encore connue.

⁹ CCAM : classification commune des actes médicaux. La liste de ces actes est consultable sur le site internet de l'Assurance Maladie : www.ameli.fr.

L'EGB résulte d'un sondage sur les bases de données administratives des régimes d'assurance maladie ; les principales caractéristiques administratives et socio-démographiques des bénéficiaires (âge, sexe, date de décès, caisse d'affiliation), qu'ils aient ou non perçu des remboursements de soins, sont extraites de ces bases.

Ces informations sont ensuite associées aux prestations versées par l'Assurance maladie (remboursements de soins, indemnités journalières...) et aux séjours hospitaliers recensés par le PMSI. L'EGB est géré par la CNAMTS au Centre national de traitement informatique (CENTI) d'Evreux.

Il s'agit donc de rapprocher des éléments provenant de bases de données différentes, d'une part pour constituer un échantillon représentatif de la population protégée, consommatrice ou non, d'autre part pour enrichir les informations du SNIIRAM qui ne concernent que les données de remboursement et comportent peu d'information sur la situation administrative des bénéficiaires.

● Intérêts et limites de l'EGB

L'EGB est un échantillon représentatif des bénéficiaires protégés par les régimes d'assurance maladie. Son taux de sondage, proche de 1/100^e, permet de rassembler une population suffisamment importante (près de 500 000 personnes du régime général) pour répondre à la plupart des questions portant sur le comportement sanitaire de la population. Ainsi, il est possible d'étudier le montant du reste à charge des patients, le niveau moyen de leur consommation de soins suivant différents critères (âge, sexe, prise

Encadré 2

Qui peut accéder à l'EGB ?

En vertu de l'article 4 de l'arrêté SNIIRAM du 20 juin 2005 et de l'arrêté du 26 septembre 2006, les médecins-conseils, les agents habilités des caisses des différents régimes de base d'assurance maladie, la CNSA¹⁰ et les URCAM¹¹ ont accès à l'ensemble des données de l'échantillon. En revanche, seuls les médecins-conseils de l'Assurance maladie (et les personnels placés sous leur responsabilité) peuvent croiser les données individuelles dites sensibles. Il s'agit des variables suivantes : le code commune, la date de soins, le mois et l'année de naissance d'un bénéficiaire. En effet, le croisement de plus d'une de ces variables avec d'autres données est considéré par la CNIL comme une levée d'anonymat.

Les données de l'échantillon sous forme individualisée sont également consultables par :

- le ministère chargé de l'assurance maladie et de la santé ainsi que les agences sanitaires placées sous sa tutelle, notamment l'InVS¹², l'Afssaps¹³ et l'INCa¹⁴ ;
- le ministère de l'économie, des finances et de l'industrie ainsi que le ministère de l'agriculture ;
- les ARH¹⁵ ;
- la Haute Autorité de santé ;
- le Haut Conseil pour l'avenir de l'assurance maladie ;
- l'Union nationale des professions de santé ;
- l'Institut des données de santé ainsi que tout autre organisme de recherche habilité par celui-ci ;
- les chercheurs habilités de l'IRDES, du CETAF¹⁶, de l'INSERM¹⁷ et du CNRS¹⁸.

Enfin, « *aucun organisme de recherche, université, école ou autre structure d'enseignement lié à la recherche poursuivant un but lucratif*¹⁹ » ne peut accéder aux informations du SNIIRAM et *a fortiori* à celles de l'échantillon.

en charge pour affection de longue durée (ALD), affiliation à la couverture maladie universelle complémentaire (CMUC), le suivi d'une population atteinte d'une ALD telle que le diabète (13 000 personnes dans l'EGB en 2007), l'insuffisance respiratoire chronique grave (2 500 personnes) ou la maladie d'Alzheimer (2 000 personnes).

Le mode de tirage de l'EGB, basé sur le NIR des individus, facilite largement la réalisation d'études longitudinales ; la mise en place de cet échantillon

remédie ainsi à un des principaux inconvénients de l'Échantillon permanent d'assurés sociaux (EPAS), premier panel de bénéficiaires mis en place par la CNAMTS, que l'EGB est amené à remplacer (encadré 3). En effet, le NIR d'une personne reste identique pendant toute sa vie, un bénéficiaire de l'EGB sera donc identifié de la même façon de sa naissance jusqu'à sa mort. Lorsque cet échantillon sera ouvert à l'ensemble des régimes de sécurité sociale, les personnes qui le composent ne le quitteront que pour cause

¹⁰ CNSA : Caisse nationale de solidarité pour l'autonomie.

¹¹ URCAM : union régionale des caisses d'assurance maladie.

¹² InVS : Institut de veille sanitaire.

¹³ Afssaps : Agence française de sécurité sanitaire des produits de santé.

¹⁴ INCa : Institut national du cancer.

¹⁵ ARH : agences régionales de l'hospitalisation.

¹⁶ CETAF : Centre technique d'appui et de formation des centres d'exams de santé.

¹⁷ INSERM : Institut national de la santé et de la recherche médicale.

¹⁸ CNRS : Centre national de la recherche scientifique.

¹⁹ Arrêté du 20 juin 2005 relatif à la mise en œuvre du SNIIRAM, Article 4.

de décès ou de départ à l'étranger. Il sera donc aisé de suivre le parcours de soins de différentes cohortes d'individus.

