

Santé travail

Programme Coset : Cohortes pour la surveillance épidémiologique en lien avec le travail

Premier bilan de la phase pilote pour la mise en place de la cohorte d'actifs relevant du régime agricole au moment de l'inclusion - cohorte Coset-MSA

Sommaire

Abréviations	2	5. Premiers résultats	22
1. Présentation générale du programme Coset	3	5.1 Refus de participation à l'enquête postale	22
1.1 Objectifs du programme Coset	3	5.2 Refus d'accès aux données issues des systèmes d'information existants	23
1.2 Présentation du dispositif général	4	5.3 Taux de participation	23
1.3 La cohorte Coset-MSA	5	5.4 Remplissage des questionnaires (annexe 2)	24
2. Collaborations	5	5.5 Codage des épisodes professionnels	25
2.1 CCMSA	5	5.5.1 Test de codage de la profession sur échantillon	25
2.2 Equipe Constances	6	5.5.2 Codage du fichier complet	26
2.3 Plateau informatique du CESP (ancien CRI – IFR69) - Inserm	6	5.6 Résultats de l'appariement des données	28
		5.6.1 Données du Sniiram	28
		5.6.2 Données de la CCMSA	30
3. Travaux préparatoires	6	6. Bilan et conclusions	32
3.1 Choix du mode de recrutement	6	6.1 Ressources humaines	32
3.2 Préparation et test des questionnaires	7	6.2 Adaptations du protocole	33
3.3 Préparation des protocoles	8	6.3 Adaptation des cahiers des charges	33
3.3.1 Dispositif d'inclusion	9	Références bibliographiques	35
3.3.2 Accès aux données des systèmes d'information existants	9	Annexes	36
3.3.3 Enquête complémentaire	9		
3.4 Choix des cinq caisses pilotes par la CCMSA	10		
4. Mise en œuvre de la phase pilote	10		
4.1 Calendrier	10		
4.2 Communication et information	11		
4.2.1 Tirage au sort	11		
4.2.2 Recherche des adresses postales	12		
4.2.3 Préparation des plis : édition et impression des documents	14		
4.2.4 Routage et gestion des plis non distribuables (PND)	14		
4.2.5 Saisie et numérisation des questionnaires	16		
4.2.6 Activité du numéro vert	17		
4.3 Echanges de fichiers et dispositif de confidentialité	18		
4.4 Extraction des données passives	19		
4.4.1 Données du Sniiram	20		
4.4.2 Données de la CCMSA	20		
4.5 Enquête complémentaire	20		

Programme Coset :

Cohortes pour la surveillance épidémiologique en lien avec le travail

Premier bilan de la phase pilote pour la mise en place de la cohorte d'actifs relevant du régime agricole au moment de l'inclusion

Cohorte Coset-MSA

Rédaction du rapport

Béatrice Geoffroy-Perez, Laetitia Bénézet, Gaëlle Santin, Laurène Delabre, Pauline Delézire, Juliette Chatelot, Département santé travail (DST), Institut de veille sanitaire (InVS)

Membres du comité de pilotage

Caisse centrale de la Mutualité sociale agricole (CCMSA) : M. Alain Pelc, (Dr Jean-Pierre Grillet), Dr Yves Cosset, (M. Joël Roy), M. Nicolas Viarouge, (M. Jean-Charles Chérie), M. Florian Bremaud

Institut national de la santé et de la recherche médicale (Inserm) : Dr Marcel Goldberg

DST-InVS : Dr Juliette Chatelot, Dr Béatrice Geoffroy-Perez, Dr Ellen Imbernon

Les auteurs tiennent à remercier chaleureusement toutes les autres personnes qui ont contribué à la réalisation de cette étude de faisabilité.

En premier lieu, l'ensemble des membres de l'équipe Constances avec lesquels la collaboration a permis de mettre au point un questionnaire coordonné et d'initier une réflexion méthodologique essentielle, relative aux problèmes de représentativité dans un programme de surveillance longitudinale.

Pour leur implication dans le dispositif de confidentialité, le Plateau informatique du Centre de recherche en épidémiologie et santé des populations (CESP) et en particulier MM. Frédéric Robergeau et Denis Hémon.

A la MSA : M^{mes} et MM. Annie Adjemian, Hervé Treppoz, Jean-Paul Larrat, Marc Parmentier, Géraldine Richard-Hamelin ainsi que les directeurs des caisses impliquées lors de la préparation de l'étude (tests des questionnaires) ou lors de sa mise en œuvre (communication, échanges informatiques), les médecins du travail, les médecins de prévention, les professionnels de la communication et les correspondants informatiques locaux.

A la Caisse nationale d'Assurance maladie des travailleurs salariés (CnamTS) : MM. Yvon Merlière, Laurent Duchet et Alain Diaz.

A l'InVS : M^{me} Stéphanie Gauvin (DST), M. Javier Nicolau (DS) ainsi que le service de communication (Scom) et le service juridique et financier (SFLE).

Relecteurs : Dr Marie Zins, M. Jean-Luc Marchand.

Abréviations

ALD	Affection de longue durée
AQ	Questionnaire auto-administré
AT-MP	Accidents du travail et maladies professionnelles
CCMSA	Caisse centrale de la Mutualité sociale agricole
CESP	Centre de recherche en épidémiologie et santé des populations
CITI	Centres de traitement informatique interrégionaux (MSA)
CnamTS	Caisse nationale d'Assurance maladie des travailleurs salariés
Cnav	Caisse nationale d'assurance vieillesse
Cnil	Commission nationale de l'informatique et des libertés
Coset	Cohortes pour la surveillance épidémiologique en lien avec le travail
CRI	Centre de ressources informatiques (Inserm)
DERS	Direction des études répertoires et statistiques
DMOI	Direction de la maîtrise d'ouvrage institutionnelle
DST	Département santé travail (InVS)
Foin	Fonction d'occultation de l'information nominative
Inserm	Institut national de la santé et de la recherche médicale
InVS	Institut de veille sanitaire
MCO	Médecine chirurgie obstétrique et odontologie
MEE	Matrice emplois-expositions
MSA	Mutualité sociale agricole
NAF	Nomenclature d'activités française
NIR	Numéro d'inscription au répertoire d'identification des personnes physiques (Institut national de la statistique et des études économiques)
NPAI	N'habite pas à l'adresse indiquée
PCS	Professions et catégories socioprofessionnelles
PMSI	Programme de médicalisation des systèmes d'information
PND	Pli non distribuable
RSA	Résumé de sortie anonymisé
SARA	Système d'accès aux retraites agricoles
Scom	Service communication (InVS)
SI	Systèmes d'information
SNGC	Système national de gestion des carrières (Cnav)
Sniiram	Système national d'informations interrégimes de l'assurance maladie
TM	Ticket modérateur
TMS	Trouble musculosquelettique
UVSQ	Université de Versailles – Saint-Quentin

1. Présentation générale du programme Coset

Le Département santé travail (DST) de l'Institut de veille sanitaire (InVS) a souhaité se doter d'un dispositif d'observation longitudinal destiné à améliorer la surveillance épidémiologique des risques professionnels à l'échelle de la population active française : le programme Coset (Cohortes pour la surveillance épidémiologique en lien avec le travail). Ce programme doit concerner l'ensemble de la population active en France, qu'elle soit effectivement en activité ou bien temporairement dans une période d'inactivité professionnelle, les salariés (du privé et du public) et les non-salariés (exploitants agricoles, travailleurs indépendants), quels que soient le statut (non-salarié ou conjoint collaborateur), la catégorie socioprofessionnelle, le secteur d'activité et le type de contrat de travail.

1.1. Objectifs du programme Coset

L'objectif principal du programme Coset est la connaissance et la surveillance de la morbidité et de la mortalité de la population active en France. En d'autres termes, il s'agit de pouvoir décrire l'état de santé des actifs selon leur activité professionnelle.

Cet objectif principal peut se décliner en objectifs spécifiques :

- décrire à l'instant « t » la morbidité (et mortalité) des actifs français selon le type d'emploi (profession, catégorie socioprofessionnelle et secteur d'activité) ;
- décrire l'évolution de la morbidité (et mortalité) des actifs français dans le temps selon le type d'emploi (profession, catégorie socioprofessionnelle et secteur d'activité) ;
- décrire et surveiller les liens entre la morbidité (mortalité) des actifs français et des expositions professionnelles qu'elles soient d'origine physique, chimique ou psychosociale ou organisationnelle ;
- calculer des fractions de morbidité (mortalité) attribuables aux facteurs d'exposition professionnelle ;
- faciliter la mise en place d'études *ad hoc* en cas de repérage de problèmes émergents ou mal documentés sur le plan scientifique.

