

## **Projet d'appariement de l'Échantillon Démographique Permanent au Système National de Données de Santé**

### **1. Intitulé du projet**

Appariement EDP-SNDS et production de l'EDP-Santé

### **2. Services producteurs**

La Bureau État de Santé de la population de la DREES et la division Enquêtes et Etudes Démographiques (EED) de l'INSEE pilotent, en collaboration avec la CNAMTS, un projet d'appariement des données de l'Échantillon Démographique Permanent (EDP) produit par l'INSEE et de celles du Système National des Données de Santé (SNDS) produit par la CNAMTS. La DREES, à l'initiative du projet, facilite les échanges entre la CNAMTS et l'INSEE, deux institutions avec qui ses partenariats sont fréquents en tant que service statistique ministériel rattaché au Ministère des Solidarités et de la Santé. La DREES se chargera également de mettre en place les conditions juridiques permettant cet appariement, dans le cadre de la Loi Informatique et Libertés et de l'article 193 de la loi de modernisation de notre système de santé.

### **3. Services réalisant l'opération**

Ce projet ne nécessite pas de collecte de nouvelles données, dans la mesure où il repose sur l'appariement de deux bases de données déjà existantes, constituées elles-mêmes par le chaînage de différents fichiers administratifs et d'enquête. L'INSEE et la CNAMTS seront en charge de réaliser les opérations d'appariement avec leurs bases respectives, à savoir l'EDP et le SNDS. En effet, le projet doit se concrétiser sous deux formes :

- un EDP-Santé agrégé disponible au Centre d'Accès Sécurisé à Distance (CASD) dans lequel seront appariées les données de l'EDP à des indicateurs de santé individuels agrégés par types de soins issus du SNDS,
- un EDP-Santé détaillé disponible sur le portail SNDS, dans lequel seront appariées les données de l'EDP aux données brutes du SNDS.

L'intérêt de ce double produit est de faciliter l'accès à des utilisateurs non familiers du SNDS ; nous revenons plus en détail sur cette question dans la partie 7.

Le circuit détaillant les étapes de la production de ces deux appariements est disponible (Annexe 4), et a été validé par les institutions concernées. Nous détaillons ici les rôles principaux joués par l'INSEE, la CNAMTS et le CASD.

#### **A. Le rôle de l'INSEE**

L'INSEE est le producteur de l'EDP, un panel d'individus sélectionnés en fonction de leur jour de naissance débuté en 1968 (4 jours par an jusqu'en 2004, puis 16 jours par an, soit 4,4 % de la population). Il collecte des informations puisées dans diverses sources, enrichies au fil des années.

L'EDP est donc déjà le produit de l'appariement de données en provenance des cinq sources de données suivantes<sup>1</sup> :

- les bulletins d'**État civil**
- les données issues des cinq **recensements** exhaustifs de 1968 à 1999 et des enquêtes annuelles de recensement (EAR) depuis 2004
- le **fichier électoral** donnant les inscriptions électorales actuelles et passées depuis 1945
- les informations issues du **panel d'actifs « tous salariés »** depuis 1967
- les **données socio-fiscales** issues de la base Fideli (Fichier Démographique des Logements et des Individus) et de la base FiLoSoFi (Fichier Localisé Social et Fiscal) depuis 2011.

C'est l'équipe actuellement en charge de la production annuelle de l'EDP qui se chargera de produire la table des identifiants correspondant au champ EDP (Numéros d'Identification au Répertoire de l'INSEE), auxquels sera associé un identifiant anonyme pour l'appariement. L'INSEE pourrait passer l'algorithme de hachage FOIN<sup>2</sup>, ou envoyer ces informations à un tiers de confiance qui sera chargé de cette opération.

Dans un second temps, cette même équipe intégrera les données agrégées issues du SNDS à l'EDP pour constituer l'EDP-Santé agrégé.

L'INSEE aura accès à l'EDP-Santé agrégé.

## B. Le rôle de la CNAMTS

La CNAMTS est gestionnaire du SNDS, instauré par l'article 193 de la loi de modernisation de notre système de santé. Celui-ci regroupera à terme :

- les données de **l'Assurance Maladie** (consommations de soins en ville et en établissement remontées dans le SNIIRAM<sup>3</sup>),
- les données **hospitalières** du PMSI<sup>4</sup>,
- les données sur les **causes médicales de décès** du CépiDC-Inserm
- les données relatives au **handicap** (remontées des Maisons départementales des personnes handicapées vers la CNSA<sup>5</sup>)
- un échantillon de données des **organismes complémentaires**.

Les deux premières catégories de données sont déjà disponibles et constituent la première version du SNDS. Les causes médicales de décès devraient alimenter le SNDS au deuxième semestre 2017. Il est prévu que le SNDS contienne vingt ans d'historique de données (et dix années supplémentaires archivées), mais actuellement seules dix années sont disponibles.