Cet échantillon devrait permettre de reconstituer l'ensemble du parcours de soins des patients sur une longue période, car les données de consommation de soins du SNIIRAM, concernant la médecine de ville et l'hospitalisation privée, peuvent être enrichies des données de remboursement des hospitalisations dans des établissements « publics »²⁰, anciennement sous dotation globale, ainsi que des informations hospitalières médicalisées du PMSI (encadré 4). Cette reconstitution sera facilitée par l'évolution du SNIIRAM, qui concerne également l'EGB, puisque, depuis février 2009, le SNIIRAM contient des informations sur les soins externes dispensés dans les hôpitaux « publics ».

En revanche, l'EGB connaît les trois mêmes limites que tous les systèmes d'informations de l'Assurance maladie :

- il ne contient aucune autre donnée socio-démographique que l'âge, le sexe, la date de décès, la caisse d'affiliation et l'affiliation à la CMUC. Le niveau d'études, le niveau de revenus ou la catégorie socio-professionnelle des assurés, dont l'influence sur la consommation de soins a été largement démontrée²¹, sont absents de l'EGB. Ces informations doivent être recueillies au moyen d'enquêtes spécifiques telle que l'enquête Santé et protection sociale (ESPS) réalisée tous les deux ans par l'Irdes²² ;
- par construction, l'échantillon ne contient que des informations

Encadré 3

L'EPAS

En 1979, la CNAMTS mettait en place un Échantillon permanent d'assurés sociaux (EPAS) qui, en l'absence d'un système exhaustif d'informations sur les remboursements de l'Assurance maladie, a permis de réaliser de nombreuses études sur le comportement sanitaire et le niveau des dépenses de santé des ressortissants du régime général. L'EPAS, dont le taux de sondage est plus faible que celui de l'EGB (1/600^e, correspondant à 80 000 personnes), ne sera plus alimenté à partir du 1^{er} mars 2010 car l'EGB est amené à le remplacer. En effet, le mode de tirage de l'EGB, basé sur le NIR du bénéficiaire, permet de pallier aux deux limites majeures de l'EPAS²³ :

- **effet de grappe** : l'EPAS résulte d'un sondage en grappes sur le NIR des ouvriers de droits ; lorsqu'un ouvrier de droits fait partie de l'échantillon, tous ses ayants droit en font également partie (conjoint ou concubin, enfants et dans une moindre mesure, ascendants). Il est impossible de mesurer les effets de grappe de l'EPAS, tel que le lien entre le niveau socio-économique des parents (niveau d'étude, revenu...) et la consommation de soins de leurs enfants, car aucune donnée socio-économique n'existe dans les fichiers de l'Assurance maladie.
- **études longitudinales difficiles** : l'identification des bénéficiaires de l'EPAS à partir du NIR des ouvriers de droits entraîne une forte déperdition d'informations au cours du temps, car une personne qui change de statut au regard de la sécurité sociale quitte généralement l'EPAS (enfant qui commence à travailler et devient son propre ouvrier de droits par exemple). On estime que près de la moitié des bénéficiaires de l'EPAS a quitté l'échantillon au bout d'une douzaine d'années²⁴, ce qui rend difficile la réalisation d'études longitudinales.

concernant les soins effectivement remboursés ; ni l'automédication, ni la consommation de médicaments prescrits mais non remboursés ne peuvent être mesurées. De même, il est impossible d'évaluer le manque de suivi par les patients des prescriptions de leur médecin puisque, par exemple, les ordonnances prescrites mais non délivrées ne sont pas recensées dans l'EGB. Les raisons de ce manque de suivi (raisons financières, négligence de la part du patient...) ne peuvent être connues que par des enquêtes spécifiques ;

- par ailleurs, si ces bases de données comportent des informations de nature médicale, elles ne contiennent aucune information clinique (examen

du malade par le médecin, résultats d'examen biologiques ou d'imagerie médicale) alors qu'elles pourraient éclairer les études épidémiologiques ou les analyses des pratiques.

L'EGB comporte enfin une limite liée à sa nature même d'échantillon : bien qu'il regroupe plus de 530 000 personnes, il ne permet pas de réaliser des études se rapportant à des examens médicaux peu fréquents ou à des maladies rares (34 personnes dans l'EGB prises en charge pour ALD au titre de la mucoviscidose en 2007 par exemple).

● Mode de tirage

Les bénéficiaires de l'EGB sont ceux dont le numéro de clé de contrôle

²⁰ Il s'agit des établissements publics et des hôpitaux privés participants au service public hospitalier.

²¹ Exemples : Raynaud D. Les déterminants individuels des dépenses de santé. Etudes et Résultats n°182, DREES, juillet 2002 ; Païta M, Ricordeau P, de Roquefeuille L, Studer A, Vallier N, Weill A. Les affections de longue durée des bénéficiaires de la CMU complémentaire, Points de repère n°8, CNAMTS, août 2007.

²² Cette enquête réalisée depuis 1988 à partir de l'EPAS, sera menée dès 2010 à partir de l'EGB.

²³ Lenormand F. Le Système d'information de l'assurance maladie, le SNIIRAM et les échantillons de bénéficiaires. Courrier des statistiques n° 113-114, mars-juin 2005.

