

La cohorte *CONSTANCES*

<https://www.constances.fr/>

Résumé du protocole

(Mise à jour Mai 2022)

UMS 011 - Cohortes épidémiologiques en population

16 av P. Vaillant-Couturier, F-94 800 Villejuif

Tel (33) 1 77 74 74 37 – fax (33) 1 77 74 74 03

SOMMAIRE

1	Objectifs de <i>CONSTANCES</i> : une cohorte généraliste en population.....	3
2	Méthodes et éléments essentiels du protocole.....	3
2.1	Mise en place et suivi de la cohorte : vue d'ensemble	3
2.2	Composition de la cohorte	4
2.3	Principales données recueillies	5
2.4	Sources de données	6
2.5	Modalités d'inclusion	7
2.6	Modalités de suivi	8
2.7	Représentativité et effets de sélection	9
2.8	Aspects réglementaires.....	11
2.9	Références.....	11
2.10	Publications issues de Constances	11
3	L'accès à la cohorte <i>CONSTANCES</i>	11

1 OBJECTIFS DE *CONSTANCES* : UNE COHORTE GÉNÉRALISTE EN POPULATION

L'objectif du projet *CONSTANCES* est de mettre en œuvre une importante cohorte épidémiologique en population destinée à contribuer au développement de la recherche en santé, et à fournir des informations à visée de santé publique. Réalisé dans le cadre d'un partenariat étroit avec la Caisse nationale d'assurance maladie (CNAM) qui en assure la majorité du financement, et avec la Caisse nationale d'assurance vieillesse (CNAV) le projet concerne la population adulte affiliée au Régime général de Sécurité sociale, qui couvre plus de 85 % de la population française. *CONSTANCES* est une « *Infrastructure nationale en biologie et santé* », labellisée et financée par les Investissements d'avenir, largement accessible à la communauté de la santé publique et de la recherche. L'Unité de service "Cohortes en Population", mixte entre l'Université de Paris, l'INSERM, l'Université Paris Saclay, l'Université de Versailles St Quentin (UMS 11), dirigée par le Pr. Marie Zins, est en charge de ce projet.

La cohorte *CONSTANCES*, visant à la représentativité, d'effectif important, se caractérisant par une couverture large de problèmes de santé et de déterminants et une ouverture vers des utilisateurs diversifiés, s'inscrit dans le cadre des cohortes « généralistes ». *CONSTANCES* est une infrastructure épidémiologique ouverte, à l'instar des grands instruments scientifiques (comme un télescope ou un accélérateur de particules, par exemple, ou un laboratoire de génotypage équipé de séquenceurs), qui ne sont pas construits pour répondre à une question spécifique, mais qui sont conçus pour aider à analyser une large gamme de problèmes scientifiques, et qui sont accessibles à la communauté des chercheurs spécialisés et de la santé publique.

2 MÉTHODES ET ÉLÉMENTS ESSENTIELS DU PROTOCOLE

La cohorte *CONSTANCES* s'appuie sur deux dispositifs existants à l'échelle nationale sur lesquels elle s'appuie largement grâce au partenariat avec la CNAM : (i) les Centres d'exams de santé de la Sécurité sociale, présents sur l'ensemble du territoire national, bénéficiant d'un plateau technique permettant le recueil de données de biomédicales dans des conditions particulièrement favorables ; (ii) les bases de données nationales du Système national des données de santé-SNDS géré par la CNAM et de la Caisse nationale d'assurance vieillesse, permettant l'accès à des données de santé et socioprofessionnelles.

2.1 MISE EN PLACE ET SUIVI DE LA COHORTE : VUE D'ENSEMBLE

CONSTANCES est une cohorte épidémiologique prospective. L'échantillon qui constitue la cohorte vise à être représentatif de la population âgée de 18 à 69 ans à l'inclusion couverte par le Régime général de Sécurité sociale (RGSS) ; l'effectif total est de 220 000 sujets, et sa structure est proportionnelle à la population du RGSS pour le sexe, l'âge et la catégorie sociale. *CONSTANCES* a été mise en place en 2012, après une importante phase pilote en 2009-2010. Le recrutement des participants s'est étalé pour l'essentiel de 2012 à 2019, et s'est terminé en 2021, date à laquelle elle comptait environ 220 000 volontaires.



