



Conseil national
de l'information statistique

Paris, le 3 décembre 2020 n° 141 / H030

**AVIS DU CNIS SUR UNE DEMANDE D'ACCÈS À DES DONNÉES
AU TITRE DE L'ARTICLE 7 BIS DE LA LOI DU 7 JUIN 1951 MODIFIÉE**

MIS À JOUR LE 21 SEPTEMBRE 2021

Au cours de sa réunion du 2 décembre 2020, la commission « Démographie et Questions Sociales » a examiné la demande d'accès à des sources administratives :

Formulée par la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) du Ministère des Solidarités et de la Santé

- des données concernant SI-VIC (SI de Suivi et de Dénombrement de Victimes) détenues par la Direction Générale de la Santé

La commission émet un **avis favorable** à cette demande d'accès.

**Le président de la commission
Jean-Philippe Vinquant**

Demande d'accès au titre de l'article 7bis de la loi n° 51-7111 du 7 Juin 1951 modifiée à des données concernant SI-VIC par la DREES.

1. Service demandeur

Direction de la recherche, des études, de l'évaluation et des statistiques (DREES)
Service statistique ministériel au sein du Ministère des solidarités et de la santé

2. Organisme détenteur des données demandées

Direction générale de la santé du ministère des solidarités et de la santé, responsable du traitement SI-VIC.

3. Nature des données demandées

La base SI-VIC (SI de Suivi et de Dénombrement de Victimes) est un système d'information mis en œuvre par la DGS en 2016 à la suite des attentats de Paris de 2015 avec l'objectif d'identification et de suivi des victimes des situations sanitaires exceptionnelles (attentats, accident majeurs).

En mars 2020, SI-VIC devient le pivot du dénombrement hospitalier pour la Covid-19. Depuis, tous les établissements de santé (>2 000) qui reçoivent des patients atteints de la Covid-19 saisissent dans SI-VIC.

La base de données contient des informations sur le patient (nom, prénom, sexe, date de naissance, adresse et commune de résidence), sur le lieu d'hospitalisation (établissement, service), sur le type d'hospitalisation (hospitalisation conventionnelle, soins critiques, soins de suite et réadaptation) et sur le mode de sortie (retour à domicile ou décès).

La DREES est destinataire des données pseudonymisées et souhaite, pour ce projet, accéder aux données nominatives.

4. Objectifs généraux des traitements prévus avec les données demandées

La DREES est déjà destinataire, en tant que sous-traitant de la DGS dans le cadre de l'appui à la gestion de la crise sanitaire, d'une partie des informations pseudonymisées enregistrées dans l'outil SI-VIC afin notamment de produire des statistiques sur le dénombrement des patients hospitalisés pour Covid-19 et leurs parcours hospitaliers.

L'accès aux données nominatives de SI-VIC est demandé pour permettre l'enrichissement des données de SI-VIC avec des données issues des traitements dont l'INSEE a la responsabilité, notamment le Fichier démographique d'origine fiscale sur les logements et les personnes (FIDELI) et les déclarations sociales (DADS/DSN/panel tous actifs occupés).

L'objectif de cet appariement est de caractériser le profil socio-démographique des patients au-delà de leur sexe et de leur âge et de comparer les caractéristiques de ces patients à celles des personnes présentes dans FIDELI. Une connaissance plus fine des caractéristiques des patients hospitalisés (type de ménage, conditions de logement, situation professionnelle notamment) répondrait à une demande sociale forte, et serait susceptible d'éclairer les autorités sanitaires dans la conduite des politiques de lutte contre l'épidémie Covid-19. La possibilité de comparer les caractéristiques de ces patients à l'ensemble des caractéristiques des personnes en France, grâce au rapprochement avec l'ensemble des personnes de FIDELI (appariement élargi), permettra d'apporter un nouveau regard sur ces résultats. Cet appariement élargi permet en effet d'analyser les déterminants des hospitalisations en lien avec le Covid et pas uniquement les caractéristiques des patients conditionnellement au fait qu'il soit dans SI-VIC (ce qui peut créer des biais d'interprétation)

Les premiers résultats de l'enquête EpiCoV montrent l'importance des conditions d'habitat sur la circulation du virus, en particulier le nombre de personnes avec qui on habite. Cette enquête montre également des taux d'infection chez les professionnels de santé deux fois plus élevés que chez les autres actifs. L'enquête met en évidence que l'exposition au virus dépend des facteurs d'exposition (lieu de résidence, conditions de logement, nécessité de travailler hors du domicile) et aux effets du confinement sur les conditions de vie (recours au télétravail, situation d'emploi, situation financière).

Les résultats issus de EpiCoV donnent une information sur le risque en population générale, mais ne permettent pas de produire des informations sur les formes graves de la maladie. D'autres études ont déjà pu éclairer un certain nombre de questions liées aux inégalités sociales de santé, et ont par exemple pu mettre en évidence des inégalités d'accès à la prévention ou de reste à charge.

L'appariement de données de SIVIC avec les données de FIDELI et DADS/DSN permettra de mettre en relation ces inégalités sociales, en termes de mode de vie, de situation professionnelle avec les parcours des patients avec une forme grave de Covid-19, nécessitant une hospitalisation, avec pour une partie d'entre eux un passage par une unité de soins critiques, et pouvant amener à un décès. Cette analyse suppose de disposer non seulement des données SI-VIC appariées à celles de FIDELI, mais également des caractéristiques des personnes de FIDELI non présentes dans SI-VIC.

