



Conseil national  
de l'information statistique

## CONSEIL NATIONAL DE L'INFORMATION STATISTIQUE

### Commission

« Services publics et services aux publics »

Réunion du 4 octobre 2017

#### Demandses d'accès à des sources administratives au titre de l'article 7bis de la loi de 1951

**Formulées par** la Direction de la Recherche de l'Évaluation, des Études et des Statistiques (DREES) du Ministère des Affaires Sociales et de la Santé.

- Accès aux fichiers fiscaux : fideli (base géolocalisée pour le tirage de l'échantillon), plfc (caractéristiques du logement), fip (numéros de téléphone), pote (caractéristiques des personnes), filosofi (variables de revenus pour l'enrichissement des données) de la DGFIP
- Données brutes du Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) détenu par l'Institut de recherche et documentation en économie de la santé (Irdes)
- Données d'ESPS 2012 et EHIS-ESPS 2014 appariées aux données brutes du SNIIRAM détenu par l'Institut de recherche et documentation en économie de la santé (Irdes)

## **Demande d'accès au titre de l'article 7bis de la loi n° 51-7111 du 7 Juin 1951 modifiée à des données issues des fichiers fiscaux par la DREES.**

### **1. Service demandeur**

La DREES : Direction de la recherche, de l'évaluation, des études et des statistiques du Ministère de la Santé

### **2. Organisme détenteur des données demandées**

La DGFIP : Direction générale des Finances publiques

### **3. Nature des données demandées**

Accès aux fichiers fiscaux : fideli (base géolocalisée pour le tirage de l'échantillon), plfc (caractéristiques du logement), fip (numéros de téléphone), pote (caractéristiques des personnes), filosofi (variables de revenus pour l'enrichissement des données)

### **4. Objectifs généraux des traitements prévus avec les données demandées**

L'accès aux fichiers fiscaux est demandé pour constituer la base de sondage de l'enquête santé européenne EHIS 2019. Il s'agit d'une enquête généraliste portant sur les 15 ans ou plus, réalisée en métropole tous les 6 ans, et pilotée au niveau européen par Eurostat. Pour cette édition, il a été décidé d'étendre le terrain aux 5 DROM de manière à être représentatif de chacun.

L'intérêt de tirer l'échantillon de l'enquête santé européenne dans cette base plutôt que dans le recensement, réside dans la possibilité d'utiliser les numéros de téléphone que comprennent ces fichiers. En effet, dans ces fichiers, un numéro de téléphone est disponible pour la moitié des ménages français.

Deux protocoles sont envisagés pour cette enquête. La décision finale afférente au choix du protocole dépendra du coût de ces protocoles et du budget à disposition. Un appel d'offre public à option décrivant simultanément les deux protocoles vient d'être lancé.

Le premier protocole prévoit d'interroger par téléphone, les personnes pour lesquelles on dispose d'un numéro de téléphone et en face à face, les autres. Bien évidemment, disposer de numéros de téléphone dans la base de sondage permettra d'interroger plus de personnes par téléphone et ainsi de réduire le coût de la collecte.

Le second protocole envisagé est mixte. Les enquêtés sont interrogés en face à face sur leur situation sociodémographique. Par la suite, un auto-questionnaire papier sur leur santé leur est donné, à charge pour eux de le renvoyer complété. Dans ce protocole, disposer de numéros de téléphone permettra aux enquêteurs de fixer des rendez-vous avec les enquêtés, limitant par là même les déplacements pour entrer en contact avec le ménage, et donc le coût de la collecte.

Le tirage sera effectué dans les fichiers fiscaux pour toute l'enquête sauf à Mayotte où le tirage aura lieu dans les données cartographiques pour disposer d'informations les plus fraîches possibles sur le bâti.

Les fichiers fiscaux seront également utilisés pour enrichir l'enquête de variables relatives notamment aux revenus et à la situation matrimoniale légale.