Les principales étapes de la mise en place et du suivi de la cohorte, sont résumées ici.

- **Sélection des sujets éligibles** : les sujets éligibles sont tirés au sort dans le Répertoire national inter-régimes de l'assurance maladie, une base de données gérée par la CNAV, qui inclue toutes les personnes enregistrées à la Sécurité sociale en France.
- **Invitation à participer** : elle s'effectue conjointement à l'invitation à venir passer un examen de santé dans un Centre d'examens de santé de la Sécurité sociale (CES). Dans un premier courrier, les personnes tirées au sort reçoivent une lettre présentant le projet *CONSTANCES*, une plaquette d'information sur le CES, ainsi qu'un coupon-réponse leur permettant de donner l'accord pour participer à la cohorte dans le cadre d'un bilan de santé.
- **Inclusion des sujets volontaires dans la cohorte** : les personnes ayant donné leur accord pour participer à *CONSTANCES* sont convoquées dans leur CES par un courrier accompagné de questionnaires à compléter à domicile. Outre les questionnaires complétés à domicile, les sujets bénéficient d'un examen de santé permettant le recueil de données de santé et complètent des questionnaires complémentaires. Un consentement éclairé est proposé aux participants, qu'ils doivent signer s'ils acceptent de participer.
- **Suivi actif** : un autoquestionnaire (postal ou par Internet) à compléter au domicile des sujets est proposé chaque année, ainsi qu'une invitation à venir au CES tous les 4 ans pour un nouvel examen de santé.
- **Suivi passif d'événements socioprofessionnels et de données de santé** : les données de santé ainsi que les causes de décès sont régulièrement extraits du SNDS, et les principaux événements socioprofessionnels depuis les bases de données de la CNAV.

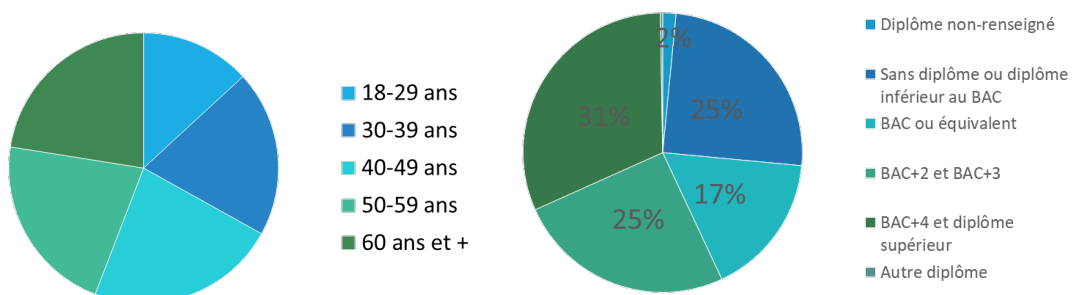
2.2 COMPOSITION DE LA COHORTE

Population source et structure de la cohorte

La population source est celle des assurés sociaux du Régime général de sécurité sociale (RGSS) géré par la Caisse nationale d'assurance maladie (CNAM), qui concerne plus de 85 % de la population française, soit plus de 55 millions de personnes.

La cohorte *CONSTANCES* est un échantillon aléatoire visant à la représentativité pour les variables d'âge, de sexe et de PCS de la population générale française relevant du RGSS âgée de 18 à 69 à l'inclusion, d'effectif proportionnel à la population générale pour ces mêmes variables.

La cohorte, composée de 54 % de femmes, constitue un échantillon largement diversifié en termes d'âge et de caractéristiques sociales.



Distribution des classes d'âge et des niveaux d'éducation

Effectif – Puissance

La grande dimension de *CONSTANCES* permet de pouvoir répondre à de multiples questions dans des domaines très variés, en menant des études épidémiologiques bénéficiant d'une bonne puissance statistique pour les maladies et/ou les facteurs de risque suffisamment fréquents. À titre d'exemple, on dénombrait parmi les participants fin 2021 environ 34 000 cas de cancer enregistrés dans le SNDS, dont plus de 5 300 incidents pendant le suivi de la cohorte ; pour les événements cardiovasculaires, les chiffres sont respectivement de 16 800 et 6 700. Ces chiffres, qui correspondent à la situation en 2021, ne font que croître avec la durée de suivi de la cohorte. Concernant les principaux facteurs de

