



Cohorte CONSTANCES

www.constances.fr

Avancement au 15 avril 2014

UNITÉ COHORTES ÉPIDÉMIOLOGIQUES EN POPULATION

UMS 011 Inserm-Université Versailles Saint-Quentin



TABLE DES MATIÈRES

1	RAPPEL - LE PROJET CONSTANCES	4
1.1	Objectifs.....	4
1.2	Éléments essentiels du protocole.....	4
1.2.1	Phase pilote	4
1.2.2	Mise en place et suivi de la cohorte : vue d'ensemble	5
1.2.3	Puissance statistique	6
1.2.4	Inclusion	7
1.2.5	Suivi longitudinal	7
1.2.5.1	Participation active des sujets au suivi	7
1.2.5.2	Recueil « passif » d'événements socioprofessionnels et de données de santé	8
1.2.6	Principales données recueillies aux différentes sources	8
1.2.7	Biobanque	9
1.2.8	Contrôle de qualité et validation des événements de santé	9
1.2.9	Représentativité et effets de sélection	10
1.2.10	Aspects opérationnels de l'inclusion et du suivi	12
1.2.10.1	Constitution des cohortes (participants et non participants)	12
1.2.10.2	Invitations	12
1.2.10.3	Convocations et questionnaires à compléter à domicile	12
1.2.10.4	Inclusion des participants par les CES et recueil de données.....	13
1.2.10.5	Examens précliniques	13
1.2.10.6	Transmission des données.....	13
1.2.10.7	Stockage des données.....	13
1.2.10.8	Suivi passif (interrogation des bases de données nationales).....	13
1.2.10.9	Suivi actif – Interrogation des participants par auto-questionnaire	14
2	TRAVAUX RÉALISÉS – ÉTAT D'AVANCEMENT.....	15
2.1	Autorisations – Avis	15
2.2	Aspects d'ordre technique et scientifique	15
2.2.1	Planification des procédures de recueil de données	15
2.2.2	Inclusions.....	16
2.2.2.1	Circuit de l'inclusion.....	16
2.2.2.2	Bilan des premières inclusions.....	16

2.2.2.3	Principales caractéristiques des participants.....	17
2.2.3	Prise en compte des effets de sélection – Résultats préliminaires	26
2.2.3.1	Facteurs liés à la participation	26
2.2.3.2	Redressements pour l’estimation de prévalences en population.....	28
2.2.4	Procédures de fidélisation.....	29
2.2.5	Le système d’information CONSTANCES.....	29
2.2.5.1	Flux de données	30
2.2.5.2	Architecture générale	33
2.2.5.3	Les applications WEB	34
2.2.6	La Biobanque de CONSTANCES	35
2.2.7	La plateforme de validation des évènements de santé	35
3	ORGANISATION DU PROJET	36
3.1	Les instances de gouvernance	37
3.2	Modalités d’accès à l’infrastructure	38
4	PARTENARIATS	38
4.1	Partenariats avec des organismes publics.....	38
4.2	Partenariats Publics- Privés	39
5	PUBLICATIONS-COMMUNICATIONS	40
5.1	Publications dans des revues à comité de lecture	40
5.2	Communications.....	40
5.3	Autres publications.....	42

1 RAPPEL - LE PROJET CONSTANCES

1.1 OBJECTIFS

L'objectif du projet CONSTANCES est de mettre en œuvre une importante cohorte épidémiologique destinée à contribuer au développement de la recherche épidémiologique et à fournir des informations à visée de santé publique. Réalisé dans le cadre d'un partenariat avec la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) et la Caisse nationale d'assurance vieillesse (Cnav) avec le soutien de la DGS, cette cohorte a vocation à constituer une infrastructure accessible à la communauté de recherche. La cohorte CONSTANCES n'est possible que grâce à des dispositifs existants à l'échelle nationale sur lesquels elle s'appuie largement : (i) les Centres d'examen de santé (CES) de la Sécurité sociale, présents sur l'ensemble du territoire national, bénéficiant d'un important plateau technique permettant le recueil de données biomédicales ; (ii) les bases de données nationales de l'Assurance maladie, de la Caisse nationale d'assurance vieillesse et du CépiDc, permettant l'accès à des données de santé et socioprofessionnelles de façon longitudinale.

La cohorte CONSTANCES, 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 accélérateur de particules, par exemple), 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 scientifique.

C'est pourquoi la cohorte CONSTANCES a vocation à faire l'objet d'un suivi longitudinal sans limite de temps, à la fois pour pouvoir étudier les effets à très long terme de facteurs de risque, et pour tenir compte de l'évolution des connaissances et des techniques, qui suscitent constamment de nouvelles questions scientifiques auxquelles CONSTANCES permettra d'apporter un éclairage.

Cependant, bien que conçue comme une cohorte généraliste à vocation très large, certaines thématiques ont été privilégiées dans la conception du protocole et le choix des données recueillies. Il s'agit essentiellement des risques professionnels, des déterminants sociaux et des inégalités sociales de santé, du vieillissement et des maladies chroniques, de la santé des femmes, de l'environnement physico-chimique et socio-économique ainsi que des interactions génétique-environnement.

1.2 ÉLÉMENTS ESSENTIELS DU PROTOCOLE

1.2.1 Phase pilote

Le projet CONSTANCES a comporté en 2009-2010 une phase pilote portant sur environ 3 500 sujets, dans sept CES (Lille, Bordeaux, Saint Briec, Tours, Pau, Toulouse et Rennes), qui a permis de tester l'ensemble des procédures et d'apporter quelques modifications. Le rapport de cette phase est disponible sur le site de CONSTANCES.

1.2.2 Mise en place et suivi de la cohorte : vue d'ensemble

La population source est celle des assurés sociaux du Régime général de sécurité sociale (RGSS), qui concerne plus de 85 % de la population française, soit environ 50 millions de personnes.

L'échantillon qui doit constituer la cohorte est représentatif de la population couverte par le Régime général de Sécurité sociale-RGSS (incluant les principales Sections locales mutualistes-SLM) âgée de 18 à 69 ans à l'inclusion ; l'effectif total visé est de 200 000 sujets, et sa structure est proportionnelle à la population du RGSS pour le sexe, l'âge et la catégorie sociale.

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

Sélection des sujets éligibles : les sujets éligibles sont tirés au sort par sondage stratifié avec probabilités inégales, en surreprésentant en fonction des variables usuelles (âge, sexe, PCS) les individus ayant une probabilité de non volontariat plus forte. Le tirage au sort est fait par la Cnav dans le Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (RNIAM), apparié au Système National de Gestion des Carrières (SNGC).

Invitation à participer : elle s'effectue conjointement à l'invitation à venir passer un examen de santé dans un CES. Dans un premier courrier, les personnes tirées au sort reçoivent une lettre et une plaquette de présentation 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 dans sujets volontaires dans la cohorte : les personnes ayant donné leur accord pour participer à CONSTANCES sont convoquées dans leur CES par un courrier précisant le jour et lieu de l'examen, ainsi que des auto-questionnaires à compléter à domicile.

Recueil des données à l'inclusion : outre les questionnaires préalablement complétés à domicile, les sujets bénéficient de l'Examen de santé périodique (EPS) usuel des CES réalisé selon des Procédures opératoires standardisées développées spécifiquement pour CONSTANCES afin d'assurer la qualité des données de santé recueillies : examen clinique, analyse de sang, mesure de la tension artérielle, du poids, de la taille et du rapport taille/hanches, électrocardiogramme et spirométrie, examen de la vue et de l'audition. En plus de l'EPS usuel, les sujets de CONSTANCES complètent d'autres questionnaires (expositions professionnelles vie entière, santé des femmes). Pour ceux qui sont âgés de 45 ans et plus une batterie de tests fonctionnels cognitifs et physique est également réalisée. Un consentement éclairé est proposé en fin de parcours aux participants, qu'ils doivent signer s'ils acceptent de participer.

Suivi actif : un autoquestionnaire, postal ou par Internet, est proposé chaque année ; une invitation à venir au CES tous les 5 ans est prévue pour l'ensemble des sujets de la cohorte.

Suivi passif d'événements socioprofessionnels et de données de santé : les principaux événements socioprofessionnels sont régulièrement extraits des bases de données de la Cnav (Déclarations Annuelles des Données Sociales, Données Nominatives Trimestrielles,

chômage, absences pour maladie, RMI, maternité). Les données de santé concernent les causes de décès par l'accès à la base de données du CépiDc-Inserm ; les principaux événements de santé seront extraits du Système national d'information inter régimes de l'assurance maladie (SNIIRAM) : données de remboursement, Affections de longue durée (ALD), diagnostic principal et diagnostics associés, actes diagnostiques et thérapeutiques pour chacun des séjours hospitaliers.

1.2.3 Puissance statistique

On a estimé le nombre d'événements de santé attendus dans la cohorte CONSTANCES à plus ou moins long terme, sous l'hypothèse d'une cohorte ayant une structure d'âge et de sexe identique à celle de la population générale française de 18 à 69 ans au recensement. On a calculé, à terme de 5, 10 et 15 ans, le nombre attendu d'événements pour les problèmes pour lesquels on dispose de données nationales de référence fiables : décès, incidence des cancers, des cardiopathies ischémiques et des maladies d'Alzheimer. On constate que le nombre de ces événements graves est élevé et permettra de nombreuses études dotées d'une bonne puissance.

Nombre attendu d'événements de santé

	Suivi 5 ans			Suivi 10 ans			Suivi 15 ans		
	H	F	Total	H	F	Total	H	F	Total
Décès toutes causes	4 131	2 133	6 264	9 727	5 502	15 229	16 983	10 736	27 719
Cancers incidents	3 162	2 220	5 381	7 036	4 855	11 892	11 444	7 823	19 267
Cardiopathies ischémiques (35-64 ans)	681	138	819	1 418	290	1 708	2 178	452	2 630
Maladies d'Alzheimer	265	240	505	793	1 007	1 800	1 548	2 469	4 018

Un autre problème conditionnant la puissance d'une étude longitudinale est celui des « perdus de vue ». Il est impossible d'estimer précisément le nombre de perdus de vue dans CONSTANCES au fil des années. On peut cependant se baser sur l'expérience du suivi d'autres cohortes, comme Gazel qui est le modèle sur lequel CONSTANCES est conçue. Dans Gazel, composée à l'origine de plus de 20 000 agents d'EDF-GDF qui sont suivis depuis 1989, la participation active par autoquestionnaire est élevée : au bout de 24 ans, seuls 3,1 % des sujets n'ont jamais renvoyé leur questionnaire annuel après avoir participé à l'inclusion ; le nombre de véritables perdus de vue, c'est-à-dire qu'on ne peut plus suivre dans les bases de données, est quant à lui infime : environ 0,9 % après 24 années de suivi. On peut penser que pour CONSTANCES, qui repose sur des méthodes similaires, on aura également un suivi efficace.

Des simulations tenant compte de la fréquence de l'exposition, du risque relatif attendu, de la durée de l'étude après la fin de la période d'inclusion et de l'incidence annuelle dans le groupe non exposé, ont montré que la puissance sera suffisante et les nombres de sujets nécessaires permettront de faire des tests dans des sous populations et d'étudier des interactions pour des problèmes relativement fréquents, qu'elle est satisfaisante dans d'autres situations selon que les analyses seront effectuées séparément chez les hommes et chez les femmes. Par contre, elle sera insuffisante pour des pathologies et des prévalences d'exposition peu fréquentes, et pour des risques relatifs faibles, un effectif de 200 000 sujets n'étant pas suffisant pour tester des

hypothèses spécifiques, particulièrement lorsqu'il s'agira de prendre en compte des interactions avec divers facteurs, et qu'il sera nécessaire de mettre des données en commun avec d'autres cohortes ; c'est une des raisons des partenariats qui ont été entrepris avec d'autres équipes, françaises ou internationales (cf. plus loin).

1.2.4 Inclusion

Ce sont 17 CES qui ont accepté de participer au projet. Ces Centres sont de taille importante, disposent d'un personnel suffisant et motivé, d'un plateau technique de qualité, et offrent une répartition géographique permettant de représenter les principales régions françaises. Les CES participants sont implantés dans 16 départements différents.



On procédera de façon graduelle à l'inclusion de l'ensemble de la cohorte sur une période de 5 ans, qui correspond au cycle quinquennal de venue des consultants des CES.

1.2.5 Suivi longitudinal

1.2.5.1 Participation active des sujets au suivi

Un autoquestionnaire postal annuel est envoyé au domicile des sujets ; il est possible de disposer de la mise à jour régulière de l'adresse postale des participants grâce aux services que La Poste, qui a mis en place des procédures adéquates ; en cas d'échec, d'autres sources seront utilisées, comme le signalement de changement d'adresse par les sujets eux-mêmes. Pour les sujets qui l'acceptent, le questionnaire annuel peut être complété par Internet.

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 de la création d'un site Internet.

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

Les sujets inclus dans CONSTANCES font l'objet d'un suivi dit « passif » (car n'impliquant pas une intervention des sujets eux-mêmes) d'événements socioprofessionnels et de données de santé par appariement régulier avec des bases de données nationales.

Les bases de données de la Cnav permet l'accès à des **données socioprofessionnelles** issues des DADS (Déclarations Annuelles des Données Sociales) transmises chaque année par les employeurs, des DNT (Données Nominatives Trimestrielles) par les employeurs de personnel de maison, ainsi que des informations de périodes d'activité / non activité des individus relevant de l'UNEDIC (chômage), de la CNAMTS (absences pour maladie, incluant les dates de début et fin d'arrêt de travail), de la Cnaf (RMI, maternité, ...), des régimes particuliers ou spéciaux (SNCF, EDF, RATP, ...), et de certains autres régimes.

Les **causes de décès** sont obtenus selon la procédure décrite dans le Décret 98-37 autorisant l'accès à la base de données du CépiDC-Inserm.

Pour les **événements de santé**, l'accès au SNIIRAM (Système national d'information inter régimes de l'assurance maladie) constitue une solution efficace. L'intérêt du SNIIRAM dans une optique épidémiologique est lié au fait qu'il contient des données individuelles médicalisées, structurées et codées de manière standardisée. La base de données de remboursement de l'Assurance maladie est adaptée aux objectifs d'analyse des pratiques de prescription, mais elle ne comporte pas d'information sur la nature des maladies traitées, et exclut par définition l'automédication et les prestations non présentées au remboursement. 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 plus coûteux. Les diagnostics sont codés selon la CIM 10 et les actes selon la Classification Commune des Actes Médicaux (CCAM).

1.2.6 Principales données recueillies aux différentes sources¹

Les données qui doivent être 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.

De façon générale, on s'est efforcé de choisir des variables déjà utilisées dans d'autres enquêtes, à la fois parce qu'il s'agit de mesures validées et du fait qu'il sera ainsi possible de disposer de données de référence pour certaines analyses. On a également, chaque fois que

¹ Catalogue des données complet : <http://www.constances.fr/fr/l-espace-scientifique/base-documentaire/autres-documents.html>

cela a été possible et pertinent, utilisé des échelles publiées dans la littérature, dont les qualités métrologiques sont établies.

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, origines géographiques du sujet et des parents, conditions de vie matérielles.