### **5. Nature des travaux statistiques prévus**

Production des données de l'enquête santé européenne EHIS 2019.

## **6. Place dans le dispositif statistique existant sur le même sujet**

Du fait de l'arrêt de l'enquête santé et protection sociale (ESPS) de l'Irdes en 2014, aucun dispositif existant ne permet de répondre à la demande d'enquête santé européenne d'Eurostat.

Par ailleurs, l'enquête EHIS devient, pour la France, la seule enquête généraliste sur la santé, représentative de la population d'au moins 15 ans et qui permet d'observer l'état de santé, le recours aux soins, les déterminants de santé et la couverture complémentaire santé, à intervalles réguliers et avec un protocole stable. Les autres enquêtes sur la santé, comme le Baromètre santé de Santé publique France, sont souvent thématiques ou font des focus sur des sous-populations et ne permettent pas de construire des indicateurs sur longue période.

## **7. Périodicité de la transmission**

Le tirage de l'échantillon de l'enquête sera effectué par l'Insee début 2019. L'enrichissement des données de l'enquête avec les fichiers fiscaux sera réalisé par le pôle revenus fiscaux de l'Insee dès que les données de la validité 2019 seront disponibles.

## **8. Diffusion des résultats**

*Cette enquête répond à une obligation européenne. Les données de l'enquête seront transmises à Eurostat en 2020 puis diffusées via le réseau Quetelet aux chercheurs.*

**Demande d'accès au titre de l'article 7bis de la loi n° 51-7111 du 7 Juin 1951 modifiée à des données concernant l'appariement de l'enquête santé européenne EHIS-ESPS 2014 aux données brutes du Système national d'information interrégimes de l'Assurance maladie (SNIIRAM) détenu par l'Institut de recherche et documentation en économie de la santé (Irdes)**

### **1. Service demandeur**

La Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), pour le bureau État de santé de la population de la sous-direction de l'Observation de la santé et de l'assurance-maladie.

### **2. Organisme détenteur des données demandées**

L'enquête santé européenne EHIS-ESPS 2014 (European health interview survey – enquête santé et protection sociale) a été réalisée par la DREES et l'Institut de recherche et documentation en économie de la santé (Irdes). Un tiers de confiance, le centre d'accès sécurisé distant aux données (CASD), a permis techniquement l'appariement des données de l'enquête avec les données médico-administratives du Système national d'information interrégimes de l'Assurance maladie (SNIIRAM, y compris les différents volets – médecine, chirurgie, obstétrique : MCO, recueil d'information médicalisé en psychiatrie : RIM-P, soins de suite et de réadaptation : SSR, hospitalisation à domicile : HAD – du Programme de médicalisation des systèmes d'information, PMSI). L'Irdes est à l'origine de cette demande d'appariement qui a fait l'objet d'une délibération (n°2013-323 en date du 24/10/2013) de la Commission nationale de l'informatique et des libertés (CNIL). L'Irdes réalise cet appariement de données, une fois que le CASD les a rendues anonymes, et est ainsi détenteur des données demandées.

Il existe deux formats de données issues de cet appariement.

- Le premier contient des données du SNIIRAM regroupées par grands postes de consommation de soins. Conformément à la délibération de la CNIL, l'Irdes met librement à disposition des chercheurs ces données.
- Le second contient les données brutes du SNIIRAM. L'accès à cet appariement avec les données brutes de consommations de soins nécessite une autorisation de la Cnil. Pour toute demande extérieure d'accès à ces données, et une fois l'autorisation CNIL nécessaire obtenue, l'Irdes se charge de demander aux régimes de l'Assurance maladie leur accord pour transmettre les données appariées.

