

SANTÉ
TRAVAIL

MARS 2025

MÉTHODES ET REPÈRES

SURVEILLANCE ÉPIDÉMIOLOGIQUE

DE LA SANTÉ DES PERSONNES

SANS EMPLOI EN FRANCE - SESSE

Protocole à partir de l'EDP-Santé

Résumé

Surveillance épidémiologique de la santé des personnes sans emploi en France – SESSE

Protocole à partir de l'EDP-Santé

Une santé dégradée peut être un facteur de risque de perte d'emploi et peut empêcher l'entrée ou le retour à l'emploi. À l'inverse, le chômage peut avoir un impact négatif sur la santé. Les personnes sans emploi présentent un sur-risque de mortalité, de problèmes cardiovasculaires, de santé mentale et sont plus susceptibles d'adopter des comportements à risque pour leur santé comme la consommation excessive d'alcool, de tabac et de cannabis. En France, selon les définitions, le nombre de personnes sans emploi varie de 2 à 5 millions. Santé publique France initie un programme afin d'une part de surveiller l'état de santé de cette population et d'autre part de leur proposer des mesures de prévention et de promotion de leur santé.

Le programme Surveillance épidémiologique de la santé des personnes sans emploi -SESSE vise à décrire les profils sociodémographiques et l'état de santé des personnes sans emploi, en les comparant aux actifs en emploi et à la population générale en France, y compris dans les territoires d'outre-mer. Deux sources de données importantes ont été identifiées : l'enquête Baromètre Santé publique France basée sur une sélection aléatoire et l'Échantillon démographique permanent EDP-Santé - Échantillon démographique permanent apparié au Système national de données de santé (SNDS). L'EDP-Santé est un large échantillon longitudinal représentatif de la population contenant toutes les bases de données du panel des salariés, les données socio-fiscales, d'état civil, les données du recensement de la population et qui intègre également les données de l'assurance maladie dont les données hospitalières, les causes médicales de décès.

Ce protocole décrit les objectifs, la méthodologie, les sources de données, les analyses statistiques, les procédures réglementaires, la confidentialité des données et propose un calendrier prévisionnel d'analyses utilisant l'EDP-Santé.

MOTS-CLÉS : PERSONNES SANS EMPLOI, CHÔMAGE, PROTOCOLE, ÉCHANTILLON DÉMOGRAPHIQUE PERMANENT, SYSTÈME NATIONAL DE DONNÉES DE SANTÉ, SURVEILLANCE ÉPIDÉMIOLOGIQUE

Citation suggérée : Goulard H. Surveillance épidémiologique de la santé des personnes sans emploi en France – SESSE. Protocole à partir de l'EDP. Saint-Maurice : Santé publique France, 2025. 41 p. Disponible à partir de l'URL : <https://www.santepubliquefrance.fr>

ISSN : 2647-4816 ; ISBN-NET : 979-10-289-0986-4 ; RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE - DÉPÔT LÉGAL : MARS 2025

Abstract

Epidemiological surveillance of the health of unemployed people in France – SESSE

Poor health can be a risk factor for job loss and can stop entry or return to employment. Conversely, unemployment can negatively impact health. Unemployed individuals are at risk for mortality, cardiovascular issues, mental health and are more likely to engage in unhealthy behaviors like excessive alcohol, tobacco, and cannabis use.

In France, according to the definition, the number of unemployed individuals ranges from 2 to 5 million and Santé Publique France aims to better understand this population, identifying those at higher risk in order to propose preventive health measures. The SESSE program seeks to describe the sociodemographic profiles and health status of unemployed people, comparing with the employed people and the general population across France, including its overseas territories.

Two important data sources were identified : the French Health Barometer Survey based on a random selection and the EDP-Santé – permanent Demographic sample matched with the National Healthcare Data System (SNDS) a large, representative longitudinal sample of the population containing all employees databases, annual declaration of social data, civil register, population census data and that also incorporates health insurance data, hospital data, and the medical causes of death.

This protocol describes objectives, methodology data sources, statistical analyses, regulatory procedures, data privacy, and a provisional schedule for analysis using EDP-Santé.

KEY WORDS: UNEMPLOYED PEOPLE, PROTOCOL, PERMANENT DEMOGRAPHIC SAMPLE, NATIONAL HEALTHCARE DATA SYSTEM, EPIDEMIOLOGY SURVEILLANCE

Sommaire

Résumé.....	2
Abstract.....	3
Équipe Projet (traitement, mise en œuvre, protection des données).....	5
Abréviations	6
1. CONTEXTE ET JUSTIFICATION	7
1.1 Importance de l'enjeu sanitaire au vu du nombre de personnes potentiellement concernées en France	7
1.2 Contexte général	7
1.3 S'intéresser à la surveillance sanitaire de cette population	8
1.4 Justification du respect de l'éthique et de l'intérêt public	9
1.5 Publication des résultats et valorisation	10
2. OBJECTIFS DE L'ÉTUDE BASÉE SUR L'EDP SANTÉ	11
3. MÉTHODOLOGIE DE L'ÉTUDE BASÉE SUR L'EDP SANTE.....	12
3.1 Conception de l'étude	12
3.2 Sources de données.....	12
3.2.1 Données de l'Échantillon démographique permanent EDP	12
3.2.2 Données de l'EDP-Santé.....	13
3.2.3 Données de santé du SNDS dans l'EDP-Santé	13
3.3 Description et justification de la population d'étude	15
3.3.1 Les personnes sans emploi.....	15
3.3.2 Différentes variables pourront être construites.....	16
3.3.3 Trajectoire de chômage.....	16
3.4 Population de l'étude	16
3.4.1 Période d'étude.....	16
3.4.2 Nombre attendu	16
3.5 Traitements et analyses des données	17
3.6 Analyses descriptives	17
3.7 Modélisation.....	18
3.7.1 L'expérience du chômage.....	18
3.7.2 Pathologies étudiées	19
3.8 Limites de l'étude.....	19
3.8.1 Limites des données de l'EDP	19
3.8.2 Limites des données du SNDS.....	19
3.8.3 Données de comportements de santé.....	19
4. RÔLE DES DIFFÉRENTS ACTEURS ET COMITOLOGIE.....	21
5. DÉMARCHES RÉGLEMENTAIRES	22
6. PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES	24
6.1 Respect des droits des personnes concernées	24
6.1.1 Information et droits des personnes	24
6.1.2 Droits d'accès, d'opposition, de rectification des personnes.....	24
6.2 Confidentialité et sécurité des données.....	25
6.2.1 Gestion du risque de ré-identification	25
6.2.2 Support de données et sécurité.....	25
6.2.3 Plateforme CASD	26
6.2.4 Circuit des données en cas d'appariement	27
6.2.5 Durée de conservation des données.....	28
7. CALENDRIER PRÉVISIONNEL ET FAISABILITÉ DU PROJET	29
8. RÉFÉRENCES BIBLIOGRAPHIQUES.....	30
7. ANNEXES	33
Annexe 1. Listes des tables et variables nécessaires au projet.....	33
Annexe 2. Information collective	40

Équipe Projet

Responsable du traitement

Santé publique France (Agence nationale de santé publique), 12 rue du Val d'Osne, 94415 Saint Maurice Cedex, représentée par sa directrice générale, Caroline Semaille.

Mise en œuvre du traitement

Direction Santé Environnement Travail, représentée par son directeur Sébastien Denys.

Équipe Projet

Hélène Goulard, Santé publique France, responsable du projet, chargée de projets scientifiques à la Direction Santé Environnement Travail, chargée de la rédaction du protocole, de la définition des objectifs, du pilotage des analyses et de la valorisation des résultats. Elle a suivi la formation spécifique concernant l'exploitation des données du Système national des données de santé (SNDS) et dispose d'une expérience dans l'exploitation de ces données via l'utilisation de la cohorte ESPri de retraités de l'ex Régime des indépendants dont le suivi épidémiologique est réalisé via des extractions du SNDS (Cnil N °911.447).

helene.goulard@santepubliquefrance.fr

Aurélie Fouquet, statisticienne à la Direction Appui, Traitements et Analyses de données (Data) de Santé publique France pour le traitement des données

Santé publique France possède une expérience dans l'analyse de l'échantillon démographique permanent (EDP) et dans l'analyse des données du Système national des données de santé.

L'agence utilise notamment les données du Système national des données de santé pour produire des indicateurs de santé et réaliser des études décrivant les incidences ou prévalences de plusieurs pathologies et d'en étudier les facteurs de risque. Santé publique France fait partie des organismes disposant d'un accès permanent aux données individuelles bénéficiaires via le SNDS et plusieurs personnes sont pour cela formées et habilitées à travailler à partir des bases de données du SNDS.

Déléguée à la protection des données

Clothilde Hachin, Santé publique France ; mail : dpo@santepubliquefrance.fr ; tel 01 41 79 57 58

Les analyses prévues à partir de l'EDP-Santé reposent notamment sur l'exploitation de deux bases de données constituées d'une part par l'Insee (l'échantillon démographique permanent EDP), d'autre part par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) (l'EDP-Santé) par appariement déterministe de l'Échantillon démographique permanent, géré par l'Insee, aux données du SNDS, géré par la Caisse nationale assurance maladie (Cnam).

ABRÉVIATIONS

ALD	Affections longue durée
BDD	Bases de données
BIT	Bureau international du travail
CASD	Centre d'accès sécurisé aux données
CépiDc	Centre d'épidémiologie sur les causes médicales de décès
Cesrees	Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé
CIM	Classification internationale des maladies
Cnam	Caisse nationale d'assurance maladie
Cnil	Commission nationale de l'informatique et des libertés
Cnam	Caisse nationale d'assurance maladie
Cnav	Caisse nationale d'assurance vieillesse
Cosmop	Programme de surveillance épidémiologique de la mortalité par causes selon l'activité professionnelle
CSS	Comité du secret statistique
DADS	Déclaration annuelle de données sociales
Dares	Direction de l'animation de la recherche, des études et des statistiques
DCIR	données consommations inter-régime
DPO	Data protection officer - Délégué à la protection des données
Drees	Direction de la recherche, des études, de l'évaluation et des statistiques (ministère en charge de la santé)
EAR	Échantillon annuel de recensement
EDP	Échantillon démographique permanent
EHIS	European Health Interview Survey
ESPrI	Programme Épidémiologie et surveillance des professions indépendantes
FOIN	Fonction d'occultation des identifiants nominatifs
HAD	Hospitalisation à domicile
Insee	Institut national de la statistique et des études économiques I
Inserm	Institut national de la santé et de la recherche médicale
ISS	Inégalités sociales de santé
MCO	Médecine, chirurgie, obstétrique
NIR	Numéro d'inscription au répertoire national d'identification des personnes physiques
OMS	Organisation mondiale de la santé
PMSI	Programme de médicalisation des systèmes d'information
PSE	Personnes sans emploi
RGPD	Règlement général sur la protection des données
RIM-P	Recueil d'information médicalisée en psychiatrie
RPS	Risques psychosociaux
RSA	Revenu de solidarité active
SESSÉ	Surveillance épidémiologique de la santé des personnes sans emploi
SIP	Santé et Itinéraire professionnel
SMR	Standardized mortality Ratio
SNDS	Système national des données de santé
SSR	Soins de suite et réadaptation

1. CONTEXTE ET JUSTIFICATION

1.1 Importance de l'enjeu sanitaire au vu du nombre de personnes potentiellement concernées en France

En France, le chômage est défini comme de l'inactivité forcée des personnes due au manque de travail, d'emploi. Le terme de « chômeur » est polysémique avec différentes définitions proposées selon les objectifs, qu'ils soient d'ordre économique, démographique ou épidémiologique et selon les sources de données disponibles. **Les « personnes sans emploi », PSE, terme que nous emploierons dans ce document correspond aux personnes sans emploi hors personnes retraitées, au foyer, en congés parental, en arrêt maladie, étudiant et préretraitées.** Elles sont celles plus susceptibles de présenter des problèmes de santé.

Selon que les statistiques des « demandeurs d'emploi » soient issues de France Travail¹, ou du Bureau international du travail - BIT, le nombre de personnes concernées s'étend de 2,3 à 5,1 millions. En moyenne au 2^e trimestre 2024, en France hexagonale, le nombre de personnes inscrites à France Travail et tenues de rechercher un emploi (catégories A, B, C) s'établit à 5 112 700. Parmi elles, 2 808 700 personnes sont sans emploi (catégorie A) et 2 303 900 exercent une activité réduite (catégories B, C). Selon France Travail, en France métropolitaine, au 2^e trimestre 2024, 388 500 ont moins de 25 ans et plus de 2 227 400 sont à la recherche d'un emploi depuis plus de deux ans.

Au 2^e trimestre 2024, le nombre de chômeurs au sens du BIT s'établit pour la France hexagonale à 2,3 millions de personnes. Pour autant, selon l'Insee et l'enquête emploi, en moyenne sur l'année 2023, 2,0 millions de personnes appartiennent au halo² autour du chômage parce qu'elles ne recherchent pas d'emploi ou ne sont pas disponibles immédiatement dans les quinze jours. Ainsi, il existe plusieurs catégories de personnes sans emploi, selon que la personne recherche, souhaite travailler, soit inscrite à France Travail ou non, indemnisée ou non et disponible ou non immédiatement, etc.