---

<sup>1</sup> Le document de travail de Stéphane Jugnot (INSEE Document de Travail N°F1406) donne plus de précisions sur les informations disponibles dans l'EDP. L'appariement avec les données fiscales et le panel « Tous salariés » est cependant postérieur à la rédaction de ce document. Des compléments rapides sur ces sources récentes sont disponibles à partir de ce lien : <https://www.insee.fr/fr/metadonnees/source/s1166>

<sup>2</sup> L'Insee expertise actuellement la question de passer l'algorithme de hachage pour envoyer les NIR à la CNAMTS sans recourir au tiers de confiance.

<sup>3</sup> Système national d'information inter-régimes de l'Assurance-maladie

<sup>4</sup> Programme de médicalisation des systèmes d'information

<sup>5</sup> Caisse nationale de solidarité pour l'autonomie

Pour le projet EDP-Santé agrégé, le département Études et Statistiques de la CNAMTS sera en charge de recevoir la table des identifiants cryptés du champ EDP, de les crypter une nouvelle fois, et d'extraire les données du SNDS correspondantes. La version envoyée à l'INSEE sera composée d'indicateurs agrégés produits par la CNAMTS ainsi que de l'identifiant anonyme. La production de ces indicateurs agrégés par types de soins est une opération sur laquelle l'expertise de la CNAMTS sera croissante étant donné la multiplication des appariements d'enquêtes au SNDS.<sup>6</sup> La DREES, également familière de ces différents appariements, pourra constituer un groupe de réflexion sur le contenu final de ces indicateurs.

Pour le projet EDP-Santé détaillé, la CNAMTS est chargée de réceptionner les données de l'EDP et de les apparier au SNDS afin de constituer le produit EDP-Santé détaillé disponible sur le portail SNDS. Cet appariement est réalisé sur l'identifiant anonyme grâce à un tiers de confiance qui pourrait être joué par le CASD ou l'Insee.

### C. Le rôle du CASD<sup>7</sup>

Dans le cas où l'Insee ne passe pas l'algorithme de hachage, le CASD joue le rôle de tiers de confiance et reçoit de la part de l'INSEE la table des identifiants du champ EDP (NIR) et l'identifiant anonyme correspondant. Il se charge de les crypter via le premier algorithme de hachage qui permet d'obtenir l'identifiant dans le SNDS, dénommé FOIN 1 et envoie ensuite cette table à la CNAMTS.

Il constitue également le dernier maillon du circuit de production et diffusion de l'appariement et reçoit de la part de l'INSEE le produit EDP-Santé agrégé, qu'il met à disposition des chercheurs.

## 4. Historique du projet et bilan d'exécution de projets similaires

Ce projet est nouveau dans le format proposé. Il s'inscrit néanmoins dans la continuité d'une réflexion collective qui a émergé depuis plusieurs années face au constat de l'absence de données fiables pour étudier certaines inégalités sociales dans le domaine de la santé.

Au début des années 2000, le projet de la cohorte COSMOP<sup>8</sup> piloté par l'Institut de Veille sanitaire (InVS) s'inscrivait, à plus petite échelle, dans une démarche similaire, et reposait sur l'appariement de l'échantillon démographique permanent aux données sur les causes médicales de décès du CépiDC (Inserm). Ce programme de surveillance de la mortalité avait pour objectif de décrire de façon régulière la fréquence des différentes causes de décès par profession.

## 5. Bilan d'exécution de projets similaires

Cette opération du programme COSMOP n'a finalement été produite qu'une fois (avec l'EDP version 2002) mais a déjà permis de produire des études sur les causes de mortalité par secteur d'activité et profession<sup>9</sup>.

Un rapport très détaillé sur la faisabilité et la mise en place du programme de surveillance COSMOP soulève les différentes limites liées à ce projet : taille d'échantillon encore trop restreinte, profession

---

<sup>6</sup> À titre d'exemple, l'enquête européenne de santé EHIS 2014 a été appariée et sera disponible sous Quételet avec des indicateurs agrégés issus du SNDS. Très récemment, l'enquête CARE sur les seniors a également été appariée au SNDS, ce qui a nécessité un important travail de la CNAMTS dans la constitution d'indicateurs.

<sup>7</sup> À l'heure où nous rédigeons cette note, l'équipe du projet n'a pas rencontré le CASD pour confirmer son implication dans le circuit et obtenir certaines informations, cette rencontre étant prévue en septembre.

<sup>8</sup> Cohorte pour la surveillance de la mortalité par profession.