Données de santé : antécédents personnels et familiaux (cancer, cardiovasculaire, psychiatrie, diabète, pathologie respiratoire), échelles de santé autodéclarée (santé perçue, qualité de vie, santé mentale, échelles spécifiques : cardiovasculaire, TMS, respiratoire), pathologies (liste de pathologies déclarées, diagnostic des 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 sont recueillies les données suivantes : examen clinique (poids, taille, rapport tour de taille-tour de hanches, tension artérielle,), examen paraclinique (vision, audition, spirométrie), investigations biologiques (glycémie à jeun, bilan lipidique, bilan hépatique, créatinémie, numération formule sanguine, examen urinaire par bandelette). Pour les personnes âgées de 45 ans et plus, des examens spécifiques sont prévus : évaluation des capacités fonctionnelles (échelle IALD,), fonctions cognitives (MMSE, Trail Making Test A - B, Code de Wechsler, Digital Finger Tapping Test, évocation sémantique et lexicale, test de mémoire de Grober & Busckhe), fonctionnement physique (vitesse de marche, équilibre, Hand Grip Test).

Comportements : consommations de tabac et d'alcool (consommations historiques et actuelles), habitudes alimentaires et 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 bénignes, endométriose et algies pelviennes chroniques, infertilité et délai de procréation, vie sexuelle et maladies sexuellement transmissibles.

1.2.7 Biobanque

Une Banque de matériel biologique doit être constituée à l'occasion du bilan de santé. Actuellement, son financement est assuré pour la conservation de 10 aliquots de sang et d'urine pour 100 000 sujets. Il est envisagé des extensions en nombre de sujets prélevés et pour d'autres spécimens biologiques (selles, prélèvements spéciaux pour omique...) selon les demandes des chercheurs et la disponibilité de financements additionnels.

1.2.8 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 paracliniques** 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 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 d'un contrôle systématique. Dans un premier temps, trois types de pathologies sont plus particulièrement analysées : les événements cardiovasculaires ischémiques, les cancers et les pathologies neurodégénératives signalés dans les sources disponibles sont systématiquement vérifiés auprès des services d'hospitalisation et/ou des médecins traitants.

1.2.9 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 la prévalence 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 (cas de CONSTANCES), la relation exposition - maladie n'est *a priori* pas différente entre les sujets inclus et ceux qui ne le sont pas. Les modalités de sélection des participants de CONSTANCES ne généreront donc *a priori* pas de biais, ou seulement des biais minimes, lorsqu'il s'agit d'études analytiques. 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 (âge, sexe, CS et secteurs économiques) ; les résultats de CONSTANCES pourront donc être généralisés.

Le volontariat entraîne inévitablement des effets de sélection, même lorsqu'on procède à un tirage au sort d'un échantillon dans une base de sondage appropriée. On rencontre en effet à l'inclusion des non participants et des non volontaires, qui sont une source potentielle de biais. Pour y remédier, on s'efforce habituellement de recueillir un minimum de données sur les non participants (âge, sexe, PCS essentiellement), afin de procéder ultérieurement à des redressements pour estimer les paramètres d'intérêt. Cette approche connaît cependant certaines limites. Ainsi, il n'est pas toujours possible de recueillir les données de redressement pour les sujets non participants. Il n'est pas toujours clair de savoir si ces données sont suffisantes pour contrôler les biais potentiels, car on sait par exemple qu'au sein de la même

catégorie socio-économique existent de larges différences à bien des égards, notamment en termes de santé, de comportements, de modes de vie, de réseaux sociaux, etc. Finalement, on est rarement en situation de contrôler complètement les biais de sélection potentiels, faute de disposer de données pertinentes recueillies à la fois pour les participants et l'ensemble des non participants et non volontaires.

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, des méthodes originales ont été développées pour CONSTANCES.

Tirage au sort : le tirage au sort est stratifié avec probabilités d'inclusion inégales, établies à partir des données d'autres enquêtes réalisées dans les CES.

Redressement pour non-participation : un échantillon de non participants de taille deux supérieure à celle de la cohorte, pour lesquels on dispose de caractéristiques sociodémographiques, grâce aux fichiers de la Cnav (sexe, âge, statut d'activité, PCS), ainsi que de nombreuses informations de santé et de recours aux soins grâce aux données du SNIIRAM a été constitué. Afin d'étudier les effets de sélection à l'inclusion, les données de la Cnav et du SNIIRAM des sujets tirés au sort sont extraites pour les 2 années précédant l'invitation à participer. Ainsi, on dispose pour les participants et l'échantillon des non participants, de données issues des fichiers du SNIIRAM et de la Cnav. Ces données 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. Cette méthode repose sur utilisation d'un modèle de régression logistique dans lequel la variable à expliquer est la variable binaire « participation ». La sélection des variables explicatives et de leurs interactions à inclure dans le modèle de prédiction constitue une étape importante et doit être effectuée de manière à ce qu'on puisse avoir confiance dans le modèle de prédiction pour tous les individus.

Redressement pour attrition : on peut considérer qu'aux erreurs près, aucune des personnes incluses dans CONSTANCES ne sera définitivement perdue de vue, puisque les participants sont suivis passivement à travers les fichiers du SNIIRAM et de la Cnav. Il y aura cependant une attrition due au non renvoi de l'auto-questionnaire annuel. Ainsi, pour que les analyses portant sur les variables de l'autoquestionnaire puissent être menées dans de bonnes conditions de validité, il convient d'effectuer un redressement pour attrition. L'inclusion dans CONSTANCES s'étendra sur 5 ans. Pour la vague 1, l'année (année 2) suivant l'année d'inclusion (année 1), on distinguera les participants ayant répondu à l'autoquestionnaire des participants n'y ayant pas répondu. On disposera pour tous les participants des données recueillies lors de l'inclusion dans CONSTANCES (année 1) ainsi que des données issues des fichiers du SNIIRAM et de la Cnav correspondant à l'année de l'inclusion (année 1), ou à l'année précédant l'inclusion (année 0). Les coefficients de redressement pour attrition l'année 2 pourront ainsi être calculés par une méthode semblable à celle du calcul des coefficients de redressement pour non-participation. Pour les années suivantes, on distinguera également les participants ayant répondu à l'autoquestionnaire des participants n'y ayant pas répondu.

Cependant, il est probable qu'un certain nombre de participants à CONSTANCES soit non répondants une année et répondants pour des années suivantes. Des solutions pourront être dérivées de la méthodologie des non réponses partielles : exclusion des non répondants, utilisation de variables indicatrices de données manquantes, imputations simples ou multiples, ou repondérations. Le choix de la méthode la mieux adaptée doit être défini au cas par cas selon le problème spécifique étudié. Le suivi passif et longitudinal des participants dans les bases de données du SNIIRAM et de la Cnav, que ceux-ci abandonnent ou non la cohorte pour non renvoi d'auto-questionnaire, permettra d'actualiser les coefficients de redressement pour attrition.

1.2.10 Aspects opérationnels de l'inclusion et du suivi

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 CNAMTS (site Améli), ceux des CPAM et des SLM participantes ; on a rappelé à cette occasion la possibilité d'exercer un droit de refuser de participer à l'étude.

1.2.10.1 Constitution des cohortes (participants et non participants)

L'équipe CONSTANCES fournit à Cnav la liste des CPAM participant à CONSTANCES ainsi que le programme de tirage au sort ; les probabilités de participation sont empiriquement définies sur la base d'enquêtes déjà réalisées à partir d'un échantillon aléatoire invité à se rendre dans un CES. La Cnav extrait, *via* le RNIAM, les NIR des bénéficiaires affiliés aux CPAM CONSTANCES, les apparie aux informations issues de ses fichiers sur la PCS, et réalise le tirage au sort. Chaque sujet tiré au sort se voit attribuer un numéro individuel non identifiant « CONSTANCES ». La Cnav adresse à un Tiers de confiance (Plateau informatique du CESP-U1018) et à la CNAMTS les données nécessaires pour les opérations suivantes. La CNAMTS génère un nouveau numéro individuel non identifiant, différent du précédent, pour l'appariement à ses bases de données.

1.2.10.2 Invitations

Le Tiers de confiance envoie à un imprimeur-routier les données permettant d'adresser les invitations aux sujets tirés au sort. 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 nationales. 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 fait l'objet d'une déclaration à la CNIL, et qu'elles disposent d'un droit d'accès, de modification, de rectification, et de suppression des données personnelles les concernant.

1.2.10.3 Convocations et questionnaires à compléter à domicile

Les personnes qui donnent leur accord pour participer à la cohorte reçoivent une convocation au CES, accompagnée d'une lettre d'explication, et de deux questionnaires à remplir à domicile et à amener lors de leur venue au CES : autoquestionnaire Modes de vie et santé et Calendrier professionnel.

1.2.10.4 Inclusion des participants par les CES et recueil de données

Questionnaires : en complément des questionnaires remplis à domicile, trois autres questionnaires sont complétés au cours de la visite au CES : questionnaire Santé des femmes et questionnaire Expositions professionnelles, questionnaire Médecin rempli par le médecin à l'issue de l'examen clinique.

1.2.10.5 Examens précliniques

Les modalités de passation de ces examens 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 pour les 45 ans et plus. Des procédures de contrôle de qualité sont régulièrement mises en œuvre.

À la fin de leur parcours au CES, les participants volontaires signent un consentement éclairé.

1.2.10.6 Transmission des données

Pour les participants, les données de résultats de l'examen de santé, et les questionnaires remplis à domicile et au CES, identifiées par le numéro individuel non identifiant CONSTANCES, sont adressées par chaque CES à l'équipe CONSTANCES.

1.2.10.7 Stockage des données

Une fois informatisées, les données sont associées à un Numéro de cohortiste permanent (NCP) non identifiant et intégrées dans les bases de données sur un serveur accessible uniquement aux personnes habilitées.

1.2.10.8 Suivi passif (interrogation des bases de données nationales)

Circuit des données avec la Cnav : pour chaque vague d'inclusion, l'équipe CONSTANCES adresse à la Cnav la liste des numéros individuels non identifiants CONSTANCES et le statut de participation correspondant. La Cnav extrait de son système d'information le statut vital et les données socioprofessionnelles qu'elle adresse à l'équipe CONSTANCES. Finalement, après les vagues d'inclusion, la Cnav détruit toutes les données ne concernant ni la cohorte de participants ni la cohorte de non participants. Dans le cadre du suivi, la Cnav adresse périodiquement à l'équipe CONSTANCES les nouvelles données socioprofessionnelles et statuts vitaux des participants et non participants.

Circuit des données avec la CNAMTS : pour chaque vague d'inclusion, l'équipe CONSTANCES adresse à la CNAMTS la liste des numéros individuels non identifiants CONSTANCES et le statut de participation correspondant. La CNAMTS applique un algorithme de génération du numéro permettant l'accès au SNIIRAM (procédure FOIN) pour chacun des participants, et extrait de son système d'information les données médico-administratives qu'elle adresse à l'équipe CONSTANCES. Pour le suivi, la CNAMTS adresse périodiquement à l'équipe CONSTANCES les nouvelles données médico-administratives des participants et non participants.

Suppression de données nominatives : après l'inclusion, le Tiers de confiance détruit toutes les données ne concernant pas les participants (y compris celles des personnes appartenant à la cohorte de non participants, dont le suivi doit être strictement anonyme). Ainsi, après l'inclusion, le Tiers de confiance dispose uniquement des données concernant les participants afin d'assurer le suivi de ces personnes.

Recueil des causes de décès : pour le recueil des causes médicales des décès survenus dans les cohortes de participants et de non participants sont recueillies selon les modalités fixées par le Décret n° 98-37 du 16 janvier 1998, qui autorise l'accès aux données relatives au décès des personnes inscrites au RNIPP (Insee) dans le cadre des recherches dans le domaine de la santé. Deux étapes successives doivent être réalisées : (1) la recherche du statut vital par interrogation du RNIPP (Insee) pour savoir si une personne est décédée ; (2) la recherche des causes de décès dans la base de données du Service d'Information sur les causes médicales de décès (CépiDc) de l'Inserm. La Cnav intervient dans ce circuit en fournissant les données nominatives des participants et des non participants au Tiers de confiance, qui effectue les requêtes nécessaires.

1.2.10.9 Suivi actif – Interrogation des participants par auto-questionnaire

Chaque participant complète tous les ans un auto-questionnaire de suivi accompagné d'une fiche de suivi destinée à mettre à jour les données nominatives (changement de nom d'usage, d'adresse). Il peut répondre sur questionnaire-papier ou par Internet.

Envoi des auto-questionnaires : l'équipe CONSTANCES adresse au Tiers de confiance la liste des numéros individuels non identifiants CONSTANCES à contacter. À partir de la base de données nominatives, le Tiers de confiance adresse à un imprimeur-routeur les données nécessaires aux envois postaux.

Traitement des retours de l'auto-questionnaire : l'équipe CONSTANCES est destinataire de l'ensemble des auto-questionnaires de suivi qui sont mis dans une chaîne de traitement (scannérisation, reconnaissance, vidéocodage, archivage...). En fin de traitement, les données sont intégrées à la base de données CONSTANCES. Pour les sujets complétant le questionnaire par Internet, un circuit *ad hoc* permet d'intégrer les réponses dans la base de données.

Gestion et traitement des adresses : après l'inclusion au CES, le Tiers de confiance est détenteur des données permettant d'écrire aux participants. L'équipe CONSTANCES génère un numéro de transfert spécifique « La Poste », assure la correspondance entre ce numéro et le numéro individuel non identifiant CONSTANCES qu'elle transmet au Tiers de confiance. Celui-ci envoie un fichier à La Poste afin qu'elle effectue les traitements de normalisation des adresses, de correction des PND et d'enrichissement du fichier (dont géocodage = X, Y et code Iris). Au cours du suivi, l'envoi des auto-questionnaires annuels est effectué en utilisant le dispositif Alliage mis en place par La Poste, qui permet d'identifier rapidement les Plis Non Distribuables (PND) et d'éliminer les retours physiques.