### **3. Nature des données demandées**

L'enquête santé européenne EHIS-ESPS 2014 est une enquête de la DREES et de l'Irdes menée auprès des ménages résidant en logement ordinaire en France métropolitaine. Elle porte sur l'état de santé, physique et mental, le recours aux soins, les déterminants de santé pour les 15 ans et plus, et recueille également des informations socio-économiques sur les ménages enquêtés. L'échantillon interrogé est constitué de ménages dont l'un des membres appartient à l'échantillon généraliste de bénéficiaires (EGB) de l'Assurance maladie.

Les informations identifiantes, dont le numéro de Sécurité sociale, sont extraites des bases locales de l'Assurance maladie. Ce numéro permet, avec l'autorisation de la CNIL, d'apparier les données de l'enquête aux données du SNIIRAM en ayant recours au CASD comme tiers de confiance. Cet appariement est rétrospectif et prospectif : aux réponses à l'enquête santé européenne menée en 2014 peuvent être associées les consommations de soins des années 2010 à 2016. Pour l'instant, les 5 premières années ont été appariées, les deux suivantes le seront en 2018.

C'est l'accès aux données d'EHIS ESPS 2014 appariées avec les 7 années d'historique des données brutes du SNIIRAM, qui fait l'objet de la présente demande.

La demande d'avis au titre de l'article 7bis constitue un préalable à la démarche de demande d'autorisation auprès de la Cnil pour l'équipe de la DREES en charge de ce projet d'étude, afin d'en ancrer l'intérêt dans les missions de la statistique publique.

#### **4. Objectifs généraux des traitements prévus avec les données demandées**

Ce projet d'étude statistique entend contribuer à la connaissance des inégalités sociales en matière de santé mentale et de recours aux soins en santé mentale, répondant ainsi à un besoin d'information en la matière jugé prioritaire au niveau du ministère chargé de la santé.

L'objectif est d'exploiter les données brutes du SNIIRAM en matière de santé mentale (classe de traitements consommés, détails et chronologie de la prise en charge, etc.), et de les mettre en regard des données recueillies par questionnaire auprès des ménages (recours à des professionnels de santé, déclaration d'une dépression, présence de symptômes dépressifs), pour étudier les inégalités sociales en matière de santé mentale.

Les principaux éléments de ce projet sont repris ci-dessous. Il est décrit précisément en annexe.

L'accès à ces données permet :

- 1) De disposer d'informations médico-administratives sur les parcours de soins des personnes déclarant une dépression ou présentant des symptômes dépressifs repérés dans l'enquête EHIS-ESPS 2014 à l'aide du Patient health questionnaire (le module PHQ-8), et de s'interroger ainsi sur d'éventuelles disparités sociales dans leurs parcours de soins ;
- 2) De caractériser socio-économiquement les personnes repérées dans le SNIIRAM comme prises en charge pour un problème de santé mentale (grâce à la codification en affection de longue durée – ALD - ou au diagnostic posé lors d'une hospitalisation ou à l'utilisation de médicaments traceurs pour certaines pathologies mentales) ;
- 3) Sous un angle plus méthodologique, l'exploitation de cet appariement permettra également de mieux cerner l'apport des données du SNIIRAM sur le champ de la santé mentale et ses limites. En effet, la DREES étudie l'opportunité d'une enquête sur la santé mentale en population générale à l'horizon 2020-2021 et souhaite donc repérer ce qui est utile et pertinent dans les bases de l'Assurance maladie et ce qui doit être complété par des enquêtes.

#### **5. Nature des travaux statistiques prévus**

Il sera, tout d'abord, nécessaire de s'approprier les données du SNIIRAM relatives à la santé mentale afin de classer les types de recours, de consommations de soins, d'étudier les motifs d'hospitalisation et de caractériser la temporalité des parcours de soins. La codification en affection de longue durée, le diagnostic posé lors d'une éventuelle hospitalisation, les médicaments traceurs de certaines pathologies mentales, les informations sur le nombre et la fréquence des recours aux soins, sur la durée des hospitalisations, durant les 7 années couvertes par l'appariement, seront ainsi mobilisées.