<sup>9</sup> Geoffreoy-Perez, B., *Analyse de la mortalité et des causes de décès par secteur d'activité de 1968 à 1999 à partir de l'échantillon démographique permanent*, Institut de Veille Sanitaire, Rapport Septembre 2006.

pas toujours bien relevée via le recensement. Celles-ci seront néanmoins dépassées dans le cadre de l'appariement EDP-SNDS, notamment grâce au quadruplement du champ EDP depuis 2004, ainsi qu'à la plus grande profondeur de données disponibles, le projet étant réalisé treize ans plus tard. Enfin, l'intégration du panel DADS à l'EDP permettra également de pallier le manque de finesse dans la description des parcours professionnels.

Le rapport souligne également la nécessité de renouveler ces opérations afin d'étudier l'évolution dans le temps de la relation entre les secteurs d'activité professionnelle et les causes de mortalité, et de consolider les résultats établis. Le projet de l'EDP-Santé fournirait donc l'opportunité de prolonger ce projet.

## **6. Objectifs généraux du projet et principaux thèmes que l'appariement permettra d'aborder**

L'objectif poursuivi par l'appariement de l'EDP et du SNDS est d'obtenir une source inédite permettant, à des fins statistiques et de recherche, de croiser des données socio-économiques et des données médico-administratives sur la santé et le recours aux soins. Cette base comblerait un vide statistique pour l'étude des inégalités sociales et territoriales de santé, de recours mais aussi de prise en charge des soins. L'EDP-Santé offrirait en effet un échantillon suffisamment large pour des études sur des catégories sociales ou des territoires particuliers, ainsi que des données fiables et précises sur la consommation de soins car non altérées par des biais de mémoire, de déclaration ou de non-réponse, et ce, pour un coût moindre par rapport à celui de la collecte d'une enquête. Les données rassemblées contiendront alors des informations sur l'état de santé, le recours aux soins et le montant des dépenses en soins, ainsi que sur les causes de décès des individus, à mettre en regard d'informations sur la situation démographique, sociale et économique, ainsi que les trajectoires spatiales et professionnelles de chacune de ces personnes. En cela, cet appariement pourrait alors ouvrir la voie à de nombreux sujets d'études dans différents champs disciplinaires tels que l'épidémiologie sociale, la sociologie, l'économie de la santé, ou la démographie. En outre, si ces données peuvent contribuer à améliorer la connaissance par le développement des recherches et études, elles pourront également constituer de précieux outils pour éclairer les politiques publiques en aidant à la prise de décision via la construction et le suivi dans le temps d'indicateurs ou l'évaluation de mesures envisagées (ex ante) ou déjà mises en places (ex post).

Pour mesurer l'étendue des sujets qu'il sera possible d'aborder grâce à l'EDP-Santé, il nous faut rappeler l'ensemble des thèmes couverts par chacun des deux produits initiaux, EDP et SNDS. C'est au croisement de ces dimensions qu'émerge alors une quantité d'études à réaliser dans le champ de la santé et du social. En tant que producteurs de ces données, et pour nous assurer de leur utilisation, nous avons consulté des chercheurs sur le projet (Annexe 2) et obtenu de nombreuses réponses collectives ou individuelles qui sont reportées en annexe 3. Sans viser l'exhaustivité, on peut les regrouper selon les thématiques suivantes :

### **A. Déterminants sociaux et territoriaux de la santé et du recours aux soins**

Il s'agit de la thématique principale que permettra d'éclairer cet appariement. En effet, le SNDS contient d'une part des informations sur l'état de santé des individus – déduites des affections de longue durée (ALD), des consommations de soins, des diagnostics hospitaliers, ou encore des causes de décès – et des informations sur le recours et les parcours de soins d'autre part, décrivant le détail du passage par le système de soins en France, de l'acte de prévention à la boîte de médicaments délivrée en pharmacie. S'il est actuellement possible de dresser le constat de ces différents

phénomènes, d'en étudier le niveau (prévalence de pathologies<sup>10</sup>, classement des causes de mortalité, recours à tel professionnel de santé), ou d'en dessiner une évolution, l'appariement avec l'EDP permettra de différencier ces derniers selon des critères sociaux ou territoriaux pour interroger la question des inégalités sociales et territoriales de santé, mais également d'étudier les caractéristiques d'une sous-population donnée. La richesse des informations contenues dans l'EDP rendra possible ces études selon différents aspects :

i. Situation familiale et santé

L'EDP permet de suivre avec précision les événements modifiant l'état-civil tels que les naissances ou les mariages, mais également d'identifier les séparations et les évolutions de la composition du ménage grâce aux informations fiscales. Ces informations croisées avec l'état de santé permettront d'instruire des sujets tels que l'effet des ruptures conjugales sur la santé des adultes et des enfants (cf. annexe X).