²⁴ Cf. article de F. Lenormand déjà cité.

du NIR est égal à une valeur donnée, confidentielle. Une étude préalable au tirage²⁵ a permis, grâce à deux tests statistiques du Chi², de s'assurer que :

- la population vivante en France est équidistribuée selon la clé de contrôle du NIR des individus. Le numéro de clé de contrôle du NIR varie de 1 à 97. Les bénéficiaires de l'EGB sont ceux dont la clé de contrôle du NIR est égal à une seule de ces valeurs, le taux de sondage de l'EGB est donc de 1/97^e ;
- la répartition par tranche d'âge de 5 ans et par sexe de la population vivante en France est indépendante de la clé de contrôle du NIR des individus et par conséquent représentative de la répartition nationale. Cette vérification était essentielle compte tenu de l'influence déterminante de l'âge et du sexe de la population sur les dépenses de santé.

Jusqu'à présent, l'EGB ne couvre que le régime général, à l'exception des fonctionnaires et des étudiants²⁶ mais à terme, il devrait concerner 1/97^e des personnes couvertes par la sécurité sociale en France²⁷, quel que soit leur régime d'affiliation.

L'EGB est un échantillon vivant : à la fin de chaque trimestre calendaire²⁸, les informations de l'ensemble des bénéficiaires de l'EGB (déjà présents ou nouveaux entrants) sont extraites des bases de données administratives de l'Assurance maladie. Chaque trimestre, l'échantillon s'enrichit des nouveau-nés, des étrangers venant travailler en France et de leurs ayants droit ; tant qu'il ne couvre que le régime général, il s'enrichit également des personnes nouvellement affiliées à ce régime pour cause de change-

Encadré 4

Appariement des données de l'EGB avec le PMSI

Le Programme de médicalisation des systèmes d'information (PMSI) est issu de la loi du 31 juillet 1991 portant réforme hospitalière et imposant aux établissements de santé, publics et privés, de procéder à l'analyse de leur activité (article L. 710-5) : « Dans le respect du secret médical et des droits du malade, [les établissements de santé] mettent en œuvre des systèmes d'information qui tiennent compte notamment des pathologies et des modes de prise en charge en vue d'améliorer la connaissance et l'évaluation de l'activité et des coûts et de favoriser l'optimisation de l'offre de soins ». Les informations recueillies sont transmises à l'autorité de tutelle de l'établissement de santé.

Actuellement, seules les données du PMSI des séjours hospitaliers des années 2005 à 2007 pour les soins de courte durée (MCO : médecine, chirurgie et obstétrique) dans les établissements de santé, publics ou privés, peuvent être appariées avec celles du SNIIRAM et de l'EGB. Les données du PMSI MCO des années 2007 et 2008 seront accessibles courant 2009.

Le PMSI MCO fournit des informations administratives et médicales pour chaque séjour³⁰.

Les principales informations dites **administratives** concernent l'identification du patient (âge, sexe et lieu de résidence), le séjour (durée, année et mois de sortie, provenance et destination du patient), l'unité médicale fréquentée (réanimation, soins intensifs, surveillance continue, néonatalogie, soins palliatifs, anesthésie ou chirurgie ambulatoire...) et, le cas échéant, le nombre de séances effectuées.

Les principales informations **médicales** concernent les diagnostics (principal, relié, associé) codés selon la CIM-10, les actes médicaux techniques codés selon la CCAM, l'indice de gravité simplifié (IGS II) pour les passages en réanimation, le poids des nouveau-nés et le Groupe homogène de malades (GHM).

Les données d'hospitalisation dans des structures ayant une activité en soins de suite ou de réadaptation (SSR), d'hospitalisation à domicile (HAD) et de psychiatrie devraient pouvoir elles aussi être appariées avec le SNIIRAM et l'EGB courant 2009.

Grâce à un identifiant anonyme, commun au SNIIRAM et au PMSI, le parcours hospitalier d'un patient peut être reconstitué et les informations peuvent être appariées aux données de remboursements de l'EGB. Ainsi, les informations purement financières d'un séjour hospitalier peuvent être enrichies d'informations médicales.

ment dans leur vie professionnelle²⁹ et de leurs ayants droit.

En revanche, les personnes décédées (la date de décès figure dans le référentiel de l'EGB), les étrangers retournant dans leur pays d'origine ou les bénéficiaires quittant le(s) régime(s)

couvert(s) par l'EGB ne seront plus protégés par ces régimes et sont considérés comme sortants ; leurs informations sont néanmoins conservées dans la base de données de l'échantillon pour permettre de réaliser des études longitudinales et de les caractériser.

²⁵ Cette étude a été réalisée à partir d'informations du répertoire national d'identification des personnes physiques, transmises à la CNAMTS par l'INSEE.

²⁶ Les fonctionnaires et les étudiants sont assurés par des sections locales mutualistes (SLM) faisant office de centres de sécurité sociale. Au 31 décembre 2008, ces organismes assurent la protection sociale de plus de 7,3 millions de personnes (Mutuelle générale de l'éducation nationale MGEN, Mutuelle de la fonction publique MFP, La mutuelle des Etudiants LMDE, Mutuelle générale de la police MGP...).