1.2 Contexte général

Une santé dégradée peut être un facteur favorisant la perte d'emploi et freinant l'entrée ou le retour à l'emploi et réciproquement le fait d'être sans emploi peut s'avérer être un facteur de dégradation de la santé. Les raisons de la surmortalité des personnes sans emploi sont multiples : un état de santé déficient qui provoque la perte d'emploi, les conséquences financières et psychologiques du chômage de longue durée, le rôle de l'effet propre du chômage de type « catalyseur » qui révèle ou accentue les effets sur la santé. En France, Mesrine et col. en 2000 ont observé que dans les cinq ans qui suivent une période de chômage, la mortalité des femmes sans emploi est environ le double de celle des femmes actives occupées et la mortalité des hommes sans emploi est le triple de celle des hommes actifs en emploi en tenant compte du groupe socioprofessionnel, du diplôme et du statut matrimonial (1). Les résultats d'une étude canadienne indiquaient que le fait de toucher des allocations de chômage était associé à une mortalité plus faible parmi les personnes sans emploi à long terme. Élargir l'accès à l'assurance-chômage peut améliorer la santé de la population et réduire les inégalités de santé liées à la perte d'emploi (2).

En France, Méneton et col. avaient mis en évidence qu'être chômeur au début du suivi était associé à un risque d'accidents cardiovasculaires cinq ans plus tard (infarctus du myocarde et accidents vasculaires cérébraux) presque deux fois plus élevé par rapport à des personnes de même âge et de même sexe occupant un emploi. Indépendamment des facteurs de risque connus (alcool, tabac,

¹ <https://statistiques.francetravail.org/stmt/publication>

² Selon l'Insee, correspond aux personnes disponibles pour travailler mais ne réalisant pas de réelles démarches pour trouver un emploi et celles qui ne sont pas disponibles immédiatement (par exemple en formation) mais qui souhaitent travailler.

sédentarité), il existerait un effet propre au chômage qui suggère que le lien serait médié par les troubles du sommeil et la dépression, connus pour augmenter le risque cardiovasculaire (3).

D'après le Baromètre de Santé publique France, les comportements individuels néfastes pour la santé sont plus fréquents chez les chômeurs par rapport aux actifs comme la consommation excessive d'alcool, de tabac, de cannabis (4). Ces conduites addictives ont des effets notamment sur le risque de perdre son emploi (5).

Les facteurs tels que le stress, l'anxiété ou la dépression sont plus élevés chez les chômeurs notamment chez les jeunes (6). On note également que lorsque le taux de chômage augmente au sein d'une population, le taux de suicide augmente également (7). Dans l'étude Cohorte surveillance de la mortalité professionnelle - Cosmop, le risque de suicide chez les chômeurs était 2,2 fois plus important que pour les actifs occupés (8). Le chômage subi peut conduire à un stress post-traumatique (9).

Une revue de littérature scientifique soulignait que les jeunes au chômage étaient particulièrement vulnérables à des problèmes de santé notamment mentale (10-12) et que les jeunes au chômage, plutôt les hommes, devraient être repérés à un stade précoce pour leurs troubles (13). De plus, l'effet du chômage sur la santé mentale était mis en évidence pour les hommes quand la durée du chômage dépassait six mois et plutôt en deuxième partie de carrière (14).

D'après le Baromètre de Santé publique France, selon la situation professionnelle, la prévalence de l'épisode dépressif caractérisé était la plus élevée chez les chômeurs, et ce quelle que soit la période d'étude. La prévalence a augmenté de plus de 5 points (12,1 à 16,7 %) chez les chômeurs sur la période 2005-2017 pour atteindre 22,0 % chez les chômeurs - en 2021 (15, 16).

Enfin, le renoncement aux soins serait multiplié par trois pour un homme au chômage et par plus de deux pour une femme au chômage par rapport aux actifs en emploi (17).

1.3 S'intéresser à la surveillance sanitaire de cette population

Depuis deux ans, Santé publique France s'attache à approfondir les connaissances sur la santé des actifs sans emploi afin de mieux caractériser et identifier les populations les plus à risque nécessitant des actions de prévention pour améliorer leur santé. Trois catégories d'âge seront étudiées en priorité car selon Brysden et col. les problématiques de santé selon la situation professionnelle ne sont pas les mêmes les 16-24 ans pour lesquels la sortie de l'école, du départ de chez les parents, l'entrée dans le travail sont les déterminants liés à la santé entre actifs en emploi et sans emploi ; les 25-44 ans pour lesquels le sommeil, le soutien social, la vie en couple, les revenus, le stress au travail sont les déterminants ; et enfin les 45-65 ans pour lesquels la pénibilité au travail, la santé globale dégradée, les exclus de l'emploi sont les déterminants de la santé en lien avec la situation vis-à-vis de l'emploi (18).

Cette étude s'inscrit dans des enjeux transversaux de l'agence concernant pour l'un la surveillance de la santé en lien avec la situation professionnelle (Direction Santé Environnement Travail) et pour l'autre les Inégalités sociales de santé (ISS) et de l'équité en santé (Direction Scientifique et International) notamment pour mieux prendre en compte les déterminants sociaux sur lesquels agir (19). Le Conseil scientifique de l'agence a souligné et soutenu l'intérêt de cette étude en matière de santé publique.

Le programme dans son ensemble se déroulera selon 3 axes : (annexe 2)

- L'Axe 1 interviendra après une synthèse de la littérature sur le sujet. Il consistera à : i) recenser les sources de données existantes, les populations de personnes sans emploi qui nous intéressent, les connaissances sur leur état de santé et ii) à définir les indicateurs épidémiologiques à retenir.

- L'Axe 2, qui a pour but de publier des données sur la santé des actifs sans emploi, est basé sur deux sources de données qui ont été priorisées :
 - A- Le Baromètre de Santé publique France
 - B- L'Échantillon démographique permanent apparié au Système national des données de santé (EDP-Santé – 3,1 millions de personnes).
- L'Axe 3 en lien avec la direction de la prévention et promotion de la santé (DPPS) a pour objectif de mettre en place une des interventions dont l'efficacité a déjà été démontrée par des évaluations scientifiques. C'est dans ce cadre que l'intervention JOBS (*Job Opportunity and Basic Skills*), a été repérée dans la littérature scientifique internationale. Ce programme est celui de l'Université du Michigan déployé en Finlande, en Irlande, en Allemagne et en Chine. Il a démontré son efficacité dans le cadre d'essais randomisés et contrôlés en particulier en réduisant la dépression et la détresse psychologique. Cette intervention a été adaptée et réévaluée au contexte finlandais par l'équipe du Pr Jukka Vuori (20-22). La mise en place de cet axe est en cours de réflexion et les résultats des analyses du Baromètre de Santé publique France permettront d'identifier et de cibler les populations les plus à risque.

Ce rapport concerne la première partie de l'exécution opérationnelle de la feuille de route de l'axe stratégique 2 partie B.

1.4 Justification du respect de l'éthique et de l'intérêt public

Le protocole, objet du rapport, s'intéresse à la population des personnes sans emploi en France pour lesquelles il n'existe pas à ce jour de système de surveillance spécifique de leur santé. Ce dispositif a pour objectif de produire régulièrement des indicateurs épidémiologiques nationaux afin de quantifier le fardeau sanitaire dans cette population et de proposer des recommandations pour la prévention.

Le repérage de catégories de personnes plus vulnérables à certaines pathologies ou causes de décès n'a aucunement pour objectif de les stigmatiser. Au contraire, il vise à orienter l'action des pouvoirs publics et à mieux prioriser les mesures de prévention et de promotion de la santé.

N'impliquant pas la personne humaine, cette étude ne nécessite pas l'avis d'un comité de protection des personnes. Seules les données des bases de données administratives existantes et médico-administratives seront utilisées. Dans le cadre du présent protocole, les données seront traitées collectivement sans intention de réidentifier un individu particulier. À ce titre, le traitement des données dans le CASD empêche toute sortie de résultats contrevenant au secret statistique.

Les missions de Santé publique France incluent la production d'indicateurs de surveillance, l'amélioration des connaissances sur l'état de santé de la population, les comportements favorables ou non pour la santé, les risques pour la santé afin de proposer des recommandations pour la prévention et la promotion de la santé notamment par une entrée par population spécifique, ici les personnes sans emploi. La surveillance de la santé des personnes sans emploi s'inscrit dans ses missions de réduction des Inégalités sociales de santé et dans la Stratégie nationale santé 2018-2022. Une des limites des autres études comme les Baromètres de Santé publique France, l'enquête S.I.P. (Santé Itinéraire Professionnel) ou les cohortes Coset MSA et Indépendants, de Constantes concerne la taille insuffisante des échantillons disponibles en France qui ne permet pas les analyses détaillées par catégories de personnes sans emploi. Il serait nécessaire de mettre en place des études spécifiques de type S.I.P. pour disposer d'un nombre suffisant (10 000 participants) de personnes décrivant leur situation vis-à-vis de l'emploi et leur santé déclarée ou bien d'exploiter des études en population générale dont une caractéristique vis-à-vis de l'emploi est d'être « sans emploi, au chômage ». Ces études souffrent malheureusement de limites en matière d'effectifs, réduisant ainsi leur puissance statistique pour mener des analyses par catégorie et identifier des leviers d'action.

Après un travail d'interview auprès d'experts sur le sujet et d'inventaire des bases de données et des enquêtes existantes en France, nous avons identifié en priorité l'EDP-Santé comme une des sources de données existantes pouvant répondre à nos objectifs. Ce projet apportera les connaissances sur cette population hétérogène au niveau national mais également à un niveau géographique plus fin, à partir d'un échantillon représentatif par âge et sexe de la population française. Ce projet produira par ailleurs des indicateurs à destination de la population et des décideurs publics qui n'existent pas actuellement. Une comparaison des résultats avec les données du Baromètre de Santé publique France en population générale est également prévue.

1.5 Publication des résultats et valorisation

Les résultats de l'étude ont pour but de contribuer à guider l'action des pouvoirs publics en identifiant les populations les plus à risque, pour lesquelles des interventions de prévention pourraient être déployées (projet transversal de Santé publique France en lien avec la Direction de la prévention-promotion de la santé).

Conformément aux recommandations, les résultats de l'étude seront rendus publics et seront présentés aux ministères concernés et autres instances publiques. Ils pourront également être soumis sous forme d'articles à comité de lecture pour publication et par ailleurs publiés sur le site de Santé publique France sous différents formats possibles (rapports institutionnels, synthèses, articles dans le *Bulletin épidémiologique hebdomadaire* ou dans la revue *La Santé en action*). Ces résultats pourront faire l'objet de présentation à des colloques et des congrès. Les principaux indicateurs pourront également être mis à disposition via le site open data de Santé publique France. Pour répondre à ses missions de surveillance, la production régulière d'indicateurs dans le temps est envisagée à travers des demandes régulières d'accès.

2. OBJECTIFS DE L'ÉTUDE BASÉE SUR L'EDP SANTÉ

Afin de repérer les personnes sans emploi les plus à risque et de proposer des mesures de prévention associées, les objectifs principaux sont :

- de décrire les différents profils sociodémographiques et l'état de santé des « personnes sans emploi » en France en les comparant à d'autres populations ; il s'agit de décrire leurs caractéristiques individuelles, leurs caractéristiques socio-éco-culturelles, leurs conditions de vie, leurs emplois récents précédents s'ils existent, leur protection sociale, etc. par catégorie d'âge, zone géographique et sexe.
- de produire des indicateurs de mortalité globale et par cause pour les « personnes sans emploi » et les comparer à ceux de la population générale, par catégories d'âge, par zone géographique et par sexe.
- d'étudier les risques relatifs de décès par cause selon le fait d'avoir déjà travaillé ou non au cours de sa vie, d'avoir eu une ou plusieurs périodes de chômage dans les années récentes, en tenant compte d'autres facteurs sociodémographiques et économiques, de conditions de vie par catégorie d'âge, zone géographique et sexe.

Si les effectifs le permettent, un focus sera réalisé sur la santé mentale et notamment le suicide par les calculs des risques relatifs de décès par suicide chez les personnes sans emploi, qui seront calculés en les ajustant sur les principaux facteurs de risque et de confusion disponibles.

- de produire des indicateurs de morbidité pour les pathologies pour lesquelles des algorithmes SNDS ont été développés et validés : diabète, maladies cardiovasculaires, santé mentale par catégorie d'âge, zone géographique et sexe.

Les objectifs secondaires sont :

- de décrire la consommation et le recours aux soins chez les hommes et les femmes selon la catégorie de personnes sans emploi/actifs en emploi par catégorie d'âge, zone géographique : consommations de soins annuels en fonction de sa couverture sociale et de ses revenus et du milieu urbain ou rural, de l'APL (Accessibilité potentielle localisée).

Un focus sera réalisé sur la santé mentale : à partir des algorithmes en développement par Santé publique France ; et/ou à partir de la cartographie des pathologies³ développée par la Cnam. Les fréquences de prise en charge pour des troubles addictifs, des traitements chroniques par antidépresseurs ou régulateurs de l'humeur, hypnotiques, neuroleptiques, anxiolytiques seront décrits.

- d'identifier les déterminants du recours aux soins global chez les femmes et les hommes sans emploi et par pathologie, catégorie d'âge et zone géographique.

Les résultats obtenus et le suivi de l'évolution dans le temps des indicateurs épidémiologiques doivent contribuer à identifier les catégories de personnes sans emploi à risque, pour leur proposer des mesures de suivi et de prévention spécifiques.