L'exploitation de données de cohortes couvrant l'ensemble de la population au travail présente de nombreux intérêts pour la surveillance épidémiologique des risques professionnels.

Les données étudiées dans le cadre du programme Coset permettront en premier lieu de décrire l'état de santé de la population en relation avec les facteurs professionnels en France. Cette description de la morbidité pourra être déclinée selon la profession et le secteur d'activité en utilisant des niveaux assez fins des classifications d'emplois dans la mesure où l'effectif surveillé dans le cadre du programme sera important. Ainsi des différences de morbidité selon l'activité professionnelle pourront être mises en évidence lorsqu'elles existent.

Le suivi dans le temps des indicateurs de morbidité permettra d'observer l'évolution des différences selon l'activité professionnelle (persistance, augmentation, réduction), et notamment de repérer une dégradation éventuelle pour une ou plusieurs catégories d'emplois. La détection précoce de tels phénomènes devrait permettre d'alerter les pouvoirs publics afin qu'ils puissent mettre en œuvre des mesures correctives plus rapides.

Les analyses statistiques des données issues de ce programme permettront également de calculer, pour certaines pathologies, la fraction de risque attribuable à différents facteurs professionnels [1].

Ce programme devra donc être à même de fournir des estimations extrapolables à l'ensemble de la population cible.

Il sera également intéressant de comparer l'étude de la morbidité avec les études de mortalité déjà existantes ou mises en œuvre parallèlement au DST de l'InVS. Ainsi des hypothèses sur les facteurs à la fois liés à l'activité professionnelle et à l'état de santé, comme par exemple les comportements de santé, pourront être formulées et faire secondairement l'objet d'études plus ciblées.

Coset est conçu comme un outil généraliste sans objectif spécifique avec un recueil d'information très large en termes de données professionnelles (parcours, nuisances subies) et en termes de

pathologies étudiées. Un tel dispositif permettra si besoin de documenter l'impact de procédés nouveaux sur la santé ainsi que les phénomènes émergents directement ou grâce au recueil complémentaire d'information sur des populations ciblées de la cohorte, ce qui représente un gain considérable en termes de temps et de moyens. De la même manière, ce dispositif peut faciliter la mise en place d'études *ad hoc* afin de documenter des modifications de risques observés dans certains secteurs d'activité ou professions particuliers.

Enfin, un autre intérêt de la mise en place de ce programme concerne l'utilisation et la valorisation de données déjà existantes. Il est en effet prévu d'utiliser les informations issues des bases de données alimentées par l'assurance maladie et l'assurance vieillesse (Système national interrégimes de l'assurance maladie (Sniiram)). Dans le domaine des expositions professionnelles, des matrices emplois expositions sur les risques chimiques, physiques (et psychosociaux) existent ou sont en cours d'élaboration au sein du DST (programme Matgéné [2]) ; ces matrices pourront être appliquées aux cursus professionnels des personnes suivies dans le programme Coset par l'intermédiaire de leur catégorie socioprofessionnelle et de leur secteur d'activité pour évaluer les expositions des personnes. Non seulement les outils existants seront valorisés mais Coset permettra l'économie de recueils de données rétrospectives difficiles, contraignants et pas toujours fiables sur les diverses expositions au travail.

1.2. Présentation du dispositif général

Le programme Coset est conçu comme un dispositif de surveillance longitudinal multiprofessions multisecteurs. Le caractère longitudinal permettra de prendre en compte les problèmes de temporalité dans la survenue des événements de santé et des changements professionnels, mais aussi de l'ensemble des expositions des individus pour analyser des risques de pathologies d'origine multifactorielle, et d'étudier des problèmes de santé qui surviennent à distance de l'exposition et même au-delà de la retraite.

La population couverte par ce programme sera un premier temps celle des actifs affiliés à l'un des trois principaux régimes de Sécurité sociale (Régime général, *via* la cohorte Constances, Régime agricole et Régime des indépendants) qui couvrent 95 % des actifs en France.

Pour le Régime général, le programme Coset s'appuiera sur les données des actifs de la cohorte Constances, en cours de mise en place par la Plateforme de recherche "Cohortes en population" Institut national de la santé et de la recherche médicale (Inserm) Unité 1018 - Université de Versailles St-Quentin (UVSQ) qui prévoit de recruter un échantillon de 200 000 personnes sur cinq années (actifs ou non) [3]. Le champ d'études du projet Constances est beaucoup plus large que celui de Coset en termes de population et d'objectifs. Un partenariat a été initié entre l'InVS et la Plateforme de recherche "Cohortes en population" Inserm Unité 1018 - UVSQ pour l'utilisation par le DST des données nécessaires à la surveillance épidémiologique des risques professionnels collectées dans la cohorte Constances.

Pour les actifs affiliés au Régime agricole et au Régime des indépendants, il est prévu de recruter et de suivre un échantillon d'environ 35 000 personnes dans chaque régime. Ces deux volets sont mis en œuvre par le DST de l'InVS en partenariat avec les régimes concernés.

Quel que soit le régime, des informations sur la santé et les conditions de travail ainsi que des informations nécessaires pour recontacter la personne dans le cadre du suivi de cohorte seront recueillies à l'inclusion par questionnaire auto-administré (AQ). En ce qui concerne les actifs affiliés au Régime général, des données cliniques et paracliniques d'intérêt recueillies lors du bilan de santé initial seront également disponibles.

Quelle que soit leur cohorte d'inclusion, les personnes seront suivies selon des modalités compatibles ; le suivi des actifs inclus au Régime agricole ou au Régime des indépendants sera coordonné avec le suivi des cohortistes Constances (ceci afin de minimiser les perdus de vue du fait de changement de statut professionnel et de régime de protection sociale éventuel). Un nouvel autoquestionnaire sera envoyé aux participants chaque année afin de suivre l'évolution de leur état de santé et de leur parcours professionnel (suivi « actif »). Ce type de suivi permettra d'adapter le questionnaire en fonction de problématiques de santé émergentes. En parallèle, lors de l'inclusion et au fil du suivi, il sera réalisé un recueil passif (ne nécessitant pas d'intervention du sujet) d'informations extraites de systèmes d'informations existants et permettant de documenter, d'une part les événements de santé et d'autre part les événements professionnels (suivi « passif »). Ce recueil

de données, appelées *données passives* dans la suite de ce document, concerne à la fois les participants et un sous-échantillon de non-participants.

Le dispositif mis en place dans le cadre de ce programme permettra de recueillir un ensemble d'informations communes au sein des différents régimes de Sécurité sociale, donc d'appréhender les différences interrégimes.

1.3. La cohorte Coset-MSA

Pour la surveillance de travailleurs du Régime agricole dans le programme Coset, le DST de l'InVS a programmé le recrutement et le suivi d'une cohorte de personnes tirées au sort parmi les actifs affiliés au régime agricole à l'inclusion. Il s'est pour cela associé à la Caisse centrale de la Mutualité sociale agricole (CCMSA). Après une phase préparatoire destinée à mettre au point le dispositif d'inclusion et les outils de recueil, une étude pilote a été mise en œuvre à partir d'octobre 2009, afin de tester sur une population restreinte le dispositif en vue de son déploiement à l'échelle nationale.

Cette phase pilote comportait plusieurs étapes : d'une part étudier la faisabilité du tirage au sort, de l'inclusion et du recueil de données par questionnaire postal, d'autre part tester les procédures d'accès aux données passives et le dispositif de confidentialité, et enfin étudier de manière approfondie les biais de sélection liée à la non-participation des personnes afin d'élaborer des stratégies de prise en compte de la non-réponse.

Le présent rapport fait la synthèse des étapes préparatoires à l'étude et des différents enseignements tirés de la phase pilote Coset-MSA. Seuls les aspects opérationnels et les premiers résultats sur la participation sont abordés dans ce rapport. Un autre rapport sera consacré à l'exploitation des données recueillies dans le cadre de cette étude et à la non-réponse.

2. Collaborations

2.1. CCMSA

Les protocoles d'étude ont été préparés en collaboration constante avec la CCMSA.

La mise au point du questionnaire et l'organisation des tests de questionnaires dans les consultations de médecine du travail ont été réalisées en collaboration avec l'Echelon national santé travail de la Direction de la santé à la CCMSA.

La faisabilité de recruter des personnes au travers des examens de prévention et les tests de questionnaires auprès des assurés a été étudiée en collaboration avec le Service prévention de la Direction de la santé et l'Echelon national santé travail de la CCMSA qui ont assuré les échanges avec les caisses MSA volontaires (Alpes du Nord, Haute-Normandie, Touraine, Maine-et-Loire, Portes de Bretagne – Morbihan, Deux-Sèvres, Picardie, Côtes-Normandes).