²⁷ Pour une population protégée de l'ordre de 64,3 millions de personnes (estimation de la population résidente en France au 1^{er} janvier 2009, source : INSEE), l'EGB devrait recenser près de 663 000 personnes.

²⁸ Fin mars, fin juin, fin septembre et fin décembre de chaque année.

²⁹ Personnes affiliées antérieurement au Régime social des indépendants et exerçant nouvellement une activité salariée par exemple.

³⁰ Source : site internet de l'Agence technique de l'information sur l'hospitalisation (ATIH) : www.atih.santé.fr.

● Respect de l'anonymat des informations

Afin de garantir la confidentialité des informations, l'alimentation de l'EGB repose sur un double cryptage des identifiants des bénéficiaires, dans les bases administratives des régimes d'une part, dans le SNIRAM d'autre part. Les organismes qui concourent à cette alimentation disposent d'un seul algorithme d'anonymisation et ne peuvent jamais détenir les deux algorithmes simultanément.

Cependant, des études avec recueil d'informations nominatives sont possibles à partir de l'EGB. Un tel travail suppose une autorisation de la CNIL et la mise en place complexe d'un circuit d'information, permettant, en amont des dispositifs d'anonymisation, de gérer à la fois les données de remboursement et le volet nominatif de l'enquête. La mise en place d'un tel circuit, à titre exceptionnel, implique qu'un organisme tiers de confiance extérieur à l'Assurance maladie soit habilité à gérer les informations nominatives relatives à l'enquête ainsi que les identifiants anonymes du SNIRAM. Ce dispositif a été retenu pour l'enquête ESPS de 2010.

● Représentativité

➤ Estimation de la population protégée

La comptabilisation des effectifs de la population protégée par les différents régimes de sécurité sociale est un problème récurrent et complexe. Il faut en effet tenir compte en permanence des déménagements des assurés (changements de CPAM), des changements de régime (affiliation temporaire aux régimes étudiants par exemple), des séjours prolongés à l'étranger, des immatriculations provisoires et de toute autre situation

Encadré 5

Le RNIAM

Le répertoire national inter-régimes de l'assurance maladie (RNIAM), géré par la Caisse nationale d'assurance vieillesse (CNAV), est un référentiel centralisé qui recense les bénéficiaires des différents régimes d'assurance maladie, les régimes de retraite ne sont pas pris en compte. Il est alimenté quotidiennement par chacun de ces régimes et permet de connaître, à une date donnée, le régime et la caisse d'affiliation de chaque personne couverte par la sécurité sociale en France, sans doubles comptes. Dans ce répertoire, les personnes sont identifiées par leur propre NIR.

En effet, dans le cas d'une double affiliation comme celle d'un enfant sur le compte de ses deux parents, le RNIAM ne recense qu'un seul rattachement, soit sur le compte de son père, soit sur celui de sa mère. L'affiliation sur le compte de l'autre parent figure dans les fichiers du régime de sécurité sociale de ce dernier, avec une mention spécifiant qu'elle est différente du rattachement au RNIAM. En outre, dès qu'une personne change de caisse ou de régime de sécurité sociale, cette mutation est immédiatement inscrite au RNIAM.

inhabituelle. L'EGB résulte d'un sondage au 1/97^e sur le NIR des bénéficiaires ; il ne couvre actuellement que le régime général hors SLM. A titre d'exemple, au cours de l'année 2008, le nombre de bénéficiaires de l'EGB vivants au 1^{er} janvier 2008 et présents dans l'échantillon³¹ est de 483 422, ce qui, en multipliant par 97, permet d'estimer la population protégée par le régime général hors fonctionnaires et étudiants (hors SLM) à 46 891 934 personnes.

Le répertoire national inter-régimes de l'assurance maladie (RNIAM - encadré 5) fournit des effectifs de la population protégée par les différents régimes de sécurité sociale, considérés comme données de référence. Cependant, bien que le RNIAM dénombre la population protégée par l'Assurance maladie sans doubles comptes, il fournit des effectifs surévalués car il ne tient pas compte des fins de droits autres qu'un changement de régime au sein de la sécurité sociale française ou un décès. Les Français travaillant à l'étranger ou les étrangers qui ont travaillé en France et prennent leur retraite dans leur

pays d'origine, sont toujours inscrits au RNIAM alors qu'ils ne relèvent plus de l'assurance maladie française.

Au 1^{er} janvier 2009, pour l'ensemble des régimes, le taux de surévaluation des effectifs du RNIAM par rapport à la population résidente en France estimée par l'INSEE³² est de 3,32 %. Si l'on applique ce taux à la population du régime général hors SLM, l'effectif moyen³³ de personnes protégées par ce régime au cours de l'année 2008 est estimé à 47 452 974 personnes. L'écart entre cet effectif estimé et celui calculé à partir de l'EGB est de 1,2 %. Cet écart s'explique par le fait que l'EGB est constitué à partir d'un référentiel centralisé du régime général, le référentiel des individus (RFI), distinct du RNIAM, même s'il lui est relié quotidiennement. Une évolution de la procédure de sélection des bénéficiaires de l'EGB visant à mieux y intégrer toutes les personnes certifiées et affiliées au régime général au sens du RNIAM d'une part, la prise en compte de tous les motifs de fin de droits au régime général dans le RNIAM à partir de 2010 d'autre part, devraient permettre de réduire cet écart.