Les informations socio-économiques proviennent principalement de l'enquête santé européenne EHIS-ESPS 2014. À partir de celle-ci, des profils de réponse à différentes questions portant sur la santé mentale (déclaration d'une dépression, présence de symptômes dépressifs) pourront également être construits. L'enquête permet de plus de disposer du recours aux psychothérapies, type de recours qui ne figure pas nécessairement dans les données de l'Assurance maladie car il n'est pas toujours pris en charge.

Conformément à ses objectifs, le projet analysera ensuite les différences sociales de recours aux soins sous les deux angles principaux présentés précédemment.

L'analyse combinera donc des statistiques descriptives et des traitements économétriques visant à isoler et comparer les effets de différentes variables de contexte (sexe, âge, composition du ménage, statut vis-à-vis du marché du travail, niveau de vie, conditions de logement, etc.) sur le recours aux soins.

#### **6. Place dans le dispositif statistique existant sur le même sujet**

L'exploitation de la grande richesse de ces données appariées serait l'occasion de réaliser une contribution originale à l'étude des inégalités sociales en matière de santé mentale.

En effet, Il existe peu d'enquêtes pouvant être appariées aux données médico-administratives. Un précédent concerne les enquêtes Handicap Santé (2008-2009) qui ont été appariées au SNIIRAM.

Les exploitations qui en ont été faites par la DREES soulignent l'intérêt de ces traitements pour étudier les disparités sociales en matière de recours aux établissements de santé publics ou privés (Pradines et al., *Études et résultats* n° 976, 2016). Cependant, cet appariement ne permettait pas d'étudier les passages par des établissements dans le champ de la santé mentale.

L'appariement d'EHIS-ESPS 2014 et des données brutes du SNIIRAM constitue donc un pont novateur qui permettra de compléter utilement les publications existantes, qui s'appuient sur l'une ou l'autre des sources, et notamment d'une part les travaux de la Caisse nationale d'Assurance maladie (CNAM) sur les consommations de soins, et d'autre part les travaux sur l'enquête EHIS-ESPS 2014, dont les premiers résultats publiés comportent d'ores et déjà des éléments de cadrage sur la fréquence des symptômes dépressifs en France et en Europe (Pisarik et al., *Études et résultats* n° 998, 2017).

## **7. Périodicité de la transmission**

Cet appariement est constitué pour une enquête ponctuelle. La transmission s'effectuera donc de manière ponctuelle et les 7 années d'historique seront conservées.

L'appariement est réalisé par l'Irdes sur la base d'un identifiant anonyme fourni par le CASD. Une fois l'appariement réalisé un nouvel identifiant anonyme est créé par l'Irdes. Les données transmises ne comporteront donc que des identifiants anonymes.

S'agissant de données médico-administratives sensibles issues du SNDS, les données devront être conservées sur un serveur garantissant un niveau de sécurité conforme au référentiel de sécurité pour les données du SNIIRAM.

## **8. Diffusion des résultats**

Seuls des statistiques très agrégées ou des résultats de modélisations seront diffusés. Ils feront l'objet d'une publication du type 4-6 pages de la DREES ou d'un article scientifique, et d'éventuelles communications dans des séminaires.

Un document méthodologique sera préparé afin de capitaliser sur les enseignements de cette exploitation et de détailler les choix de traitements retenus pour rendre compte de la richesse de ces données.

**Demande d'accès au titre de l'article 7bis de la loi n° 51-7111 du 7 Juin 1951 modifiée à des données concernant l'appariement des enquêtes santé ESPS 2012 et EHIS-ESPS 2014 aux données brutes du Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) détenu par l'Institut de recherche et documentation en économie de la santé (Irdes)**

### **1. Service demandeur**

La Direction de la recherche, des études, de l'évaluation et des statistiques (DREES), pour le bureau État de santé de la population de la sous-direction de l'Observation de la santé et de l'assurance-maladie, qui réalise une étude en collaboration avec l'Institut de recherche et documentation en économie de la santé (IRDES).