2 TRAVAUX RÉALISÉS – ÉTAT D'AVANCEMENT

2.1 AUTORISATIONS – AVIS

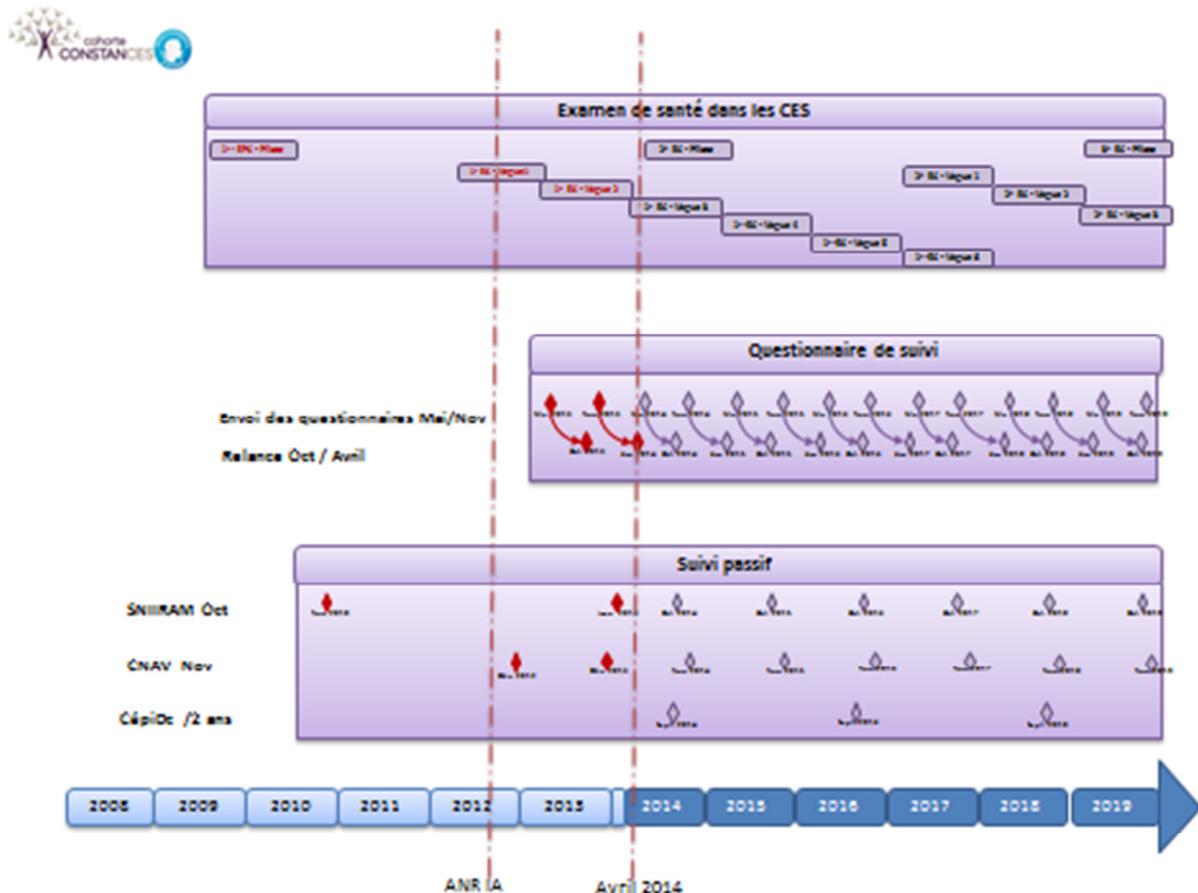
CONSTANCES a reçu les autorisations et avis favorables suivants :

- Label d'intérêt général et de qualité statistique du CNIS, le 10 octobre 2008 ;
- Comité de qualification institutionnel (CQI) Inserm, le 18 février 2009 ;
- CCTIRS-CNIL : après les autorisations pour le pilote et la constitution des échantillons, la cohorte CONSTANCES a obtenu un avis favorable du CCTIRS le 25 novembre 2010 et l'autorisation de la Cnil le 3 mars 2011 ;
- CPP Ile de France XI : 11 avril 2013 (avis pour la réalisation de tests neuropsychologiques).

2.2 ASPECTS D'ORDRE TECHNIQUE ET SCIENTIFIQUE

2.2.1 Planification des procédures de recueil de données

CONSTANCES est un projet dont les premières phases préparatoires ont commencé depuis longtemps. Des données sont recueillies depuis 2012 à différentes sources avec des périodicités variables, comme l'illustre la figure suivante (en rouge : étapes déjà réalisées).



2.2.2 Inclusions

2.2.2.1 *Circuit de l'inclusion*

Le lancement effectif de la cohorte a débuté mi-2012. Le recrutement et l'inclusion des volontaires impliquent une série complexe de processus impliquant différents partenaires, qui vont du tirage au sort à l'intégration des données dans la base CONSTANCES.

Selon les sources de données (questionnaires, données des bases nationales, données collectées dans les CES), ils génèrent différents flux qui sont détaillés plus loin (*Système d'information CONSTANCES*).

2.2.2.2 *Bilan des premières inclusions*

Le **nombre d'inclusions** est en augmentation progressive. La première année, la CNAMTS n'a autorisé que 20 000 inclusions, afin de donner aux CES participants le temps de s'adapter aux contraintes des procédures de qualité strictes de CONSTANCES. Actuellement (avril 2014) un peu plus de 37 000 sujets ont été inclus dans les CES.

Le **taux de participation** attendu avait été estimé *a priori* à environ 10 %. Globalement, le taux observé est proche de ce chiffre, avec comme attendu une importante variation selon le régime d'affiliation, les adhérents du Régime général strict (RGSS) participant nettement moins que ceux des SLM (tableau suivant, données de septembre 2013 ; ces taux de participation sont sous-estimés car certains CES remplissent les tableaux de suivi de participation avec retard).

Régime	Taux participation (%)
CAMIEG	13,7
LMDE	4,7
MFP	13,1
MG	14,3
MGEN	14,5
MNH	9,6
RGSS	5,7

Les CES jouent évidemment un rôle crucial dans les inclusions, car ils assurent le premier contact avec les volontaires et l'explication du consentement, et ils réalisent les examens paracliniques et cliniques de l'inclusion ; 593 agents des CES ont reçu une formation spécifique et sont répertoriés.

Un **contrôle de qualité** de l'examen paraclinique est assuré par la société ClinSearch, qui par des visites sur site d'Assistants de recherche épidémiologique (2 jours/mois sur site dans chaque CES), assure la formation des personnels des CES, le monitoring de l'examen de santé et le respect des Procédures opératoires standardisées (POS) qui ont été développées pour CONSTANCES. Le contrôle de qualité de la biologie est assuré par les associations Asqualab et EuroCell. Des rapports mensuels sont adressés à l'équipe CONSTANCES. Les POS et les vidéos de formation des personnels sont disponibles sur le site CONSTANCES.

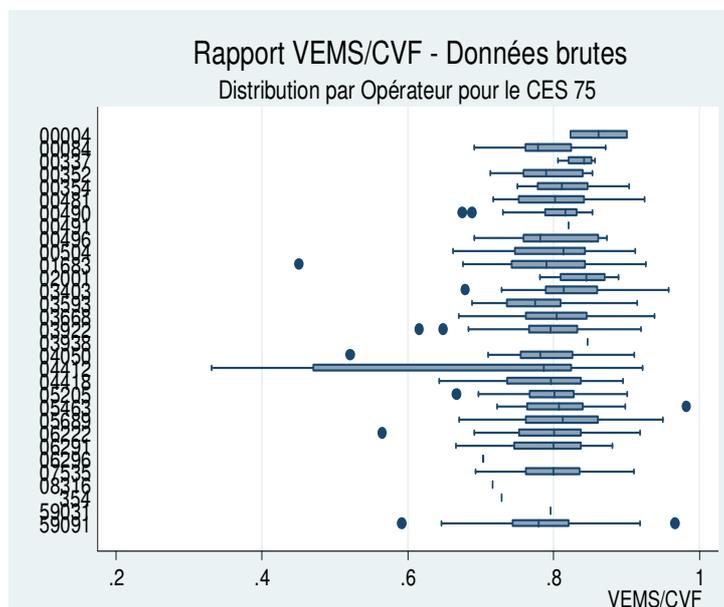
Les analyses de contrôle de qualité montrent que, grâce aux efforts de formation et de monitoring, les données recueillies dans les 17 CES participants sont d'excellente qualité. Ainsi, les résultats de l'examen de la fonction respiratoire (spirométrie, examen particulièrement

sensible à l'opérateur), sont très homogènes d'un CES à l'autre, comme le montre le tableau suivant qui présente les moyennes du coefficient VEMS/CV, qui est l'indicateur principal de la fonction respiratoire (résultats ajustés sur âge, sexe et statut tabagique).

Qualité des spirométries : résultats VEMS/CV

CES	Moyenne	IC à 95%	Moyenne	
			Standardisée*	IC à 95%
Marseille	0,796	(0,688 0,904)	0,796	(0,758 0,833)
Angoulême	0,769	(0,648 0,891)	0,773	(0,733 0,812)
Saint-Brieuc	0,776	(0,657 0,896)	0,780	(0,734 0,827)
Nîmes	0,795	(0,670 0,920)	0,802	(0,757 0,847)
Toulouse	0,823	(0,704 0,942)	0,822	(0,773 0,870)
Bordeaux	0,807	(0,687 0,927)	0,804	(0,762 0,846)
Rennes	0,807	(0,694 0,920)	0,808	(0,772 0,844)
Saint-Nazaire	0,786	(0,667 0,906)	0,789	(0,757 0,822)
Orléans	0,802	(0,699 0,905)	0,799	(0,770 0,827)
Nancy	0,817	(0,699 0,935)	0,808	(0,765 0,851)
Lille	0,781	(0,653 0,909)	0,782	(0,740 0,823)
Pau	0,818	(0,699 0,936)	0,820	(0,778 0,863)
Lyon	0,840	(0,710 0,970)	0,829	(0,789 0,869)
Paris	0,794	(0,656 0,932)	0,790	(0,742 0,838)
Poitiers	0,792	(0,658 0,925)	0,796	(0,740 0,852)

Le monitoring individuel de chaque opérateur permet de repérer les écarts et de mettre en place des opérations correctrices (cf. figure ci-dessous).



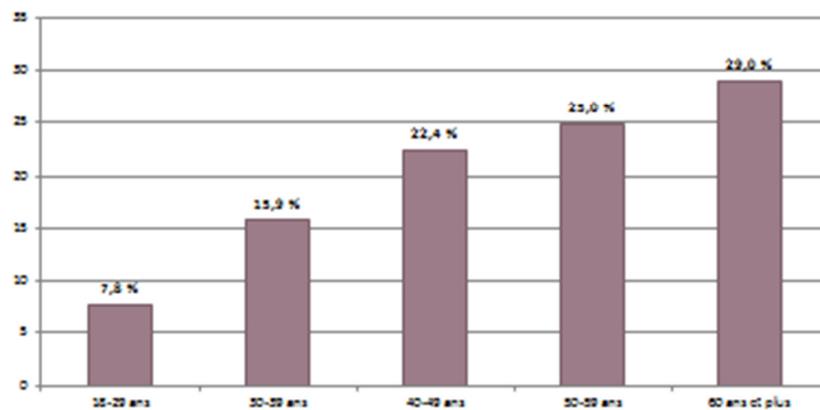
2.2.2.3 Principales caractéristiques des participants

Remarque : du fait de la périodicité variable des flux de données (cf. plus haut) et de l'important travail de vérification et de contrôle des données avant intégration dans la base de données

centrale, l'ensemble des données d'un participant n'est donc consolidée qu'après un délai de plusieurs mois après son inclusion. C'est pourquoi les données présentées ici correspondent aux sujets inclus à la date 28 Février 2014.

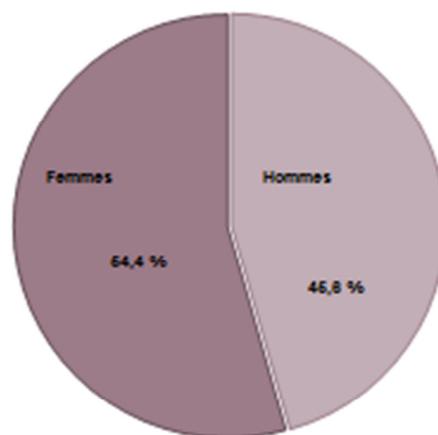
Âge : on observe une sous-représentation des classes d'âge les plus jeunes (18-30 ans) ; des actions spécifiques sont en préparation pour augmenter leur participation : compte CONSTANCES sur les réseaux sociaux, collaboration en cours avec des organismes travaillant sur des populations jeunes (Inpes, INJEP, LMDE, MFP-S notamment).

Répartition des effectifs par classe d'âges



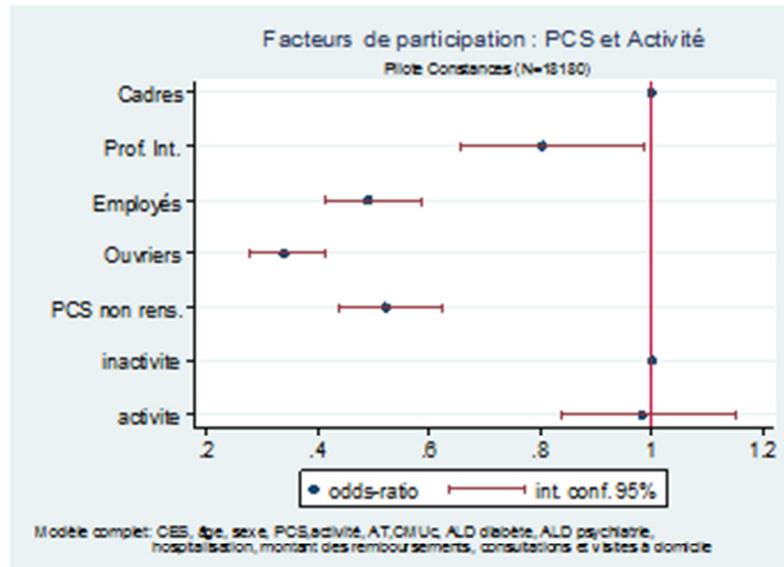
Sexe : comme attendu, on observe une meilleure participation des femmes.

Répartition par sexe



Effectif total de 19 003 individus : 8 701 hommes et 10 302 femmes

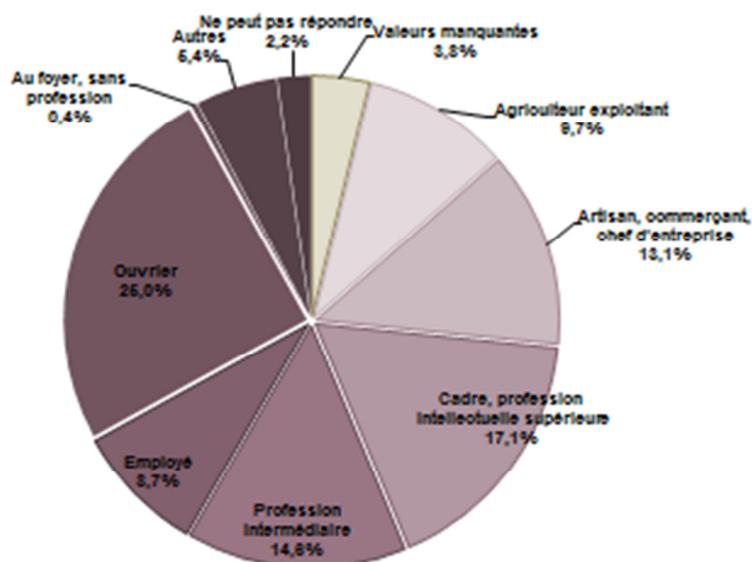
Situation sociale : comme attendu, les ouvriers participent moins que les cadres et que les employés (données issues du SNGC de la Cnav sur les participants et non participants)



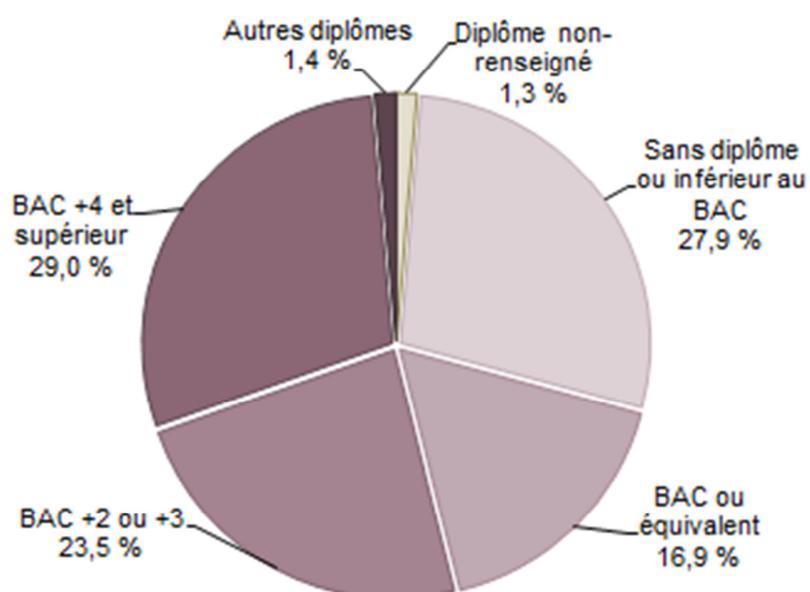
Symposium international sur les questions de méthodologie, 16-18 octobre 2013 Ottawa

On observe cependant une importante diversité sociale de l'échantillon de participants : l'origine sociale (PCS du père) est variée, et toutes les catégories sociales, les niveaux d'éducation et les situations familiales sont représentés dans la cohorte, qui comporte un nombre non négligeable de personnes en difficulté financière qui leur a fait renoncer à certains soins de santé.