³¹ Pour le calcul de cet effectif, on a exclu les personnes décédées ou ayant quitté l'EGB avant 2008, les personnes nées avant 1896 (année de naissance du doyen des français) ou après 2008, ainsi que les ressortissants de Mayotte.

³² Source : INSEE, estimations de population au 1^{er} janvier 2009 (résultats provisoires arrêtés fin 2008).

³³ Moyenne de la population au 31 décembre 2007 et au 31 décembre 2008.

Le tableau 1 permet de constater que la répartition en pourcentage par âge et sexe des bénéficiaires de l'EGB protégés au cours de l'année 2008 est, dès à présent, très proche de celle des effectifs moyens du RNIAM au cours de cette même année, puisque l'écart n'est jamais supérieur à 0,4 point par tranche d'âge et 0,3 point par sexe.

➤ Estimation des dépenses de santé

Actuellement, les dépenses de santé retracées dans le SNIIRAM concernent la médecine de ville et l'hospitalisation privée ; les informations relatives aux séjours hospitaliers dans des établissements publics ou participants au service public hospitalier, anciennement sous dotation globale, sont également intégrées à l'EGB. Elles ne couvrent que les séjours remboursés en 2006 et 2007 et aucune information n'est encore disponible pour 2008. Le PMSI public pourrait également être utilisé pour estimer les dépenses de santé dans ces établissements à partir de l'échelle nationale de coût³⁴ (ENC). Mais actuellement, les informations du PMSI accessibles par l'EGB ne couvrent que les courts séjours (médecine, chirurgie et obstétrique - MCO) entre 2005 et 2007.

Pour l'année 2008, le tableau 2 fournit l'estimation par poste de dépense du taux de recours aux soins (nombre de consommateurs du poste de dépense divisé par l'effectif de la population protégée), puis de la dépense moyenne remboursée par personne consommante (personnes ayant perçu au moins un remboursement de soins dans l'année 2008 pour le poste considéré) et par personne protégée (effectif global de 483 422 personnes - Cf. paragraphe précédent).

Des publications à venir traiteront de façon beaucoup plus détaillée du niveau de la consommation de soins

Tableau 1

Répartition en pourcentage de la population protégée de l'EGB et des effectifs moyens de bénéficiaires du régime général (RG) hors SLM certifiés au RNIAM en 2008

	EGB			RNIAM RG hors SLM		
	Hommes	Femmes	Total	Hommes	Femmes	Total
0 - 9 ans	6,9%	6,7%	13,6%	6,9%	6,6%	13,4%
10 - 19 ans	6,2%	5,9%	12,0%	6,2%	5,8%	12,0%
20 - 29 ans	6,1%	6,1%	12,2%	6,3%	6,3%	12,6%
30 - 39 ans	7,5%	7,6%	15,1%	7,5%	7,7%	15,2%
40 - 49 ans	7,0%	7,5%	14,5%	7,1%	7,5%	14,6%
50 - 59 ans	6,0%	6,6%	12,6%	6,0%	6,7%	12,7%
60 - 69 ans	4,2%	4,7%	8,9%	4,2%	4,6%	8,7%
70 - 79 ans	2,8%	3,9%	6,7%	2,8%	3,8%	6,6%
80 ans et plus	1,4%	3,1%	4,4%	1,3%	2,8%	4,0%
Ensemble	48,1%	51,9%	100,0%	48,2%	51,8%	100,0%

Champ : France entière
Source : CNAMTS

Tableau 2

Estimation par poste de dépense du taux de recours aux soins et de la dépense moyenne remboursée par consommant et par personne protégée en 2008

Poste de dépense	Taux de recours	Dépense remboursée moyenne	
		par consommant	par personne protégée
Total soins de ville	92,00%	1 099,6	1 011,6
dont Honoraires privés	89,29%	304,9	272,2
dont Auxiliaires médicaux	30,49%	359,1	109,5
dont Biologie	45,18%	104,2	47,1
dont Pharmacie	85,57%	366,6	313,7
dont LPP (a)	40,42%	158,2	63,9
dont Transports	7,22%	587,9	42,4
dont IJ (maladie et AT) (b)	10,47%	1 518,7	159,0
Total établissements sanitaires privés (c)	10,04%	1 684,5	169,1
dont MCO (court séjour) (d)	9,62%	1 387,9	133,5
dont Psychiatrie	0,17%	4 813,3	8,3
dont SSR (e)	0,44%	5 824,6	25,5

- (a) LPP : liste des produits et prestations (essentiellement de l'appareillage et des prothèses)
 (b) IJ : indemnités journalières pour maladie ou accidents du travail
 (c) hors honoraires des professionnels de santé
 (d) MCO : médecine, chirurgie, obstétrique
 (e) SSR : soins de suite et de réadaptation

Champ : France entière
Source : CNAMTS - EGB

³⁴ A terme l'estimation de ces dépenses reposera sur la facturation individuelle des séjours hospitaliers.

des ressortissants du régime général, évalué à partir de l'EGB. Il convient au préalable de présenter des éléments permettant de juger de la fiabilité des estimations obtenues et, par conséquent, de la représentativité et des limites de cet outil.