### **2. Organisme détenteur des données demandées**

L'Enquête santé et protection sociale (ESPS) est une enquête par sondage en population générale qui est réalisée depuis 1988 par l'IRDES. Les données demandées correspondent aux deux derniers volets de cette enquête : ESPS 2012 et EHIS-ESPS 2014<sup>1</sup>.

Depuis 2010, l'enquête est directement appariée aux Données de consommation inter-régimes (DCIR), issues du Système national de l'information inter-régimes de l'Assurance maladie (SNIIRAM), enrichies des données du Programme de médicalisation des systèmes d'information (PMSI).

Pour les appariements, l'Irdes est à l'origine des demandes qui font l'objet de délibérations (n°2012-055 du 16 février 2012 et n°2013-323 en date du 24 octobre 2013) de la Commission nationale de l'informatique et des libertés (CNIL). L'Irdes réalise ces appariements de données, une fois que le CASD les a rendues anonymes, et est ainsi détenteur des données demandées.

Il existe deux formats de données issues de ces appariements.

Le premier contient des données du SNIIRAM regroupées par grands postes de consommation de soins. Conformément à la délibération de la CNIL, l'Irdes met librement à disposition des chercheurs ces données.

Le second contient les données brutes du SNIIRAM. L'accès à cet appariement avec les données brutes de consommations de soins nécessite une autorisation de la Cnil. Pour toute demande extérieure d'accès à ces données, et une fois l'autorisation CNIL nécessaire obtenue, l'Irdes se charge de demander aux régimes de l'Assurance maladie leur accord pour transmettre les données appariées.

### **3. Nature des données demandées**

L'Enquête santé et protection sociale (ESPS) est l'enquête de référence sur la santé, l'accès aux soins et la couverture maladie en France. Menée auprès des ménages résidant en logement ordinaire en France métropolitaine, elle recueille des données sur l'état de santé, physique et mental, le recours aux soins, les déterminants de santé pour les 15 ans et plus<sup>2</sup>, et recueille également des informations socio-économiques sur les ménages enquêtés. L'échantillon interrogé est constitué de ménages dont l'un des membres appartient à l'échantillon généraliste de bénéficiaires (EGB) de l'Assurance maladie. C'est l'accès aux données d'ESPS 2012 et EHIS-ESPS 2014 appariées avec les données brutes du SNIIRAM, qui fait l'objet de la présente demande.

La demande d'avis au titre de l'article 7bis constitue un préalable à la démarche de demande d'autorisation auprès de la Cnil pour l'équipe de la DREES en charge de ce projet d'étude, afin d'en ancrer l'intérêt dans les missions de la statistique publique.

---

<sup>1</sup> En 2014, l'enquête est coproduite par l'IRDES et la DREES et constitue le support de l'Enquête santé européenne (EHIS). Elle se dénomme ainsi EHIS-ESPS.

<sup>2</sup> En 2012, l'enquête a également été menée auprès des moins de quinze ans, mais ces données ne sont pas utilisées dans le cadre de ce projet.

#### **4. Objectifs généraux des traitements prévus avec les données demandées**

Ce projet d'étude statistique entend comparer la morbidité déclarée à la morbidité repérée à travers le recours au système de soins et la dimension sociale des écarts de morbidité entre ces deux mesures.

En effet, en santé publique, le recours aux résultats d'enquête déclarative est très courant pour obtenir des informations sur l'état de santé d'une population à un moment donné. Il s'agit en effet d'une méthode moins contraignante, aussi bien financièrement qu'en matière d'organisation et de sollicitation des répondants, qu'un dispositif d'observation médicale qui reposerait sur des examens par un professionnel de santé.