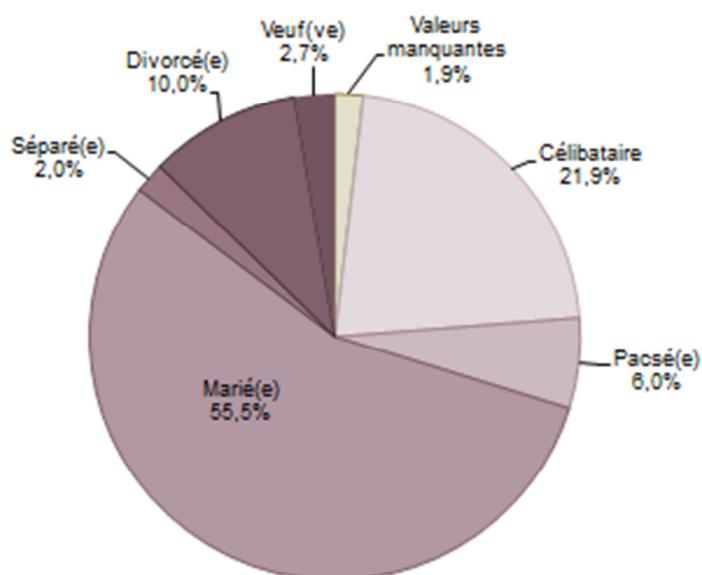
PCS de la père durant l'adolescence



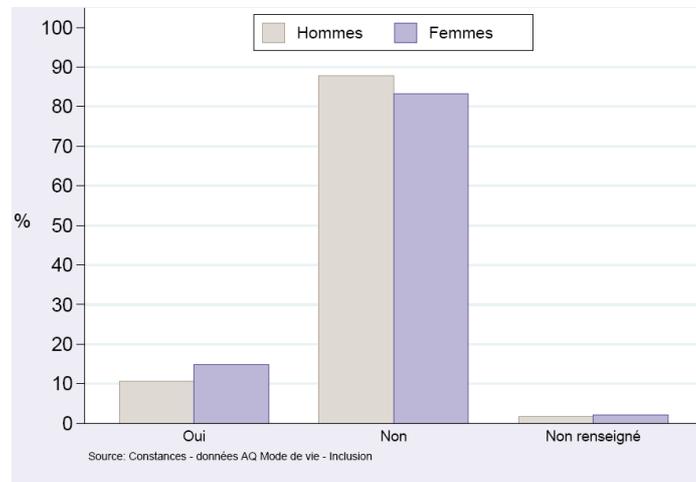
Répartition selon le diplôme obtenu



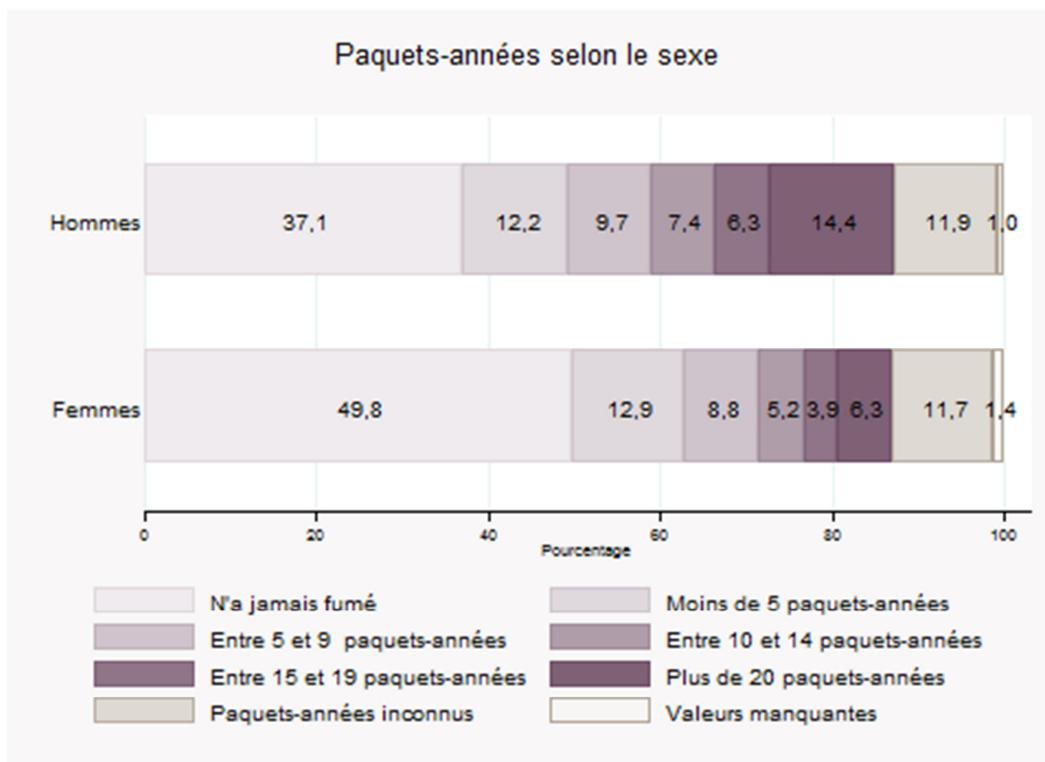
Situation matrimoniale



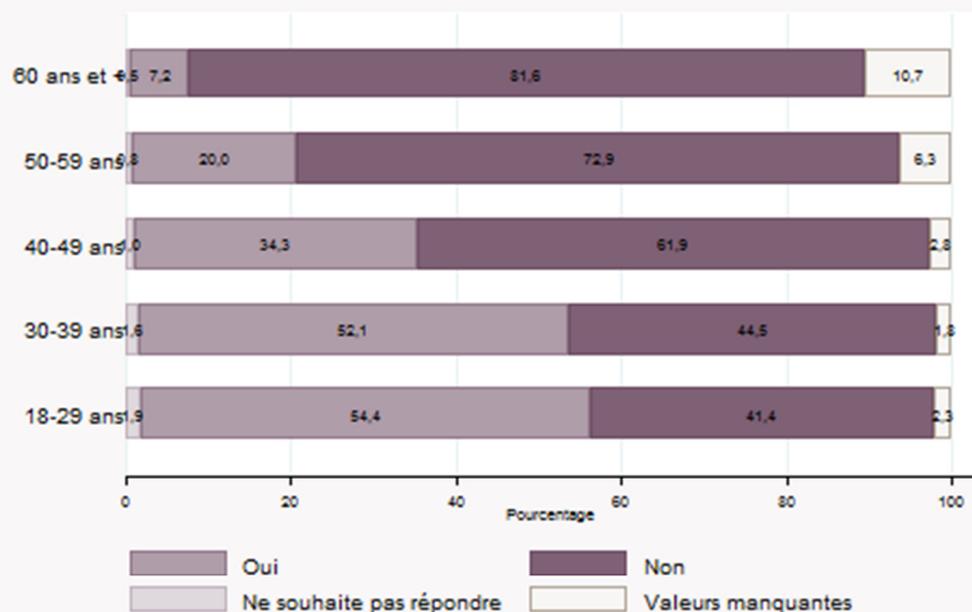
Renoncement à certains soins de santé pour des raisons financières dans les 12 derniers mois



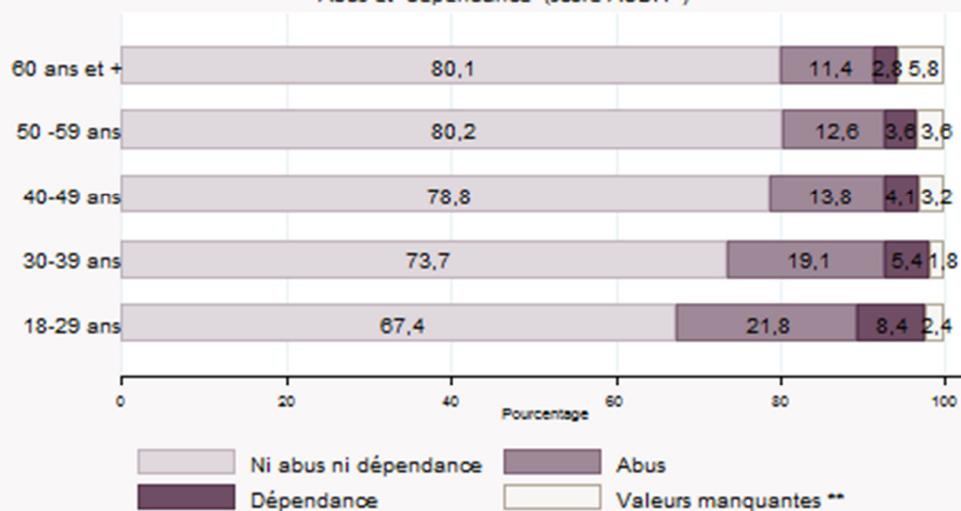
Comportements (tabac, alcool) : sont présentées ici les données de statut tabagique et de consommation de cannabis pour illustrer les données concernant les habitudes de vie : on observe une distribution diversifiée du statut tabagique et du nombre de paquets années. Cette diversité est également observée pour les consommations de cannabis et d'alcool, pour lesquelles nous présentons des données concernant l'abus et la dépendance (score AUDIT).



Consommation de cannabis selon l'âge



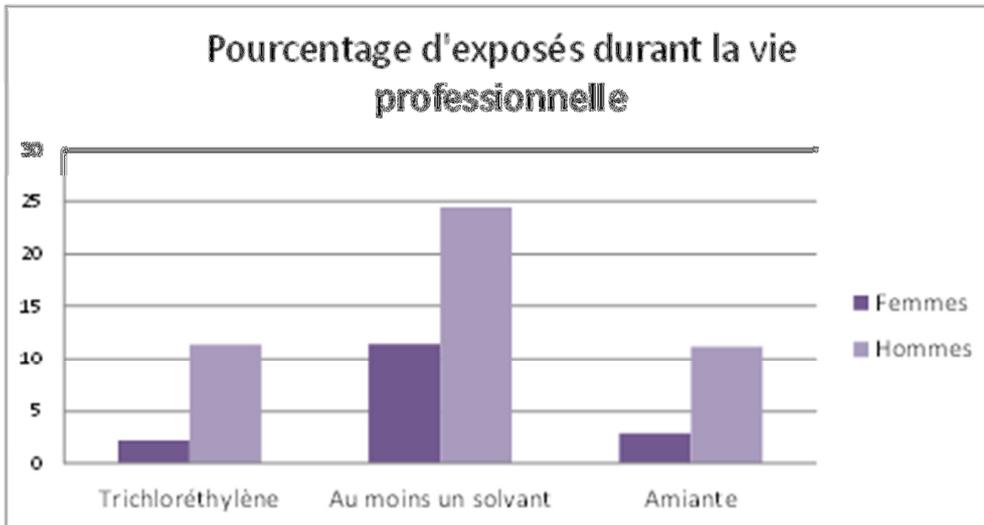
Consommation de boissons alcoolisées selon l'âge Abus et dépendance (score AUDIT*)



* Hommes => Abus : 8-12 Dépendance : 13 et +
Femmes => Abus : 7-11 Dépendance : 12 et +

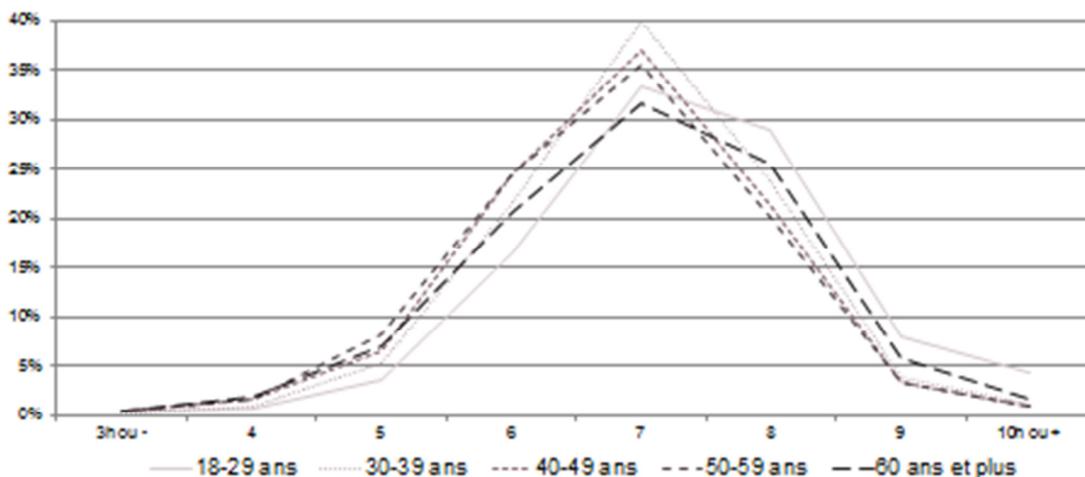
** Les valeurs manquantes prennent en compte les non-réponses et les scores non-calculables comptant moins de 4 réponses aux 10 items. 123 hommes et 370 femmes ont déclaré ne jamais avoir bu d'alcool au cours de leur vie et n'ont donc pas été pris en compte dans le calcul de l'Audit.