risque, on compte 28 % des participants qui sont, selon le score AUDIT, en situation d'abus vis-à-vis de la consommation d'alcool, et 4,8 % en situation de dépendance ; si 42,5 % des volontaires n'ont jamais fumé, 19,4 % sont des fumeurs actuels et 38,1 % des ex-fumeurs. On dénombre 23 % des hommes et 10 % des femmes qui sont ou ont été exposés à des solvants en milieu de travail durant leur carrière professionnelle ; 20 % des hommes et 15 % des femmes ont ou ont eu des horaires de travail atypiques, notamment horaires alternants et travail de nuit. Grâce à ces effectifs importants, de nombreuses études dotées d'une bonne puissance ont déjà été réalisées concernant des aspects très diversifiés de la santé en relation avec des facteurs de risque de nature variée.

Un autre problème conditionnant la puissance d'une étude longitudinale est celui des « perdus de vue ». Grâce à l'appariement avec les bases de données du SNDS et de la CNAV, le nombre de « vrais » perdus de vue est négligeable (personnes qui demanderont à sortir de la cohorte, erreurs d'enregistrement, sujets quittant la France de façon définitive). Concernant la participation active par questionnaire annuels, elle se situe à environ 70 % en moyenne chaque année, et depuis l'origine jusqu'à 2021, seuls moins de 0,5 % avaient souhaité quitter la cohorte.

2.3 PRINCIPALES DONNÉES RECUEILLIES

Les données qui sont recueillies de façon systématique pour l'ensemble des participants sont destinées à fournir un corpus permettant de décrire et de suivre dans le temps l'évolution de phénomènes constituant une base d'information et une référence sur l'état de santé, la morbidité générale et la mortalité, le statut socio-économique et professionnel, l'environnement familial et social et du lieu de vie, les facteurs de risque personnels et environnementaux.

Caractéristiques sociodémographiques, statut et situation sociale

Situation et activité professionnelle, niveau d'études, niveau de revenus, situation matrimoniale, composition du ménage, statut socio-économique des parents et du conjoint, conditions de vie matérielles, origine géographique du sujet et de ses parents.

Données de santé

Antécédents personnels et familiaux, échelles de santé auto-déclarée (santé perçue, qualité de vie, santé mentale, échelles spécifiques : cardiovasculaire, troubles musculosquelettiques, santé respiratoire), pathologies (liste de pathologies déclarées, diagnostic des Affections de longue durée-ALD et des hospitalisations, absence au travail, handicaps, limitations, incapacités et traumatismes, recours aux soins et prise en charge), date et cause médicale de décès.

Dans les CES : examen clinique (poids, taille, rapport tour de taille-tour de hanches, tension artérielle, fréquence cardiaque), examen paraclinique (vision, audition, spirométrie, électrocardiogramme), investigations biologiques (régulation glycémique, bilan lipidique, bilan hépatique, créatinémie, numération formule sanguine, examen urinaire).

Pour les personnes âgées de 45 ans et plus, des examens spécifiques sont réalisés :

- Évaluation des capacités fonctionnelles : échelle IALD (*Instrumental Activities of Daily Living*), capacité à utiliser les nouvelles technologies.
- Fonctions cognitives : MMSE ; Trail Making Test A - B ; Code de Wechsler ; Évocation lexicale sémantique et alphabétique ; Test de mémoire RL–RI 16 de Grober & Busckhe.
- Fonctionnement physique : vitesse de marche ; *Digital Finger Tapping Test* ; test de l'équilibre (station unipodale) ; *Hand Grip Test*.

Comportements de santé

Consommations de tabac et d'alcool, fréquentiel d'habitudes alimentaires, activité physique, usage du cannabis, orientation sexuelle.

Facteurs professionnels

Histoire professionnelle, expositions professionnelles à des agents chimiques, physiques et biologiques, contraintes posturales, gestuelles et organisationnelles, stress au travail.

Problèmes de santé spécifiques des femmes

Traitements de la ménopause, ostéoporose et fractures ostéoporotiques, troubles sphinctériens et de la statique périnéale, mastopathies, endométriose et algies pelviennes chroniques, infertilité et délai de procréation, vie sexuelle et maladies sexuellement transmissibles.