³ Outil d'analyse des pathologies et des dépenses associées. Elle couvre une cinquantaine de pathologies, traitements chroniques et épisodes de soins - <https://www.assurance-maladie.ameli.fr/etudes-et-donnees/par-theme/pathologies/cartographie-assurance-maladie>

3. MÉTHODOLOGIE DE L'ÉTUDE BASÉE SUR L'EDP SANTE

3.1 Conception de l'étude

L'étude correspond à une étude de cohorte historique avec une analyse longitudinale sur une période comprise entre le 1er janvier 2008 et le 31 décembre 2018 ou la dernière date de disponibilité des données (DDD) possible au moment de la mise à disposition. Les données de l'EDP-Santé, issues de l'appariement de l'EDP et du SNDS, permettront de produire les indicateurs de surveillance de l'état de santé des personnes sans emploi.

Le programme de surveillance doit se reposer sur un échantillon de la population suffisamment important pour permettre l'étude de facteurs de risque de pathologies ou de décès en population.

L'échantillon d'étude sera constitué, de l'ensemble des individus de l'EDP-Santé de 16 à 65 ans ; issus des 3,1 millions de personnes EDP vivantes en 2008 ou nées après, parmi lesquels un pourcentage de personnes aura connu ou non une période sans emploi.

3.2 Sources de données

3.2.1 Données de l'Échantillon démographique permanent (EDP)

L'EDP est un panel sociodémographique, créé en 1968 par l'Insee, afin d'étudier les parcours professionnels, résidentiels ou familiaux des personnes résidant en France. Il contient les informations de sources administratives et enquêtes pour des individus nés certains jours de l'année (4 jours à la création de l'échantillon en 1968 élargis en 2004 à 16 jours par an, soit 4,4 % de la population après l'an 2000). C'est à partir du NIR (Numéro d'inscription au répertoire national d'identification des personnes physiques) non restitué dans la base étude de l'EDP que l'appariement des cinq sources suivantes est réalisé⁴:

- *Les données des bulletins d'état civil de naissance, de mariage, de décès*

Données d'état civil. Une information sur l'occupation des personnes vis-à-vis de l'emploi est disponible dans des bulletins de l'état civil sous forme de « sans emploi » ou « sans profession », pour les parents dans les bulletins de naissance d'enfant, et les époux dans les bulletins de mariage, lorsque les personnes n'ont jamais travaillé.

- *Les informations issues du panel d'actifs « tous salariés » :*

Les données concernant les salaires, durée de paie, catégorie socioprofessionnelle, secteur d'activité etc. permettent d'identifier l'absence d'activité salariée au cours d'une période donnée. La notion de « chômage indemnisé », **est disponible à partir de 2008** dans les données sociales du Panel d'actifs.

- *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)*

Les données socio-fiscales permettent à partir de 2011 (données de 2010) une approche précise de la situation sociale des individus ainsi que de la situation vis-à-vis de l'emploi. Les informations qui nous intéressent sont issues du volet traitements, salaires, prime pour l'emploi, pensions et

⁴ Insee. Document de Travail N°F1406 Manuel utilisateur de l'EDP 2016-2018 - <https://www.insee.fr/fr/metadonnees/source/s1166>

rentes de la déclaration de revenu pour autres revenus imposables de préretraités ou chômage, puis demandeur d'emploi, demandeur d'emploi de plus d'un an.

- *Les données des enquêtes annuelles de recensement (EAR)*

Les données des enquêtes annuelles de recensement : l'information disponible sur le « chômage » indemnisé ou non avec la notion de recherche ou non d'emploi et de son ancienneté apparaît dans les données des EAR. Les données issues des EAR depuis 2004 contiennent des informations pour l'étude sur :

- Le type d'activité au moment du recensement (inactif, **chômeur**, retraité, actif occupé) et les conditions d'emploi (**recherche ou non, ancienneté de recherche, temps partiel, chômage indemnisé**) ;
- Il faudra pouvoir distinguer les pré-retraites des chômeurs par une variable ou un ensemble d'autres variables (TYPACT de EAR) ;
- Les données sensibles de type pays de naissance de l'individu seront demandées afin d'étudier l'origine comme un déterminant ou une variable d'ajustement pour l'étude des inégalités sociales de santé dans cette population.

- *Les données du fichier électoral*

Les données du fichier électoral ne seront pas utilisées.

3.2.2 Données de l'EDP-Santé

L'EDP-Santé a été constitué par l'appariement de l'EDP dans sa version élargie avec les données du SNDS (remboursements des soins en ville et en établissements de santé, actes et diagnostics des séjours d'hospitalisation du PMSI et causes médicales de décès).

Cet appariement à l'échelon individuel a été réalisé en 2019 par la Drees en collaboration avec l'Insee, grâce au NIR conservé dans la base étude de l'Insee (23).

L'échantillon constitué, baptisé EDP-Santé, a pour champ les individus encore vivants en 2008 ou nés ultérieurement. Ainsi, 3,3 millions d'individus EDP ont été appariés au SNDS (dont 3,1 millions présents dans les sources fiscales). Il couvre à ce jour la période 1er janvier 2008 – 31 décembre 2018.

L'autorisation de mise en œuvre comportait un appariement et exploitation pour cinq années par la Drees. Cet échantillon est exploité par la Drees pour la production des indicateurs de la Loi de santé.

3.2.3 Données de santé du SNDS dans l'EDP-Santé

En plus des données de l'EDP, l'EDP-Santé contient les données du SNDS.

Créé en 1999 pour évaluer les dépenses et soins de santé, le Système national d'information inter-régimes de l'assurance maladie (Sniiram), est un ensemble de bases de données informationnelles des régimes d'assurance maladie français qui n'a cessé de croître et d'évoluer pour devenir, en 2016, le SNDS, qui par ailleurs a vocation à évoluer et à intégrer progressivement de nouvelles informations dans les années à venir (données sociales et médicales et complémentaire santé) (24). Aucune donnée n'est disponible concernant la situation professionnelle.

Les données SNDS disponibles sont les suivantes :

1. Depuis 2008, les données de l'Assurance maladie (consommations de soins en ville et en établissement remontées dans le Sniiram) ;
2. Depuis 2008, les données hospitalières du Programme de médicalisation des systèmes d'information ; Médecine Chirurgie Obstétrique, Hospitalisation à domicile, Soins de suite et réadaptation, RIM-P
3. Depuis 2006, les données sur les causes médicales de décès du CépiDC-Inserm
4. Depuis 2012, les tops pathologies proposés par la Cnam qui correspondent à certaines pathologies traitées

Les informations remontées correspondent à l'ensemble des données du SNDS sont de quatre ordres :

- **Le bénéficiaire** (sexe, **mois et année de naissance**, rang de naissance, lieu de résidence, régime, couverture maladie universelle complémentaire, aide à la complémentaire santé) ;
- **Les prestations, dépenses et remboursements** (soins de ville, en établissements de santé, et montants associés) :
 - *Prestations de soins de ville (consultations, visites...),*
 - *Prescriptions de médicaments,*
 - *Actes techniques,*
 - *Prélèvements biologiques,*
 - *Dispositifs médicaux (aides techniques),*
 - *Autres prestations (cures, transports...),*
 - *Indemnités journalières (maladies, accidents du travail et maladies professionnelles, maternité)*
 - *Séjours hospitaliers,*
 - *Soins hospitaliers (hors séances)*
- **L'état de santé du bénéficiaire** : l'historique des affections de longue durée (ALD) renseigne sur certaines pathologies chroniques des bénéficiaires, ainsi que les diagnostics hospitaliers (PMSI-MCO, SSR, RIM-P, HAD, ou la consommation de médicaments ou d'actes très spécifiques à des pathologies).
- **La cartographie des pathologies traitées de la Cnam** :
 - Cette cartographie proposée par la Cnam est composée de variables en oui/non, qui indiquent le repérage d'une prise en charge pour pathologie pour un individu une année donnée. Il existe 56 tops pathologies, regroupés en 13 grands groupes de pathologies : Maladies cardioneuro-vasculaires (dont la maladie coronaire, l'insuffisance cardiaque...) ; Facteurs de risque vasculaire traités ; Diabète ; Cancers ; Maladies neurologiques ou dégénératives ; Maladies psychiatriques ; Traitements psychotropes ; Maladies respiratoires chroniques ; Maladies inflammatoires ou rares ou VIH/Sida ; Maladies du foie ou du pancréas ; Insuffisance rénale chronique terminale traitée ; Maternité, avec ou sans pathologies ; Autres affections de longue durée. Les tops sont calculés chaque année et selon plusieurs versions de l'algorithme d'identification. La première année de disponibilité de la cartographie est 2012.
 - Par ailleurs, certains algorithmes sont validés par des équipes de recherche ou par Santé publique France, par exemple le diabète traité (25, 26), la maladie de Parkinson (27) certaines pathologies cardiovasculaires, le cancer du poumon (validé mais non publié) à partir des codes de la Classification internationale des maladies (CIM), de diagnostics d'hospitalisations ou de reconnaissance en affection de longue durée (ALD), et/ou de prise de médicaments. Plusieurs algorithmes sont en cours de développement concernant des maladies psychiatriques notamment (28, 29).
 - Les tables et variables issues de l'EDP et du SNDS utiles sont détaillées en Annexe 2.

3.3 Description et justification de la population d'étude

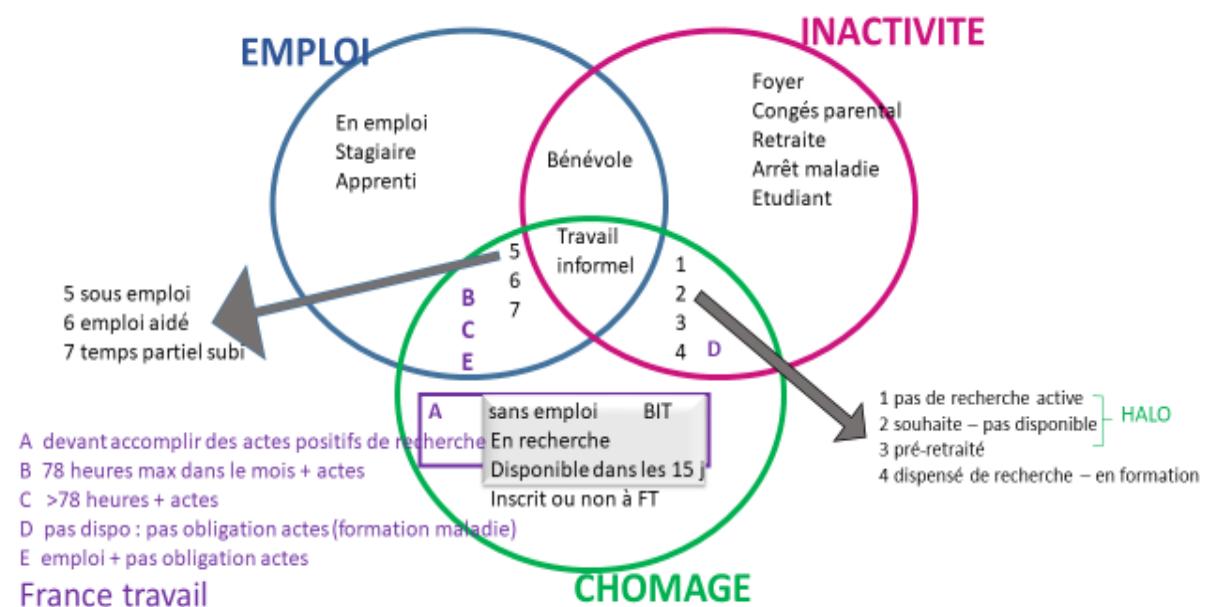
Le ciblage de la population répond en premier lieu aux individus entrant dans le champ des échantillons de population concernés. Pour l'EDP-Santé, il s'agit des individus vivants en 2008 ou nés après.

La population d'étude sera constituée de l'ensemble des individus EDP présents en 2008 susceptibles d'être au chômage ou de l'avoir été.

Parmi eux, un pourcentage sera inactif, actif avec ou sans emploi.

3.3.1 Les personnes sans emploi

Le sujet des définitions de qualification autour de l'emploi est complexe ; elles varient dans le temps et selon les objectifs, institutions, sources comme le montre le schéma ci-dessous.



Surveillance épidémiologique de la Santé des Personnes Sans Emploi – Santé publique France 2023

Plusieurs sources de données pour définir les personnes sans emploi sont disponibles dont les principales incluent les données socio-fiscales, les recensements exhaustifs, les EAR, le Panel d'actifs (cf. sources de données p. 13). La qualité des données des différentes sources n'est a priori pas connue tant que les premiers descriptifs n'auront pas été réalisés. À l'issue du premier tri à plat de données, des tests de cohérence et mesures des données manquantes (cf. traitement et analyses des données) seront réalisés. La définition d'être sans emploi pourra être redéfinie, reconstruite à partir des différentes variables des sources (cf. Annexe 1).

Les données de recensement seront utilisées pour documenter les écarts entre la notion de chômage ou recherche d'emploi déclarés et l'information sur le chômage indemnisé identifié par les données d'origine administrative (Panel et données fiscales).

Les données du Panel d'actifs seront utilisées en première intention pour définir les personnes sans emploi (et en miroir les actifs en emploi). En effet, nous disposerons de notion de chômage indemnisé et d'emploi salarié une année donnée, à partir de 2008 et jusqu'à la date de dernière année disponible, permettant de savoir si une année donnée la personne était en chômage exclusif ou non. Cela permet de reconstituer le parcours de l'individu par rapport à sa situation de chômage sur la période d'étude. Selon la qualité des dates de début et de fin issues du panel, et grâce aux

informations supplémentaires issues des données socio-fiscales de 2010, il sera possible de préciser la situation de l'individu à un moment donné et en cumulé.