Le dispositif de tirage au sort, les modalités de recueil des informations passives, les circuits de données, le dispositif de confidentialité ont été élaborés en collaboration étroite avec la Direction des études répertoires et statistiques (DERS) et la Direction de la maîtrise d'ouvrage institutionnelle (DMOI) de la CCMSA. Ces mêmes Directions en ont également assuré la mise en œuvre.

La CCMSA a également assuré la coordination avec les Directions des caisses participant au pilote. Une réunion téléphonique a été organisée en mars 2009 rassemblant les correspondants des cinq caisses pilotes (médecins du travail, chargés de communication et correspondants informatiques). Cette réunion avait pour objet de présenter l'étude, ses objectifs et son déroulement, de préciser la contribution attendue de la part des caisses participantes et de permettre un échange sur les modalités d'organisation, principalement sur les aspects informatiques pour la collecte des adresses postales et les aspects de communication.

La collaboration avec la CCMSA, effective dès le début de la préparation de l'étude a été formalisée le 15 juillet 2008 par la signature d'une convention de partenariat pour une durée initiale de trois ans. Une fois les protocoles de la phase pilote mis au point, un avenant précisant la contribution des partenaires sur le dispositif retenu a été rédigé et signé le 29 juin 2010. La convention initiale a été reconduite en juillet 2011 pour une durée de trois ans.

2.2. Equipe Constances

La collaboration avec l'équipe Constances (Plateforme de recherche "Cohortes en population" Inserm Unité 1018 – UVSQ) a été effective durant toute la phase pilote de l'étude Coset-MSA. Des échanges réguliers sur l'avancement des projets respectifs sont intervenus durant toute la phase de préparation et de réalisation de la phase pilote. Ces échanges ont porté essentiellement sur l'harmonisation des questionnaires et des méthodes mises en œuvre.

Une réflexion commune a par ailleurs été initiée dans le cadre de la collaboration avec l'équipe Constances et développée au sein d'un groupe de travail méthodologique rassemblant des statisticiens de l'Inserm, de l'Institut national des études démographiques et de l'InVS. L'objet de ce groupe de travail est d'appréhender les problèmes de représentativité dans les études longitudinales. Ce groupe de travail s'est particulièrement penché sur les solutions envisagées pour mesurer les biais de non-réponse et les prendre en compte autant que possible lors de l'analyse afin de produire des indicateurs sanitaires extrapolables à l'ensemble de la population source. Ces solutions s'appuient, d'une part sur les données passives disponibles à la fois chez les répondants et les non-répondants et d'autre part sur les données recueillies par interview auprès d'un échantillon de non-répondants.

2.3. Plateau informatique du CESP (ancien CRI – IFR69) - Inserm

Le suivi à long terme par questionnaire ou au travers d'informations passives contenues dans des systèmes d'informations de la protection sociale suppose de conserver des données directement identifiantes pour les personnes appartenant à la cohorte Coset-MSA.

Afin de permettre le respect de la confidentialité de ces données, l'InVS a mis en place un circuit faisant intervenir un organisme tiers, dit de confiance, dont le rôle est de mettre en correspondance les données issues de différentes sources sans que les informations issues de questionnaires ne soient reliées aux informations identifiantes. Le Plateau informatique du Centre de recherche en épidémiologie et santé des populations (CESP) a accepté de jouer ce rôle. Le bilan des échanges et des ressources mobilisées par cette activité sera fait à travers les échanges concernant la cohorte pilote.

3. Travaux préparatoires

3.1. Choix du mode de recrutement

Pour le recrutement d'actifs du Régime agricole, plusieurs modalités d'inclusion ont été envisagées :

- Recrutement *via* la médecine du travail

Le recrutement des actifs agricoles par l'intermédiaire des visites systématiques de médecine du travail n'était pas pertinent puisque son champ est restreint. En effet, les exploitants agricoles sollicitent la médecine du travail principalement lors de leur installation ou lorsqu'ils rencontrent un problème de santé ou bien encore à la suite d'un accident du travail. Il était par ailleurs difficile de recruter par ce moyen des travailleurs saisonniers (même les saisonniers réguliers).

- Recrutement *via* les examens de prévention

Le recrutement de volontaires pour la cohorte Coset lors des examens de prévention, a également été envisagé. Chaque affilié à la MSA (y compris les exploitants qui peuvent être assurés par un autre organisme) se voit proposer, tous les 10 ans un examen de prévention comprenant un autoquestionnaire et des examens biométriques (audiométrie, visiométrie, spirométrie, contrôle de la tension artérielle, tonométrie...) et un examen biologique, définis en fonction de l'âge. Ces examens sont réalisés par des laboratoires nationaux partenaires (4 ou 5 sur le territoire national) qui se rendent sur place dans les cantons. Ce dispositif accessible à tous les bénéficiaires rencontre un large succès auprès des affiliés puisque chaque année ce sont entre 20 et 25 000 salariés et entre 30 et 40 000 non-salariés qui répondent aux invitations des caisses. Ce dispositif mis en œuvre au niveau départemental est conçu et standardisé au niveau national, laissant au choix des caisses locales différentes options dans l'organisation des invitations. La possibilité de s'appuyer dessus pour la constitution de l'échantillon d'actifs affiliés à la MSA de la cohorte Coset a donc été explorée dans un premier temps (au cours de l'année 2007) avec une inclusion envisagée selon un cycle d'invitation de

cinq années. Cependant, les différences importantes de procédure d'invitation aux examens de prévention selon les régions (cycles d'invitation de 5 ou 10 ans, sélection des bénéficiaires invités tantôt sur des critères d'âge, tantôt sur le canton de résidence) ont conduit à écarter ce dispositif puisqu'il ne permettait pas de couvrir l'ensemble de la population cible, en termes d'âge et de répartition géographique.

- Recrutement par autoquestionnaire postal

Il a donc été décidé fin 2007 d'organiser le recrutement des personnes par tirage au sort parmi les actifs couverts par le Régime agricole, puis envoi postal d'un autoquestionnaire d'inclusion [4]. Outre une mise en œuvre plus rapide, ce dispositif permettait un recrutement plus large, une meilleure prise en compte des biais de sélection ainsi qu'un allègement des contraintes pour les éventuels participants (pas de déplacement). Un tel dispositif n'exclut pas de recontacter les personnes au cours du suivi pour un examen médical.

3.2. Préparation et test des questionnaires

Afin d'assurer la comparabilité des questions avec celles qui seront recueillies pour les salariés du Régime général, le questionnaire est construit à partir de celui mis au point dans le cadre du projet Constances. Ce dernier a été développé partie par partie dans des groupes de travail associant des spécialistes de chaque domaine [5]. C'est dans ce cadre que le DST a contribué à la mise au point de la partie professionnelle relative aux conditions de travail actuelles, à l'historique des expositions, ainsi qu'au repérage de troubles musculosquelettiques (TMS).

Les questions de Constances d'intérêt pour la surveillance des risques professionnels concernaient principalement la vie professionnelle, les comportements et l'état de santé des personnes (notamment santé générale, consommation d'alcool et de tabac, conditions de travail actuelles, historique professionnel et historique des expositions professionnelles). Ces questions ont été reprises dans le questionnaire Coset-MSA.

L'équipe Coset du DST a ensuite adapté le questionnaire pour permettre l'utilisation de matrices cultures-expositions développée dans le cadre du programme Matphyto [6] afin de calculer des indicateurs d'expositions aux produits phytosanitaires. Des items ont également été ajoutés dans l'historique professionnel afin de permettre le codage automatisé des professions déclarées à l'aide du logiciel Sicore® développé par l'Institut national de la statistique et des études économiques.

Enfin, un groupe de travail a été constitué avec des membres de l'Echelon national santé travail de la Mutualité sociale agricole afin de compléter ou adapter le questionnaire à la situation des travailleurs du régime agricole.

Les premières versions du questionnaire ont été testées entre septembre et novembre 2007 auprès des exploitants agricoles lors d'examens de prévention d'une part et auprès des salariés agricoles lors d'examens de médecine du travail d'autre part.

Les objectifs du test étaient les suivants :

- vérifier que les questions étaient compréhensibles ;
- vérifier que les questions étaient adaptées à la population cible ;
- estimer le temps de remplissage du questionnaire ;
- ajouter des questions omises ;
- évaluer la perception globale du questionnaire et plus généralement de l'étude Coset.