Biobanque

Une banque de matériel biologique est constituée à l'occasion du bilan de santé. Elle comporte pour chaque sujet 26 aliquots de sang (sérum, plasma hépariné au lithium, plasma EDTA et buffy coat) et d'urine. Elle concerne environ 60 000 participants.

2.4 SOURCES DE DONNÉES

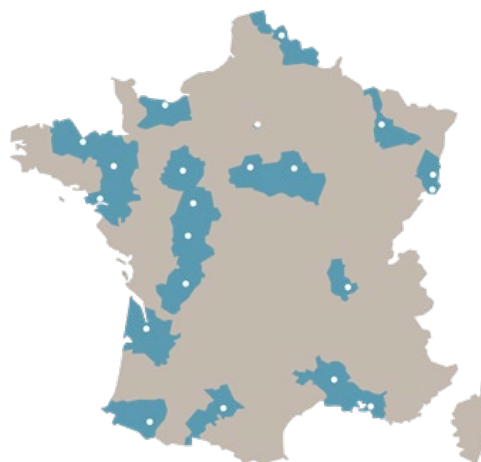
À l'inclusion et pendant le suivi, les participants font l'objet de recueils de données actifs et passifs.

Recueils actifs

Questionnaires : à l'inclusion, plusieurs questionnaires sont proposés : Mode de vie et santé, Santé des femmes, Calendrier professionnel, Expositions professionnelles, Médecin (rempli par le médecin à l'issue de l'examen clinique). Pendant le suivi, un questionnaire annuel est proposé, sous forme papier ou internet, selon le choix des volontaires.

Examens paracliniques : dans le cadre du partenariat avec la CNAM, les examens sont réalisés lors de l'inclusion et de façon périodique pendant le suivi, dans des Centres d'examens de santé de la Sécurité sociale (CES) répartis sur le territoire français.

Ce sont actuellement 21 CES qui participent au projet. Ces Centres sont de taille importante, disposent d'un personnel compétent et régulièrement formé, d'un plateau technique de qualité, et offrent une répartition géographique permettant de représenter les principales régions françaises. Les CES participants sont implantés dans 20¹ départements métropolitains. Il n'existe aucune structure équivalente en France, et seule la participation des CES permet la mise en œuvre de *CONSTANCES*.



Répartition des CES participant à *CONSTANCES*

¹ Deux CES sont situés à Paris

Recueil passif : appariement aux bases de données administratives nationales

Un recueil dit passif (car n'impliquant pas une intervention des sujets eux-mêmes) permet l'extraction des données des bases administratives nationales

SNDS : il contient des données de santé individuelles médicalisées, structurées et codées de manière standardisée. La base de données de remboursement de l'Assurance maladie comprend de façon détaillée toutes les prestations faisant l'objet d'un remboursement (consultations de professionnels de santé, médicaments et dispositifs médicaux). La base de données des ALD (Affections de Longue Durée) concerne tous les affiliés exonérés du ticket modérateur, et code l'affection exonérante selon la Classification Internationale des Maladies (CIM 10). La base de données du PMSI (Programme de médicalisation des systèmes d'information des hôpitaux) comprend pour chacun des séjours hospitaliers le recueil du diagnostic principal, du ou des diagnostics associés, de l'âge, du sexe et des actes diagnostiques et thérapeutiques. Les diagnostics sont codés selon la CIM 10 et les actes selon la Classification Commune des Actes Médicaux (CCAM).

CNAV : les bases de données de la CNAV concernent les événements socioprofessionnels. La CNAV ayant pour rôle d'assurer le droit au paiement de la retraite, elle collecte de façon prospective depuis le premier emploi les données sociales provenant des employeurs et de différents organismes et régimes gestionnaires des prestations sociales : type d'emploi et de contrat, salaire, périodes d'activité/non activité des personnes (chômage, absence pour maladie, maternité...).

2.5 MODALITÉS D'INCLUSION

Constitution de la cohorte

Préalablement au tirage au sort, une information générale a été faite dans les départements concernés par *CONSTANCES*, via les médias locaux, le site de la CNAM, ceux des Caisses départementales et des SLM participantes ; on a rappelé à cette occasion la possibilité d'exercer un droit de refuser de participer à l'étude. Le recrutement des volontaires s'étant étalé sur plusieurs années, les opérations suivantes ont été réalisées pour chaque vague annuelle d'inclusion.