Des coefficients d'extrapolation entre les dépenses remboursées regroupées par poste, calculées dans le SNIIRAM et dans l'EGB, ont été estimés pour l'année 2007³⁵. Pour chaque poste, le coefficient d'extrapolation est égal au rapport entre la dépense exhaustive du SNIIRAM et celle échantillonnée de l'EGB. Des intervalles de confiance à 95 % ont également été calculés pour ces coefficients, l'amplitude de ces intervalles permet d'apprécier la fiabilité des estimations obtenues à partir de l'EGB. Ces intervalles de confiance ont été calculés par la méthode du Bootstrap (encadré 6).

Les valeurs des coefficients d'extrapolation et les bornes des intervalles de confiance ont été définies pour chaque poste de dépense, dans l'ensemble de la population puis par tranche d'âge, pour la France métropolitaine d'une part, la France entière d'autre part. L'ensemble de ces informations sera communiqué à tous les utilisateurs de l'EGB et sera actualisé régulièrement.

Dans cette publication, ne seront présentés que les principaux coefficients d'extrapolation entre les dépenses de l'EGB et celles du SNIIRAM concernant l'ensemble des ressortissants du régime général hors SLM pour la France entière.

● Dépense moyenne par consommant

Le tableau 3 fournit les valeurs et les bornes des intervalles de confiance des coefficients d'extrapolation des montants remboursés moyens par

Tableau 3

Coefficients d'extrapolation entre les dépenses moyennes remboursées par consommant calculées sur l'EGB et le SNIIRAM

Poste de dépenses	Coefficients d'extrapolation	Intervalle de confiance à 95 %	
		borne inférieure	borne supérieure
Total soins de ville	1,00	0,99	1,01
dont Honoraires privés	0,98	0,98	0,99
dont Auxiliaires médicaux	0,99	0,97	1,01
dont Biologie	1,00	0,99	1,01
dont Pharmacie	0,99	0,97	1,00
dont LPP	0,99	0,97	1,01
dont Transports	1,02	0,99	1,05
dont IJ (Maladie + AT)	0,99	0,97	1,00
Total établissements sanitaires privés	0,99	0,97	1,02
dont MCO (court séjour)	0,98	0,95	1,01
dont Psychiatrie	1,05	0,94	1,15
dont SSR	1,03	0,99	1,09

Champ : Régime général hors SLM - France entière - 2007

Source : CNAMTS

Encadré 6

Intervalles de confiance par la méthode du Bootstrap

Lorsque la loi de probabilité de l'estimateur d'un paramètre est inconnue dans un échantillon donné, appelé « échantillon initial », la méthode du Bootstrap³⁶ permet de construire des intervalles de confiance pour ce paramètre. Il s'agit d'une méthode de rééchantillonnage qui consiste à construire un grand nombre d'échantillons de la taille de l'échantillon initial, obtenus par tirage aléatoire avec remise des individus de l'échantillon initial (échantillons Bootstrap), et à calculer l'estimateur du paramètre dans chaque échantillon Bootstrap.

Lorsque le nombre d'échantillons Bootstrap est suffisamment élevé, la distribution empirique de l'estimateur peut souvent être approximée par une loi normale, centrée autour de l'estimateur du paramètre dans l'échantillon initial. Une fois que cette normalité approximative a été vérifiée, il est aisé de construire des intervalles de confiance à partir de la loi normale.

Ainsi, par exemple, un intervalle de confiance à 95 % sera égal à la valeur de l'estimateur du paramètre dans l'échantillon initial $\pm 1,96$ écart-types de la distribution empirique du paramètre dans les échantillons Bootstrap.

Une autre méthode de calcul, plus directe, consiste à retenir comme bornes de l'intervalle de confiance à 95 %, les quantiles d'ordre 0,025 et 0,975 de la distribution empirique de l'estimateur du paramètre dans les échantillons Bootstrap. Lorsque la normalité de cette distribution est vérifiée, les résultats de ces deux méthodes de calcul coïncident.

500 échantillons Bootstrap ont été tirés pour construire les intervalles de confiance des coefficients d'extrapolation entre les dépenses remboursées dans l'EGB et celles du SNIIRAM ; les bornes des intervalles de confiance ont été déterminées par la méthode directe de calcul (quantiles d'ordre 0,025 et 0,975).

³⁵ Les coefficients d'extrapolation de 2008 seront disponibles ultérieurement.

³⁶ Saporta G. Probabilités, analyse des données et statistique, 2^e édition révisée et augmentée : Editions TECHNIP ; 2006, p. 308-383.