ii. Situation socio-économique et santé

Le recensement, les données sociales du panel des salariés (DADS) ainsi que les données fiscales permettent une approche précise de la situation sociale des individus, avec des informations sur le diplôme, la profession, et donc plus largement la catégorie sociale, le niveau de vie ou encore le pays d'origine. Beaucoup de travaux sont attendus sur les inégalités de santé selon ces caractéristiques. Ainsi, en matière de prévention, il sera possible de cibler les populations les plus exposées au risque de certaines maladies, mais aussi celles qui recourent le moins aux actes préventifs, de dépistage ou de vaccination par exemple. Concernant le recours aux soins, ces données permettront aussi d'identifier des spécificités dans le choix des parcours de soins, dont les conséquences sont significatives aussi bien en termes d'organisation du système de soins que d'état de santé. Une étude est par exemple prévue sur les profils sociaux des patients passés par les urgences.

iii. Travail et santé

La finesse des données issues du panel des salariés sur les carrières professionnelles offrira la possibilité d'approfondir les liens entre la sphère professionnelle et celle de la santé. Ainsi, dans la continuité du programme COSMOP, il sera possible d'étudier les causes de mortalité par secteur d'activité. La santé mentale, approchée sous sa forme médicalisée, pourra être interrogée en lien avec la trajectoire professionnelle, grâce au caractère longitudinal des données, de même que la prise en charge de ces situations.

iv. Trajectoires spatiales et santé

Enfin, l'EDP contient également des informations suivies sur les différents lieux d'habitation, offertes par le croisement des données d'État civil, des recensements et des données fiscales. Cela permettrait de reconstituer des trajectoires de mobilité spatiale des individus et d'étudier d'éventuels liens avec l'état de santé et la survenue de certaines pathologies spécifiques, mais également l'accès à des soins ou des professionnels de santé donnés. La couverture nationale de la base couplée avec sa grande taille offrira d'ailleurs la possibilité de suivre l'état de santé dans certaines régions ou départements, ou selon certaines trajectoires. Ainsi, une étude pourrait être

---

<sup>10</sup> Le SNDS ne contient pas de résultats d'examen, hormis les diagnostics à l'entrée et la sortie d'une hospitalisation. Néanmoins, la CNAMTS propose des indicateurs identifiants un certain nombre de pathologies chez les individus (les « tops pathologies »). Ils sont construits sur des informations disponibles à la fois dans ces diagnostics hospitaliers, mais également dans les motifs d'ALD ou dans la consommation de soins « marqueurs » (la consommation d'un traitement spécifique au traitement du diabète par exemple).

conduite sur l'état de santé et le recours aux soins des habitants ultra-marins en tenant compte des migrations en provenance ou à destination de la métropole et leurs liens avec les parcours de soins.

## B. Différentiels sociaux de mortalité

L'EDP-Santé permet, comme indiqué plus haut, des études sur les causes de décès en lien avec les parcours professionnels, et également sur les différentiels sociaux de mortalité par cause de décès, et différence de causes de mortalité par génération de naissance, sujet que souhaite par exemple étudier l'Insee, suite à la mise en évidence d'une stagnation de la mortalité par âge pour les générations 1941-1953 pour les hommes et 1941-1946 pour les femmes, alors que les générations précédentes et suivantes ont connu une baisse de mortalité d'une génération à l'autre<sup>11</sup>.

## C. Système de protection sociale et d'assurance maladie

### i. Assurances maladie et dépenses de santé

Les données fines du SNDS offrent une grande qualité d'informations sur les dépenses de soins, distinguant la partie prise en charge par l'assurance maladie de celle restant à charge de l'assuré. À l'heure actuelle, ces données sont disponibles pour la couverture de l'assurance maladie obligatoire (AMO). Elles ont vocation à être enrichies par des données de remboursement des assurances maladie complémentaires<sup>12</sup>. L'appariement permettra ainsi d'étudier quelles populations restent effectivement exposées à des restes à charge après AMO ou AMC élevés, malgré les dispositifs dédiés (ALD, CMU-C, etc.), et pour quels types de soins. L'apport de l'EDP sera de pouvoir étudier le poids des dépenses de santé, par poste de soins, selon des variables sociodémographiques (niveau de vie, PCS,...). Ces éléments font régulièrement l'objet de demandes internationales (OCDE, Eurostat).

Dans une approche dynamique, les données appariées amélioreraient la capacité d'évaluation de politiques publiques, et permettraient d'évaluer par exemple l'effet de la prise en charge des dépenses de santé sur les parcours de soins selon les caractéristiques de la population.

### ii. Arrêts de travail et indemnités journalières

Les mécanismes des arrêts de travail et le versement des Indemnités Journalières (IJ) sont des phénomènes complexes à étudier : ils touchent à la fois aux champs de la santé, du travail et de l'entreprise. Grâce aux données de carrière issues des DADS et aux données de la CNAMTS sur le versement des indemnités journalières ainsi que les consommations de soins du bénéficiaire, l'EDP-Santé permettra ces études sur les facteurs déterminants et les mécanismes individuels de recours aux arrêts de travail en intégrant des caractéristiques de l'emploi actuel ainsi que des antécédents individuels tels que le chômage ou de précédents arrêts de travail.