3.3.2 Différentes variables pourront être construites

1. Chômage indemnisé exclusif (sans aucun emploi salarié) ou non l'année d'observation;
2. Chômage indemnisé exclusif ou non au cours des un/deux ans précédent l'année d'observation ;
3. Nombre cumulé d'années de chômage indemnisé exclusif ou non entre 2008 et année X.

3.3.3 Trajectoire de chômage

En fonction de la précision apportée par les données issues des différentes sources, une ou plusieurs variables caractérisant des types de trajectoires de chômage sur la période d'étude pourront être créées selon l'enchaînement des périodes de chômage indemnisé, des périodes de travail salarié et les autres périodes que pourront documenter les données socio-fiscales ou les données des recensements. On peut en effet faire l'hypothèse que deux années de chômage consécutives par exemple n'ont pas les mêmes effets sur la santé qu'une année de chômage alternée par une année sans chômage.

Les fichiers de variables nécessaires selon les sources sont détaillés en Annexe 1.

3.4 Population de l'étude

Parmi les 3,1 millions d'individus EDP présents dans les données fiscales, 99,5 % sont par ailleurs identifiés dans le SNDS, offrant ainsi une très bonne couverture en matière de données de santé.

3.4.1 Période d'étude

La période d'observation devrait démarrer en 2010 avec une profondeur des données jusqu'en 2008 (surveillance dynamique dans les prochaines années avec des extractions régulières de l'EDP/EDP-Santé) sauf pour l'étude des causes de décès où elle se fera entre 2014 et la dernière date disponible (du fait de la qualité des données des causes > 80% après 2014).

3.4.2 Nombre attendu

D'après l'Insee, pour une année donnée, les actifs représentent environ 74 % de la population. Parmi les 3,1 millions d'individus de l'échantillon EDP-Santé, on peut s'attendre à 2 294 000 actifs dont 2 064 600 d'actifs en emploi et 229 400 actifs sans emploi (taux de chômage 10 %) chaque année environ en transversal. Entre 2008 et la dernière date disponible, on observera des actifs sans emploi qui ont retrouvé un emploi et ceux qui l'ont perdu entre 2008 et la dernière date disponible, et des personnes dont la situation n'a pas changé. Il est difficile a priori d'estimer précisément les effectifs selon la qualification retenue *in fine* autour de la notion de sans-emploi. Néanmoins, nous pouvons affirmer que les effectifs seront suffisants afin d'assurer une surveillance spécifique de cette catégorie de la population.

3.5 Traitements et analyses des données

Une première description des variables brutes sera réalisée avant les analyses proprement dites afin de mesurer la qualité globale des données, le pourcentage de données manquantes selon les sources et pour les variables utiles pour nos analyses.

1. Le tri à plat des variables par année : tri à plat des variables selon les domaines et notamment sur les variables permettant de caractériser le non-emploi et le chômage.
2. Des tests de cohérence intra et inter sources seront effectués sur : âge, sexe, statut marital, nombre d'enfants ; la situation vis-à-vis de l'emploi selon les sources pour repérer les personnes sans emploi.

Les données incohérentes seront exclues des analyses et des méthodes d'imputation pourront être utilisées si les données manquantes sont trop importantes pour des variables importantes des analyses (30).

Les logiciels statistiques utilisés sont SAS et R et leur utilisation se fera au regard des expériences des personnes intervenant dans la gestion des bases de données et pour les analyses. Les analyses se feront via le CASD.

3.6 Analyses descriptives

Les tests de comparaison de moyenne, de pourcentage pourront être utilisés pour comparer les données entre PSE et des PSE par rapport aux actifs en emploi par l'utilisation du Chi2 de Pearson ou les tests de Student.

Les analyses transversales décrivant les profils sociodémographiques et les caractéristiques de santé ou de consommation de soins seront réalisées en comparant les populations de chômeurs aux actifs en emploi à différents moments de la période d'étude afin de décrire l'évolution de ces caractéristiques.

Pour étudier le profil et les caractéristiques des personnes sans emploi en France entre 2008 et la dernière date de disponibilité, le profil des catégories de personnes sans emploi par groupe d'âge, zone géographique et par sexe, en termes de fréquence, prévalence selon les caractéristiques individuelles, caractéristiques socio-éco-culturelles, de conditions de vie et leur protection sociale sera produit.

Pour la mortalité globale des personnes sans emploi de 2008 à la dernière date de disponibilité des taux de référence Insee dans la population, la mortalité des personnes sans emploi, une année donnée, sera dans un premier temps comparée à celle de la population générale en France mise à disposition par l'Insee⁵, par le calcul de ratios standardisés de mortalité (SMR) par âge et sexe (pertinence du fait de l'effectif). Les ratios standardisés de mortalité peuvent être calculés sur des périodes de cinq ans correspondant au nombre de sujets observés divisé par le nombre de sujets attendus si la mortalité était celle dans la population française soit 2008-2012 et 2013-2017. Leurs intervalles de confiance seront calculés.

Le taux brut de mortalité est le rapport du nombre de décès de l'année pour 100 000 personnes-années.

Les taux de mortalité bruts et standardisés par âge, par région, par département pourront être calculés chez les hommes et les femmes.

⁵ :<https://www.insee.fr/fr/statistiques/2383440>

Pour la mortalité par causes des personnes sans emploi de 2014 (année à partir de laquelle la qualité des données sur les causes est > 80 %) à la dernière date de disponibilité, les taux de mortalité par causes de décès seront calculés chez les personnes sans emploi, calculés par période et par classe d'âge lorsque les effectifs le permettent.

Pour étudier la morbidité et le recours aux soins des personnes sans emploi de 2008 à la dernière date de disponibilité, le recours aux soins chez les hommes et les femmes sans emploi sera décrit par année par rapport aux actifs en emploi : pourcentage de consommations de soins annuels en fonction de sa couverture sociale et de ses revenus et du milieu urbain ou rural, de l'APL (accessibilité potentielle localisée), etc.

3.7 Modélisation

À l'issue des analyses descriptives de la population des chômeurs, selon les pathologies repérables via le SNDS, la fréquence des pathologies, et les hypothèses soulevées, les liens entre chômage et survenue d'événements seront modélisés.

Pour étudier les liens entre la ou les catégories de personnes sans emploi et la survenue du décès ou d'un événement de santé, un modèle de survie avec des variables dépendantes du temps pour le chômage et la survenue d'événements de santé comme le décès sera utilisé.

3.7.1 L'expérience du chômage

Les analyses longitudinales seront stratifiées par sexe et ajustées sur l'âge et permettront de respecter la relation temporelle entre l'événement étudié et les facteurs étudiés. Les facteurs étudiés comme possibles déterminants de l'état de santé seront pris en compte par la construction de variables dépendantes du temps calculées à chaque année d'observation.

Dans le cas de l'analyse en longitudinal, à partir d'un t0 en 2010, il s'agira de caractériser les individus (personnes-années) selon leur exposition au chômage exclusif ou non (en instantané et en cumulé) avant la survenue des événements de santé. Les expositions dont on veut étudier l'effet sont classiquement des variables fixes, définies à la date d'origine du suivi (t0). Elles concernent les caractéristiques intrinsèques des sujets (par exemple, le sexe), ou les expositions présentes à t0 (par exemple, le chômage). Au cas présent, il s'agit d'étudier l'effet d'une exposition susceptible de survenir ou de varier après t0 telle que la situation vis-à-vis de l'emploi. Il existe une méthode permettant de prendre en compte des variables dépendantes du temps « à changements multiples » comme le fait d'être au chômage qui pourrait évoluer durant la période. Les variables dépendantes du temps à changements multiples sont relatives à des expositions qui varient dans le temps, ici le chômage. Leur codage pourra être quantitatif, qualitatif à plusieurs classes, ou binaire (31, 32).

Le modèle de Cox permet de prendre en compte les changements temporels dans les covariables explicatives offrant ainsi une approche pour étudier la dynamique des événements au fil du temps (31, 32). Le modèle de Cox sera utilisé en priorité et d'autres modèles à effets aléatoires ou fixes pourront également être utilisés (33).

Pour l'étude des liens entre le chômage et les événements de santé, on modélisera à partir de 2010, la survenue de la maladie en lien avec le chômage chez les personnes indemnes de la pathologie dans les deux ans précédents.

De même, pour étudier le chômage consécutif à l'état de santé, l'EDP-Santé permet d'identifier des personnes sans notion de chômage au cours des deux ans précédents et de décrire les trajectoires des personnes après survenue d'une pathologie donnée.

Il existe également des méthodes telles que les modèles structurels marginaux qui tiennent compte de la « causalité inverse » et des variations temporelles décrites par Robins *et al.* que nous pourrions utiliser (34).

3.7.2 Pathologies étudiées

Les pathologies qui seront étudiées sont celles repérables dans le SNDS, suffisamment fréquentes et dont l'histoire de la maladie est compatible avec des délais de survenue après l'exposition au chômage suffisante.

Les effets du chômage sur la santé mentale peuvent s'installer dès la perte d'emploi, se prolonger et s'aggraver après plusieurs années (9). Pour les maladies cardiovasculaires (MCV), on pourra étudier la survenue de ces pathologies après plusieurs années de suivi (cinq, sept, dix ans).

Pour les cancers, du fait des temps de latence, la période n'est probablement pas suffisamment étendue pour étudier le rôle du chômage dans la survenue de cancers. En revanche, sous réserve de disposer des effectifs suffisants, l'EDP-Santé pourrait permettre d'étudier la perte d'emploi et le retour à l'emploi chez des personnes atteintes de cancer (35, 36).

3.8 Limites de l'étude

S'agissant d'une population qui n'a jamais été explorée dans l'EDP-Santé, les études sont d'abord exploratoires.

Les limites sont d'abord celles de l'utilisation d'une variété de bases de données complexes réunissant plusieurs sources de données à la temporalité différente. Les limites des sources ont été explorées en détail dans le rapport de la Drees⁶.

3.8.1 Limites des données de l'EDP

Certaines personnes ne sont pas couvertes par les analyses telles que les personnes non-résidentes en France, décédées à l'étranger dont les données de décès ne sont pas remontées en France et les personnes nées à l'étranger (caractérisées par un sous-enregistrement des décès lié à des mobilités résidentielles de retour) (37). Ils seront peu nombreux.

3.8.2 Limites des données du SNDS

Les limites sont celles des bases de données médico-administratives connues (erreurs de codage, problème de chaînage de données, faux positifs et faux négatifs, biais de classification). Le chaînage entre les entrepôts de données SNDS s'est beaucoup amélioré ces dernières années (38).

On ne pourra pas analyser les personnes qui ne sont pas identifiées au SNDS du fait d'un problème de NIR ou de non-consommation de soins une année donnée. À noter, suite au travail de la Drees lors de l'appariement, la surreprésentation des personnes âgées.

Une fois exclues certaines personnes, du fait d'incohérences sur les données ou de périodes complètes de données manquantes, nous vérifierons que l'échantillon d'analyse est bien représentatif de la population source par âge, sexe, région.

3.8.3 Données de comportements de santé

Les données des bases collectées dans un objectif économique, administratif et/ou de gestion, ne comportent pas de données individuelles sur les facteurs de risque comme notamment la consommation de tabac, d'alcool, de cannabis, l'activité physique, etc. qui peuvent être des facteurs de risques ou de confusion dans le lien entre chômage et pathologie. Lorsque l'on décrit la mortalité

⁶ <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/edp-sante-enrichissement-de-lechantillon-demographique-permanent-par>

et les causes de décès cette limite n'est pas bloquante. Pour les objectifs analytiques cherchant à mieux caractériser le rôle du chômage sur la santé, il existe des méthodes pour prendre en compte les fréquences de comportement de santé (tabagisme, alcool, drogues, etc.) issues d'autres sources de données comme le Baromètre de Santé publique France (39-48). Des méthodes ont été décrites pour en tenir compte en ajustant les résultats sur des facteurs de confusion non mesurés (49-51).

4. RÔLE DES DIFFÉRENTS ACTEURS ET COMITOLOGIE

Les analyses sont conduites en accord avec la stratégie de la Direction Santé Environnement Travail de Santé publique France dans le cadre de ses missions de surveillance de la santé en lien avec le travail et de l'unité qualité des milieux de vie et du travail et santé des populations.

Santé publique France a la charge de :

- La conception de l'étude ;
- Le financement de l'étude ;
- La réalisation des démarches réglementaires ;
- La coordination des différents acteurs ;
- L'application de l'information collective ;
- L'analyse statistique des données ;
- L'interprétation des résultats ;
- L'écriture de rapport d'étude pour chaque période de suivi ;
- La diffusion des résultats et leur communication.

L'Insee met à disposition les données utiles au projet et leur documentation. La mise à disposition de ces données est encadrée par le Comité du secret statistique (CSS)⁷. Les résultats de l'exploitation feront par ailleurs l'objet d'échanges avec les chercheurs exploitant également ces données dans le cadre du groupe utilisateurs animé par la direction de la démographie de l'Insee.

La Drees a réalisé l'appariement entre l'EDP et les données du SNDS. La Drees a été sollicitée dès le démarrage du projet d'exploitation de l'EDP-Santé, afin de définir les modalités d'accès aux données de l'échantillon apparié, à la fois sur le plan réglementaire et technique. Cette mise à disposition des données est encadrée par une convention non financière conclue entre la Drees et Santé publique France.