Huit caisses départementales MSA, sur les 9 sollicitées par la CCMSA, se sont portées volontaires (trois services de médecine du travail et cinq caisses pour les examens de prévention).

Selon les caisses, les questionnaires ont été remplis par les personnes à leur domicile ou sur place lors de l'examen de prévention ou de la consultation de médecine du travail.

Deux binômes de l'équipe Coset du DST de l'InVS se sont déplacés sur le lieu des consultations médicales ou des examens de prévention afin de proposer de remplir le questionnaire si cela n'avait pas été fait à domicile, relire avec la personne le questionnaire rempli, recueillir ses impressions et documenter les motifs de non-réponse.

Afin de ne pas perturber le déroulement des examens prévus par un délai de remplissage trop long, le questionnaire Coset a été découpé en trois parties (historique professionnel, santé, emploi et situation sociale actuels) et chacune des trois parties a été testée séparément. Au total, 128 personnes (dont 77 non-salariés) ont accepté de remplir un questionnaire dont 46 questionnaires santé, 46 questionnaires emploi actuel et 34 historiques professionnels, les refus ayant été plus nombreux pour cette dernière partie. Certaines questions du projet de questionnaire ont été adaptées suite aux remarques formulées lors de ces tests. Les principales modifications effectuées sont listées ci-dessous :

- introduction de questions-filtres pour certains items ;
- ajouts de questions ou de modalités de réponses ;
- adaptation des libellés de la Nomenclature d'activités française (NAF) peu explicites avec l'aide des hygiénistes industriels du DST de l'InVS.

Par ailleurs, dans le cadre de la collaboration entre le programme Coset et le projet Constances, des échanges réguliers ont permis d'adapter certaines questions communes en fonction des résultats des tests de questionnaires réalisés par l'équipe Constances.

Des échanges avec le projet Matphyto, développé au DST au sein de l'unité d'évaluation des expositions professionnelles, ont également permis d'adapter l'historique professionnel en vue de son croisement ultérieur avec la matrice cultures-expositions.

3.3. Préparation des protocoles

Trois protocoles ont été mis au point sur la période 2008-2009 pour la mise en œuvre des différentes étapes de la phase pilote de la cohorte Coset-MSA. Ces protocoles concernaient respectivement :

- les modalités de recrutement des volontaires parmi les actifs affiliés au régime agricole [4] ;
- les modalités d'accès aux informations individuelles contenues dans les systèmes d'information existants et permettant de documenter passivement l'état de santé et la carrière des personnes [7] ;
- la réalisation d'une enquête par interview auprès d'un échantillon de non-répondants au questionnaire postal afin d'étudier au mieux des biais de non-réponse dans la cohorte [8].

Ces protocoles sont consultables sur le site Internet Coset. Chacun a été mis au point en collaboration étroite avec la CCMSA, puis soumis à l'avis du Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS) et à l'approbation de la Commission nationale de l'informatique et des libertés (Cnil).

	Avis CCTIRS	Avis Cnil
Inclusion d'actifs relevant de la MSA	n°09.047 Avis favorable le 15 janvier 2009	Demande 909091 Autorisation le 9 octobre 2009
Accès aux données des systèmes d'information existants dans le cadre de la phase pilote du projet Coset-MSA	n°10.157 Avis favorable le 11 mars 2010	Demande 910191 Autorisation DR 2010-321 le 19 octobre 2010
Enquête complémentaire auprès d'un échantillon de non-participants lors de la phase pilote de la cohorte Coset-MSA	n°10.137 Avis favorable le 11 mars 2010	Demande 910176 Autorisation DR 2010-148 le 24 juin 2010

3.3.1. Dispositif d'inclusion

Le protocole d'inclusion pilote prévoyait le tirage au sort de 10 000 personnes dans les bases de l'assurance retraite de la CCMSA. La population source a été définie par les personnes résidant dans les cinq départements pilotes, âgées de 18 à 65 ans au 31 décembre 2008, et ayant été assurées dans la caisse de la MSA au moins 90 jours calendaires (durée cumulée) au cours de l'année 2008, quel que soit leur type d'activité. La CCMSA correspondait ensuite avec les caisses participantes pour récupérer l'adresse postale des personnes tirées au sort et constituer ainsi un fichier de routage destiné au prestataire chargé de l'impression et du routage des plis.

Les personnes sélectionnées recevaient une plaquette et une lettre d'information, un formulaire de changement d'adresse et un autoquestionnaire par voie postale. L'envoi initial était suivi d'une relance à l'identique du premier pli ciblée sur les personnes n'ayant pas répondu à la phase initiale. L'autoquestionnaire rempli était retourné à l'aide d'une enveloppe post-réponse. La saisie (manuelle pour la phase pilote) a été confiée à un prestataire, qui après traitement, remettait les fichiers de données au DST.

Le circuit des informations retenu incluait un dispositif de confidentialité impliquant la CCMSA et le Plateau informatique du CESP en vue du suivi des personnes incluses dans l'étude : les questionnaires étaient personnalisés et saisis avec un numéro d'ordre non identifiant propre à l'étude (NC-DST). La CCMSA conservait la correspondance entre un autre numéro d'ordre propre à la MSA (NC-MSA) et les données d'identification ; le Plateau informatique du CESP conservait quant à lui la correspondance entre les numéros NC-MSA et NC-DST, ainsi que les coordonnées postales des participants.

3.3.2. Accès aux données des systèmes d'information existants

Le protocole d'accès aux données des SI décrivait les modalités de recueil d'informations individuelles dites "passives" par appariement direct aux données contenues dans les systèmes d'information existants (Système national d'information interrégimes de l'assurance maladie (Sniiram), Programme de médicalisation du système d'information (PMSI), Système national de gestion des carrières (SNGC) de la Caisse nationale d'assurance retraite (CNAV), données du régime agricole relatives aux cotisations, aux contrats de travail, à l'assurance vieillesse, au contrôle médical (reconnaissance en affection de longue durée (ALD) et en accident de travail ou maladie professionnelle (AT-MP) permettant de documenter l'état de santé ou la situation socioprofessionnelle des personnes.

L'intérêt de ces données est double :

- d'une part, l'exploitation des informations passives permettra de compléter les données recueillies par autoquestionnaire (afin de documenter l'état de santé et la situation professionnelle de la personne), voire de supprimer certaines questions lorsque les systèmes d'information suffisent et/ou sont plus pertinents pour y répondre ;
- d'autre part, ces données, disponibles à la fois chez les participants et les non-participants, doivent permettre d'effectuer des comparaisons contribuant à quantifier les effets de sélection lors du recrutement liés au volontariat et de corriger les estimations en prenant en compte la non-participation.

L'extraction test devait permettre en particulier : 1) de sélectionner le noyau d'informations optimal pour leur utilisation dans le cadre du suivi individuel des personnes et pour leur utilisation dans le cadre des comparaisons participants/non-participants ; 2) de tester les procédures et circuits informatiques envisagés pour l'extraction de ces données.

Les circuits des données s'appuient sur le dispositif de confidentialité mis en place faisant intervenir le Plateau informatique du CESP et la CCMSA.

3.3.3. Enquête complémentaire

Dans le cadre de l'étude des biais de sélection liée à la non-participation des personnes, une enquête ponctuelle a été réalisée au décours de l'inclusion pilote auprès d'un échantillon de non-répondants. Elle consistait à interroger sur un sous-ensemble d'items, des non-répondants à la phase d'inclusion du pilote afin d'appréhender les biais potentiels de non-participation d'une façon complémentaire à celle envisagée au travers des informations passives. Cette double approche doit permettre de

discuter sous plusieurs angles les biais liés à la non-participation à la cohorte. En effet, compte tenu de l'importance des objectifs descriptifs du programme Coset, de la taille finale attendue et de la longévité souhaitée de la cohorte Coset-MSA, l'étude approfondie des biais liés à la non-participation lors de cette phase pilote est essentielle.

L'enquête concernait un échantillon aléatoire de 500 non-participants à l'enquête initiale. Deux modes de recueil ont été attribués aléatoirement par l'InVS (1) un recueil prioritairement par téléphone et enquête en face-à-face pour les personnes non jointes par téléphone (350 personnes), (2) un recueil prioritairement en face-à-face et enquête par téléphone pour les personnes habitant au-delà de 20 kilomètres des limites du département concerné par l'enquête (150 personnes). Le questionnaire reprenait une partie des questions posées lors de l'enquête postale initiale, sur l'état de santé et la situation professionnelle de la personne.