- (1) La CNAV extrait du RNIAM (Répertoire national inter-régimes des bénéficiaires de l'assurance maladie, qui contient toutes les personnes en France inscrites à la sécurité sociale), les personnes éligibles (adultes âgés de 18 à 69 ans résidants dans les départements *CONSTANCES*, affiliées au RGSS) parmi lesquelles est réalisé un tirage au sort à probabilités inégales afin de surreprésenter les catégories d'âge et le statut socio-participant peu aux enquêtes. Chaque sujet tiré au sort se voit attribuer un numéro individuel non identifiant "NConstances".
- (2) En France, chaque personne dispose d'un identifiant national unique appelé NIR (« Numéro d'inscription au répertoire ») ; cet identifiant est utilisé dans les bases de données du SNDS et de la CNAV pour attribuer aux individus leurs données personnelles. La CNAV extrait les NIR des bénéficiaires des départements *CONSTANCES*. La CNAV adresse à la CNAM les données nécessaires pour les opérations suivantes.
- (3) La CNAM génère un nouveau numéro individuel non identifiant, différent du précédent, pour l'appariement au SNDS.

Tiers de confiance

Les personnes éligibles tirées au sort sont invitées à participer à *CONSTANCES* par voie postale. À cette fin, ainsi que pour les contacts ultérieurs avec les participants, un dispositif incluant un « tiers de confiance », organisme indépendant de *CONSTANCES*, qui gère les informations identifiantes des volontaires a été mis en place. Le tiers de confiance reçoit des organismes partenaires les données nominatives et les adresses des sujets tirés au sort et il établit une table de correspondance avec le numéro individuel non identifiant NConstances. Il n'a accès à aucune autre donnée concernant les participants. Tous les échanges ultérieurs avec les participants (courriers, emails...) passent par le

tiers de confiance, les volontaires n'étant identifiés dans la base de données *CONSTANCES* que par leur numéro individuel non identifiant NConstances.

Invitation à participer

Le Tiers de confiance envoie à un imprimeur-routeur les données permettant d'adresser les invitations aux sujets éligibles. Par l'intermédiaire d'un coupon-réponse, les personnes peuvent donner leur accord pour participer à la cohorte ou indiquer leur refus quant à la possibilité d'être tiré au sort et de faire partie d'un échantillon de non participants suivis passivement dans les bases de données administratives nationales ; en effet, afin de tenir compte des effets de sélection liés au volontariat, un échantillon de non participants est tiré au sort parmi ceux qui n'ont pas répondu à l'invitation (cf. plus loin : *Représentativité et effets de sélection*).

Dans tous les documents sont précisées les finalités et les modalités de la cohorte. Concernant l'information aux personnes sur le traitement de leurs données, les documents précisent que cette étude a obtenu l'autorisation de la CNIL, et qu'elles disposent d'un droit d'accès, de modification, de rectification, et de suppression des données personnelles les concernant.

Inclusion des participants par les CES

Lors de l'inclusion dans les CES, les volontaires bénéficient d'examen cliniques et paracliniques, dont les modalités de passation sont fixées par des Protocoles opératoires standardisés (POS) visant à garantir une stricte standardisation. Les examens réalisés sont les suivants : biométrie, vision, audition, spirométrie, ECG, tension artérielle, examens fonctionnels cognitifs et physiques pour les 45 ans et plus. Des procédures de contrôle de qualité sont régulièrement mises en œuvre. Les données de résultats de l'examen de santé, identifiées par le numéro individuel non identifiant NConstances sont adressées par chaque CES à l'équipe *CONSTANCES* ; les questionnaires remplis à domicile et au CES sont envoyés pour saisie à un prestataire puis transmis à l'équipe *CONSTANCES*.

Consentements éclairés

Lors de la visite au CES, un livret d'information sur le projet *CONSTANCES* est remis aux participants et un formulaire de consentement aux différents recueils de données est proposé ; les volontaires doivent signer ce consentement éclairé pour faire partie de *CONSTANCES*. Un livret d'information et un formulaire de consentement spécifiques sont également proposés pour la biobanque.