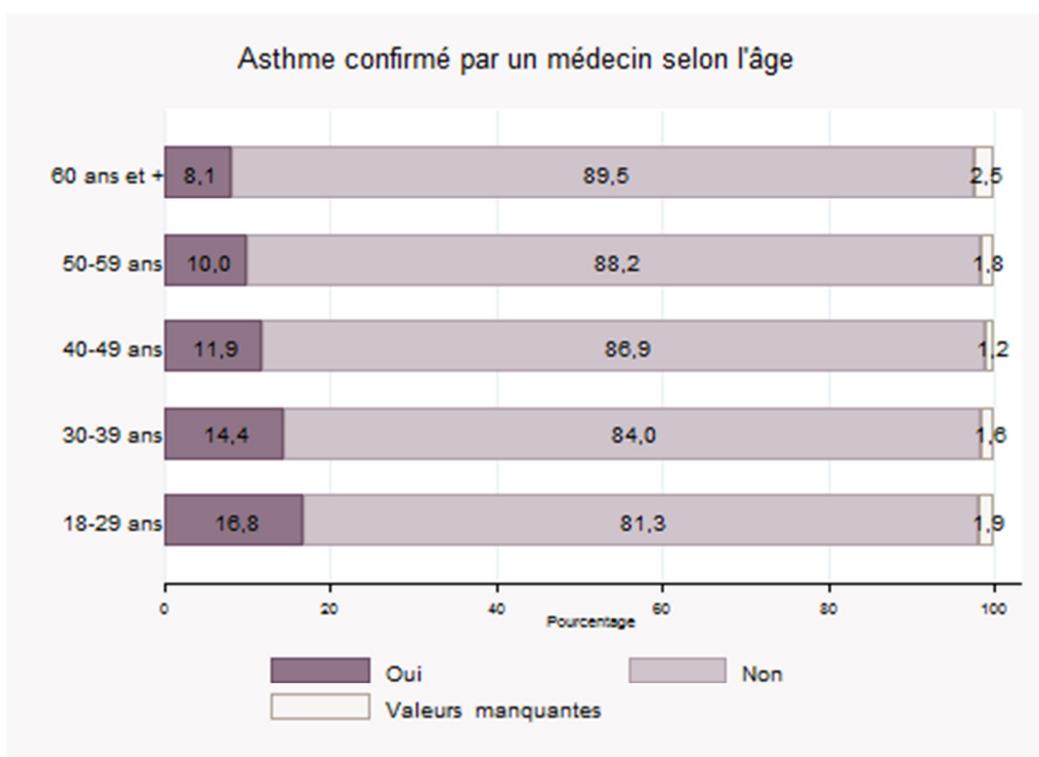
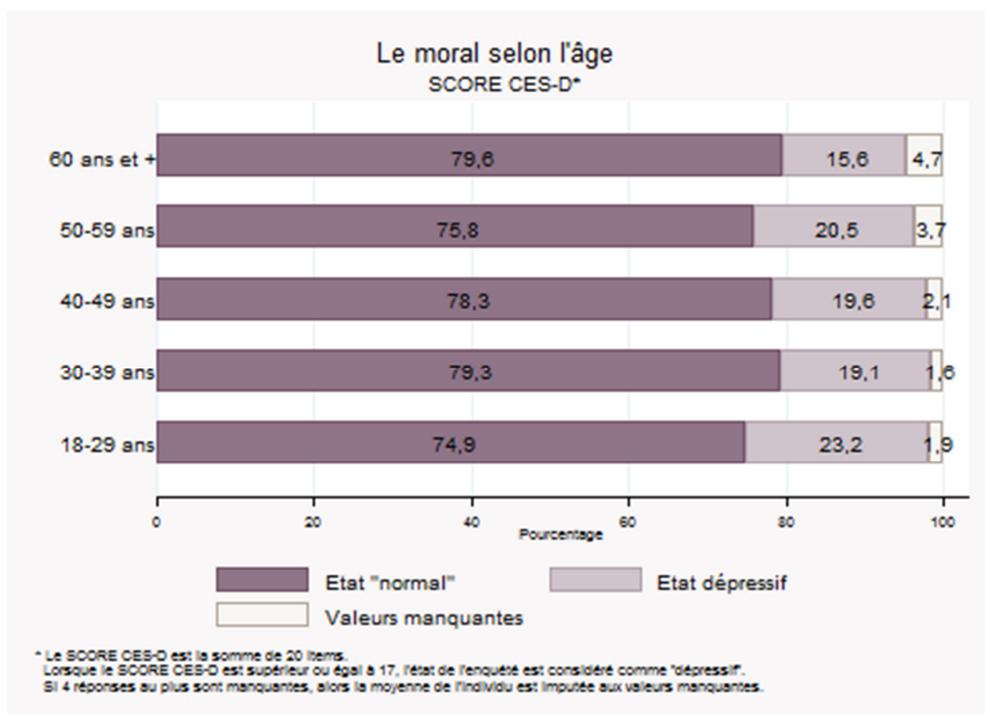
Facteurs professionnels : de nombreuses données professionnelles sont recueillies ; à titre d'exemple, la prévalence vie entière d'exposition aux solvants (exposition chimique la plus fréquente) montre une prévalence importante d'exposition à des nuisances en milieu de travail.



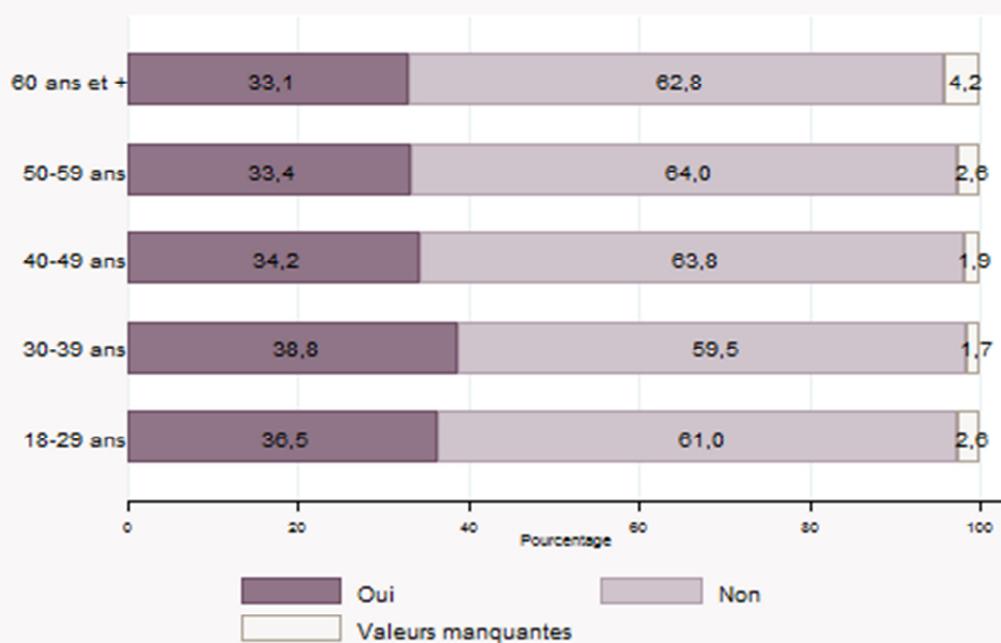
Santé : à titre d'illustration, on présente des données de distribution d'heures de sommeil, des données de prévalence de dépressivité (échelle CES-D), d'asthme et de rhinites, ainsi que de diabète. Il apparaît que les prévalences observées parmi les participants de CONSTANCES sont très proches des données nationales de référence disponibles. La prévalence d'autres pathologies chroniques est également proche de ce qui est attendu en population générale (non montré), ce qui permettra à la fois de fournir de bonnes estimations de prévalences et la réalisation d'études épidémiologiques dans ces domaines.

Nombre d'heures de sommeil selon l'âge

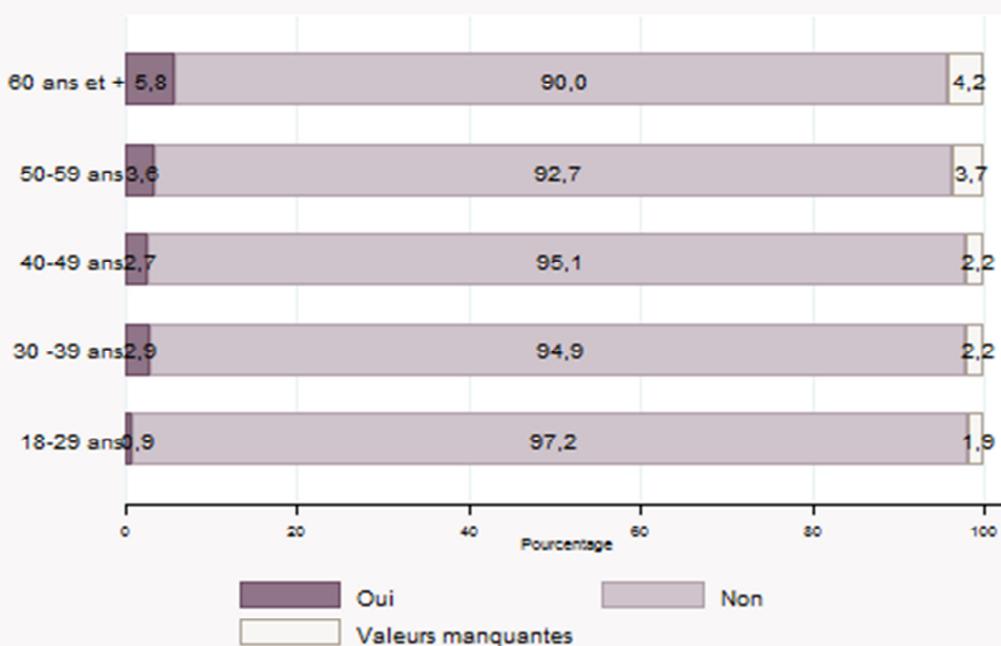




Allergies nasales selon l'âge



Diabète diagnostiqué par un professionnel de santé selon l'âge



2.2.3 Prise en compte des effets de sélection – Résultats préliminaires

Comme on l'a rappelé, certains effets bien connus de sélection à l'inclusion sont liés notamment à l'âge, au sexe et à la position sociale. Il est donc usuel, quand ces données sur les non-participants sont disponibles, de redresser les données observées selon ces critères. Cependant, ils sont insuffisants, dans un contexte épidémiologique, pour contrôler les biais de sélection potentiels, car on sait que l'état de santé joue également un rôle sur la participation. Mais il est rarement possible d'avoir des informations de santé sur les non-participants à prendre en compte dans les pondérations des échantillons.

Dans le cas de CONSTANCES, le calcul des coefficients de redressement pour non-participation repose sur la constitution de la cohorte « contrôle » de non participants pour lesquels on dispose des mêmes informations que pour les participants : caractéristiques sociodémographiques provenant de la Caisse nationale d'assurance vieillesse-Cnav (âge, sexe, statut d'activité et PCS), ainsi que de nombreuses informations de santé et de recours aux soins grâce aux données du SNIIRAM et du PMSI.

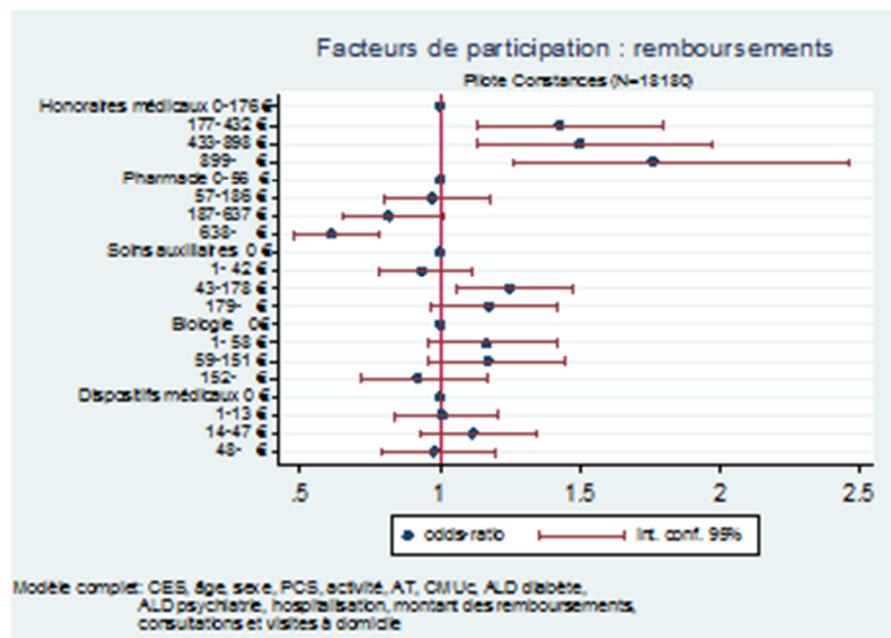
Les données de l'inclusion concernant les participants et les non participants de la phase pilote ont pu être appariées avec succès avec la base du SNIIRAM incluant le PMSI et les fichiers de la Caisse nationale d'assurance vieillesse-Cnav². Une analyse approfondie concernant les effets de sélection est en cours, dont quelques résultats préliminaires sont présentés ci-dessous. Ils illustrent les possibilités offertes par CONSTANCES pour l'estimation de la prévalence en population générale de certaines conditions de santé et de comportements grâce à l'utilisation des données concernant les non-participants.

Les résultats préliminaires présentés ici proviennent de l'analyse des données de remboursement de soins et de prise en charge de 2007-2008 (c'est-à-dire les 2 années avant l'inclusion de la phase pilote) issues du SNIIRAM-PMSI pour les participants et les non-participants (Guéguen *et al*, 2013).

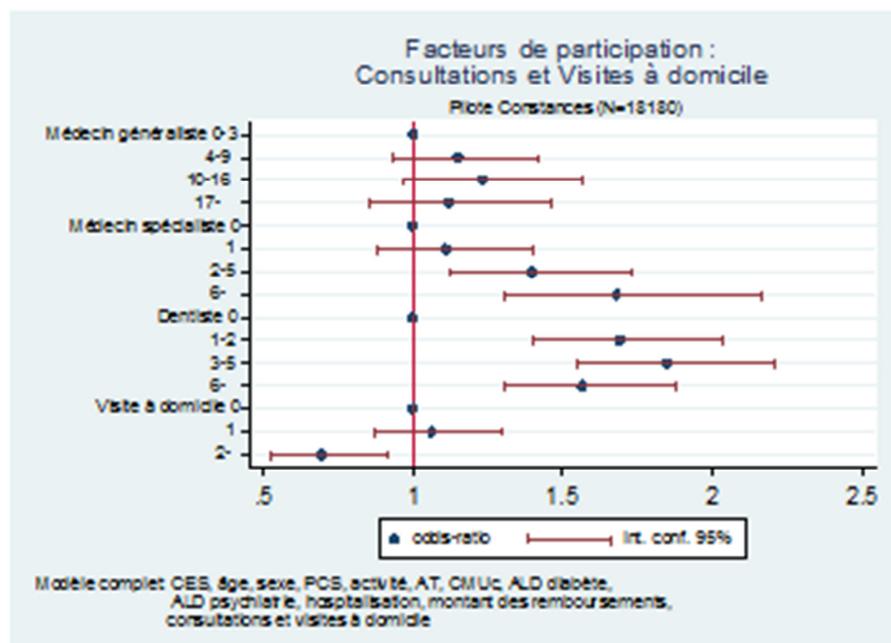
2.2.3.1 Facteurs liés à la participation

Les figures suivantes qui présentent des odds-ratios (mesure de l'association entre certaines caractéristiques et la participation à la cohorte CONSTANCES par un modèle de régression logistique dans lequel la variable à expliquer est la variable binaire « participation »), montrent que les participants sont plus fréquemment consommateurs de soins, mais que s'ils dépensent plus en honoraires médicaux et en consultations dentaires, ils dépensent moins en pharmacie, biologie, kinésithérapie ou de transports. Elles montrent également la participation en fonction des certaines consommations de soins. Les participants ont plus fréquemment recours à des soins, mais ceux-ci sont moins coûteux que les non participants, suggérant qu'ils portent attention à leur santé mais qu'ils ont moins de problèmes graves que les non participants.