Le dispositif envisagé pour le circuit des données faisait intervenir (1) le DST qui a réalisé le tirage au sort des personnes à enquêter, (2) la CCMSA qui transmettait au prestataire, les coordonnées des personnes à enquêter (3) le prestataire chargé de l'enquête. Cette enquête ne sera pas reconduite lors de l'extension nationale.

3.4. Choix des cinq caisses pilotes par la CCMSA

Cinq caisses de gestion ont été sollicitées par la Caisse centrale de la MSA pour participer à la phase pilote. Les caisses ont été choisies de manière à assurer une diversité régionale, à représenter les deux populations d'actifs affiliés (salariés et non-salariés agricoles) et en fonction de contraintes techniques informatiques. Ainsi, trois caisses concernaient principalement des exploitants agricoles et deux principalement des salariés agricoles. Par ailleurs, ces caisses avaient été choisies afin que les échanges informatiques transitent par chacun des cinq centres de traitement informatiques interrégionaux (CITI) de la MSA¹.

- Exploitants : MSA 62, 64, 71 CITI : CIMAFAF, CITIMAM, CIMAEST
- Salariés agricoles : MSA 29 ,13 CITI : GIMSAO, CIMARAL.

4. Mise en œuvre de la phase pilote

4.1. Calendrier

Les différentes étapes de mise en œuvre et leur période de réalisation sont présentées ci-après :

Etapes	Période
Tirage au sort	Juillet 2009
Recherche des adresses postales	Octobre 2009
Préparation des plis	Octobre 2009 - janvier 2010
Envoi des plis*	Février 2010 + relance en mars 2010
Saisie	Mai – août 2010
Réalisation de l'enquête complémentaire	Novembre 2010 - février 2011
Extraction des données passives	Janvier - mai 2011
Bouclage du dispositif de confidentialité	Juillet 2011

**L'envoi des questionnaires initialement prévu sur la période d'octobre et novembre 2009 a du être repoussé notamment en raison de l'attente de l'accord de la Cnil et du délai insuffisant pour éditer les questionnaires. En raison des périodes de fêtes puis d'élections à la MSA en janvier 2010, il a été choisi de décaler le démarrage de l'étude à février 2010.*

¹ NB : afin de ne pas multiplier les sollicitations des assurés, les départements Agrican [9] ont été exclus pour la phase pilote. L'extension nationale pourra inclure ces départements car elle sera suffisamment à distance de l'étude Agrican.

4.2. Communication et information

En complément de l'information individuelle par courrier et plaquette d'information, envoyés sous le même pli que le questionnaire, le protocole d'inclusion prévoyait une information générale des personnes au travers d'une campagne d'information lancée *via* les sites Internet de la CCMSA, des caisses locales de la MSA et de l'InVS.

L'information consistait en une présentation des objectifs et du déroulement de l'étude. Les personnes étaient informées de leur sélection par tirage au sort et des modalités de participation. Les supports mentionnaient également aux personnes leur liberté de participer, de se retirer de la cohorte à tout moment et de leur droit d'accès et de rectification des informations les concernant.

L'information portait également sur les données recueillies dans le cadre du suivi passif. Afin de permettre la comparaison des personnes suivies aux non-participants pour documenter les biais de sélection, il était indispensable de collecter ce type de données pour l'ensemble des personnes tirées au sort. Les personnes étaient donc informées que des données de ce type les concernant étaient susceptibles d'être collectées, de façon totalement confidentielle, qu'elles aient ou non renvoyé le questionnaire.

L'information portait également sur leur droit de s'opposer à l'exploitation anonymisée de ces données et les moyens à leur disposition pour le faire : volet à retourner à l'InVS, contact *via* le site Internet, appel téléphonique vers un numéro vert (appel gratuit).

Par ailleurs, un communiqué de presse a été émis par l'InVS et la CCMSA quelques jours avant le dépôt poste des plis (annexe 1).

Echanges InVS – CCMSA

Concernant la campagne d'information locale, les services de communication respectifs de l'InVS et de la CCMSA se sont mis en relation afin de convenir d'un plan de communication en amont du lancement de l'étude dans le but de favoriser la participation. Certains supports de communications initialement envisagés, tels que des communiqués dans des journaux locaux ou un affichage dans les caisses locales participantes ont été abandonnés, la CCMSA estimant que cela risquait d'engendrer trop de sollicitations localement auxquelles les caisses ne pourraient faire face en raison de leur implication limitée dans le dispositif. Les échanges ont donc contribué à rédiger et mettre en forme les documents d'information joints au pli à envoyer aux personnes. Une charte graphique et un logo ont également été développés au niveau du Service communication de l'InVS (Scom) en concertation avec l'équipe projet et la CCMSA.

Un numéro vert a été mis en place pour l'étude Coset-MSA, à l'attention des personnes sollicitées pour participer (voir 4.3.6 Activité du numéro vert).

Développement du dossier thématique

En attendant le développement d'un site web dédié, un dossier thématique a été développé sur le site Internet de l'InVS. Ce dossier était accessible directement *via* l'adresse www.coset.fr ou depuis le site Internet de la CCMSA et des caisses locales MSA participant à la phase pilote. Le site Internet dédié a été développé depuis et mis en ligne en mai 2011 (www.coset.fr).

Ce site permet à la fois de donner des informations sur l'étude mais également de s'opposer au tirage au sort ou à la collecte d'informations issues des systèmes d'information.

4.2.1. Tirage au sort

Le tirage au sort de l'échantillon d'actifs affiliés à la MSA a été préparé en collaboration étroite avec la DERS et la DMOI de la CCMSA. Pour sa mise en œuvre dans le cadre de la phase pilote, la direction générale de la CCMSA a accepté de recevoir dans ses locaux deux statisticiennes de l'InVS afin de réaliser le tirage au sort en collaboration avec les statisticiens de la CCMSA. L'échantillon de sujets sollicités dans le cadre de l'étude pilote a été tiré au sort parmi les sujets éligibles repérés dans le système d'information sur les retraites de la CCMSA (base SARA : système d'accès aux retraites agricoles) géré par la DMOI. Aucun refus d'être tiré au sort n'a été signalé en amont.

La première étape, réalisée par la DERS et la DMOI, a été de constituer la base de sondage, en fonction des critères d'éligibilité, par extraction des informations dans le système SARA. Les critères d'éligibilités étaient d'être âgés de 18 à 65 ans au 31 décembre de l'année précédente, d'avoir travaillé

comme relevant du régime agricole au moins 90 jours calendaires (durée cumulée), quelle que soit l'activité, et de ne pas avoir notifié de refus d'être tirés au sort (voir 4.2).

Une première requête dans le système SARA a été constituée par la DMOI sur les données consolidées de l'année 2007 afin de mettre au point et de tester le programme de tirage au sort. La requête sur les données de l'année 2008 a été effectuée par la DMOI au mois de juin 2009 pour tenir compte de la mise à jour annuelle des informations (fin de tous les reports aux comptes). La comparaison des données de l'extraction 2007 et celles de l'année 2008 a permis de s'assurer de l'achèvement complet de l'ensemble des reports aux comptes.

Le programme de tirage au sort, mis au point par l'InVS en collaboration avec la DERS, a ensuite été appliqué sur la population éligible. Le tirage au sort a été stratifié sur la caisse de gestion, le sexe du sujet, son statut (exploitant/mixte, salarié) et sa classe d'âge (trois classes : 18-34, 35-49, 50-65). L'objectif était d'obtenir 2 000 individus dans chacune des cinq caisses participantes (soient 10 000 au total). Dans le même temps, deux numéros anonymes distincts ont été générés pour la mise en place du dispositif de confidentialité.

En raison de la très bonne qualité de la base de sondage constituée, seules deux demi-journées de travail ont été nécessaires pour la réalisation du tirage au sort.

Limites de la base de sondage :

- Connues en amont du tirage au sort

Deux limites avaient été identifiées lors de l'élaboration du protocole.

L'une concerne le fait que seuls les exploitants présents au 1^{er} janvier de l'année sont présents dans SARA. Ainsi, les exploitants installés en cours d'année 2008 étaient de fait absents de la base de sondage et donc exclus du tirage au sort.

Ensuite, pour des raisons pratiques de d'identification et de suivi ultérieur, les personnes dont le NIR n'était pas connu ont été volontairement exclues du tirage au sort. Cette situation concernait environ 0,2 % des personnes de la base.

- Repérées lors de la phase pilote

En raison de la requête initiale qui portait sur 5 caisses, le critère des 90 jours travaillés dans le régime dans l'année en cumulé a été calculé sur les données extraites et ne portait donc que sur l'activité connue au sein de la caisse dans laquelle l'individu a été sélectionné. Par ailleurs, il s'est avéré (voir 5.6.2 Données de la CCMSA) que des personnes étaient susceptibles de changer de Numéro d'inscription au répertoire d'identification des personnes physiques (NIR) lorsque leur état civil est incomplet (jour et mois de naissance manquants) ce qui compromet leur suivi ultérieur.