Tableau 4

**Coefficients d'extrapolation entre les dépenses moyennes remboursées par consommant
calculées sur l'EGB et dans le SNIIRAM par tranche d'âge
et intervalles de confiance à 95 %**

Poste de dépense		Total soins de ville	Honoraires privés	Auxiliaires médicaux	Biologie	Pharmacie	LPP	Transports	IJ (Maladie + AT)
0 - 19 ans	coefficient	1,00	0,97	0,98	1,00	0,98	1,05	1,04	1,07
	borne inférieure	0,97	0,96	0,90	0,98	0,90	0,98	0,93	0,94
	borne supérieure	1,03	0,98	1,02	1,01	1,06	1,12	1,18	1,24
20 - 29 ans	coefficient	0,99	0,97	0,95	0,99	0,91	1,02	1,09	0,97
	borne inférieure	0,95	0,96	0,88	0,98	0,80	0,92	0,93	0,94
	borne supérieure	1,02	0,98	1,03	1,01	1,02	1,16	1,30	1,01
30 - 39 ans	coefficient	1,03	1,00	1,04	1,01	1,03	1,04	0,98	1,01
	borne inférieure	1,00	0,99	0,99	0,99	0,97	0,96	0,87	0,97
	borne supérieure	1,05	1,01	1,10	1,03	1,09	1,14	1,09	1,04
40 - 49 ans	coefficient	1,01	0,99	0,98	0,99	1,02	0,93	0,98	1,01
	borne inférieure	1,00	0,97	0,94	0,98	0,98	0,87	0,91	0,98
	borne supérieure	1,03	1,00	1,04	1,01	1,06	1,01	1,08	1,05
50 - 59 ans	coefficient	1,00	1,01	0,99	1,01	0,99	0,96	0,97	0,98
	borne inférieure	0,98	0,99	0,94	0,99	0,96	0,91	0,90	0,95
	borne supérieure	1,02	1,02	1,03	1,02	1,01	1,01	1,05	1,01
60 - 69 ans	coefficient	1,01	0,99	0,98	0,99	1,00	1,00	1,11	1,01
	borne inférieure	0,99	0,98	0,92	0,97	0,98	0,95	1,03	0,92
	borne supérieure	1,02	1,00	1,04	1,01	1,02	1,05	1,19	1,11
70 - 79 ans	coefficient	1,01	0,99	0,99	1,01	1,01	1,01	1,02	nc
	borne inférieure	0,99	0,98	0,94	0,99	0,99	0,96	0,94	nc
	borne supérieure	1,03	1,01	1,04	1,03	1,04	1,06	1,10	nc
80 ans et plus	coefficient	1,02	1,00	1,01	1,00	1,01	0,97	1,03	nc
	borne inférieure	1,00	0,98	0,98	0,98	0,99	0,94	0,95	nc
	borne supérieure	1,04	1,02	1,05	1,03	1,02	1,01	1,11	nc
Ensemble de la population	coefficient	1,00	0,98	0,99	1,00	0,99	0,99	1,02	0,99
	borne inférieure	0,99	0,98	0,97	0,99	0,97	0,97	0,99	0,97
	borne supérieure	1,01	0,99	1,01	1,01	1,00	1,01	1,05	1,00

Champ : Régime général hors SLM - France entière - 2007

Source : CNAMTS

consommant, pour les principaux postes de dépenses de santé en 2007. Pour chaque poste, le coefficient d'extrapolation est égal à la division de la dépense moyenne du SNIIRAM par celle de l'EGB.

Tous les coefficients d'extrapolation présentés dans le tableau 3 sont proches de 1, ce qui prouve que les dépenses moyennes par consommant calculées dans l'EGB sont très proches de celles du SNIIRAM et valide le niveau des dépenses moyennes présentées dans le tableau 2. L'EGB est donc un très bon outil pour étudier les dépenses moyennes de santé et leur évolution au cours du temps.

L'amplitude de l'intervalle de confiance est faible pour les soins de ville. Elle est plus élevée pour les soins exécutés dans des établissements sanitaires privés, en particulier en psychiatrie et pour les soins de suite et de réadaptation (SSR) où le nombre de consommateurs de l'EGB est faible (respectivement 721 et 2 005 personnes).

Quel que soit le poste, l'étude de l'évolution des dépenses moyennes de soins dispensés dans des établissements sanitaires privés doit toujours être réalisée avec précaution. En particulier, l'analyse de ces dépenses par âge n'est pas pertinente pour les tranches d'âge dont les effectifs de consommateurs de l'EGB sont trop faibles.

Le tableau 4 fournit les valeurs des coefficients d'extrapolation et les intervalles de confiance des dépenses moyennes de soins de ville par consommant et par tranche d'âge.

Ces valeurs sont calculées lorsque le nombre de consommateurs de l'EGB est supérieur ou égal à 100. A l'exception de l'ensemble des soins de ville, des honoraires privés et des actes de biologie, l'amplitude des intervalles de confiance est généralement beaucoup plus importante par tranche d'âge

Tableau 5

Coefficients d'extrapolation entre le montant global des dépenses remboursées dans l'EGB et dans le SNIIRAM

Poste de dépenses	Coefficients d'extrapolation	Intervalle de confiance à 95 %	
		borne inférieure	borne supérieure
Total soins de ville	104,84	104,11	105,60
dont Honoraires privés	104,96	104,48	105,46
dont Auxiliaires médicaux	104,03	101,64	106,31
dont Biologie	105,64	104,84	106,51
dont Pharmacie	105,21	103,86	106,55
dont LPP	103,83	101,48	105,88
dont Transports	107,32	103,91	110,80
dont IJ (Maladie + AT)	103,47	101,76	105,44
Total établissements sanitaires privés	105,18	102,53	107,76
dont MCO (court séjour)	103,66	100,79	106,50
dont Psychiatrie	112,42	99,39	128,28
dont SSR	110,02	102,99	117,51

Champ : Régime général hors SLM - France entière - 2007
Source : CNAMTS

que pour l'ensemble de la population. En outre, la valeur du coefficient d'extrapolation peut s'éloigner sensiblement de 1 dans certaines tranches d'âge (1,11 pour les transports chez les personnes âgées de 60 à 69 ans par exemple).