² L'appariement des participants inclus en 2012 vient également d'être réalisé, mais les données n'ont pas encore été analysées.

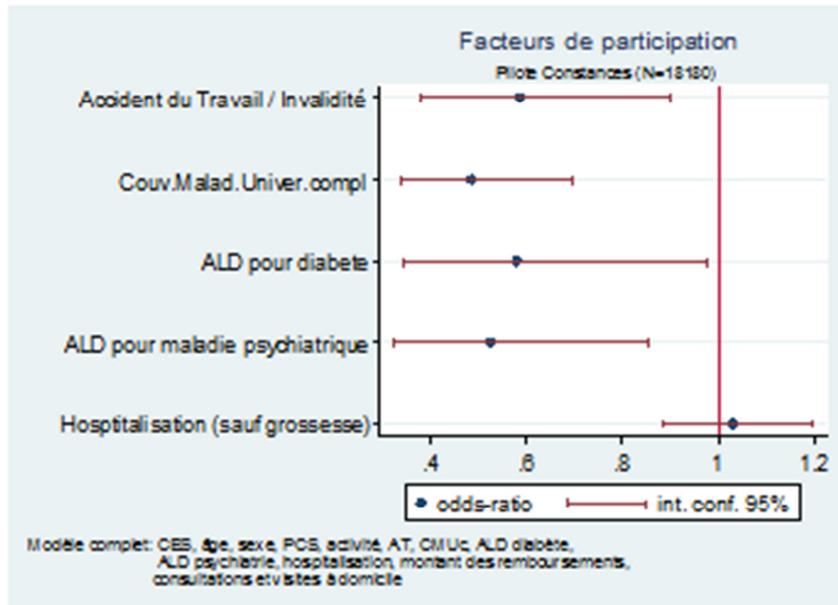


Symposium international sur les questions de méthodologie, 16-18 octobre 2013 Ottawa



Symposium international sur les questions de méthodologie, 16-18 octobre 2013 Ottawa

L'analyse des données médicalisées montrent que les non participants sont plus nombreux à être en affection longue durée (ALD) pour troubles psychiatriques et diabète.



Symposium international sur les questions de méthodologie, 16-18 octobre 2013 Ottawa

Au total, ces premiers résultats bien que préliminaires, montrent notamment des effets du statut social et professionnel et l'existence de différences de santé entre les participants et les non participants : les non participants sont plus remboursés pour des affections de longue durée, les participants vont plus souvent consulter un médecin que les non participants, mais vraisemblablement de façon préventive, car les dépenses totales des participants sont inférieures à celles des non participants. Les analyses plus précises du PMSI et notamment des actes techniques sont en cours et permettront de vérifier ces hypothèses.

2.2.3.2 Redressements pour l'estimation de prévalences en population

À partir des probabilités de participation ainsi obtenues, il a été possible de procéder à des redressements. À titre d'illustration on présente quelques résultats concernant l'estimation de la prévalence en population de certaines caractéristiques de comportement et de santé.

Dans la première colonne figurent les prévalences brutes observées parmi les participants de la cohorte CONSTANCES ; dans la seconde colonne, on a ajusté sur les variables sociodémographiques usuelles ; dans la troisième colonne, on a ajouté dans le modèle des données de santé issues du SNIIRAM (en rouge : pourcentages de modification par rapport aux prévalences brutes).

	Prévalences brutes	+ Sexe, Âge, PCS	+ SNIIRAM*
Ex-fumeur	28,6 [26,2 -30,1]	26,0 [23,5 -28,4] (- 10 %)	25,2 [22,4-28,0] (- 12 %)
Abstinent	18,3 [16,3-20,3]	20,5 [18,0-22,9] (+ 4 %)	21,3 [18,3-24,2] (+ 16 %)
Obésité	10,1 [8,5 -11,6]	10,8 [9,0 -12,7] (+ 7 %)	12,3 [9,8 -14,7] (+ 21 %)
Glycémie à jeun >7mmol	2,2 [1,5-3,0]	2,1 [1,3-2,6] (- 5 %)	3,5 [2,0-5,0] (+ 59 %)

* Prestations, ALD, PMSI

On obtient des modifications importantes de l'estimation des prévalences grâce aux redressements réalisés, qui s'expliquent par la prise en compte des effets de sélection : en effet, les ex-fumeurs (personnes certainement soucieuses de leur santé) ont une probabilité de participer plus élevée que l'ensemble de la population obtenue par tirage au sort, alors que les personnes abstinentes (la plupart pour raison de santé), les obèses et les diabétiques ont une probabilité de participation plus faible.

Il faut noter que ce travail a nécessité l'acquisition de compétences spécifiques pour la prise en main des bases de données de la Cnav et du SNIIRAM, qui sont d'une grande complexité et encore rarement utilisés à cette échelle dans ce contexte.

2.2.4 Procédures de fidélisation

Un contact régulier avec les participants est un élément important pour assurer la fidélisation tout au long du suivi de la cohorte. Ce contact prend la forme d'un « Journal de la cohorte CONSTANCES » présentant les résultats acquis, les projets associés, etc., adressé une fois par an aux participants (2 numéros déjà réalisés). Le site Internet de CONSTANCES³ a des fonctions voisines, et permet un contact direct avec les participants. L'équipe traite environ une cinquantaine de mails par mois. Un Numéro Vert est mis à disposition, et reçoit environ entre 5 et 10 appels par jour. Une Newsletter est régulièrement envoyée aux personnes s'abonnant sur le site.

Un premier envoi de questionnaire de suivi annuel a été envoyé aux participants de la première vague de 2012 ; le taux de retour est excellent : 75 % avant relance, ce qui montre un grand intérêt des volontaires pour CONSTANCES.

2.2.5 Le système d'information CONSTANCES

La base de données Constances est particulièrement complexe présentant des caractéristiques qui la rendent hautement sensible :

³ <http://www.constances.fr/>

- La volumétrie des effectifs à inclure est élevée : 200 000 volontaires sont attendus dans la cohorte, ce qui compte tenu d'un taux de participation estimé à 10 %, amène à inviter 10 fois plus et en tirer au sort encore 4 fois plus pour constituer la cohorte témoins ;
- Elle est constituée de données individuelles « à caractère personnel » au sens de la loi Informatique et libertés, c'est-à-dire qu'il est possible d'identifier les personnes concernées (données nominatives détenues par le Tiers de Confiance) ; de plus, elle contient de nombreuses données sensibles ;
- Certaines des données collectées à l'échelle individuelle proviennent d'appariements avec des bases de données nationales (SNIIRAM, Cnav, CépiDc), qui font elles-mêmes l'objet d'une protection renforcée ;
- La multiplicité des flux de données (questionnaires papier, outil WEB, fichiers en provenance des bases de données nationales ...) et les fréquences de traitement (au fil de l'eau, hebdomadaire, trimestriel et annuel) augmentent la complexité du suivi ;
- L'ouverture à la communauté dans un but de recherche : il faut maintenir un catalogue de données à jour et conserver l'historique pour garantir la fiabilité des données qui seront utilisées pendant une très longue période.

Ces caractéristiques, ainsi que la très longue durée de collecte et de conservation des données, imposent des contraintes très strictes pour la collecte et la gestion des données. Les procédures mises en place se doivent de respecter les textes législatifs et réglementaires destinés à préserver une confidentialité et une sécurité de haut niveau concernant les données à caractère personnel. Aucune transmission ou publication de données permettant d'identifier les participants n'est possible sans autorisation préalable des personnes concernées et de la Cnil.

2.2.5.1 Flux de données

La base de données Constances est enrichie via des flux de données issus des process suivants :

- « Du tirage au sort à l'inclusion » pour la constitution de la cohorte de participants, *(on détaillera uniquement ce process à titre d'illustration de la complexité du SI CONSTANCES)*
- « Suivi passif » pour le recueil des données à travers les systèmes d'information nationaux,
- « Suivi actif » pour le suivi annuel des participants par envoi d'auto-questionnaires.

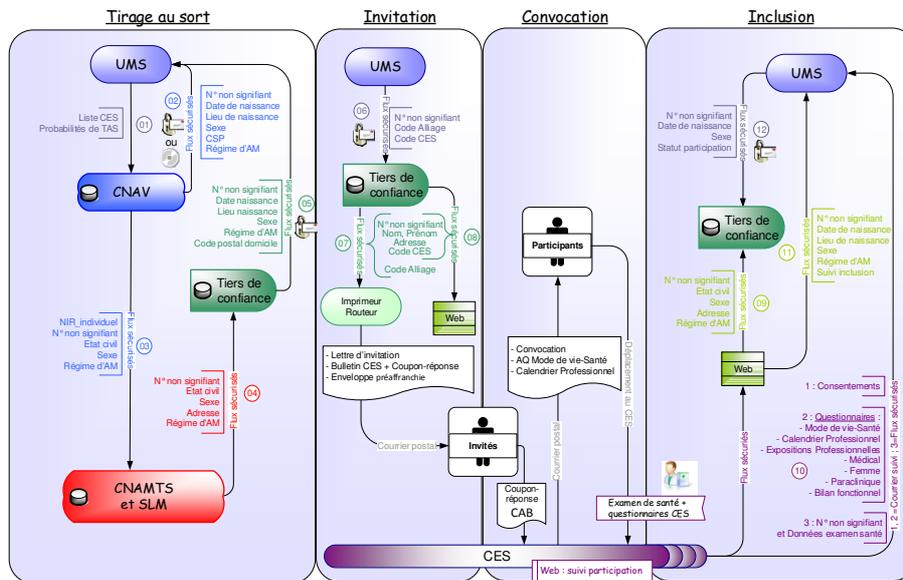
Ces trois principaux process font appel aux intervenants suivants :

- **UMS 011** : elle assure le pilotage des flux et la sécurisation des données.
- **CES** : 17 centres répartis sur toute la France accueillent les volontaires, réalisent l'examen et transmettent les données à l'UMS.
- **CNAV** : elle assure le tirage au sort et fournit annuellement des données individuelles pour les sujets inclus dans Constances.
- **CNAMTS** : elle fournit annuellement les adresses des tirés au sort des affiliés au RGSS et les données individuelles.
- **SLM/CAMIEG** : elles fournissent également les adresses des tirés au sort de leurs affiliés.
- **CépiDc** : il fournit les causes de décès.

- **Tiers de Confiance (TC)** : il assure le stockage des adresses et envoie trimestriellement les listes des invitations à l'imprimeur/routeur.
- **Imprimeur/Routeur (IR)** : il envoie aux invités des documents (questionnaires, coupons-réponse de participation à Constances...).
- **La Poste** : traite les PND (flux quotidien) ainsi que les déménagements (flux Optimis, à la demande).
- Sujets participants à la cohorte : ils se rendent en CES pour réaliser un examen de santé et répondent aux questionnaires de suivi 1 fois par an.

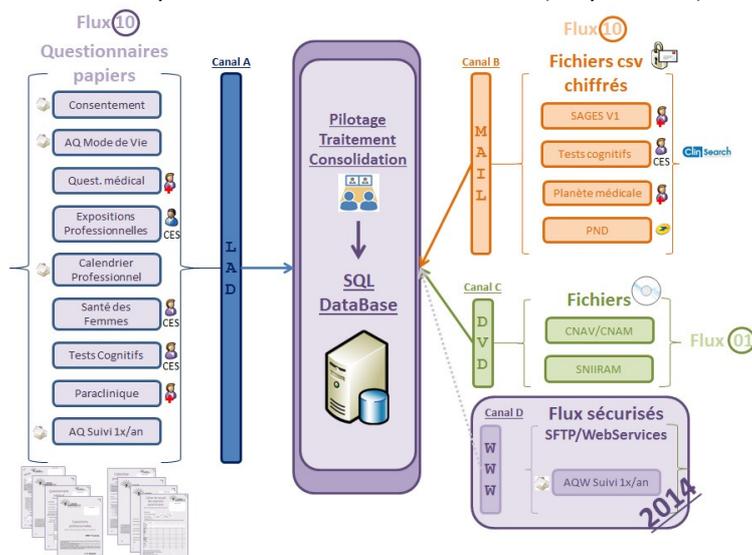
Le schéma ci-après décrit l'enchaînement des traitements de données lors de l'inclusion des volontaires dans Constances :

- **TAS :**
 - **Flux 01** : l'UMS envoie à la CNAV le plan de sondage (probabilités de TAS par CES).
 - **Flux 02** : la CNAV renvoie à l'UMS les données non nominatives à l'UMS.
 - **Flux 03** : la CNAV transmet à la CNAMTS/SLM cette population tirée au sort afin d'en extraire les adresses.
 - **Flux 04** : la CNAMTS et les SLM transmettent les adresses au TC.
 - **Flux 05** : le TC transmet à l'UMS les données non nominatives extraites par la CNAMTS/SLM.
- **Invitations :**
 - **Flux 06** : l'UMS réalise le tirage au sort et envoie un fichier pour les invitations mensuelles.
 - **Flux 07** : le TC envoie à l'IR les adresses pour les invitations.
 - **Flux 08** : le TC alimente l'application WEB de suivi des sujets en CES.
- **Convocation :**
 - le participant se présente au CES pour réaliser son examen.
 - le CES saisit dans l'application de suivi la réalisation (ou non) de l'examen.
- **Inclusion :**
 - **Flux 09** : le TC récupère de l'application WEB les informations nominatives des ajoutés au CES.
 - **Flux 10** : le CES envoie à l'UMS le consentement et les questionnaires des participants.
 - **Flux 11** : l'UMS récupère de l'application WEB les statuts des Examens de Santé.
 - **Flux 12** : l'UMS actualise les statuts de participation vers le TC.