Enfin, seule la caisse d'affiliation de l'employeur était disponible dans SARA. Ainsi la caisse associée à une personne tirée au sort correspondait au département du siège de l'employeur pour les salariés et au siège de l'exploitation pour les exploitants, et non pas au département de résidence de la personne (voir 4.2.2).

Pour l'extension nationale

La base de sondage sera constituée selon ses critères définitifs, intégrant le critère des 90 jours travaillés qui portera cette fois sur l'ensemble de l'activité de l'individu indépendamment de la caisse d'affiliation de l'employeur. Elle exclue de fait les exploitants installés au cours de l'année précédente et devra également exclure les personnes dont le NIR n'est pas connu. Autant que possible, il sera également nécessaire d'exclure du tirage au sort les personnes dont l'état civil est incomplet.

4.2.2. Recherche des adresses postales

La recherche des adresses au niveau des caisses locales a été effectuée dans les suites du tirage au sort. Le format des adresses retenu est celui utilisé en routine par la CCMSA.

Civilité et identité du destinataire	38 caractères
Complément adresse postale	38 caractères
N° de voie, complément, type et libellé de la voie	38 caractères
Lieu-dit, commune de résidence, boîte postale	38 caractères
Code postal et libellé bureau distributeur	38 caractères

Il était prévu par le protocole qu'un échange entre le prestataire chargé du routage avec les services permettant la normalisation des adresses postales soit réalisé avant l'envoi. Cependant le format des adresses utilisés étant normalisé et la MSA ayant des fichiers d'adresses de très bonne qualité, cette normalisation n'a finalement pas été nécessaire.

Lors de la recherche des adresses au niveau des caisses de gestion, 232 adresses (soit 2,3 %) n'ont pas été retrouvées (tableau 1). Ces adresses manquantes concernaient majoritairement les salariés agricoles de la caisse du Finistère.

I Tableau 1 I

Répartition des adresses non retrouvées

Caisse	Salariés	Non-salariés	Total
13	4		4 (2 %)
29	157	5	162 (70 %)
62	16		16 (7 %)
64	35	2	37 (16 %)
71	13		13 (6 %)
Total	225	7	232 (100 %)

En réalité, la caisse renseignée dans SARA s'est avérée être la caisse à laquelle est affiliée l'entreprise. Celle-ci peut être différente de la caisse d'affiliation du salarié (la caisse de gestion de l'affilié n'est pas disponible dans SARA), et ce pour plusieurs raisons :

- les entreprises avec plusieurs établissements ont la possibilité de cotiser dans une seule caisse ;
- les salariés peuvent travailler et résider dans des départements dépendant de caisses de gestion différentes.

Une réorganisation est en cours à la MSA avec la fusion de caisses départementales ; au 31 décembre 2010, seules 35 caisses subsistent (il en existait 46 au moment du tirage au sort). En revanche, la fusion des systèmes informatiques des caisses n'est pas à effet immédiat, c'est pourquoi dans certains cas des adresses hors département ont pu être retrouvées (c'est le cas des résidents dans les Alpes-Maritimes déclarés dans les Bouches-du-Rhône). Dans le cas du département du Finistère, c'est par une recherche manuelle dans les fichiers du département des Côtes-d'Armor qu'il a été possible de retrouver 118 adresses qui manquaient. Ces 118 adresses ne faisaient pas partie du fichier de routage initial et n'ont été intégrées qu'au fichier de relance postale.

Par ailleurs, pour 27 adresses, l'identification du destinataire était impossible puisqu'après recherche des adresses au niveau des caisses départementales, les informations concernant l'identité du destinataire ne correspondaient pas à une personne physique mais à une institution (foyer, organisme de tutelle...). Ces adresses ont pu être repérées car le champ « identité du destinataire » ne commençait pas par la civilité M^{me}, M. ou M^{lle}. Elles ont donc été reprises manuellement par la CCMSA qui a ajouté à ce champ les nom et prénom qui figuraient dans la base de sondage.

Pour l'extension nationale

A la date de mise en œuvre de l'étude à l'échelle nationale, la totalité des fusions de bases informatiques prévues sera achevée mais le problème se posera toujours pour les salariés résidant et travaillant dans des départements gérés par des caisses MSA différentes.

La CCMSA a proposé de conserver le dispositif tel quel mais de mettre en place une "bourse" aux identifiants pour les individus pour lesquels l'adresse ne serait pas remontée lors de la première requête. La recherche pourrait se faire de façon progressive en commençant au niveau de la caisse elle-même afin de retrouver les adresses lorsque les bases départementales ne sont pas fusionnées, puis au niveau de la région informatique voire de la France entière. Une procédure sera proposée par la CCMSA et pourra être testée et mise en place au niveau des centres informatiques.

Un temps de vérification du fichier de routage par la CCMSA devra être prévu lors de l'extension nationale. En cas d'absence du champ « identité du destinataire » dans les adresses postales,

il faudra prévoir une procédure automatisée afin de compléter les champs manquants par les nom et prénom de l'état civil de la personne.

4.2.3. Préparation des plis : édition et impression des documents

Les étapes d'édition, d'impression et de routage ont été confiées à un prestataire unique sélectionné dans le cadre d'un appel d'offres conformément à la loi sur les marchés publics.

Ces prestations se rapportaient à plusieurs documents : un questionnaire de 40 pages, une lettre d'information, une plaquette d'information, un formulaire de changement d'adresse postale, une notice d'aide pour remplir le questionnaire et une enveloppe T pour la réexpédition du questionnaire rempli. La lettre, le questionnaire et le formulaire de changement d'adresse devaient être personnalisés par le numéro de confidentialité attribué lors du tirage au sort.

La réalisation des prestations a été considérablement retardée par plusieurs facteurs :

- le lancement du marché est intervenu au cours de l'été ce qui a différé le début effectif des prestations ;
- le prestataire a sous-estimé la charge de travail liée à la mise en page du questionnaire et n'a pas mis en œuvre dès le début le suivi qualité qui aurait dû être réalisé lors de cette étape. Ceci a occasionné un très lourd travail de reprise par l'InVS avec des allers-retours de versions non conformes ;
- l'accord Cnil a été obtenu tardivement.

En octobre 2009, le travail avec le prestataire a été réorganisé avec mise au point d'un nouveau planning de réalisation et de nouvelles règles de validation ce qui a permis le lancement effectif de l'étude au début du mois de février 2010.

Plusieurs défauts d'édition ou d'impression ont cependant été relevés *in fine* : erreurs de couleur, décalages de schémas, râteaux de réponse incomplets, feuillets manquants ou en double dans certains exemplaires de questionnaire.

A noter que l'impression du numéro utilisé par le prestataire pour le rapprochement des éléments en plus gros caractères que le numéro de participant a généré une confusion pour les personnes sollicitées pour participer.

Pour l'extension nationale

Pour l'inclusion à l'échelle nationale, un délai d'au moins trois mois devra être réservé à la préparation des documents (phase prépresse) et au moins deux réunions devront être programmées avec le prestataire en plus des échanges par courrier électronique. Dans la mesure du possible, il faudra éviter la période de vacances d'été pour lancer le marché ou prévoir un délai supplémentaire.

Il sera important de bien préciser, dans la lettre d'information, le champ des personnes concernées pour éviter en particulier la non-participation des personnes qui ne sont plus en activité ou qui ne travaillent pas dans le secteur primaire.

D'un point de vue technique, les questionnaires pour l'extension nationale devront être édités selon les contraintes nécessaires à leur numérisation et saisie par lecture automatique de documents, ce qui impose des contraintes techniques supplémentaires lors de la mise en page et de l'édition des documents.

Par ailleurs, le numéro de participant et son code-barres devront être mieux visibles sur le questionnaire afin d'éviter toute confusion avec le numéro utilisé par le routeur pour le rapprochement de documents. L'augmentation de la taille du numéro de questionnaire sous le code-barres en facilitera également la lecture lors de la numérisation des documents (voir 4.2.5 Saisie et numérisation des questionnaires).

4.2.4. Routage et gestion des plis non distribuables (PND)

Routage

Le taux de participation obtenu dans cette enquête reflète la qualité du rapprochement des éléments de pli et des enveloppes portant l'adresse postale. De plus, à travers les questionnaires retournés, il a

été possible de vérifier que pour 90 des 2 389 questionnaires reçus, le sexe ou la date de naissance étaient différents de ceux de la personne tirée au sort. Compte tenu de la possibilité que ce soit une autre personne qui décide de participer en lieu et place de la personne concernée, le faible pourcentage de discordances observé parmi les répondants (4 %) témoigne lui aussi de la qualité du rapprochement des adresses postales et des éléments de plis personnalisés.