Néanmoins, ces résultats déclaratifs reposent sur une hypothèse forte : celle de la bonne connaissance de leur état de santé par les individus. Ainsi, les méthodes déclaratives apportent une réponse adéquate pour la mesure de la santé perçue, mais en matière de déclaration des pathologies il existe des biais de déclaration qui pourraient entraîner une mauvaise approximation de la prévalence des pathologies. En outre, dans le cas des données de santé, peut aussi émerger un souhait de dissimuler certaines informations considérées comme sensibles ou personnelles. Cette situation ressort plus particulièrement dans le cas de situations stigmatisantes et entraînent un biais de mesure même dans le cas d'informations largement connues par une majorité de la population. A l'inverse, une pathologie peut être déclarée par un enquêté sans que celui-ci n'ait recours aux soins correspondant.

Ce projet d'étude s'appuierait sur la comparaison entre la morbidité déclarée dans l'enquête Santé et Protection Sociale par les répondants, parmi une liste de 15 maladies, et les « top pathologies », indicateurs repérant la présence de pathologies *via* le SNIIRAM. Ces tops reposent sur des algorithmes développés par l'Assurance maladie permettant d'inférer pour chaque individu différentes pathologies à partir des affections de longue durée (ALD), des consommations de médicaments traceurs et des diagnostics en cas d'hospitalisation, informations collectées dans le SNIIRAM. Ils permettent d'identifier une sous-population d'enquêtés qui présentent certaines pathologies et de se pencher sur la déclaration de pathologies par cette population. Ainsi, ce travail sera ciblé sur la non-déclaration d'une pathologie repérée et non l'inverse, qui consisterait à s'intéresser au non-recours au système de soins dans le cas de la déclaration d'une pathologie dans l'enquête.

La compilation des deux volets d'enquête, 2012 et 2014, permettra d'obtenir un plus grand nombre d'observations pour étudier les écarts de déclaration de pathologies plus rares dans la population.

#### **5. Nature des travaux statistiques prévus**

L'analyse combinera des statistiques descriptives croisant les pathologies repérées à travers le SNIIRAM avec les pathologies déclarées par les enquêtés, ainsi que des traitements économétriques visant à isoler et comparer les effets de différentes variables de contexte (sexe, âge, niveau de diplôme, composition du ménage, statut vis-à-vis du marché du travail, niveau de vie, catégorie socio-professionnelle, etc.) sur les écarts entre la prévalence de pathologies déclarées et de celles mesurées par le recours aux soins.

#### **6. Place dans le dispositif statistique existant sur le même sujet**

Cette étude prolongera ainsi les travaux de recherche menés dans les années 2000 (Dauphinot et al., 2008, *Revue Française des Affaires Sociales*, Devaux et al., 2008, *Revue Française des Affaires Sociales*) qui avaient mis en évidence la sous-déclaration des problèmes de santé et le biais de déclaration lié à l'indicateur de maladies chroniques, pouvant notamment être lié au statut socio-économique.

#### **7. Périodicité de la transmission**

Cet appariement est constitué de manière ponctuelle pour chacune des deux enquêtes. La transmission s'effectuera donc de manière ponctuelle et les 7 années d'historique seront conservées.

L'appariement est réalisé par l'Irdes sur la base d'un identifiant anonyme fourni par le CASD. Une fois l'appariement réalisé un nouvel identifiant anonyme est créé par l'Irdes. Les données transmises ne comporteront donc que des identifiants anonymes.

S'agissant de données médico-administratives sensibles issues du SNIIRAM, les données devront être conservées sur un serveur garantissant un niveau de sécurité conforme au référentiel de sécurité pour les données du SNIIRAM.

#### **8. Diffusion des résultats**

Seuls des statistiques très agrégées ou des résultats de modélisations seront diffusés. Ils feront l'objet d'une publication du type 4-6 pages de la DREES ou d'un document de travail méthodologique, et d'éventuelles communications dans des séminaires.