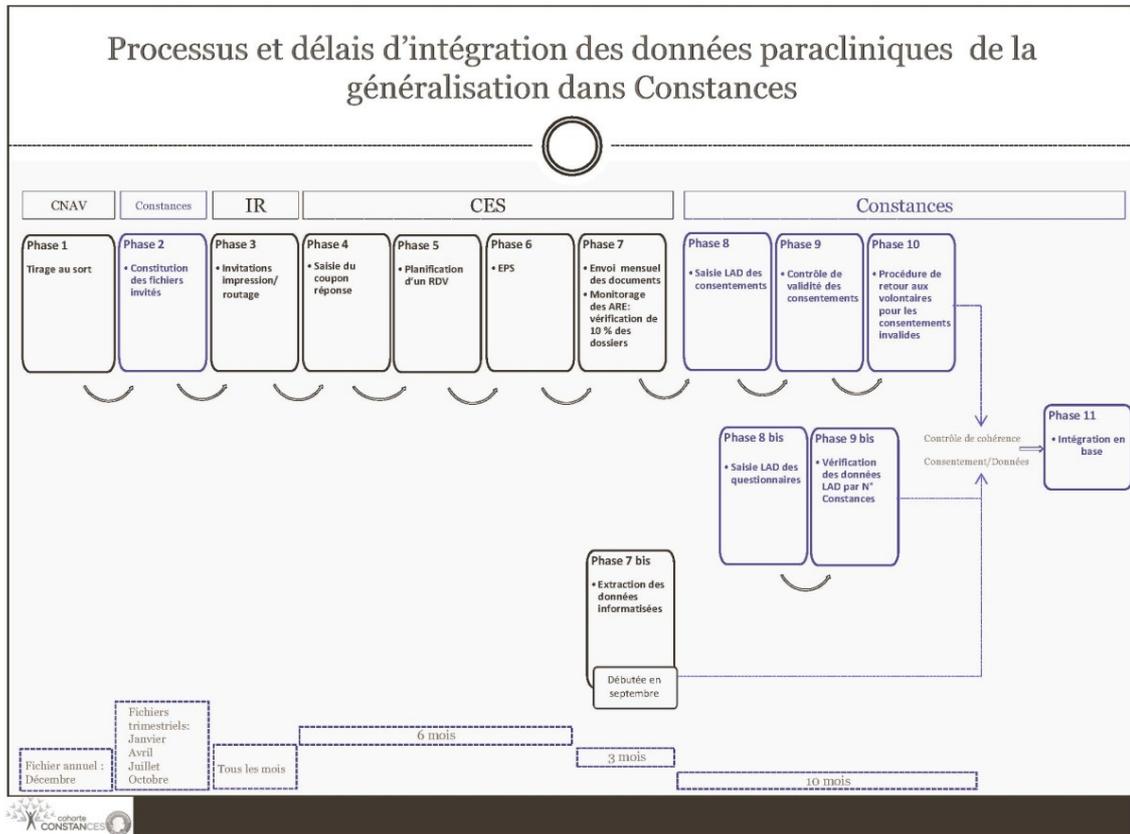


Le schéma suivant décrit les flux de données traités lors la phase d'inclusion :

- **Canal A** : les documents papier envoyés par les CES sont traités en LAD (Lecture Automatisée de Documents), et remontent automatiquement en base de données après vérification opérateur si besoin.
- **Canal B** : certains flux arrivent par fichiers cryptés et sont intégrés automatiquement dans la base (par exemple : les PND ou bien les données issus des tests cognitifs).
- **Canal C** : les données volumineuses en provenance de la Cnav et de la CNAMTS arrivent sous format de DVD, les fichiers sont également cryptés. Une intervention opérateur est nécessaire pour déposer ces fichiers en entrée de la base.
- **Canal D** : les flux WEB vont faire leur apparition à horizon mi 2014 avec la dématérialisation des questionnaires de suivi annuel (cf. plus loin).



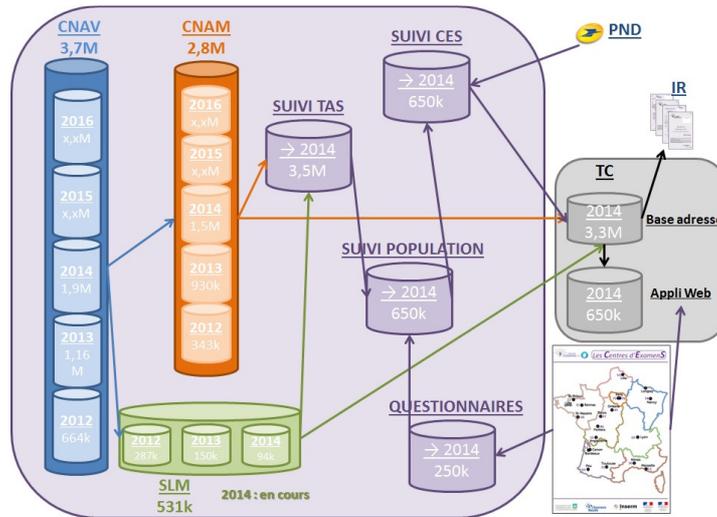
Une période assez longue sépare le moment du tirage au sort et l'intégration des données dans la base CONSTANCES du fait des nombreuses opérations à réaliser, comme illustré par le schéma suivant.



2.2.5.2 Architecture générale

La base de données Constances est en réalité un ensemble de bases qui sont reliées entre elles par des flux logiques dont les traitements sont asynchrones, c'est ce qui fait toute la complexité du système. L'inclusion des 200 000 volontaires sur une durée de 5 ans apporte une difficulté supplémentaire quant à la gestion de la base.

Le schéma ci-dessous cartographie de manière succincte les connexions entre les différentes bases qui constituent le système d'information Constances.



2.2.5.3 Les applications WEB

Un important effort de dématérialisation des flux de données a été mis en œuvre et se poursuit pour améliorer le suivi des participants et remplacer progressivement le recueil de données sur documents papier. Ainsi, le transfert par Internet des données de l'examen neuropsychologique est déjà opérationnel, et des applications pour le transfert des résultats de l'examen paraclinique et pour le suivi des inclusions par WEB sont en cours de déploiement dans les 17 CES. Un marché, préparé en liaison avec la Cnil, a été notifié en décembre 2013 pour le développement d'une plateforme pour la saisie en ligne par les volontaires des questionnaires de suivi annuel. Cette application devrait être opérationnelle courant 2014. Enfin, nous étudions actuellement la mise en place d'un portail unifié pour l'accès à la base CONSTANCES.

Les Applications Web dans Constances

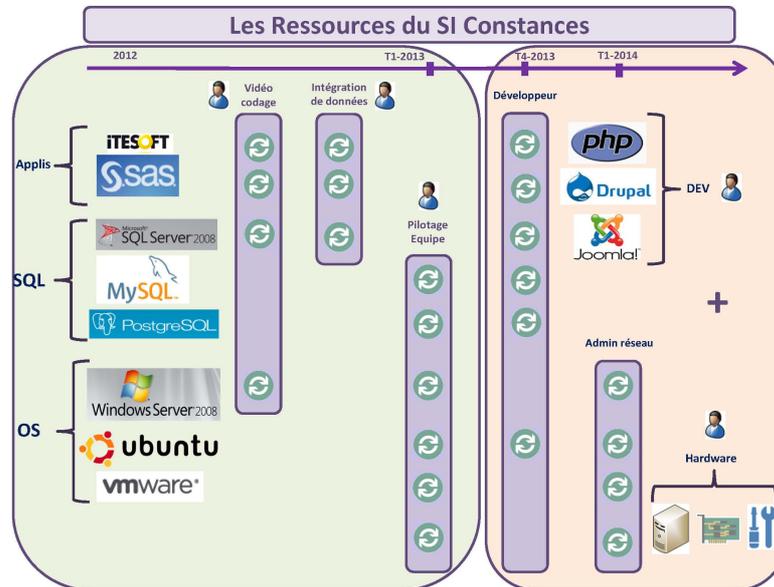
1 - Site Web Constances.fr

2 - Appli de suivi des CES

3 - Questionnaire de suivi en ligne

4 - Le futur Portail

Un SI aussi complexe que celui de CONSTANCES fait appel à diverses ressources et outils informatiques, comme schématisé ci-dessous.



Finalement, nous prévoyons de demander une certification ISO à moyen terme, lorsque les développements en cours seront achevés.

2.2.6 La Biobanque de CONSTANCES

La biobanque est un élément majeur de la cohorte. Nous avons procédé, en lien avec l'Infrastructure nationale BIOBANQUES, à une analyse détaillée des caractéristiques scientifiques et techniques de la biobanque, que nous envisageons de mettre en place à partir de 2014. Cette étude systématique de différents scénarii a fait l'objet d'une publication aux éditions Lavoisier⁴.

Nous finalisons actuellement avec l'aide de l'Infrastructure BIOBANQUES un appel d'offres pour « Prestations de services et de fournitures pour la biobanque CONSTANCES ». Le choix de la structure retenue pour l'installation et le fonctionnement de la biobanque de CONSTANCES sera fait en lien avec l'Infrastructure BIOBANQUES.

La biobanque de CONSTANCES fait partie du consortium européen BBMRI (*Biobanking and Biomolecular Resources Research Infrastructure*).

2.2.7 La plateforme de validation des évènements de santé

Pour les données de santé extraites des bases de données nationales se pose un important problème de validation. On sait en effet que, pour des raisons diverses, les données concernant des problèmes de santé réunies au sein du SNIIRAM ne sont pas toutes fiables. Ceci est

⁴ Henny J, Goldberg M, Zins M. Guide pour la constitution d'une Biobanque associée aux études épidémiologiques en population générale. Paris : Inserm-Lavoisier, Collection Tec & Doc, novembre 2012.

notamment le cas pour les diagnostics médicaux précis ; or les contraintes d'un suivi épidémiologique de qualité nécessitent une validation rigoureuse. Une attention particulière est donc portée à la validation des diagnostics extraits des bases de données médico-administratives, qui feront l'objet d'un contrôle systématique. La procédure prévue repose sur l'utilisation de plusieurs sources : déclarations des individus (*via* les questionnaires), données collectées par les CES, données fournies par les bases médico-administratives. Les données collectées aux diverses sources citées sont considérées comme des signalements de pathologies potentielles, qui devront être confirmées par d'autres moyens, dans le cadre d'une validation directe individuelle.

Une validation directe individuelle des événements de santé sera déclenchée à partir des signalements trouvés dans les sources citées, notamment dans le SNIIRAM qui offre l'avantage de l'exhaustivité et qui est indépendante des sujets, donc insensible aux biais de déclaration. Les données extraites régulièrement des bases du SNIIRAM concernent les Affections de longue durée (ALD), les Résumés de sortie anonymisés (RSA) des hôpitaux qui contiennent notamment des données médicales (diagnostic principal et ensemble des diagnostics associés et des actes pratiqués), les maladies professionnelles et les données exhaustives détaillées de consommation de soins. Aucune de ces sources ne permettra, ni de confirmer, ni de dater avec une validité suffisante des événements de santé incidents durant le suivi. Il est donc indispensable de pouvoir accéder à des documents médicaux complémentaires permettant à des experts médicaux indépendants, sur la base des documents collectés, d'aboutir à un diagnostic précis. Les données médicales recueillies seront mises en forme et soumises à des Comités d'experts indépendants pour validation des événements (Comités externes de validation) à partir de grilles préétablies de définition des pathologies et de critères de définition. L'élaboration des critères de validation d'événements cliniques est en cours, notamment dans le cadre d'un Work Package du Consortium européen de cohortes *BBMRI-Large Prospective Cohorts* auquel CONSTANCES participe (*WP6-Harmonisation of clinical endpoint data*). Dans un premier temps, la procédure de validation portera sur trois groupes de pathologies : cancer, cardio-vasculaires et neurodégénératif ; ultérieurement, ces procédures seront étendues à d'autres pathologies.

La quasi-totalité des participants a donné son consentement écrit pour permettre l'accès à leurs médecins et à leurs dossiers médicaux. Du fait du nombre important d'événements médicaux attendus, les opérations de contact avec les sujets, leurs médecins et les hôpitaux sont confiées à un organisme externe ; une société a été sélectionnée après un marché public et la procédure sera mise en œuvre courant 2014.

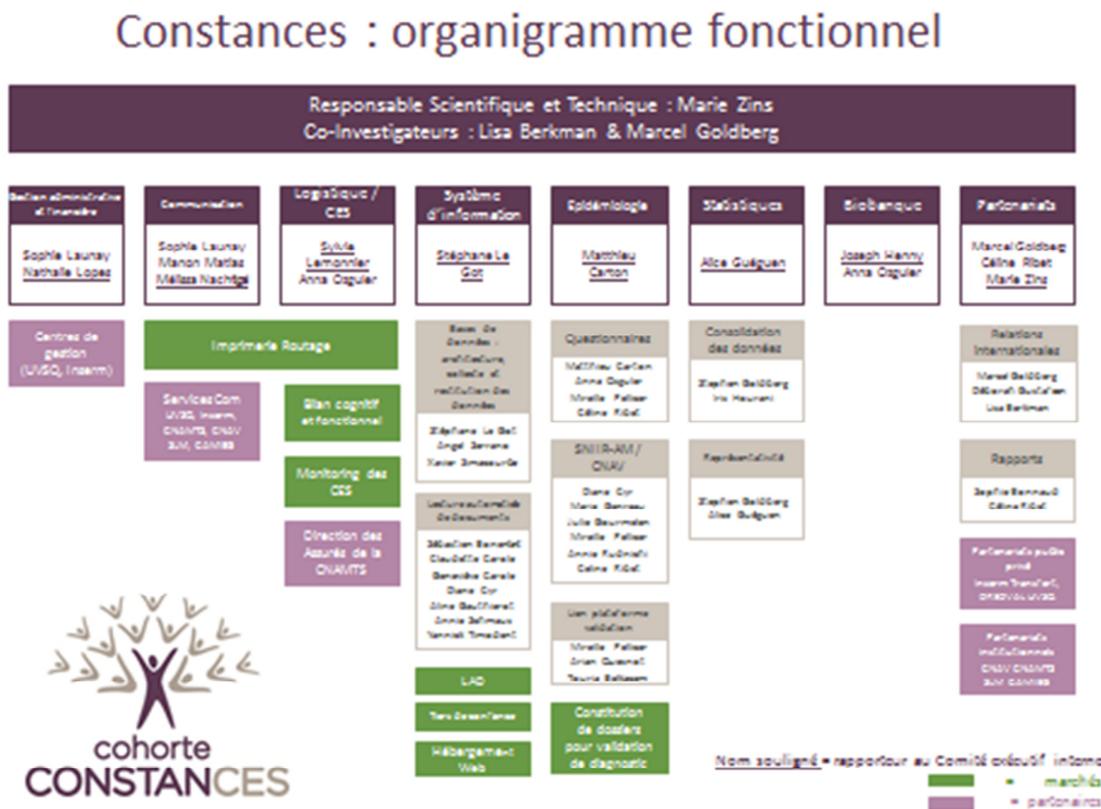
3 ORGANISATION DU PROJET

CONSTANCES a été lauréat de l'appel à projets Infrastructures 2011 des Investissements d'avenir, et a été labellisé à ce titre « Infrastructure nationale en biologie et santé ». Le contrat correspondant entre l'ANR et l'Université de Versailles Saint Quentin en Yvelines (UVSQ), établissement coordinateur du projet, a été signé en octobre 2012. En octobre 2013, par avenant au contrat ANR-UVSQ, l'Inserm est devenu un partenaire de l'Infrastructure.

Dans ce contexte, les instances de gouvernance de CONSTANCES ont été modifiées en tenant compte des recommandations du jury et du comité de pilotage de l'appel à projets Infrastructures.

3.1 LES INSTANCES DE GOUVERNANCE

Équipe en charge de CONSTANCES : elle est structurée selon les principales activités de l'infrastructure, selon l'organigramme suivant.



CONSTANCES est sous la responsabilité scientifique et technique de l'Unité mixte de service Inserm-UVSQ « Cohortes épidémiologiques en population » (UMS 011).