## D. Handicap et dépendance

---

<sup>11</sup> Blanpain et Buisson, Insee, 2016, document de travail n°F1606

<sup>12</sup> La loi prévoit l'intégration au SNDS d'un échantillon de données des organismes complémentaires. Ce projet d'appariement des données des Assurances maladies (ADAM) est également piloté par la DREES. Cet échantillon pourrait contenir des individus communs avec l'EDP (car sélectionnés en partie sur les mêmes dates de naissance) pour permettre des études sur les restes à charge après remboursement de l'assurance maladie complémentaire.

Certains aspects de la dépendance et à moyen terme du handicap seront également couverts par l'EDP-Santé.

Concernant la dépendance, le SNDS informe notamment sur le passage ou la présence d'une personne en établissement d'hébergement pour personnes âgées dépendantes (EHPAD). L'appariement apportera ainsi des informations socio-économiques sur une population hors du champ de la plupart des enquêtes en ménages ordinaires (personnes vivant en institution). Il permettra également de faire des études sur l'entrée en institution et ses liens avec les configurations familiales ou les situations sociales.

La loi prévoit également l'intégration de données sur le handicap, *via* les Maisons départementales des personnes handicapées, projet qui devrait être mis en œuvre à l'horizon 2020. La prise en charge des dispositifs et matériels médicaux, et les restes à charge associés sont une préoccupation forte vis-à-vis des inégalités liées à la situation sociale souvent défavorable de la population handicapée. Les données du SNDS permettront d'étudier les dépenses dans ces dispositifs au regard des revenus des individus. En outre, l'Irdes mène en parallèle un projet d'étude de la Faisabilité de l'Identification des Situations de Handicap (FISH) visant à évaluer la possibilité de repérer les personnes à risques de handicap à partir des données détaillées du SNDS. L'issue de ce projet pourrait permettre à terme, de bénéficier dans le SNDS de l'identification de sous-populations souffrant de certaines formes de handicap et d'ouvrir un large champ d'étude à la fois sur leur santé et recours aux soins, mais aussi sur leur situation socio-économique, grâce à l'EDP-Santé.

En définitive, la mise en place de ce projet permettra d'explorer un vaste champ de problématiques par la simple description d'inégalités, par la mise au jour des mécanismes qui y ont conduit, ou encore par des analyses sur longue période permettant de mettre en évidence des causalités entre événements sociaux et événements de santé.

## **7. Origine de la demande et utilisateurs potentiels des résultats**

Le projet EDP-Santé est une demande conjointe du ministère en charge de la santé, et de l'INSEE, puisque le projet a été inscrit à son agenda 2025. Par ailleurs, ce projet a été évoqué au cours de différentes occasions (Séminaire de la Drees sur les « Inégalités sociales de santé », 2015) et rapports auprès de Commissions (rapport auprès du Conseil Économique, Social et Environnemental sur les espérances de vie en bonne santé en février 2017, par exemple). La diversité des institutions à l'origine de cette demande illustre le vaste champ des utilisations potentielles, que ce soit dans le domaine de la recherche pure, ou dans celui de l'action publique, les deux pouvant bien évidemment être très liés.

Ainsi, en matière d'utilisateurs potentiels, le monde de la recherche et des études est le premier concerné. Qu'il s'agisse de laboratoires de recherche en sciences sociales, en épidémiologie ou d'agences de santé publique, les attentes vis-à-vis de ce projet sont nombreuses, notamment pour la réalisation d'études approfondies sur les inégalités sociales et spatiales de santé. En témoignent les retours lors de la diffusion d'une note de consultation sur les besoins vis-à-vis du projet (Annexes 3 et 4).

Objet de fortes attentes, l'EDP-Santé verra son accès limité et bien encadré :

- D'une part pour des raisons juridiques : ces données comporteront des informations sensibles, notamment sur l'état de santé des personnes, et un risque de réidentification très

fort de ces individus étant donné la taille de l'échantillon et le niveau de détail des informations contenues. L'accès à chacune des deux bases de données (EDP et SNDS) est d'ailleurs soumis à des protocoles différents, l'un relevant de l'accès à des données de statistique publique, l'autre pour celui de l'accès aux données de santé (voir 20.), permettant de sécuriser leur utilisation.