Recueil passif : interrogation des bases de données nationales

Extraction des données de la CNAV : pour chaque vague d'inclusion, la CNAV extrait de son système d'information le statut vital et les données socioprofessionnelles des participants et non participants qu'elle adresse à l'équipe *CONSTANCES*. Quelle que soit leur année d'inclusion, **les données des participants sont extraites depuis leur premier emploi.**

Extraction des données de la CNAM : Pour chaque vague d'inclusion, la CNAM applique un algorithme de génération du numéro permettant l'accès au SNDS pour chacun des sujets, et extrait de son système d'information les données médico-administratives des participants et non participants. Quelle que soit leur année d'inclusion, **les données des participants sont extraites depuis l'année 2007**, première année de disponibilité des données du SNDS.

2.6 MODALITÉS DE SUIVI

Traçage des sujets

Outre le signalement de changement d'adresse par les sujets eux-mêmes, la mise à jour régulière de l'adresse postale des participants est réalisée par La Poste, qui a mis en place des procédures adéquates. Les participants sont par ailleurs tracés dans les bases de données du SNDS et de la CNAV.

Participation active des sujets au suivi

Un autoquestionnaire annuel (postal ou Internet) est à compléter à domicile des sujets. Ceux-ci sont également invités tous les 4 ans dans leur CES pour un examen de santé de suivi. Il est crucial de

maximiser le taux de participation personnelle au suivi de la cohorte. À cette fin, un contact régulier avec les participants prend notamment la forme d'un « *Journal de la cohorte CONSTANCES* » présentant les résultats acquis, les projets associés, etc., adressé régulièrement aux participants, et d'un site Internet, incluant une newsletter trimestrielle.

Recueil « passif » d'événements socioprofessionnels et de données de santé

Les volontaires inclus dans *CONSTANCES*, ainsi que les non participants, font l'objet d'un suivi de données de santé et d'événements socioprofessionnels par appariement avec des bases de données nationales de la CNAV et du SNDS. Chaque année, la CNAV et la CNAM adressent à l'équipe *CONSTANCES* les nouvelles données socioprofessionnelles, les statuts vitaux et les données de santé des participants et non participants

Contrôle de qualité et validation des événements de santé

Les auto-questionnaires font l'objet des contrôles habituels : pourcentages de non réponse, de données manquantes, délai de retour, etc.

Pour les données recueillies au cours de la visite d'inclusion au CES, un contrôle de qualité systématique et permanent a été mis en place afin d'évaluer la précision, la reproductibilité, la concordance, la validité interne et externe des données recueillies, et d'étudier les facteurs de variabilité. Des Attachés de recherche épidémiologique font régulièrement des visites sur site pour vérifier le respect des procédures opératoires standardisées.

Pour les données extraites des bases de données nationales, une attention particulière est portée à la validation des diagnostics extraits des bases de données médico-administratives, qui font l'objet de divers contrôles.

Stockage des données

Les données sont associées pour chaque sujet à leur numéro permanent non identifiant et intégrées dans les bases de données de *CONSTANCES*. Celles-ci sont stockées sur l'infrastructure du « Centre d'accès sécurisé aux données » (CASD : www.casd.eu/), qui dispose de toutes les certifications de conformité aux référentiels de sécurité s'appliquant aux données personnelles.

Périodicité des recueils de données

La périodicité du suivi est variable selon les sources.

L'autoquestionnaire est complété annuellement, ce qui permet un suivi serré, de recueillir des données très nombreuses sans demander aux volontaires un effort trop important, de réagir rapidement pour mettre en place de nouvelles études, et de fidéliser les participants, car un trop long délai entre deux questionnaires est un facteur d'abandon. Certaines données sont recueillies chaque année (état de santé et morbidité déclarée, événements de vie et caractéristiques du lieu de résidence, tabac, alcool, etc.) ; d'autres font l'objet d'un recueil dont la périodicité est plus longue, selon un calendrier préétabli (échelles de santé, questionnaires spécifiques d'un domaine de santé ou de facteurs de risque, comme les habitudes alimentaires).

Le suivi dans les bases de données administratives nationales est permanent, puisque celles-ci enregistrent pour l'essentiel des événements en continu.