Porteurs du projet : aux actuels PI et Co-PI, s'est adjointe à titre de Co-PI le Professeur Lisa Berkman, directrice du *Harvard Center for Population and Development Studies* de la *Public Health School* de l'Université Harvard.

Comité exécutif : la gestion scientifique, technique et administrative de CONSTANCES est assurée par un Comité exécutif composé du PI, des deux co-Pis et des responsables des domaines d'activité de l'infrastructure (noms soulignés dans l'organigramme).

Comité scientifique interne : composé de chercheurs extérieurs à l'équipe CONSTANCES, il a pour rôle d'aider celle-ci dans les décisions et les choix scientifiques courants, notamment concernant les données à collecter. Sa composition permet de couvrir les principales thématiques abordées par la cohorte.

Instances de gouvernance externes :

Comité de pilotage institutionnel : du fait de la création tardive de l'UMS 011 qui donne un cadre institutionnel clair au projet, la mise en place de cette instance qui doit associer les organismes partenaires de CONSTANCES (CNAMTS, Cnav, DGS, Inserm et UVSQ) n'est pas encore effective. Un accord de consortium est en cours de finalisation.

Conseil scientifique international indépendant : le Conseil scientifique international est composé de 14 personnalités scientifiques, dont la moitié d'étrangers, couvrant les principales thématiques de CONSTANCES.

Comité d'éthique : actuellement, c'est le Comité d'éthique de l'Inserm qui joue ce rôle.

3.2 MODALITÉS D'ACCÈS À L'INFRASTRUCTURE

Toutes les équipes de recherche, françaises ou étrangères, publiques et privées, sont éligibles pour utiliser l'infrastructure CONSTANCES. Les demandes d'accès aux données de la cohorte se font dans le cadre d'appels à propositions (AAP) qui seront diffusés deux fois par an. Le texte de l'appel à projet a été rédigé ainsi qu'une Charte à destination des utilisateurs (versions française et anglaise). Les procédures d'évaluation des demandes ont été fixées (évaluation par le Conseil scientifique international, accord par le Comité de pilotage institutionnel).

Un premier AAP « réservé » a été lancé en mai 2013 à destination des chercheurs ayant participé à la préparation du protocole de CONSTANCES. Fin mars 2014, 14 dossiers, réunissant 33 équipes ont été proposés ; ils sont actuellement en cours d'examen par le Conseil scientifique. D'ores et déjà, plusieurs projets ont obtenu des financements de divers organismes (Anses, ANSM, IReSP). Fin 2014, il sera mis à disposition des chercheurs une première base correspondant à environ 25 000 sujets, incluant l'ensemble des données recueillies, y compris celles du SNIIRAM et de la Cnav. Le premier AAP « public » sera lancé fin 2014.

4 PARTENARIATS

4.1 PARTENARIATS AVEC DES ORGANISMES PUBLICS

Des partenariats avec divers organismes publics et agences de santé ont été établis, ou sont en cours de discussion. Ils portent essentiellement sur la possibilité pour ces organismes d'utiliser l'infrastructure CONSTANCES pour leurs besoins spécifiques.

Parmi les agences de santé, une convention a été signée avec l'ANSM ; une convention cadre et une convention « fille » avec le Département Santé Travail sont en cours de signature. Des discussions sont en cours avec l'INRS, la HAS et l'INCa. Parmi les organismes à vocation sanitaire et sociale, des discussions sont actuellement en cours avec la MGEN, la MFP et la MFP-Services, la Camieg, l'INJEP.

4.2 PARTENARIATS PUBLICS- PRIVÉS

Nous travaillons avec Inserm-Transfert (filiale de l'Inserm) pour la mise en place de partenariats public- privés (PPP) avec les entreprises de santé, ainsi qu'avec de la Direction de la valorisation de l'UVSQ pour les autres secteurs économiques.

L'établissement de PPP avec des industriels du secteur de la santé est largement entamé : un projet de contrat-type a été élaboré et cinq entreprises ont déjà signé un accord de confidentialité et une lettre d'intention ; les négociations sont très avancées avec plusieurs autres compagnies.

4.3 PARTENARIATS INTERNATIONAUX

La cohorte CONSTANCES participe à plusieurs consortiums internationaux associant des cohortes en population visant à harmoniser les données et à faciliter le partage de celles-ci.

Il existe une collaboration privilégiée avec la Cohorte nationale allemande dont les objectifs et le protocole sont très voisins de ceux de CONSTANCES⁵. CONSTANCES fait par ailleurs partie du consortium *BBMRI-Large-Scale Prospective Cohorts*⁶ qui réunit les principales cohortes en population européennes, comme le montre la figure ci-dessous.



⁵ www.nationale-kohorte.de

⁶ www.globalsciencecollaboration.org/public/site/PDFS/biobanks/Perola%20M.%20BBMRI-Large%20Prospective%20Cohorts.pdf

5 PUBLICATIONS-COMMUNICATIONS

5.1 PUBLICATIONS DANS DES REVUES À COMITÉ DE LECTURE

Carton M, Leclerc A, Plouvier S, Herquelot E, Zins M, Goldberg M, Roquelaure Y, Descatha A. Description of musculoskeletal disorders and occupational exposure from a field-pilot of a large population-based cohort (CONSTANCES). *J Occup Environ Med*. 2013, 55:859-61.

Zins M, Berkman LF, Goldberg M and the CONSTANCES team. The CONSTANCES cohort, an epidemiological research infrastructure. Methods and results of the pilot phase. *Epidemiology Biostatistics and Public Health*. 2013, 10: e8921-1-e8921-1 DOI: 10.2427/8921.

Ribet C, Bonenfant S, Serrano A, Coeuret-Pellicier M, Goldberg M, Zins M. Flux sécurisés de données à caractère personnel dans la cohorte CONSTANCES. *J Gestion Econ Med*. 2012, 30: 395-409.

Fortier I, Doiron D, Little J, Ferretti V, Stolk RP, Knoppers BM, Hudson TJ, Burton PR, on behalf of the International Harmonization Initiative. Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. *Int. J. Epidemiol*, 2011, doi: 10.1093/ije/dyr106.

Cambon-Thomsen A, Thorisson GA, Mabile L, Andrieu S, Bertier G, Boeckhout M, Carpenter J, Dagher G, Dalgleish R, Deschênes M, di Donato JH, Filocamo M, Goldberg M, Hewitt R, Hofman P, Kauffmann F, Leitsalu L, Lomba I, Mabile L, Melegh B, Metspalu A, Miranda L, Napolitani F, Oestergaard MZ, Parodi B, Pasterk M, Reiche A, Rial-Sebbag E, Rivalle G, Rochaix P, Susbielle G, Tarasova L, Thomsen M, Thorisson GA, Zawati MH, Zins M; BRIF workshop group. The role of a Bioresource Research Impact Factor as an incentive to share human bioresources. *Nat Genet*. 2011, 43: 503-504.

Zins M, Bonenfant S, Carton M, Coeuret-Pellicier M, Guéguen A, Gourmelen J, Nachtigal M, Ozguler A, Quesnot A, Ribet C, Rodrigues G, Serrano A, Sitta R, Brigand A, Henny J, Goldberg M. The CONSTANCES Cohort: an Open Epidemiological Laboratory. *BMC Public Health* 2010; 10:479.

5.2 COMMUNICATIONS

Zins M. La cohorte CONSTANCES, une infrastructure pour la recherche et la santé publique. Atelier Epilepsie - ITMO Neurosciences, Paris 19 décembre 2013.

Zins M. The CONSTANCES cohort: an infrastructure for research and public health. Developmental determinants in chronic diseases and ageing - From research to novel policies and value creation. Montpellier, 2-3 December 2013.

Guéguen A, Sitta R, Rudnichi A, Gourmelen J, Cyr D, Goldberg M, Zins M. Utilisation conjointe de données d'enquête et des bases de données administratives nationales pour estimer des fréquences de paramètres de santé. L'exemple de la cohorte Constances. *Statistique Canada, Symposium international de 2013 sur les questions de méthodologie - Produire des estimations fiables à partir de bases imparfaites*, Ottawa 15-18 octobre 2013.

Zins M. La cohorte Constances : appariements avec le SNIIR-AM et le SNGC (CNAV). Séminaire Systèmes d'information, Haut Conseil de la Santé Publique, Ministère de la Santé, Paris, 12 février 2013 (Conférence invitée).

Zins M., La cohorte Constances : mise en place des partenariats public privé, in 6ème Forum scientifique de pharmaco épidémiologie 2013: Paris (Conférence invitée).

Zins M, Ribet C, Coeuret-Pellicier M, Bonenfant S, Serrano A, Rodrigues G, Goldberg M. Appariements de la cohorte Constances aux bases de données médico-administratives françaises. *Congrès Adelf-Epiter*, Bruxelles, 12-14 Septembre 2012.

Sitta R, Guéguen A, Gourmelen J, Cyr D, Zins M and the Constances team. Use of linked registries in the design of cohort studies, a tool against selection bias: the Constances example. 33rd Annual Conference of the International Society for Clinical Biostatistics, 19 – 23 August 2012, Bergen.

Ribet C, Bonenfant S, Serrano A, Rodrigues G, Goldberg M, Zins M. Flux sécurisés de données à caractère personnel dans le domaine de la santé : L'exemple de la cohorte CONSTANCES. IIIe Congrès National ADELFF / ÉMOIS, Dijon, 12-12 mars 2012.

Zins M. Étude Constances : une cohorte pour la recherche et la santé publique. Conférences de Fondamental-Chenevier. Créteil, 7 Février 2012 (Conférence invitée).

Henny J, Ozguler A, Goldberg M, Zins M. The biobank of the population-based French CONSTANCES Cohort. ESBB Conference, 16-19 November 2011, Marseille.

Zins M. Constances une cohorte pour la recherche et la surveillance. Journée de l'InVS. Veille et Surveillance : quelles données pour l'action publique. 28 et 29 avril 2011. Paris.

Zins M, Goldberg M, Bonenfant S, Carole C, Carole G, Carton M, Launay S, Ozguler A, Pellicer M, Quesnot A, Ribet C, Serrano A, Sitta R. Mise en place de la cohorte Constances. Premiers résultats de la phase pilote. Congrès international d'épidémiologie, Marseille, 15-17 Septembre 2010.

Zins M. Présentation de la cohorte Constances. Conférence nationale de santé - Débat public : Comment utiliser les données de santé ? Paris, 3 février 2010.

Guéguen A, Sitta R, Bénézet, Santin G, Goldberg M, Zins M. L'apport des bases administratives et médico-administratives pour la prise en compte des effets de sélection dans la cohorte Constances. Congrès international d'épidémiologie, Marseille, 15-17 Septembre 2010.

Santin G., Bénézet L., Guéguen A., Sitta R., Zins M., Geoffroy-Perez B., Goldberg M. Stratégies pour prendre en compte la non-réponse dans les cohortes Coset MSA et Constances. Colloque francophone sur les sondages. Tanger Maroc 23-25 mars 2010

Goldberg M, Guéguen A, Sitta R, Zins M. Le point de vue de l'épidémiologie : représentativité ou universalité ? Journée Représentativité SFdS-Centre Maurice Halbwachs, Paris, École Normale Supérieure, 4 février 2010.

Goldberg M, Bonenfant S, Ribet C, Zins M. Flux sécurisés de données à caractère personnel dans le domaine de la santé. L'exemple de la cohorte Constances. Séminaire Appariements sécurisés de la SFdS, Paris, 16 novembre 2010.

Bénézet L., Santin G., Guéguen A., Sitta R., Gauvin S., Sarter H., Razafindratsima N., Zins M., Geoffroy-Perez B., Goldberg M. Stratégies pour étudier le biais de non-réponse dans les cohortes Coset et Constances. XXVeme Symposium sur les questions de méthodologie de Statistique Canada - Les enquêtes longitudinales : de la conception à l'analyse. Ottawa, 27-30 Octobre 2009.

Guéguen A, Sitta R, Bénézet, Santin G, Lanoe JL, Goldberg M, Zins M. L'apport des bases administratives et médico-administratives dans la cohorte Constances. XXVeme Symposium sur les questions de méthodologie de Statistique Canada - Les enquêtes longitudinales : de la conception à l'analyse. Ottawa, 27-30 Octobre 2009.

Goldberg M, Zins M, Guéguen A. Cohortes épidémiologiques, causalité et biais de sélection. Rencontres de statistique appliquée - Panels et cohortes. Paris, Ined, 12 novembre 2009.

M. Goldberg, Zins M. Épidémiologie en population et données hospitalières : les besoins, les problèmes, les solutions. Colloque ADELFF - EMOIS "Système d'Information Hospitalier et Épidémiologie", Saint-Malo, 3 et 4 avril 2008 (Conférence invitée). Rev Epidém et Santé Publ, 2008,56 :S5.

Zins M, Goldberg M. Bases de données d'origine administrative et recherche en épidémiologie et en santé publique : le projet PLASTICO. Séminaire Inserm-Institut Télécom., Paris, Novembre 2008.

Couris CM, Schott AM, Andreu N, Goldberg M, Zins M. Un réseau pour le développement de l'utilisation des bases de données médico-administratives pour l'épidémiologie et l'évaluation en santé. Congrès Adelf-Emois « Système d'Information Hospitalier et Épidémiologie », Saint Malo, 3, 4 avril 2008. Rev Epidém et Santé Publ, 2008,56 :S13.

Zins M. Les cohortes professionnelles : intérêt général et objectifs. 11ème Colloque de l'ADEREST, 22-23 novembre 2007, Nantes (Conférence invitée).

Zins M, Goldberg M. Les Bases de données d'origine administrative et recherche en épidémiologie et en santé publique. Le projet PLASTICO. Les journées du Réseau Quételet, 14 décembre 2007.

5.3 AUTRES PUBLICATIONS

Goldberg M, Zins M. Les cohortes généralistes en population. L'exemple des cohortes Gazel et Constances. Bull Acad Nat Médecine, 2013 (accepté).

Henny J, Goldberg M, Zins M. Guide pour la constitution d'une Biobanque associée aux études épidémiologiques en population générale. Paris : Inserm-Lavoisier, Collection Tec & Doc, novembre 2012.

Goldberg M, Zins M. Études épidémiologiques de cohorte « exposés – non exposés ». EMC (Elsevier Masson SAS, Paris), Pathologies professionnelles et de l'environnement, 16-870-A10, 2010.

Goldberg M, Guéguen A, Leclerc A, Lanoé JL, Zins M. Les cohortes épidémiologiques prospectives - Principes, difficultés, solutions. In : Guilbert P, Haziza D, Ruiz-Gazen A, Tillé Y (Eds). Méthodes de sondage : application aux enquêtes longitudinales, à la santé, aux enquêtes électorales et aux enquêtes dans les pays en développement. Paris, Dunod, 2008, pp 79-87.

Goldberg M, Zins M. épidémiologie et bases de données d'origine administrative : une voie prometteuse. J Epid Santé Publ, 2007,1: 7-13.

Cœuret-Pellicer M, Zins M. Les bases de données de l'Assurance maladie. Inserm Unité 687, Juin 2006.

Ribet C, Genet J, Zins M. Les bases de données socioprofessionnelles. Inserm Unité 687, Septembre 2006.