Cette analyse ne permettra pas de prendre en compte la présence ou pas de comorbidités, comme une pathologie chronique (ex : diabète), facteur de risque avéré pour les formes graves. L'appariement avec les données du Système national de données de santé en revanche permettrait de compléter cette analyse sur ce volet.

5. Nature des travaux statistiques prévus

La DREES reçoit quotidiennement le fichier SI-VIC pseudonymisé. Elle utilise les informations issues de SI-VIC pour réaliser des prévisions d'entrées / sorties en hospitalisation, retours à domicile et décès en date de survenue corrigée des délais d'enregistrement, et à des fins d'étude (par exemple, le Dossier de la DREES consacré au parcours d'hospitalisation des malades du Covid-19). La DREES a acquis une bonne expertise des variables d'études présentes dans le fichier.

Le schéma d'appariement envisagé est le suivant :

- .L'INSEE est directement destinataire des traits d'identités du fichier SI-VIC en provenance de l'ANS.
- .L'INSEE reconstitue un NIR pour les identifiants SI-VIC et FIDELI sur la base des traits d'identités (le NIR est présent nativement dans la DSN). Le cas échéant, un premier appariement sera réalisé entre SI-VIC et FIDELI sur la base des traits d'identité nom, prénom, sexe, date de naissance, adresse et commune de résidence, afin de compléter les traits d'identités SIVIC par le lieu de naissance issu de FIDELI (le lieu de naissance, indispensable pour la recherche du NIR, n'est pas présent nativement dans le fichier SIVIC) et de limiter en volume les recherches de NIR sur le fichier FIDELI. Ce premier appariement serait pris en charge par le missionnaire INSEE chargé du pilotage de l'appariement SI-VIC/FIDELI/DSN, en poste à l'Inspection Générale
- .Sur la base des NIR reconstitués, l'INSEE établit la correspondance identifiant SI-VIC/identifiant FIDELI/identifiant DSN.
- .L'INSEE raccroche ensuite les variables d'études FIDELI et DSN au fichier, et renvoie à la DREES un fichier avec l'identifiant SI-VIC et les variables FIDELI/DSN, dont l'identifiant FIDELI non significatif, pour les patients dans SI-VIC et dans FIDELI.
- .L'INSEE conserve par ailleurs une base avec l'ensemble des variables d'études FIDELI pour l'ensemble des individus de FIDELI, qu'ils aient été hospitalisés ou non, avec l'identifiant SI-VIC pour les individus également présents dans SI-VIC.
- .La DREES réapparie ensuite complète ensuite, pour les patients hospitalisés avec les variables d'études SI-VIC sur la base de l'identifiant SI-VIC pour constituer le fichier individuel complet puis le renvoie à l'INSEE, en conservant uniquement les variables strictement nécessaires pour l'analyse.
- .L'INSEE et la DREES reconstituent enfin la base SI-VIC FIDELI appariée avec les informations de FIDELI pour l'ensemble des individus de FIDELI, qu'ils aient été hospitalisés ou non.

Ce schéma d'appariement permet notamment d'éviter qu'un même acteur dispose à la fois des traits d'identité SI-VIC et des variables d'étude SI-VIC.

Il est proposé de conserver à l'INSEE une table de correspondance entre identité SI-VIC et NIR. Sous réserve de faisabilité juridique, cela permettrait de procéder, dans un deuxième temps et en fonction de la qualité de la reconstitution du NIR, à un appariement avec le SNDS comme cela a été fait dans le cadre de l'enquête santé.

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

Aucun dispositif existant ne permet de répondre à l'objectif énoncé.

Ces statistiques viendraient compléter d'autres données sur l'épidémie Covid-19 et seraient susceptibles d'éclairer les autorités sanitaires dans la conduite des politiques de lutte contre l'épidémie Covid-19.

En particulier, de telles données permettraient d'identifier quels sous-groupes de population, en termes de niveau de vie ou de situation professionnelle, sont les plus à risque de forme grave. L'appariement élargi permet de comparer les caractéristiques des patients de SI-VIC à ceux de la population générale et donc de distinguer si ces sous-groupes de population le plus à risque le sont du fait d'un plus fort risque d'être hospitalisé ou, d'un risque d'être hospitalisé en soins critiques conditionnellement au fait d'être hospitalisé plus important. Elles viendraient alimenter les réflexions des Directions métier du Ministère : Direction générale de la Santé, Direction générale de l'offre de soins et Direction générale de la cohésion sociale, ainsi que des Agences régionales de santé, et de l'Agence Santé publique France, sur les messages de prévention destinés à la population, l'adaptation des gestes barrières, ou encore la fragilité des populations selon les caractéristiques des territoires. Ces directions n'auraient accès qu'à des tableaux de statistiques agrégés, intégrés le plus souvent dans des notes internes produites par la Drees.

7. Périodicité de la transmission

Il est prévu de réaliser un premier appariement pour voir la faisabilité et la qualité de la reconstruction du NIR à partir des éléments nominatifs de SI-VIC. Pour cela est proposé d'utiliser les données disponibles de SI-VIC depuis le début de la première vague (1^{er} mars). En fonction des résultats, un second appariement en fin d'épidémie pourrait être envisagé.

8. Diffusion des résultats

Les résultats statistiques donneront lieu à des publications de la DREES. Si les résultats sont probants, le fichier pourrait dans un second temps être mis à disposition de la communauté des chercheurs.

Le service producteur cédant a été informé en amont de la demande.