Le Centre d'accès sécurisé aux données (CASD) est un groupement d'intérêt public proposant, sous forme de prestations techniques, une technologie de bulles sécurisées destinée à sécuriser le traitement sécurisé des données sensibles pour la recherche scientifique. Un espace de travail mis à disposition de l'utilisateur par le CASD, lui permet de travailler à distance sur les données tout en garantissant au producteur que les données ne puissent être extraites.

Aucun comité n'est associé à ce projet à ce jour. Des experts internes à Santé publique France ou extérieurs pourront être mobilisés sur des sujets ou des problèmes spécifiques, avec le cas échéant, si cela est nécessaire, la constitution d'un comité d'appui thématique.

⁷ <https://www.comite-du-secret.fr/>

5. DÉMARCHES RÉGLEMENTAIRES

Un certain nombre de démarches réglementaires a été nécessaire en lien avec la DPO (déléguée à la protection des données) de Santé publique France pour obtenir les autorisations d'exploiter les données nécessaires à ce projet, lequel implique de concilier les exigences en matière de secret statistique et de traitement de données de santé.

La demande auprès des fournisseurs de données est un préalable au dépôt de dossier au comité du secret statistique (CSS).

Le comité du secret statistique : [Loi n° 51-711 du 7 juin 1951](#)

« Le secret statistique permet d'apporter aux personnes, physiques ou morales, qui fournissent des informations utilisées pour l'établissement de statistiques l'assurance que ces informations ne seront pas utilisées d'une façon susceptible de leur porter tort.

Le secret statistique interdit à l'administration, ou à la personne chargée d'une mission de service public, dépositaire des informations :

- de communiquer à des tiers des renseignements individuels recueillis par voie d'enquête statistique
- de divulguer des informations qui lui ont été transmises par des tiers à des fins exclusives d'établissement de statistiques

Il permet ainsi d'assurer, notamment :

- aux personnes physiques que la confidentialité sur leur vie personnelle et familiale sera garantie
- aux entreprises que le secret commercial sera respecté : les informations transmises ne seront pas mises à la disposition de leurs concurrents ».

Après engagement de confidentialité de toute personne ayant accès aux données et information des producteurs de données, le protocole a été soumis au CSS qui a autorisé l'accès aux données le 30 novembre 2023 (Dossier SCSS N° 4206). Cette autorisation a été complétée par une demande auprès des archives de France pour accéder par dérogation à des archives publiques non librement communicables. Cette dérogation a été accordée le 19 décembre 2023.

Le projet prévoyant également le traitement de données de santé, des formalités préalables complémentaires ont été nécessaires auprès du *Health Data Hub*⁸. S'agissant d'une recherche n'impliquant pas la personne humaine et ne répondant pas à une méthodologie de référence, une demande d'avis a été déposée auprès du Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (Cesrees)⁹ chargé de se prononcer notamment sur la pertinence éthique et la qualité scientifique du projet et sur son caractère d'intérêt public au regard de sa finalité et de la méthodologie mise en œuvre. Après les réponses apportées par Santé publique France aux questions, un avis favorable a été rendu par ce comité le 14 mars 2024 N° 14748182.

En suivant, le dossier accompagné des avis des deux comités (comité du secret et Cesrees) a enfin été soumis à l'examen de la Commission nationale de l'informatique et des libertés¹⁰ (Cnil) (dossier N° 924117), en se plaçant sous le régime de l'autorisation, ce traitement présentant une finalité d'intérêt public. Le traitement a été autorisé le 17 juin 2024 - DR-2024-145 modifiée en date du 28 octobre 2024 (DR 2024-265).

⁸ <https://www.health-data-hub.fr/qui-sommes-nous>

⁹ <https://www.health-data-hub.fr/cesrees>

¹⁰ <https://www.cnil.fr/>

Enfin, une analyse d'impact relative à la protection des données (AIPD) a également été effectuée, rendue nécessaire pour l'instruction de ce type de demande d'autorisation et également par la mobilisation de sources d'origine administrative, par l'usage de données « sensibles », et par le rapprochement entre différentes sources de données. Cette analyse n'a pas révélé de risque élevé pour les droits et libertés des personnes physiques.

6. PROTECTION DE LA VIE PRIVÉE, SÉCURITÉ ET CONFIDENTIALITÉ DES DONNÉES

6.1 Respect des droits des personnes concernées

6.1.1 Information et droits des personnes

Une dérogation à l'obligation d'information individuelle, dans le cadre de l'article 14.5 b du règlement général sur la protection des données (RGPD) est demandée.

En effet, les données appariées par la Drees de l'EDP-Santé sont des données pseudonymisées à l'issue de différentes procédures ne permettant pas de remonter à l'individu pour le contacter. La procédure SAFE mise en place par la Cnam et les différents numéros de pseudonymisation mis en place dans des tables de correspondance séparées ne permettent pas la réidentification des personnes. Ainsi, les personnes concernées ne pourront pas être informées individuellement directement sur ce projet.

De ce fait, une information collective sera réalisée sur le projet et notamment sur les droits garantis par le RGPD sur :

1. le site de la Drees : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/ledp-sante-enrichissement-de-lechantillon-demographique-permanent-par>, dans la rubrique « Autres travaux »
2. le site de Santé publique France www.santepubliquefrance.fr informant sur l'existence de ce projet et décrivant :
 - Les contexte, objectifs, méthodes, résultats attendus
 - Les sources et le lien sur la Drees présentant l'appariement EDP-Santé : <https://drees.solidarites-sante.gouv.fr/>
 - Les droits garantis par le RGPD et la loi Informatique et libertés aux personnes concernées et les modalités d'exercice de ces droits
 - Les coordonnées du :
 - DPO : adresse électronique et postale de contact
 - Équipe projet : adresse électronique et postale de contact
 - Le calendrier prévisionnel de l'étude et la durée de conservation de données via la page internet

6.1.2 Droits d'accès, d'opposition, de rectification des personnes

Les données étant pseudonymisées, la réidentification des personnes est impossible et les droits d'accès, de rectification et d'opposition ne pourront s'exercer qu'à travers les producteurs de données selon leurs modalités : la Drees, la Cnam et l'Insee.

Concernant les sources EDP-Santé : les personnes concernées souhaitant faire valoir leurs droits devront en faire la demande auprès de Santé publique France en s'adressant à l'adresse dpo@santepubliquefrance.fr en indiquant le code EDP-Santé dans leur demande.

Cette procédure figure également dans la note d'information collective publiée sur le site de Santé publique France (cf. Annexe 2).

Après enregistrement et vérification de la demande, la DPO de Santé publique France pourra :

- Concernant les données de l'EDP-Santé, informer la Drees de l'exercice de droit d'une personne en envoyant un mail à l'adresse drees-rgpd@sante.gouv.fr avec le code EDP-Santé.
- Concernant les données provenant de la base de données du SNDS, réorienter vers la procédure détaillée sur le site dédié <https://www.snds.gouv.fr/SNDS/Protection-de-la-donnee> précisant que les personnes concernées souhaitant faire valoir leurs droits devront en faire la demande auprès du directeur de l'organisme gestionnaire du régime d'Assurance maladie obligatoire auquel la personne est rattachée, conformément aux [articles 92 à 95 du décret n° 2005-1309 du 20 octobre 2005](#).
- Concernant les données provenant de la base de données de l'EDP, réorienter vers l'adresse contact-rgpd@insee.fr

Le droit d'opposition prévu aux premier et troisième alinéas de l'article 56 de la loi n° 78-17 du 6 janvier 1978 porte sur l'utilisation des données dans les traitements mentionnés dans l'[article L. 1461-3 du code de la santé publique](#). À noter que ce droit ne s'applique pas aux traitements nécessaires à l'accomplissement des missions des services de l'État, des établissements publics ou des organismes chargés d'une mission de service public compétents.

Les personnes concernées peuvent également saisir la Cnil dans les conditions prévues par la réglementation : <https://www.cnil.fr/fr/ajir>

Pour toute question concernant spécifiquement l'utilisation des données dans le cadre de ce projet, ils pourront s'adresser à une adresse générique qui sera créée à Santé publique France SESSSE@santepubliquefrance.fr en tant que responsable du traitement.

6.2 Confidentialité et sécurité des données

6.2.1 Gestion du risque de réidentification

Les données du SNDS sont pseudonymisées et ne contiennent aucune donnée directement identifiante. L'identifiant patient utilisé pour réaliser l'appariement est dérivé du NIR de manière irréversible par applications successives des différentes fonctions de pseudonymisation FOIN1, FOIN2, FOIN3. Les opérations de pseudonymisation du NIR se déroulent via la procédure SAFE mise en œuvre par l'Insee pour la transmission des informations à la Cnam. Santé publique France n'entre jamais en possession ni des informations sources ni des numéros issus de la procédure FOIN. Les données issues des sources EDP et SNDS seront pseudonymisées par un numéro distinct généré par les producteurs de données pour permettre leur appariement.

Parmi les variables quasi-identifiantes, les suivantes sont généralisées :

- La date de naissance est tronquée pour ne conserver que l'année et le mois ;
- L'adresse de résidence est généralisée au code commune.

Le SNDS ne contient aucune donnée liée à l'affiliation ethnique, politique, religieuse, philosophique ou syndicale, ni de donnée génétique, ni de donnée biométrique, ni de donnée sur la vie ou l'orientation sexuelle.

6.2.2 Support de données et sécurité

Une convention signée entre la Drees et Santé publique France, encadrera la mise à disposition des données de l'EDP-Santé en précisant les conditions de sécurité des données, ainsi que les objectifs et le calendrier.

L'analyse des données aura lieu sur un espace projet dédié, hébergé sur la plateforme du CASD, homologuée au référentiel de sécurité pour les données du SNDS. La SD-Box assure la sécurisation de l'accès aux données via l'infrastructure centrale du CASD et l'environnement de travail de l'utilisateur en intégrant de bout en bout les prérequis les plus exigeants en matière d'homologation du poste d'accès (ISO 27001, ISO 27701, homologation au référentiel de sécurité des données de santé, PSSI-MCAS...).

Les données seront chiffrées avec GnuPG par la Drees et protégées par un mot de passe. Cette procédure garantit la sécurisation de la transmission des données et la protection des données. Puis, les données chiffrées protégées par un mot de passe seront transmises par la Drees au CASD, qui placera les données dans un espace sécurisé du CASD. Les données seront stockées et traitées sur cet espace sécurisé. Les données seront uniquement accessibles aux membres habilités de l'équipe projet.

La demande au Comité du secret a été réalisée selon le calendrier prévu en parallèle des demandes Cnil.

6.2.3 Plateforme CASD

Toutes les analyses se feront sur l'espace projet sécurisé de la plateforme du CASD accessibles aux personnes habilitées seulement. Aucune donnée individuelle ne sera enregistrée en dehors de la plateforme. Seuls les résultats anonymes seront exportés sous forme de tableaux et de graphiques.

Ainsi l'espace dédié au CASD pour Santé publique France respecte le référentiel de sécurité du SNDS et du RGDP et permet la traçabilité des accès :

- Une authentification forte des utilisateurs à base de certificats et de biométrie (empreinte digitale)
- Des points d'accès sécurisés et authentifiés (SD-Box)
- L'accès d'un utilisateur ou d'un point d'accès peut être arrêté à distance
- L'utilisateur n'a accès qu'aux applications autorisées (accès distant)
- Chaque projet dispose d'un espace dédié dont l'accès est réservé aux seuls utilisateurs habilités
- Un contrôle par les gestionnaires du CASD des informations exportées de manière à garantir le respect du secret statistique, et la conservation de toutes les opérations d'export d'informations
- La traçabilité de tous les traitements sur les données sensibles
- L'activité (en débit) d'un point d'accès est surveillée
- Les données stockées sont chiffrées

L'accès à la plateforme se fera par une cheffe projet, une chargée d'études, tous préalablement habilités à accéder aux données sur la plateforme du CASD.

6.2.4 Circuit des données en cas d'appariement

La Figure 2 résume le circuit d'appariement entre l'EDP et le SNDS.

Étape 1. Constitution de la cohorte initiale et transmission du NIR à la Cnam

Dans le cadre de sa base source, le producteur dispose pour chaque personne de son NIR en clair, de sa date de naissance et de son sexe.

À partir de la base source, le producteur constitue la cohorte initiale, comme décrit dans la partie méthodologie au chapitre 3. Selon les critères de ciblage.

Un numéro de cohorte non signifiant (ID_PROJSPF) est attribué de façon aléatoire à chaque personne concernée.

Le producteur transmet à une structure déconcentrée de la Cnam la table de pseudonymisation [ID_PROJSPF ; NIR] constituée du Numéro d'Identification au Répertoire -NIR du patient et/ou de l'ouvrant droit, du sexe, de la date de naissance ainsi qu'un numéro de cohorte non signifiant (ID_PROJSPF) pour toutes les personnes présentes dans la cohorte initiale créé spécifiquement et différent de l'identifiant communément présent dans les données sources. Le producteur devra créer un fichier normé à partir de la table de pseudonymisation [ID_PROJSPF, NIR] et le déposer sur la plateforme SAFE suivant les modalités qui lui seront communiquées ultérieurement.

Étape 2. Extraction des données du SNDS à partir du NIR

Une structure déconcentrée de la Cnam applique aux NIR en clair figurant dans le fichier normé (généré à partir de la table de pseudonymisation [ID_PROJSPF ; NIR]) les mêmes algorithmes de hachage utilisés lors de la consolidation de la base principale du SNDS. À partir de l'identifiant SNDS ainsi obtenu, la Cnam cible les personnes concernées de l'étude dans le SNDS et extrait les données requises pour le projet. Elle attribue un identifiant non signifiant spécifique au projet (ID_SNDS_PROJSPF) et consolide en parallèle une table de correspondance des identifiants pseudonymisés [ID_PROJSPF ; ID_SNDS_PROJSPF] entre l'identifiant temporaire transmis par le producteur (ID_PROJSPF) et l'identifiant présent dans l'extraction SNDS qui sera fournie (ID_SNDS_PROJSPF).