Les participants sont invités tous les 4 ans au CES pour un examen de santé de suivi.

2.7 REPRÉSENTATIVITÉ ET EFFETS DE SÉLECTION

Une des sources majeures de biais dans les enquêtes épidémiologiques provient des effets de sélection, qui peuvent concerner l'estimation de la prévalence ou de l'incidence de la maladie (ou de l'exposition à un facteur de risque) et l'estimation de l'association entre exposition et maladie. Dans les cohortes longitudinales, des effets de sélection peuvent se produire lors de l'inclusion, et tout au long du suivi du fait de l'attrition de la cohorte.

Le problème des biais lié aux effets de sélection est très différent selon qu'il s'agit d'objectifs analytiques ou descriptifs. Au sein d'une cohorte dont les procédures d'inclusion ont été les mêmes pour tous les sujets, la relation exposition - maladie n'est *a priori* pas différente entre les sujets inclus et ceux qui ne le sont pas. **Lorsqu'il s'agit d'études analytiques, les modalités de sélection des participants de CONSTANCES ne généreront donc a priori pas de biais**, ou seulement des biais minimales. Le problème de l'attrition au cours du suivi peut par contre être à l'origine de biais importants si la probabilité de ne plus être suivi diffère chez les exposés et non exposés, et/ou chez ceux qui sont ou ne sont pas devenus malades, ce qui est souvent le cas.

Pour l'étude descriptive de la fréquence des problèmes de santé et des expositions, les paramètres d'intérêt doivent être estimés sur un échantillon représentatif de la population cible. On a vérifié que la structure de la population de l'ensemble des départements où sont situés les CES CONSTANCES est pratiquement identique à celle de la France entière pour les principales caractéristiques démographiques et socioprofessionnelles ; les résultats de CONSTANCES pourront donc être généralisés à la population française.

Le volontariat entraîne inévitablement des effets de sélection, même lorsqu'on procède à un tirage au sort aléatoire d'un échantillon dans une base de sondage appropriée, car les volontaires et les non participants diffèrent pour divers facteurs d'intérêt qui sont une source potentielle de biais.

Afin d'obtenir un échantillon représentatif de la population cible et de minimiser les biais liés aux effets de sélection à l'inclusion et durant le suivi, les méthodes suivantes ont été mises en œuvre.

Tirage au sort

Le tirage au sort est fait dans une base de sondage constituée de la totalité des affiliés du RGSS des départements où sont situés les CES CONSTANCES, âgés de 18 à 69 ans à l'inclusion. Le tirage au sort est stratifié avec probabilités inégales selon âge, sexe et PCS, en surreprésentant les catégories participant peu aux enquêtes.

Redressement pour non-participation

Un échantillon aléatoire de non-participants, deux fois plus important que celui des participants, est constitué au fur et à mesure des inclusions. Pour ces sujets, on dispose de caractéristiques sociodémographiques, grâce aux fichiers de la CNAV, ainsi que de nombreuses informations de santé et de recours aux soins grâce aux données du SNDS. Ainsi, on dispose, à la fois pour les participants et l'échantillon des non participants, de données issues des fichiers du SNDS et de la CNAV. Ces données « auxiliaires » permettent d'estimer les probabilités de participation à CONSTANCES à l'aide de modèles de prédiction ; l'inverse de la probabilité de participation fournit ensuite le coefficient de redressement propre à chaque participant. Il est ainsi possible de procéder à des pondérations prenant en compte les effets de sélection à l'inclusion ; la richesse des données socioprofessionnelles et de santé utilisées permet des redressements très efficaces.

Pour le calcul de prévalences, on procède à un calage de la cohorte sur les marges de la population de référence, en utilisant l'échantillon aléatoire de non-participants pour calculer les marges pertinentes qui intègrent des données sociodémographiques et économiques et des données relatives à la santé. La qualité des estimations est donc sensiblement améliorée par le calcul de marges ayant un rapport spécifique avec la finalité santé de la cohorte. La mise au point de ces méthodes a permis à Constances de bénéficier du Label d'intérêt général et de qualité statistique du Conseil national de l'information statistique (CNIS : www.cnis.fr/).