- D'autre part, ces deux bases sont complexes. Il est donc nécessaire de prévoir un accompagnement des utilisateurs afin de les aider à s'approprier les données. Il s'agit cela dit d'une problématique déjà prise en compte par l'INSEE et par la CNAMTS respectivement pour leurs deux produits et ces deux institutions mettent progressivement en place des dispositifs d'accompagnement des utilisateurs de ces données. Ainsi, l'INSEE réunit une à deux fois par an le groupe des utilisateurs de l'EDP afin d'échanger sur les évolutions de la base et sur les sujets d'étude en cours. La CNAMTS, quant à elle, développe des produits orientés vers l'utilisation en matière d'études,<sup>13</sup> et organise des formations pour initier les utilisateurs aux données. L'Institut National des Données de Santé, créé en même temps que le SNDS, a pour rôle de « mettre à disposition les données auprès de tous les acteurs, privés comme publics, citoyens, usagers du système de santé, professionnels de santé, établissements de santé, chercheurs, entreprises »<sup>14</sup>. Enfin, c'est pour faciliter l'exploitation des données de santé qu'il a été proposé que l'EDP-Santé disponible au CASD contienne des indicateurs agrégés, simplifiant l'information brute disponible dans le SNDS.

Il existe un vivier de chercheurs qui utilisent actuellement l'un ou l'autre des produits et constituent déjà un effectif important, et dont beaucoup sont demandeurs de ces données complémentaires. Par exemple, l'agence Santé Publique France et le GIP-IRDES comptent déjà de nombreux utilisateurs du SNDS et ont partagé leur intérêt pour un enrichissement de celui-ci par des données sociodémographiques et économiques. À l'inverse, à l'INED, de nombreux chercheurs utilisent actuellement l'EDP sur les thématiques démographiques et aimeraient compléter ces travaux par des indicateurs agrégés sur la santé.

#### **8. Place dans le dispositif statistique déjà existant sur le même sujet ; apport à la connaissance du domaine par rapport à d'autres sources déjà disponibles (enquêtes ou fichiers administratifs)**

Actuellement, certains dispositifs proposent une réponse partielle à la question de l'étude des inégalités sociales de santé. Ainsi, les données médico-administratives seules contiennent un petit nombre de variables socio-démographiques, à savoir l'âge, le sexe, la commune de résidence et le fait de bénéficier ou non de la couverture maladie universelle complémentaire (CMU-C). Un *indice de défavorisation sociale* a également été constitué permettant de connaître le niveau social de la commune dans laquelle un individu réside. Cet indicateur est un agrégat de plusieurs grandes variables socio-économiques propres au territoire, mais ne décrit pas directement la situation sociale de l'individu, et ne permet donc pas d'avoir une approche aussi fine qu'avec les multiples variables de l'EDP<sup>15</sup>.

---

<sup>13</sup> Elle a par exemple récemment créé un produit exhaustif « simplifié » (le DCIRS), contenant des variables retraitées afin d'éviter les doublons ou les lignes fictives.

<sup>14</sup> <http://www.indsante.fr/page1.html>

<sup>15</sup> Un travail a été réalisé grâce à l'indice de défavorisation sociale, par une équipe de chercheurs en santé publique, mais reste au niveau de l'analyse écologique et non individuelle : <https://www.cairn.info/revue-sante-publique-2015-3-page-383.htm>

En fait, les études sur les inégalités sociales de santé sont réalisées à partir d'enquêtes déclaratives, fournissant alors un grand nombre d'informations sur la situation des individus. Celles-ci présentent néanmoins leurs limites, notamment la taille des échantillons, qui ne permet pas des études à une échelle fine (géographiquement, ou sur des sous-populations), leur coût, qui risque d'amener dans le futur à leurs réductions progressives, ainsi que les biais de déclaration et de mémoire sur les questions de santé qui peuvent compromettre la qualité des données avec une non-réponse à traiter ou une fiabilité contestable de la réponse en moyenne – c'est particulièrement le cas lorsqu'on interroge sur le recours aux soins par exemple, mais également sur la catégorie socio-professionnelle ou sur les revenus, des variables difficiles à recueillir via les enquêtes.

Ainsi, pour l'étude descriptive des inégalités sociales de santé, les enquêtes nationales sur la santé sont la source première. On peut donner l'exemple de l'enquête européenne de santé dont la dernière vague a été pilotée en 2014 par l'Irdes et la Drees et qui a permis de mettre en évidence des écarts de santé perçue ou de déclaration de symptômes dépressifs par PCS.<sup>16</sup> Cette enquête a elle-même été appariée au SNDS, afin de bénéficier d'une mesure complémentaire et plus « objective » de l'état de santé (via les pathologies, les causes de mortalité ou la consommation de soins), ainsi que du recours aux soins. L'EDP-Santé s'appuierait néanmoins sur des variables socio-économiques également objectivées dans les données fiscales et les données du panel DADS. En outre, la taille de cet échantillon permettra des croisements très fins, qui sont impossibles avec l'enquête santé européenne. Par ailleurs, le suivi annuel des données sera beaucoup plus riche que l'observation permise tous les six ans par le dispositif européen. Pour autant, l'EDP-Santé ne constituera en aucun cas un substitut à ces enquêtes qui servent à la production d'indicateurs européens harmonisés sur la santé déclarée, et permettent également d'approcher d'autres déterminants de l'état de santé comme le mode de vie, à travers les pratiques sportives, addictives (consommation de tabac ou alcool,..) ou alimentaires.