Pour des raisons d'image, il a été décidé de cibler la relance postale auprès des seules personnes n'ayant pas répondu à l'envoi initial. Cette étape a nécessité la saisie systématique par le DST de tous les identifiants à l'origine des plis et appels entrants. Certains refus exprimés par téléphone ou par mail n'ont pu être pris en compte faute de bonne communication du numéro de confidentialité. Cette phase, gérée manuellement, a demandé du temps secrétariat et a retardé la phase de relance par rapport à l'envoi initial du fait du délai nécessaire pour prendre en compte les modifications.

Identification des NPAI (n'habite pas à l'adresse indiquée) et PND (plis non distribuables)

La principale difficulté rencontrée au cours de la phase d'inclusion pilote est l'échec complet d'identification des personnes auxquelles le pli n'a pu être remis. Le cahier des charges pour l'impression et le routage des plis comportait la mise en œuvre d'un service de type Alliage® de La Poste pour la remontée dématérialisée des plis non distribués. S'il a été possible de vérifier que le prestataire avait bien souscrit un contrat auprès de la Poste correspondant à ce service, il s'est avéré impossible d'identifier le moindre NPAI. Les exemplaires de plis qui ont été retournés au DST ont d'ailleurs permis de constater que les codes-barres utilisés lors des deux envois successifs ne possédaient pas la structure des codes-barres de la charte Alliage®.

Suite à cet échec, il a été décidé en novembre 2010 de faire traiter l'ensemble du fichier de routage par le service Optimis 1 de La Poste. Ce service permet, sans envoyer de courrier, de restructurer, normaliser et valider postalement des adresses et d'identifier ainsi les plis non distribuables (PND)¹. Cette solution alternative à l'envoi d'un courrier d'information aux personnes qui n'avaient pas donné de leurs nouvelles lors des deux envois initiaux, devait permettre de repérer *a posteriori* les personnes tirées au sort et qui n'auraient pu s'opposer au recueil d'informations passives par manque d'information.

Le service Optimis permet de pointer, d'une part les adresses non valides et, d'autre part les personnes ayant déménagé depuis 6 à 36 mois (à la date de la recherche Optimis soit le 11 janvier 2011) et dont l'ordre de réexpédition du courrier est échu. Le tableau 2 présente le résultat du traitement du fichier d'adresses par Optimis en fonction de la réponse obtenue lors de l'envoi du questionnaire.

I Tableau 2 I

Identification des plis non distribuables en fonction de la réponse à l'envoi initial

	Adresse inutilisable ²		Adresse litigieuse ³		Adresse normalisée		Total
Aucune réponse	406	6 %	850	12 %	6 096	83 %	7 352
Formulaire adresse	29	23 %	10	8 %	86	69 %	125
Aq / Refus exprimé	66	3 %	250	10 %	2 090	87 %	2 406
Adresse erronée ⁴	0	0 %	1	33 %	2	67 %	3
Total	501	5 %	1 111	11 %	8 274	84 %	9 886

¹ Courrier dont l'adresse de destination est fautive ou incomplète, ou dont l'adresse est bonne mais le destinataire inconnu ou a été connu mais a déménagé depuis longtemps.

² Voies inconnues dans la localité, déménagés dont l'adresse n'est pas commercialisable...

³ Adresses non corrigées mais potentiellement utilisables : type de voie non précisé dans le fichier, avec plusieurs choix possibles (rue, avenue, place...), code Cedex invalide...

⁴ Adresse de l'employeur et non de la personne tirée au sort.

Le taux d'adresses litigieuses était du même ordre de grandeur parmi les personnes pour lesquelles des nouvelles ont été reçues sans mention de changement d'adresse, et parmi les personnes pour lesquelles on n'a eu aucun retour. Il a donc été considéré que ces personnes avaient probablement reçu l'information initiale.

Seules les 406 personnes n'ayant pas donné de nouvelle à l'envoi du questionnaire et dont l'adresse a été jugé inutilisable par La Poste ainsi que les deux personnes avec notion d'adresse erronée ont

été considérées comme n'ayant pas été correctement informées. Les informations passives ne seront pas exploitées pour ces personnes.

Cependant le décalage temporel (11 mois après l'envoi du premier questionnaire) maximise les PND puisque sont identifiés comme tels tous les individus qui ont déménagé depuis l'envoi des questionnaires ainsi que les déménagés qui ont quand même reçu le questionnaire car ils avaient souscrit un contrat de suivi de courrier qui est arrivé à échéance depuis.

Pour l'extension nationale

Le ciblage de la relance postale a été réalisé manuellement par la saisie systématique de tous les numéros de plis retournés au DST et des numéros de confidentialité des appels reçus. A grande échelle, la faisabilité de cibler la relance sera appréciée en fonction des capacités techniques, du coût humain et technique et du délai nécessaire à cette saisie.

Une attention particulière devra être apportée à la gestion des plis non distribués lors de l'enquête nationale. Il est manifeste qu'on ne peut se reposer entièrement sur le prestataire pour cette opération, même s'il s'agit d'un professionnel du routage. Afin de limiter le nombre d'intermédiaires et limiter ainsi le risque d'égarement, il est proposé que l'InVS passe contrat avec la société d'acheminement pour le suivi des courriers. Ainsi, l'InVS serait le destinataire des numéros de pli non distribués. Il est également possible de générer les numéros de plis au moment du tirage au sort afin qu'ils soient transmis au prestataire par la CCMSA dans le fichier de routage. Il faudra vérifier auprès de la société d'acheminement que c'est faisable. Le prestataire n'aurait ainsi plus qu'à générer le code-barres correspondant et l'imprimer au dessus du bloc adresse sur la lettre accompagnant le questionnaire.

Le cahier des charges prévoyait que l'InVS ajouterait des adresses tests dans le fichier de routage afin de contrôler son bon déroulement et la gestion correcte des PND. Cependant, conformément au protocole, le fichier était fourni au prestataire directement par la CCMSA sans passage par l'InVS donc ces adresses n'ont pas été ajoutées. Il sera nécessaire, lors de l'extension nationale, de fournir à la CCMSA un fichier d'adresses de test à intégrer au fichier de routage.

4.2.5. Saisie et numérisation des questionnaires

En raison du nombre relativement réduit de questionnaires reçus lors de la phase pilote, la saisie a été réalisée manuellement par un prestataire. Cela permettait en outre de tenir compte des commentaires des personnes en vue du repérage de questions mal comprises ou mal acceptées. Une variable supplémentaire a été créée pour chaque page pour indiquer si du texte a été inscrit en dehors des cases.

Le prestataire, choisi à l'issue d'une procédure de consultation, a assuré une double saisie manuelle contrôlée et vérifiée par relecture de l'ensemble des questionnaires ; par deux personnes différentes, la seconde saisie étant faite à l'aveugle du résultat de la première.

Un dictionnaire des données a été préparé par l'équipe Coset. Un fichier dérivé a été fourni au prestataire explicitant la description du format et de la position des différentes variables souhaitées dans le fichier de saisie. Ceci a non seulement facilité la programmation du masque de saisie par le prestataire mais aussi l'intégration du fichier .txt dans la base d'analyse par le DST.

Chaque questionnaire a ensuite été numérisé sous un fichier image et transmis à l'InVS sur des supports DVD (sous un format image commun). Une base de données Access® a été développée par le prestataire pour faciliter leur consultation à l'aide du numéro de questionnaire figurant sur la première page (code-barres). A noter que la taille réduite des caractères sous le code-barres du questionnaire a rendu plus difficile leur reconnaissance et a entraîné des erreurs lors de la numérisation. Il a donc été nécessaire de reprendre manuellement les numéros inconnus ou en double pour recoder les numéros.

Pour l'extension nationale

Certains atouts ou améliorations possibles ont été relevés en vue de l'extension nationale.

La saisie des questionnaires en nombre lors de l'inclusion et du suivi à l'échelle nationale est prévue par lecture automatique de document avec vidéo-codage des champs en clair.

Un dictionnaire des données sera préparé par l'équipe Coset à l'intention de la société de saisie. Il sera nécessaire de prévoir des champs plus longs pour certaines réponses en texte libre et un codage spécifique des variables en cas de choix multiples. Les valeurs manquantes pourront être conservées à vide.