La Cnam transmet ensuite cette table de correspondance des identifiants pseudonymisés [ID_PROJSPF ; ID_SNDS_PROJSPF] ainsi que l'extraction SNDS au gestionnaire de la plateforme cible dédiée au projet. Les données SNDS extraites sont supprimées à l'issue du processus mais la correspondance entre les pseudonymes pourra être conservée par la Cnam pendant la durée d'accès du projet aux données SNDS, notamment si des mises à jour des données SNDS sont nécessaires.

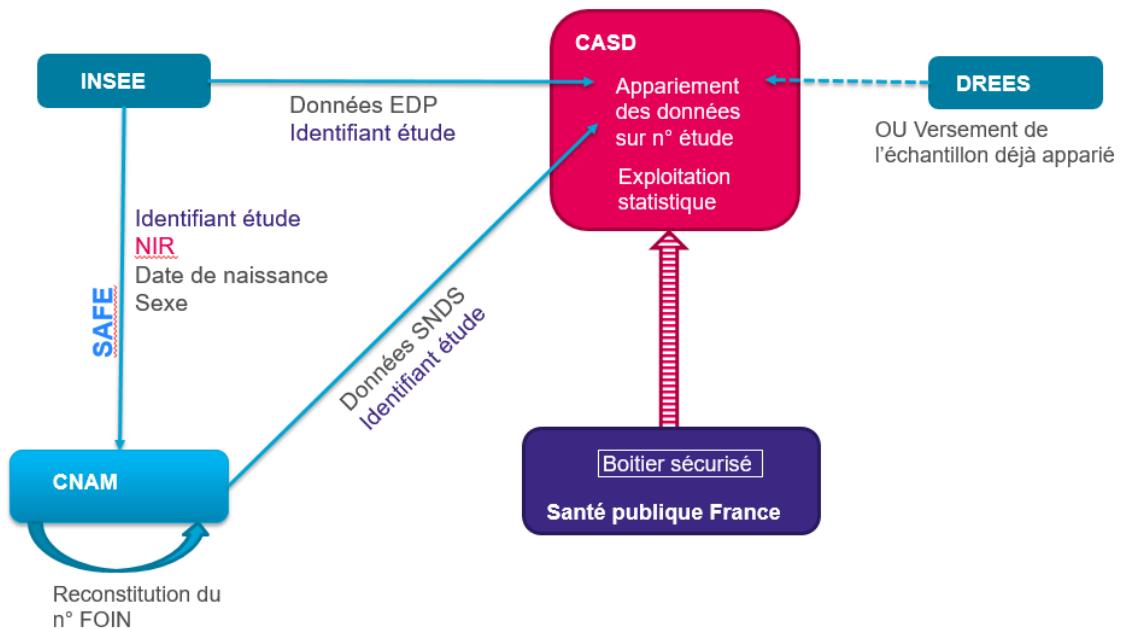
Étape 3. Appariement des jeux de données et traitement dans la plateforme cible

La cohorte initiale pseudonymisée (données EDP) et l'extraction du SNDS avec la table de correspondance des identifiants pseudonymisés [ID_PROJSPF ; ID_SNDS_PROJSPF] sont transmises respectivement par le producteur et la Cnam au gestionnaire de la plateforme cible dédiée pour le projet via des réseaux dédiés, permettant un transfert privé (hors internet) et sécurisé des données.

À partir du numéro de cohorte non signifiant (ID_PROJSPF), les jeux de données sont ensuite appariés dans cette plateforme par l'opérateur de la plateforme et un nouvel identifiant est créé de façon aléatoire pour chaque personne concernée (ID_EDPSNDS_SPF).

Les données sont stockées et traitées dans un espace projet sécurisé de la plateforme cible, accessible aux personnes habilitées seulement. Elles seront conservées pendant le nombre d'années autorisées pour la réalisation de l'étude et la publication des résultats, puis elles seront définitivement supprimées.

Figure 2. Circuit des données SESSE



6.2.5 Durée de conservation des données

Les données seront conservées :

- **En base active :**
Les données seront conservées pour la réalisation du projet, jusqu'au 30 novembre 2027.
- **En base archive :**
Les données seront archivées temporairement et resteront consultables pendant 2 années après la fin de l'exploitation, **jusqu'au 30 novembre 2029** le temps nécessaire à la valorisation complète des travaux. Ensuite, l'accès à l'espace projet via le CASD sera impossible et les données seront supprimées définitivement de l'espace projet.

7. CALENDRIER PRÉVISIONNEL ET FAISABILITÉ DU PROJET

Étapes	Dates
• Préparation des différents éléments, demande préalable aux fournisseurs	Février à Octobre 2023
• Dépôt du dossier au Cesrees	Octobre 2023
• Accord Comité du secret statistique	Novembre 2023
• Accord Cesrees	Mars 2024
• Accord Cnil	Juin 2024
• Convention Drees	Juillet-Octobre 2024*
• Mise à disposition des données	Novembre 2024*
• Mise en œuvre des analyses	Décembre 2024-Juin 2028*
• Publications de rapports/articles	Janvier 2026–Novembre 2029*
• Fin du projet	30 novembre 2029

* prévisionnel

Santé publique France dispose d'une bonne connaissance de l'utilisation des données du SNDS depuis plusieurs années sur plusieurs champs ainsi que de l'EDP (projet Cosmop) et de l'exploitation des données via le CASD. Les équipes de Santé publique France disposent de la formation, de l'expérience et de l'expertise du SNDS et de l'EDP.

La cheffe de projet dispose d'une expérience dans :

- les cohortes et l'analyses de données longitudinales ;
- les analyses de mortalité effectuées à partir des données RNIPP et des causes de décès (Cépidc) ;
- les données SNDS : formation SNDS sur architecture et EGB en 2012 et sur l'accès au portail SNDS ;
- le développement d'un algorithme SNDS sur certains cancers professionnels à partir des extractions SNDS de la cohorte ESPrl à partir des données PMSI-MCO/ALD qui a été validé avec des registres généraux de cancer (52).

8. RÉFÉRENCES BIBLIOGRAPHIQUES

1. Mesrine A. La surmortalité des chômeurs : un effet catalyseur du chômage ? . *Économie et statistique*. 2000;Insee, no 334, :pp. 33-48.
2. Shahidi FV, Parnia A. Unemployment Insurance and Mortality Among the Long-Term Unemployed: A Population-Based Matched-Cohort Study. *Am J Epidemiol*. 2021;190(10):2124-37.
3. Meneton P, Kesse-Guyot E, Mejean C, Fezeu L, Galan P, Hercberg S, et al. Unemployment is associated with high cardiovascular event rate and increased all-cause mortality in middle-aged socially privileged individuals. *Int Arch Occup Environ Health*. 2015;88(6):707-16.
4. Andler R. Une santé dégradée pour les personnes sans emploi. *La santé en action*. 2019;452:43-4.
5. Airagnes G, Lemogne C, Meneton P, Plessz M, Goldberg M, Hoertel N, et al. Alcohol, tobacco and cannabis use are associated with job loss at follow-up: Findings from the CONSTANCES cohort. *PLoS One*. 2019;14(9):e0222361.
6. Moore TH, Kapur N, Hawton K, Richards A, Metcalfe C, Gunnell D. Interventions to reduce the impact of unemployment and economic hardship on mental health in the general population: a systematic review. *Psychol Med*. 2017;47(6):1062-84.
7. Laanani M, Ghosn W, Jouglard E, Rey G. Association entre taux de chômage et suicide, par sexe et classe d'âge, en France métropolitaine, 2000-2010. *Bull Épidémiol Hebd*. 2015.
8. Cohidon C, Santin G, Geoffroy-Perez B, Imbernon E. [Suicide and occupation in France]. *Rev Epidemiol Sante Publique*. 2010;58(2):139-50.
9. Debout M. *le traumatisme du chômage*. Editions de l'Atelier. 2015.
10. Vancea M, Utzet M. How unemployment and precarious employment affect the health of young people: A scoping study on social determinants. *Scand J Public Health*. 2017;45(1):73-84.
11. Norstrom F, Janlert U, Hammarstrom A. Is unemployment in young adulthood related to self-rated health later in life? Results from the Northern Swedish cohort. *BMC Public Health*. 2017;17(1):529.
12. Hammarstrom A, Janlert U. Early unemployment can contribute to adult health problems: results from a longitudinal study of school leavers. *J Epidemiol Community Health*. 2002;56(8):624-30.
13. Lappalainen K, Manninen P, Rasanen K. Association Among Sociodemographic Factors, Work Ability, Health Behavior, and Mental Health Status for Young People After Prolonged Unemployment. *Workplace Health Saf*. 2017;65(2):65-73.
14. Blasco S, Brodaty, T. *Chômage et santé mentale en France*. eco et stat. 2016.
15. Leon C, Chan Chee C, Du Roscoët E, 2017* elgBdSpF. *la dépression en france chez les 18-75 ans : résultats du baromètre santé 2017*. *Bull Épidémiol Hebd* 2017.
16. Léon C, Du Roscoët E, Beck F. prévalence des épisodes dépressifs en france chez les 18-85 ans : résultats du baromètre santé 2021. *Bull Épidémiol Hebd*. 2021.
17. Feral-Piessens AL, Rives-Lange C, Matta J, Rodwin VG, Goldberg M, Juvin P, et al. Forgoing health care under universal health insurance: the case of France. *Int J Public Health*. 2020;65(5):617-25.
18. Brydsten A, Hammarstrom A, San Sebastian M. Health inequalities between employed and unemployed in northern Sweden: a decomposition analysis of social determinants for mental health. *Int J Equity Health*. 2018;17(1):59.
19. Fassin D, Grandjean H, Kaminski M, Lang T, Leclerc A. *Les inégalités sociales de santé*. Paris: La Découverte; 2000. 448 p.
20. Vuori J, Silvonen J, Vinokur AD, Price RH. The Typhon Job Search Program in Finland:

benefits for the unemployed with risk of depression or discouragement. *J Occup Health Psychol.* 2002;7(1):5-19.

21. Vuori J, Price RH, Mutanen P, Malmberg-Heimonen I. Effective group training techniques in job-search training. *J Occup Health Psychol.* 2005;10(3):261-75.
22. Hollederer A, Jahn HJ. Results from a Nationwide Evaluation Study of Labor Market-Integrative Health Promotion for the Unemployed: Impact of the JOBS Program Germany. *Int J Environ Res Public Health.* 2023;20(19).
23. Dubost C-L LA. L'EDP-Santé, un appariement des données socio-économiques de l'échantillon démographique permanent au Système national des données de santé. . *Les dossiers de la Drees* 2020:66.
24. Tuppin P, Rudant J, Constantinou P, Gastaldi-Menager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the systeme national d'information interregimes de l'Assurance Maladie (SNIIRAM) to the systeme national des donnees de sante (SNDS) in France. *Rev Epidemiol Sante Publique.* 2017;65 Suppl 4:S149-S67.
25. Fosse-Edorh S, Rigou A, Morin S, Fezeu L, Mandereau-Bruno L, Fagot-Campagna A. [Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes]. *Rev Epidemiol Sante Publique.* 2017;65 Suppl 4:S168-S73.
26. Quantin C, Roussot A, Besson J, Cottenet J, Bismuth MJ, Cousin F, et al. 2tude des algorithmes de reprérage de la dépression dans le sniiram par le réseau Redsiam. . *Revue française des affaires sociales* 2016.
27. Moisan F, Gourlet V, Mazurie JL, Dupupet JL, Houssinot J, Goldberg M, et al. Prediction model of Parkinson's disease based on antiparkinsonian drug claims. *Am J Epidemiol.* 2011;174(3):354-63.
28. cancer Ind. Algorithme de sélection des hospitalisations liées à la prise en charge du cancer dans les bases nationales d'activité hospitalière de court séjour «algorithme cancer ». 2013.
29. cancer Ind. Algorithme de sélection des hospitalisations liées au cancer en mco /étude de validation. . Boulogne-Billancourt : Collection les Données, INCa, Santé publique France; 2018 44 p 2018.
30. Rubin DB. *Multiple Imputation for Nonresponse in Surveys* Wiley. 1987.
31. Kooperberg C, Clarkson DB. Hazard regression with interval-censored data. *Biometrics.* 1997;53(4):1485-94.
32. Desquillet L, Meyer L. [Time-dependent covariates in the Cox proportional hazards model. Theory and practice]. *Rev Epidemiol Sante Publique.* 2005;53(1):51-68.
33. Junna L, Moustgaard H, Martikainen P. Current Unemployment, Unemployment History and Mental Health: A Fixed Effects Model Approach. *Am J Epidemiol.* 2022.
34. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550-60.
35. de Boer AG, Taskila T, Ojajarvi A, van Dijk FJ, Verbeek JH. Cancer survivors and unemployment: a meta-analysis and meta-regression. *JAMA.* 2009;301(7):753-62.
36. Chimienti M, Morlino G, Ingravalle F, Vinci A, Colarusso E, De Santo C, et al. Unemployment Status Subsequent to Cancer Diagnosis and Therapies: A Systematic Review and Meta-Analysis. *Cancers (Basel).* 2023;15(5).
37. Cambois E, Jusot F. Contribution of lifelong adverse experiences to social health inequalities: findings from a population survey in France. *Eur J Public Health.* 2011;21(5):667-73.
38. Olive F, Gomez F, Schott AM, Remontet L, Bossard N, Mitton N, et al. [Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: a longitudinal approach becomes possible]. *Rev Epidemiol Sante Publique.* 2011;59(1):53-8.