Pour l'étude plus suivie des trajectoires de santé, il existe également des cohortes, comme la cohorte Elfe pour la santé des enfants ou la cohorte Constances. Cette dernière est d'ailleurs appariée au SNDS afin de repérer les consommations de soins des enquêtés, qui font également l'objet d'examens de santé réguliers, en plus de leurs réponses à des questionnaires déclaratifs. Si ce dispositif complet permet une approche globale de l'état de santé croisant santé perçue et diagnostiquée par un professionnel de santé, situation socio-économique et comportements de santé, il s'appuie néanmoins sur un fort biais de sélection<sup>17</sup> qui ne permet pas de tirer des conclusions sur l'ensemble de la population. En outre, concernant les études sur le moyen ou plus long terme, elles sont rendues plus délicates par le biais d'attrition inhérent aux cohortes.

En ce qui concerne la connaissance du système de protection sociale et d'assurance maladie, l'EDP-Santé a vocation à remplacer le modèle de simulation OMAR (Outil de Microsimulation pour l'Analyse des Restes à charge) de la DREES. Ce modèle reconstitue le partage de la dépense individuelle entre la Sécurité sociale, l'organisme complémentaire et l'individu, en s'appuyant notamment sur une enquête auprès des organismes de couverture complémentaire. L'EDP-Santé remplacera également la base Hygie (Système d'Information sur les Indemnités Journalières), qui est actuellement la seule base permettant l'étude des arrêts de travail et du versement des indemnités journalières, par le biais d'un appariement entre les données de la CNAMTS et celles de la CNAV.

---

<sup>16</sup> Pisarik J., Rochereau T., État de santé des Français et facteurs de risque, *Études et résultats*, 2017, 998.

<sup>17</sup> Lié au mode de recrutement dans la cohorte.

## 9. Insertion dans un système d'information

Aucune enquête ou post-enquête n'est envisagée sur cet échantillon pour l'instant.

## 10. Unité(s) statistique(s)

Les deux bases de données étant centrées sur les individus, l'EDP-Santé aura également comme unité statistique les individus.

## 11. Champ de l'enquête (secteurs d'activité – préciser les codes NAF – ; taille de l'unité enquêtée ; catégories de ménages ou d'individus ; etc.)

Le champ de l'appariement correspond au croisement du champ des deux bases de données, à savoir :

- les individus présents dans l'EDP, nés lors de 16 jours particuliers de l'année et dont on dispose d'une information dans l'une des bases appariées (les principales étant historiquement l'état civil, contenant les bulletins de naissance pour les personnes nées sur le territoire, et le recensement) ;
- les individus présents dans le SNDS, à savoir ceux qui ont consommé au moins un soin sur la période concernée par les données et dont le soin a fait l'objet d'un remboursement par une caisse d'assurance maladie obligatoire, ceux qui ont été hospitalisés ou sont décédés.

Ce dernier critère est important à souligner dans la mesure où il entraînera une restriction des études aux individus intégrés au système de soins français. Ainsi, les populations très vulnérables qui n'y accèdent jamais seront hors du champ du SNDS, de même que des individus qui n'ont pas fait diagnostiquer leur état de santé, ou ceux qui auraient uniquement recours à des soins qui ne font pas l'objet d'un remboursement par l'assurance maladie. Ces populations sont cependant probablement marginales dès lors que l'appariement couvrira une période longue (10 ans au démarrage du projet). De plus, les non-consommateurs seront laissés dans la base finale afin de permettre des études ou des comparaisons avec ces populations.

## 12. Champ géographique de l'enquête

Le champ géographique couvert par l'EDP est intrinsèquement lié aux sources qui l'alimentent. Il correspond aux personnes résidant, travaillant, ou connaissant un événement d'état-civil enregistré en France, autrement dit en métropole ou dans les DOM, les DOM ayant été introduits à des dates différentes selon les sources et selon la disponibilité des informations (Mayotte comprise, pour une partie des sources).

Le SNDS est également disponible sur la France entière et l'EDP-Santé offrira donc la possibilité de faire des études sur l'ensemble de la France, ainsi qu'à des échelles géographiques très fines.

## 13. Bases de sondage, échantillonnage (taille de l'échantillon)

Le taux de sondage est celui de l'EDP, soit 4,4 % de la population. La base études de l'EDP 2015, disponible depuis le printemps 2017, contenait à titre d'exemple 3,6 millions d'individus.