Par conséquent, avant de faire une étude sur une tranche d'âge donnée, il convient d'examiner l'amplitude de l'intervalle de confiance pour juger de la pertinence des résultats obtenus et de la représentativité de l'EGB dans ce domaine.

● Dépense globale

Le tableau 5 fournit la valeur et les bornes de l'intervalle de confiance à 95 % des coefficients d'extrapolation des montants remboursés globaux par grands postes en 2007. Pour

chaque poste, le coefficient d'extrapolation est égal au rapport entre la dépense remboursée totale du SNIIRAM et celle de l'EGB.

En raison du taux de sondage de l'EGB, les valeurs des coefficients d'extrapolation des dépenses globales devraient être proches de 97. Elles sont supérieures à 97 en raison d'un chaînage incomplet entre les deux référentiels sources de l'EGB : le RFI, qui recense tous les bénéficiaires du régime général, consommateurs ou non, et le SNIIRAM qui contient les données de consommation de soins. Les informations provenant de ces deux fichiers sources étant anonymes sans possibilité de retrouver les identifiants réels des individus, les causes de ce chaînage incomplet ne sont pas identifiées à ce jour. En effet, lors de la recherche des données de consommation, aucune dépense de

santé n'est retrouvée dans le SNIIRAM pour 7 % des bénéficiaires de l'EGB extraits du RFI. La résolution de ces problèmes de chaînage permettra, à terme, de réduire les coefficients d'extrapolation qui devraient se rapprocher de 97.

Par ailleurs, l'évolution du SNIIRAM devrait également permettre d'améliorer les coefficients d'extrapolation des dépenses globales :

- en premier lieu, depuis septembre 2008, il est possible d'isoler les

dépenses non-individualisées remboursées à certains établissements, comme les maisons de retraite médicalisées, et de les exclure des dépenses du SNIIRAM prises en compte dans le calcul des coefficients d'extrapolation. Ceci devrait permettre de réduire la valeur de ces coefficients et l'amplitude de certains intervalles de confiance comme ceux des auxiliaires médicaux et des transports ;

- en second lieu, les dépenses de soins externes dispensés dans les

établissements anciennement sous dotation globale, vont être progressivement intégrées au SNIIRAM. Ceci devrait permettre, d'une part de mieux recenser la population consommatrice dans le SNIIRAM et donc dans l'EGB, d'autre part de mieux cerner le parcours de soins complet du patient.

Pour ces deux raisons, les coefficients d'extrapolation figurant dans ce document seront actualisés tous les ans.

Nous tenons à remercier madame Nathalie Delorme (département maîtrise d'ouvrage informatique stratégie et études) ainsi que l'ensemble des informaticiens de la direction de l'informatique et de l'aide au pilotage de la Caisse nationale d'assurance maladie.



Collection des points de repère

N°1 : Disparités géographiques de la santé en France : les affections de longue durée

N°2 : La progression des génériques au premier trimestre 2006

N°3 : Coût des trente affections de longue durée pour l'Assurance maladie

N°4 : L'utilisation des systèmes d'information de l'assurance maladie pour l'analyse des pratiques de prescription : quelques exemples

N°5 : Déterminants de l'évolution des indemnités journalières maladie

N°6 : La consommation d'antibiotiques : situation en France au regard des autres pays européens

N°7 : Prise en charge de la mucoviscidose en affection de longue durée

N°8 : Les affections de longue durée des bénéficiaires de la CMU Complémentaire

N°9 : Les bénéficiaires d'affection de longue durée au 31 décembre 2006

N°10 : Hypertension artérielle et facteurs de risque associés : évolutions des traitements entre 2000 et 2006

N°11 : Les disparités géographiques de consommation d'indemnités journalières en maladie

N°12 : Comparaisons européennes sur huit classes de médicaments

Collection des points de repère (suite)

N°13 : Démographie et honoraires des médecins libéraux en 2005

N°14 : Un an de codage CCAM en secteur libéral (premiers résultats sur une année complète : 2006)

N°15 : Les dispositifs médicaux : situation et évolution en 2007

N°16 : Les causes médicales de l'invalidité en 2006

N°17 : Les réformes dans l'organisation des soins primaires en Angleterre

N°18 : Les maladies chroniques psychiatriques et neurologiques des enfants et adolescents en France : prévalence des affections de longue durée en 2004

N°19 : Les données 2007 du codage de la biologie médicale

N°20 : Les personnes en affection de longue durée au 31 décembre 2007

N°21 : Évolution des actes techniques en secteur libéral en 2007

N°22 : Mieux connaître les dépenses de médicaments : une comparaison méthodique de deux sources de données GERS et CNAMTS

N°23 : Démographie et honoraires des médecins libéraux en 2006

N°24 : Remboursement de médicaments antiasthmatiques : une approche de la prévalence et du contrôle de l'asthme

Contact : laurence.de-roquefeuille@cnamts.fr