39. Andler R, Guignard R, Nguyen-Thanh V, Pasquereau A, Beck F, Deutsch A, et al. consommation de tabac et utilisation d'e-cigarette en france en 2015 : premiers résultats du baromètre cancer 2015. Bull Épidémiol Hebd 2015.
40. Andler R, Metlaine A, Richard JB, Zeghnoun A, Nguyen-Thanh V, Léger D. sommeil et consommation de substances psychoactives : résultats du baromètre de santé publique france 2017. Bull Épidémiol Hebd 2017.
41. Andler R, Quatremère G, Gautier A, Nguyen-Thanh V, Beck F, 2017* elgBdSpF. consommation d'alcool : part d'adultes dépassant les repères de consommation à moindre risque à partir des données du baromètre de santé publique france 2021. Bull Épidémiol Hebd. 2021.
42. Andler R, Richard JB, Guignard R, Quatremère G, Verrier F, Gane J, et al. baisse de la prévalence du tabagisme quotidien parmi les adultes : résultats du baromètre de santé publique france 2018. Bull Épidémiol Hebd. 2018.
43. Andler R, Richard JB, Guignard R, Quatremère G, Verrier F, Gane J, et al. la consommation d'alcool des adultes en france en 2021, évolutions récentes. Bull Épidémiol Hebd. 2024.
44. Guignard R, Beck F, Wilquin JL, Andler R, Nguyen-Thanh V, Richard JB, et al. la consommation de tabac en france et son évolution : résultats du baromètre santé 2014. Bull Épidémiol Hebd 2014.
45. Guignard R, Nguyen-Thanh V, Andler R, Richard JB, Beck f, Arwidson P. usage de substances psychoactives des chômeurs et des actifs occupés et facteurs associés : une analyse secondaire du baromètre santé 2010. Bull Épidémiol Hebd. 2017.
46. Pasquereau A, Andler R, Guignard R, Soullier N, Gautier A, Richard JB, et al. consommation de tabac parmi les adultes en 2020 : résultats du baromètre de santé publique france. Bull Épidémiol Hebd 2020.
47. Richard JBA, R. ; Cogorda, C. ;Spilka, S. ; Nguyen-Thanh, V. ; et le groupe Baromètre de Santé publique France 2017*. LA CONSOMMATION D'ALCOOL CHEZ LES ADULTES EN FRANCE EN 2017. Bull Épidémiol Hebd 2017.
48. Andler R, Richard JB, Cogorda C, Deschamps V, Escalon H, Nguyen-Thanh V. nouveau repère de consommation d'alcool et usage : résultats du baromètre de santé publique france 2017. Bull Épidémiol Hebd 2017.
49. Kriebel D, Zeka A, Eisen EA, Wegman DH. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. Int J Epidemiol. 2004;33(5):1040-5.
50. Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med. 1988;13(1):105-18.
51. Rothman K GS. Modern Epidemiology Boston: Little Brown,. 1998.
52. Goulard H, Homere J, Maurisset S, Coureau G, Defossez G, d'Almeida T, et al. Validation of an algorithm for identifying incident cancer cases based on long-term illness and diagnosis related group program data from the French National Health Insurance Information System (SNDS). Pharmacoepidemiol Drug Saf. 2024;33(1):e5709.

7. ANNEXES

Annexe 1. Listes des tables et variables nécessaires au projet

Les tables nécessaires au projet ont été listées pour l'EDP et pour le SNDS.

Surveillance épidémiologique de la santé des personnes sans emploi

Figure 1 • Structure des tables de l'EDP

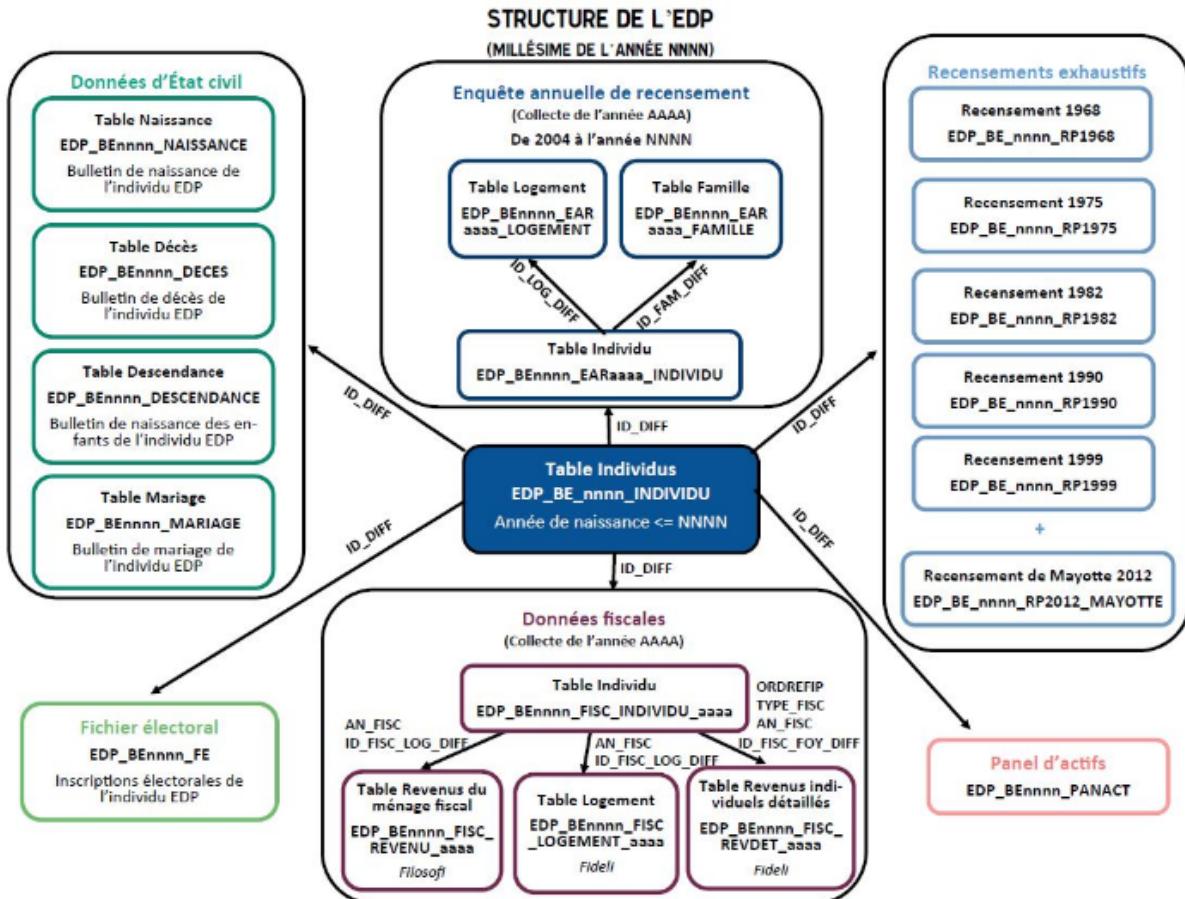
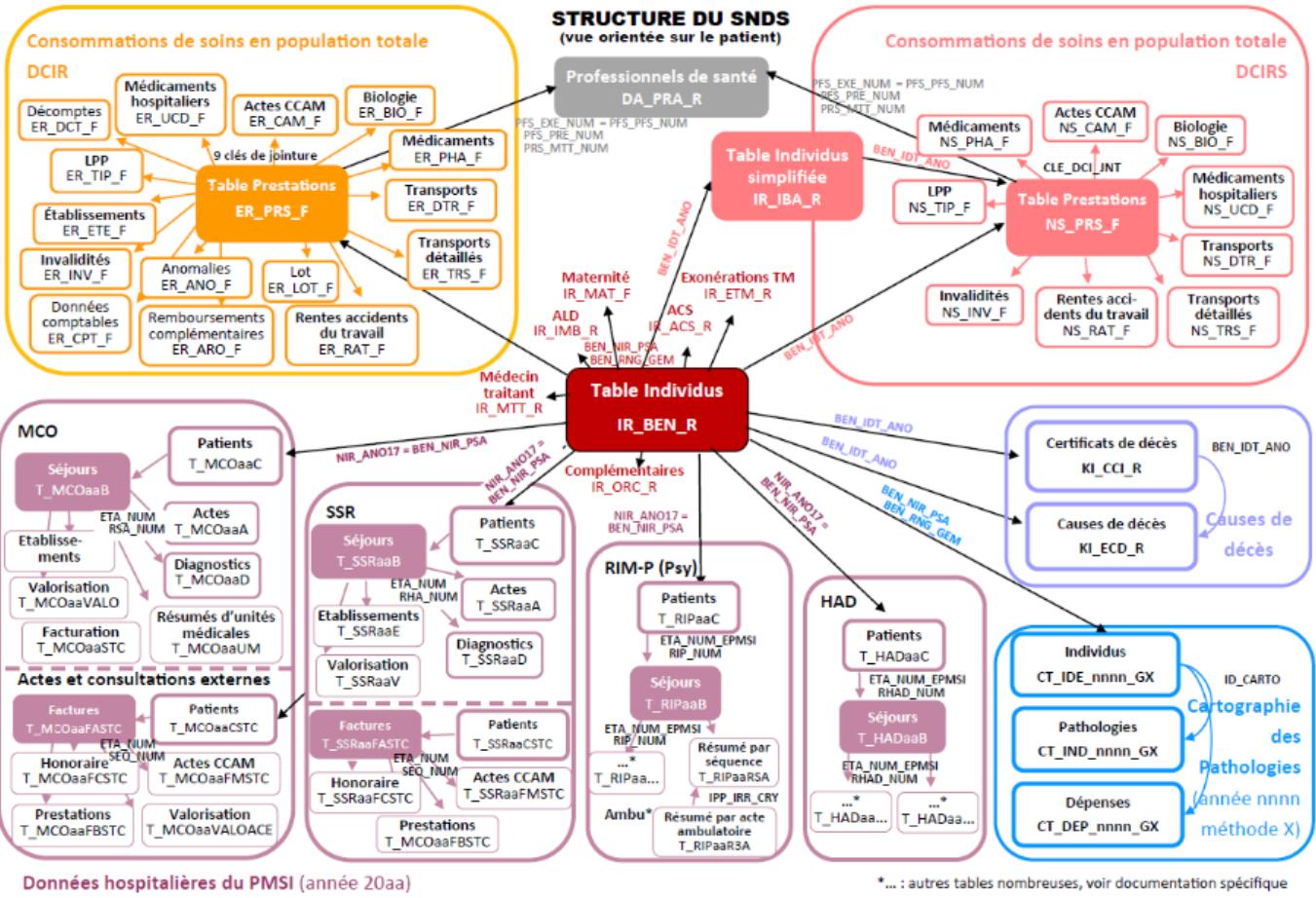


Figure 4 • Schéma de la structure du SNDS



Données résumé

- Table individu

Données fiscales

- Table individu
 - Table revenu ménage
 - Table logement
 - Table revenu individuel

Bulletins des recensements annuels EAR

- Table individu
 - Table famille
 - Table logement

Données état civil

- Table décès
 - Table descendance
 - Table naissance
 - Table mariage

Données du panel d'actifs

- Table panact

Données SNDS

- Table individus
 - Table individus simplifiés
 - Table Prestations DCIR
 - Tables Prestations DCIRS
 - Professionnels de santé
 - PMSI-MCO : séjours, Factures
 - PMSI-SSR : séjours, Factures
 - PMSI-HAD : séjours
 - PMSI-RIMP : séjours
 - Causes de décès
 - Certificats de décès
 - Cartographie des pathologies

Variables principales par source :

Données « résumé de l'individu »

Intitulé des variables principales nécessaires
Identifiant de diffusion de l'individu EDP
Date de naissance de l'individu EDP
Commune et département de naissance de l'individu EDP
Sexe de l'individu EDP
Date du décès de l'individu EDP
Commune et département du décès de l'individu EDP
Nombre d'enfants de l'individu EDP
Nombre de mariages
Nombre de bulletins individuels EAR de l'individu EDP
Existence d'un acte de décès, d'un jugement de disparition ou d'absence
Existence d'un acte de naissance ou d'un acte d'adoption de l'individu EDP
Nombre d'actes de mariage de l'individu EDP
Nombre d'années d'activité de l'individu EDP dans le panel d'actif
Inscription au Répertoire national d'identification des personnes physiques
Nombre de déclarations fiscales de l'individu EDP pour l'année aaaa

Données fiscales

Intitulé des variables principales nécessaires
Identifiants de l'individu, de diffusion du foyer, du logement
Année de déclaration fiscal
Année de perception des revenus
Nombre d'UC (Unité de consommation) du ménage
Niveau de vie du ménage
Centile de niveau de vie
Revenu disponible monétaire du ménage
Type de revenu principal dont chômage/préretraite
Allocation chômage, autres allocations
Retraites et pension
Indicatrices de présence décès, revenus, logement
Variables de pondération ou de redressement
Toutes les Variables liées au logement
Ménage pauvre 50 %
Ménage pauvre 60 %
Toutes les allocations
RSA
Minima sociaux