## 14. Mode de collecte

Il s'agit de données déjà collectées, administratives pour la plupart.

#### 15. **Date (même approximative) et périodicité de la collecte**

La production de l'appariement dépend notamment d'un certain nombre d'étapes juridiques dont les délais sont variables et notamment l'autorisation de la CNIL. Un calendrier provisoire est donc fourni (Annexe 1). Le premier appariement devrait néanmoins pouvoir être réalisé à la suite de la production de la base études de l'EDP au premier semestre 2019 (correspondant aux données EDP et SNDS jusqu'en 2017).

La première étape prévoit l'appariement des deux bases historiques, à savoir l'EDP dans sa totalité, et les dix années actuellement disponibles du SNDS, les études envisagées s'appuyant pour beaucoup sur des données longitudinales.

Par la suite, il est prévu de réaliser cet appariement en routine tous les ans, et d'actualiser et compléter la base, afin de réaliser des études sur les évolutions de l'état de santé et des situations socio-économiques.

#### 16. **Temps de réponse - Contraintes imposées à l'enquête**

*Ne s'applique pas (appariement de données administratives)*

#### 17. **Instances de concertation (comitologie) : comité scientifique, de concertation, d'orientation, de pilotage, de suivi**

La comitologie autour du projet est encore en cours de conception : au-delà des notes de consultation diffusées auprès d'une liste étendue d'organismes de recherche, l'équipe projet est actuellement centrée autour de l'INSEE et de la Drees.

Dans un second temps, une fois les procédures juridiques engagées, un comité de suivi du projet sera mis en place, ainsi que des groupes de travail thématiques notamment sur les indicateurs agrégés à conserver dans la base EDP-Santé agrégé.

#### 18. **Exploitation et diffusion des résultats**

| Type de diffusion envisagée selon la codification (1)  | Période de diffusion envisagée | Niveau géographique (2) envisagé pour les résultats | Remarques |
|--|--------------------------------|---|-----------|
| Production de nouvelles séries sur les indicateurs de santé par caractéristiques socio-économiques | 2020                           | Départemental                                       |           |

|  |           |  |  |
|--|-----------|--|--|
| Études de format court   | 2020      | France entière ou Zones géographiques ciblées (par exemple, DOM) |  |
| Utilisation des données longitudinales pour des sujets de recherche plus approfondis | 2021-2022 | France entière   |  |

### 19. Moyens humains et financiers mobilisés pour l'enquête

La DREES consacre 0,8 ETP (niveau A) sur toute la durée de la mise en place du projet (2016-2019), et l'INSEE, très impliqué dans la construction du projet consacre 0,3 ETP niveau A et 0,3 ETP niveau B.

Les coûts liés à la réalisation de l'appariement en tant que tel sont plus difficiles à chiffrer et font l'objet de demandes encore en cours auprès du CASD, de l'INSEE, et de la CNAMTS qui réaliseront une grande partie des opérations. Pour l'équipe EED de l'INSEE, cela représentera 0,1 ETP niveau A et 0,1 ETP niveau B. Si l'INSEE fait passer FOIN1, il faudra prendre en compte le coût en ETP (non estimé à ce stade).

Le CASD doit également indiquer s'il demandera une contrepartie financière pour le stockage de la base, ce qui n'est pas actuellement le cas pour l'EDP actuel.

### 20. Procédure juridique envisagée

La réalisation de cet appariement repose sur une autorisation de la CNIL. Les autres étapes de la procédure juridique sont en train d'être définies : la DREES attend une décision du Conseil d'État sur la mise en œuvre de la loi informatique et libertés, modifiée par la loi de modernisation de notre système de santé, pour les données statistiques relatives à la santé.

Actuellement, pour accéder à l'EDP, la procédure prévoit un passage devant le comité du secret statistique. En particulier, les chercheurs doivent souligner l'apport de l'EDP pour leur sujet de recherche et montrer les limites des autres sources directement accessibles. Si la dérogation est acceptée, le chercheur suit une formation d'une journée au sein du CASD à l'issue de laquelle il sera autorisé à exploiter la base depuis une SD-Box.

Concernant les données de santé, le chercheur doit faire une demande d'autorisation à la CNIL. La procédure consiste à soumettre son projet à l'INDS, guichet unique des demandes de traitement dans le domaine des données de santé, qui transfère le dossier pour avis à un comité d'expert, le CEREES. La CNIL se prononce après avis du CEREES.

L'accès à l'EDP-Santé pourrait donc reposer sur ces deux procédures, dont les attentes sont relativement proches : le chercheur proposerait simultanément son projet à l'INDS et au comité du

secret. La DREES prévoit de faire appel à l'automne 2017 à un cabinet externe d'expertise juridique pour réfléchir aux possibilités de simplification.