Redressement pour attrition

On peut considérer qu'aux erreurs d'enregistrement près, aucune des personnes incluses dans CONSTANCES ne sera définitivement perdue de vue, puisque les participants sont suivis passivement à travers les fichiers administratifs du SNDS et de la CNAV. Il y aura cependant une attrition due au non renvoi de l'auto-questionnaire annuel ou à la non-participation aux examens de santé de suivi. Comme on dispose pour tous les participants des données recueillies lors de l'inclusion dans CONSTANCES ainsi que des données issues des fichiers du SNDS et de la CNAV correspondant à

l'année de l'inclusion, des coefficients de redressement pour attrition pourront être calculés par une méthode semblable à celle du calcul des coefficients de redressement pour non-participation. Le suivi passif et longitudinal des participants dans les bases de données du SNDS et de la CNAV, que ceux-ci abandonnent ou non la cohorte pour non renvoi d'auto-questionnaire, permet d'actualiser régulièrement les coefficients de redressement pour attrition.

2.8 ASPECTS RÉGLEMENTAIRES

Toutes les procédures décrites ont fait l'objet des formalités réglementaires auprès des organismes concernés : Commission nationale de l'informatique et des libertés-CNIL, Comité de Qualification Institutionnel de l'Inserm, Comité de protection des personnes. Toutes les procédures utilisées pour la création, la gestion et l'utilisation de la cohorte sont conformes au Règlement Général sur la Protection des Données (RGPD).

2.9 RÉFÉRENCES

Les principaux aspects du protocole de *CONSTANCES* ont fait l'objet de publications.

Zins M, Goldberg M and the Constances team. The French *CONSTANCES* population-based cohort: design, inclusion and follow-up. *Eur J Epidemiol.* 2015;30(12):1317-28. doi: 10.1007/s10654-015-0096-4.

Goldberg M, Carton M, Descatha A, Leclerc A, Roquelaure Y, Santin G, Zins M; *CONSTANCES* team. *CONSTANCES: a general prospective population-based cohort for occupational and environmental epidemiology – Cohort profile.* *Occup Environ Med.* 2017;74(1):66-71. doi: 10.1136/oemed-2016-103678.

Ruiz F, Goldberg M, Lemonnier S, Ozguler A, Boos E, Brigand A, Giraud V, Perez T, Roche N, Zins M. High quality standards for a large-scale prospective population-based observational cohort: *Constances.* *BMC Public Health.* 2016;16(1):877-886. doi: 10.1186/s12889-016-3439-5.

Henny J, Nadif R, Le Got S, Lemonnier S, Ozguler A, Ruiz F, Beaumont K, Brault D, Sandt E, Goldberg M, Zins M. The *CONSTANCES* Cohort Biobank: An Open Tool for Research in Epidemiology and Prevention of Diseases. *Front. Public Health* 2020; 8:605133. doi: 10.3389/fpubh.2020.605133.

2.10 PUBLICATIONS ISSUES DE *CONSTANCES*

Les publications issues des recherches menées dans la cohorte sont consultables sur le site de *Constances* : www.constances.fr/espace-scientifique/productions-scientifiques.php#publications-comite-lecture.

3 L'ACCÈS À LA COHORTE *CONSTANCES*

Toutes les équipes, françaises ou étrangères, publiques et privées, peuvent demander à utiliser la cohorte *CONSTANCES* pour leurs travaux de recherche. Il peut s'agir d'utilisation des données disponibles aussi bien que de matériels biologiques collectés, et/ou de collecter des données supplémentaires auprès de volontaires pour un objectif spécifique. Les demandeurs doivent soumettre un dossier qui détaille les aspects scientifiques et méthodologiques du projet, ses éventuelles implications concernant la confidentialité, l'éthique et la déontologie, ainsi que son financement et un calendrier prévisionnel de réalisation. Les demandes sont examinées par le Conseil scientifique international de *CONSTANCES*.

Une Charte pour l'accès à la cohorte *CONSTANCES* précise les modalités d'ouverture de l'infrastructure : modalités de sélection des projets, conditions d'exploitation des données disponibles, modalités de recueil, d'exploitation et d'accès liées aux données recueillies par des investigations complémentaires, financement des projets, responsabilités scientifiques, protection des données, diffusion des résultats de la recherche).