Bulletins des recensements annuels EAR

Intitulé des variables principales nécessaires
Identifiant de diffusion de l'individu EDP
Tout indicateur de redressement des variables
Marqueur EDP du bulletin = 1
Ancienneté arrivée France
Année de collecte
Année arrivée France
Sexe
Date de naissance
Département de naissance
Indicateur du lieu de naissance (France / Étranger)
Département de naissance ou pays si naissance à l'étranger
Nationalité
Qualité d'immigré
Situation matrimoniale
Commune de résidence chiffrée, redressée
Catégorie de population (logement ordinaire ou autre précisé)
Diplôme
Catégorie socioprofessionnelle et CS antérieure
Condition d'emploi
Activité économique employeur
Mode de vie
Mode de cohabitation
Variables de positions professionnelles
Recherche de travail
Profession antérieure pour chômeurs ou inactifs
Statut professionnel antérieur pour chômeurs ou inactifs
Ancienneté de recherche de travail
Lieu de travail
Temps de travail
Type d'activité

Données d'état civil

Intitulé des variables principales nécessaires
Décès individu EDP :
Identifiant de l'individu EDP
Identifiant de l'événement
Type d'événement
Lieu de naissance de l'individu EDP (code département suivi du code commune)
Lieu de décès de l'individu EDP (code département suivi du code commune)
Domicile de l'individu EDP
Sexe de l'individu EDP
Catégorie socioprofessionnelle de l'individu EDP au moment de l'enregistrement du décès codé selon le libellé de l'acte de décès
Profession :
La profession qui doit être indiquée s'entend non seulement comme celle actuellement exercée par les intéressés, mais également comme celle qu'ils exerçaient en dernier lieu avant de cesser leur activité professionnelle. Il convient, dans cette hypothèse, de préciser « ancien », « en retraite », « honoraire », ou en cas de chômage, de préciser « sans emploi ». Lorsqu'une personne n'a jamais exercé une profession, il convient d'indiquer « sans profession ».
Indicatrice de redressement de la catégorie socioprofessionnelle de l'individu EDP au moment de l'enregistrement du décès selon le libellé de l'acte de décès
Enfant individu EDP :
Identifiant de l'individu EDP
Identifiant de l'événement
Type d'événement
Rôle de l'individu EDP : mère ou père
Date de naissance de l'enfant
Date et lieu de naissance des parents
Lieu de naissance de l'individu EDP (code département suivi du code commune)
Nationalité des parents
Situation matrimoniale des parents
Catégorie socioprofessionnelle du père et de la mère
Indicatrices redressement père ou mère
Mariage individu EDP :
Date de naissance des époux
Lieux de naissance des époux
Nationalité des époux
Catégorie socioprofessionnelle des époux

Données du panel d'actifs

Intitulé des variables principales nécessaires
Identifiant de l'individu EDP
Année d'activité de l'individu EDP
Année d'entrée dans le panel de l'individu EDP
Nombre d'entreprises dans lesquelles l'individu EDP a travaillé durant l'année
Salaire net en euros constants de l'individu EDP durant l'année
Durée de paie en jours, entre 1 et 360, de l'individu EDP durant l'année
Durée de paie convertie en équivalent "temps complet et plein salaire" exprimée en jour
Catégorie socioprofessionnelle de l'individu EDP durant l'année à 1 chiffre
Catégorie socioprofessionnelle de l'individu EDP durant l'année à 2 chiffres
Catégorie socioprofessionnelle de l'individu EDP durant l'année à 1 chiffre jusqu'en 1982 inclus
Catégorie socioprofessionnelle de l'individu EDP durant l'année à 2 chiffres jusqu'en 1982 inclus
Catégorie socioprofessionnelle avec qualification pour la Fonction publique d'État (FPE) uniquement
Code PCS de l'individu EDP durant l'année à 4 chiffres depuis 1993
Catégorie d'agents
Département de résidence de l'individu EDP
Commune de résidence de l'individu EDP depuis 1988
Département de travail de l'individu EDP
Commune de travail de l'individu EDP
Type de contrat de travail depuis 2005
Condition d'emploi de l'individu EDP
Contrat de travail
Nombre d'heures rémunérées de l'individu EDP durant l'année depuis 1993
Durée de paie dans l'année (exprimée en jours)
Durée de paie convertie en équivalent « temps complet et plein salaire » (exprimée en jours)
Domaine d'emploi
Catégorie socioprofessionnelle à 1 chiffre
Indicatrice de poste non annexe

Données SNDS

Cf. liste des tables citées plus haut. Les données correspondent aux données fournies par dans les extractions du SNDS.

Il n'y a pas d'informations sur le statut d'emploi dans le SNDS

Résumé des domaines couverts selon les sources de données			
Domaines	Sources dans l'EDP-Santé	Domaines	Sources dans l'EDP-Santé
Statut vis-à-vis de l'emploi ou du non-emploi		Facteurs socio-éco-culturels et de conditions de vie	
Chômage inscrit ou non à France Travail	EAR	Activités sociales	EAR
Chômage inscrit ou non à France Travail ou formation rémunérée par France Travail	Panel	Défavorisation sociale indice	SNDS
Chômage ou pré-retraite indemnisés	Socio-fiscales	Recours aux soins déclaré ou objectivé	
Recherche un emploi depuis plus d'un an	EAR & socio-fiscales	Consultation dans les 12 mois	SNDS
Recherche un premier emploi	EAR	Psychotropes/Antidépresseur	SNDS
Durée sans emploi/recherche	EAR	Séjours à l'hôpital	SNDS
Souhaite travailler	EAR & socio-fiscales	Lors du Covid-19	SNDS
Secteur et/ou profession catégorie sociale antérieure	EAR & Panel	Prévention/Vaccination	
Déjà travaillé/Dernière profession	EAR	Dépistage des cancers	SNDS
Retraité(e) ou préretraité(e)	EAR & Panel & état civil	Vaccination Covid-19/Grippe	SNDS
Pré-retraite	EAR	Protection sociale	
Formation	Socio-fiscales	Complémentaire santé solidaire	SNDS
Caractéristiques individuelles		CMU/CMU C (Couverture maladie universelle Complémentaire)	SNDS
Âge	EAR & socio fiscales & Panel & état civil	Minima sociaux	EAR
Sexe	EAR & socio fiscales & Panel & état civil	ACS (Allocation complémentaire santé)	SNDS
Statut ménage/foyer	EAR & socio fiscales & état civil	AME (Aide médicale d'État)	SNDS
Statut migratoire/ethno racial	EAR	ARE (Allocation retour emploi)	Socio-fiscales
Région/Département résidence	Recensement & EAR & socio-fiscales & Panel & état civil	RSA (Revenu solidarités active)	Socio-fiscales
Facteurs socio éco, culturels et de conditions de vie		TOTAL alloc par Unité de consommation (UC)	Recensement EAR
Nombre d'enfants	Recensement &EAR & socio fiscales & état civil	Allocation PAJE (Prestation d'accueil du jeune enfant)	Socio-fiscales
Revenus/situation financière	Recensement & EAR	Allocation AAH (Adulte handicapé)	Socio-fiscales
Diplôme/Niveau d'études	Socio-fiscales	Autres : ALD (Affection longue durée)	SNDS
Qualité du logement/véhicule	Recensement & EAR	État de santé	
		Algorithmes, Top Cnam et Décès et causes	SNDS

Annexe 2. Notice d'information collective sur l'utilisation des données concernant la surveillance épidémiologique de la santé des personnes sans emploi (SESSe)

Responsable du traitement : Santé publique France

Contexte

Une santé dégradée peut être un facteur favorisant la perte d'emploi et le fait d'être sans emploi peut s'avérer être un facteur de dégradation de la santé. Les raisons de la surmortalité des chômeurs sont multiples : un état de santé déficient qui provoque le chômage, les conséquences financières et psychologiques du chômage de longue durée, le rôle de l'effet propre du chômage qui accentue ou relève les problèmes de santé. Afin de définir les politiques de prévention adaptées à mettre en place dans cette population, disposer d'indicateurs de surveillance épidémiologique permettant de décrire les populations les plus à risque pour leur santé est nécessaire. Ainsi, les résultats obtenus et l'évolution dans le temps des indicateurs épidémiologiques de morbi-mortalité doivent contribuer à alerter sur les sous-populations à risque.

L'utilisation et le croisement de sources de données historiques existantes, relatives aux facteurs socio-démographiques, professionnels et aux événements de santé, régulièrement mises à jour et disponibles à l'échelle nationale présente de nombreux atouts dans le domaine de la surveillance épidémiologique.

Objectifs

L'objectif principal est d'étudier l'impact sanitaire d'un contexte socio-économique individuel défavorable chez les personnes sans emploi ou au chômage pour identifier les leviers possibles afin de promouvoir les interventions de prévention dans cette population ou dans des sous-groupes identifiés les plus vulnérables.

L'intérêt public de cette étude, sa qualité scientifique et sa pertinence éthique ont été confirmés par un comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (Cesrees), indépendant du responsable de traitement. Il a également reçu un avis favorable du Comité du Secret Statistique. Le projet s'appuie sur un appariement entre les données socio-économiques des individus présents dans l'Échantillon démographique permanent (EDP) de l'Insee et leurs données de soins et d'hospitalisation issues Système National des Données de santé (SNDS). L'échantillon obtenu appelé EDP-Santé a été constitué par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees).

L'EDP-Santé ne contient pas d'information permettant l'identification directe des personnes. Les données sont mises à disposition et utilisées par les personnels habilités de Santé publique France dans un espace sécurisé. Les données seront conservées pendant la durée du projet, jusqu'au 30 novembre 2029.

Finalité et base juridique du traitement

Dans le cadre de cette recherche, un traitement informatique de vos données personnelles va être mis en œuvre pour répondre à ces objectifs. Le responsable du traitement des données, qui est le gestionnaire de l'étude, est Santé publique France dont les coordonnées figurent en dernière page de ce document.

Le traitement de ces données, qui a pour finalité de produire des indicateurs de suivi de la santé des actifs sans emploi est conforme au Règlement Général européen sur la Protection des Données 2016/679 (RGPD) et à la loi du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés (Loi Informatique et liberté).

Le traitement de vos données personnelles est fondé sur la mission d'intérêt public dont est investi le responsable de traitement Santé Publique France (article 6.1.e du RGPD – licéité du traitement) et la dérogation de traiter des données de santé à des fins de recherche scientifique (article 9.2.j du RGPD – exception permettant de traiter des données de santé).

Utilisation des données

Les données seront accessibles exclusivement par les membres de l'équipe projet de Santé publique France habilités, ayant signé un engagement de confidentialité. Aucune information individuelle ne pourra être communiquée en dehors de l'équipe. Par ailleurs les résultats produits auront un caractère collectif ne permettant pas d'identifier les individus. Les données seront conservées pendant le nombre d'années autorisé pour la réalisation du projet, soit quatre ans à compter de la mise à disposition des données, c'est-à-dire jusqu'à 30 novembre 2027 et elles feront ensuite l'objet d'un archivage intermédiaire en étant consultables durant 2 années, le temps nécessaire à la valorisation scientifique complète des travaux, jusqu'au 30 novembre 2029.

Personnes concernées

Le projet s'appuie sur un appariement entre les données socio-économiques des individus présents dans l'Échantillon démographique permanent (EDP) de l'Insee et leurs données de soins et d'hospitalisation issues Système National des Données de santé (SNDS). L'échantillon obtenu appelé EDP-Santé a été constitué par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees).

L'EDP-Santé ne contient pas d'information permettant l'identification des personnes. Les données sont mises à disposition et utilisées par les personnels habilités de Santé publique France dans un espace sécurisé. Les données seront conservées pendant la durée du projet, jusqu'en 2028, puis archivées jusqu'à publication des travaux.

Conformément à l'article 66 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, ce projet de recherche a fait l'objet d'une autorisation de la Commission Nationale de l'Informatique et des Libertés (Cnil) en date du 17 juin 2024 (Décision DR 2024-145) modifiée en date du 18 octobre 2024 (DR 2024-265) autorisant l'Agence nationale de santé publique à mettre en œuvre un traitement de données ayant pour finalité une étude portant sur la Surveillance épidémiologique de la santé des personnes sans emploi, intitulée « SESSE ». (Demande d'autorisation n° 924102). La Cnil est l'autorité de contrôle chargée de surveiller l'application des règles relatives à la protection des données, afin de protéger les libertés et droits fondamentaux des personnes physiques à l'égard d'un traitement de données.

Pour exercer ces droits ou en savoir plus sur les modalités d'exercice de ces droits, vous pouvez vous adresser au délégué à la protection des données de Santé publique France :

Par mail : dpo@santepubliquefrance.fr (en indiquant le code « EDP-Santé » dans votre demande) ou par courrier à Santé publique France, déléguée à la protection des données, 12 rue du Val d'Osne, 94415 Saint Maurice Cedex

Par ailleurs, vous disposez d'un droit d'introduire une réclamation auprès d'une autorité de contrôle, en particulier auprès de la Cnil si vous considérez que le traitement de données à caractère personnel vous concernant constitue une violation du règlement général sur la protection des données et de la loi informatique et libertés <https://www.cnil.fr/fr/adresser-une-plainte>.

À l'issue de l'étude, et si vous le souhaitez, vous pourrez être informé(e) des résultats globaux qui seront référencés sur la documentation ouverte du Health Data Hub (HDH) (<https://www.health-data-hub.fr/projets>) et sur le site de Santé publique France (<https://www.santepubliquefrance.fr/>).

Pour toute autre question concernant l'utilisation des données de l'EDP-Santé dans le cadre de ce projet, merci de les adresser à sesse@santepubliquefrance.fr.