Les questionnaires seront livrés sous forme numérisée en format permettant la consultation en ligne par numéro de questionnaire et numéro de page. En effet, la numérisation des questionnaires s'est avérée très utile tant sur la réduction de l'encombrement que sur la facilité d'accès aux documents en cas de retour au document.

Une mention sera précisée dans le cahier des charges pour la recherche de discordances par confrontation systématique des numéros de la base de numérisation et du fichier de saisie.

4.2.6. Activité du numéro vert

La réception sur ce numéro a été assurée par les membres de l'équipe Coset afin de recueillir directement les réactions des personnes, notamment leurs motivations concernant leur participation ou leur refus et les difficultés d'interprétation du questionnaire. Une fois la période d'inclusion achevée, la réception des appels a été confiée au secrétariat du DST sur les plages horaires de bureau. La réception des appels, initialement possible de 10h – 18h30 du lundi au vendredi a ainsi été réduite à 10-12h et 14-18h à partir du 12 avril en raison de la diminution du volume d'appel, à distance des envois de questionnaire.

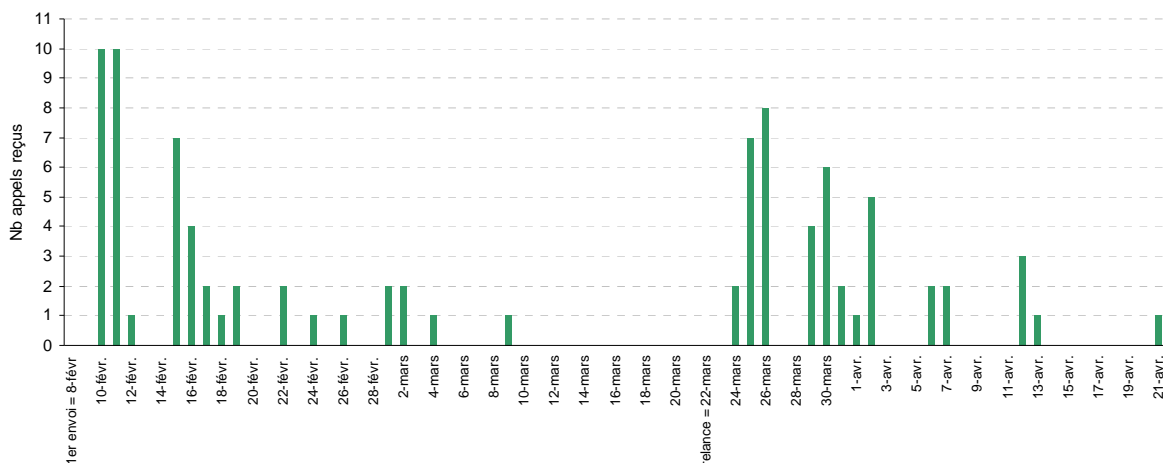
La ligne a été maintenue en vue des étapes suivantes.

Entre le 10 février et le 21 avril, 91 appels ont été réceptionnés, pour 9 886 personnes sollicitées par courrier. Si ce volume semble proportionnellement faible, il laisse cependant attendre un volume important d'appels au moment de la phase de lancement de l'étude à l'échelle nationale puisqu'il est prévu de contacter environ 175 000 personnes.

La distribution du nombre d'appels quotidiens montre qu'ils se concentrent dans les deux premières semaines consécutives aux envois, initial puis de relance (figure 1).

Figure 1

Distribution du nombre d'appels quotidien sur les deux mois et demi d'étude



Les principaux motifs d'appels émanent de personnes (tableau 3) qui ne se sentaient pas concernées par le champ de l'étude mais faisant bien partie des personnes éligibles, qui manifestaient leur refus de participer ou qui souhaitaient une assistance pour le remplissage du questionnaire (en particulier en raison d'un code APE manquant dans la notice d'aide fournie avec le questionnaire).

I Tableau 3 I

Motif des appels au numéro vert

	Effectif
Personnes pensant qu'elles n'étaient pas concernées (retraités, en arrêt de travail, double actif, conjoint collaborateur, emplois de bureau, radiés)	24
Refus	20
Aide au remplissage des questionnaires	19
Informations diverses (objectifs de l'étude, obligation de participer, confidentialité...)	16
Questionnaire déjà renvoyé (relance)	6
Autres (incapacité, décès, homonymie, adresse employeur...)	6
Total	91

Extension nationale

La gestion des réponses aux appels lors du lancement à l'échelle nationale sera confiée à une plateforme de réception et de traitement des appels. Le rythme des appels sera concentré sur les deux premières semaines après chaque envoi postal. L'équipe Coset assurera elle-même la formation des répondants et fournira une liste de questions-réponses en fonction des questions les plus fréquemment posées lors de la phase pilote.

Afin de déterminer si les plages horaires doivent être maintenues telles quelles ou étendues, il a été demandé à France Télécom l'activité du numéro vert en dehors des horaires de la permanence. Il n'a cependant pas été possible de récupérer cette information avec précision.

Les appels recueillis lors de la phase pilote permettront d'affiner le cahier des charges pour la plateforme de réception, de former les personnes et de leur fournir un guide de questions-réponses types.

4.3. Echanges de fichiers et dispositif de confidentialité

Les circuits de données s'appuient sur le système d'archivage des données déjà mis en place conformément aux protocoles. Ce dispositif fait intervenir la CCMSA et un tiers de confiance. Le dispositif est conçu de telle manière que l'appariement individuel des données puisse être réalisé i) sans que la CCMSA, le tiers de confiance ou les organismes possesseurs des données d'intérêt n'entrent jamais en possession d'informations qu'ils ne détiennent déjà et ii) sans que l'InVS n'ait accès aux informations d'identification des personnes. Tous les rapprochements de fichiers sont réalisés par l'intermédiaire des différents numéros de confidentialité et d'un numéro d'ordre détruit à l'issue de la réception et de la validation des fichiers par l'InVS (voir protocole [7]).

Le calendrier des échanges entre les différents partenaires est détaillé dans le tableau 4.

I Tableau 4 I

Calendrier des échanges de fichiers

Date	Organismes impliqués	Echange
10/2009	CCMSA → Routeur	Fichier de routage
11/2010	InVS → CCMSA	Fichier de participation
11/2010	InVS → TdC	Fichier extraction SI CCMSA
11/2010	CCMSA → CnamTS	Fichier extraction Sniiram
12/2010	CCMSA → TdC	Fichier de suivi
12/2010	TdC → CCMSA	Fichier extraction SI CCMSA
02-05/2011	CnamTS → InVS	Données Sniiram
02-06/2011	CCMSA → InVS	Données SI CCMSA
06-07/2011	CCMSA	Bouclage du dispositif (consolidation fichier archivé)

Tous les échanges de données ont été réalisés avec la solution de cryptage GPG. Chaque organisme destinataire d'un fichier a transmis sa clé publique à l'organisme à l'origine de la transmission. Il a été décidé que l'ensemble des transmissions se faisait sous forme cryptée sur support CD ou DVD lorsque le fichier à transmettre possède des données nominatives (nom, prénom, adresses ou NIR) et par mail s'il ne contient pas de telles données. Après une phase d'appropriation du logiciel de cryptage, le cryptage s'est déroulé de manière satisfaisante. Chaque partenaire a généré une clé et la partie publique de chacune des clés générées a été transmise à chaque émetteur de données.

Lors des échanges de fichiers, quelques anomalies ont été relevées : omission de cryptage, contenu du fichier non conforme (variables transmises, nombre d'individus). Par ailleurs, les différents organismes n'ont pas été systématiquement informés des transferts réalisés et de la conformité des éléments transmis. Ceci a en particulier gêné l'InVS dans son rôle de coordination des échanges et a pu entraîner un retard par rapport au calendrier initialement prévu des transmissions.

Pour l'extension nationale

Afin d'optimiser les échanges entre les partenaires, plusieurs précautions seront à prendre au moment de l'extension nationale.

Concernant le cryptage, tous les fichiers de données transmis seront cryptés avec GPG. Les envois se feront sur support CD ou DVD lorsque le fichier à transmettre possède des données nominatives (nom, prénom, adresse ou NIR), et par mail s'il ne contient pas de telles données.

A chaque échange, l'expéditeur informera par mail l'InVS systématiquement ainsi que les organismes concernés, de l'expédition et de la nature des informations transmises avec un bref descriptif du contenu (nombre de lignes dans le fichier et nom des variables transmises). A réception, l'organisme destinataire informera par mail l'InVS systématiquement ainsi que les organismes concernés, de la réception et de la nature des informations reçues avec un bref descriptif du contenu (nombre de lignes dans le fichier et nom des variables transmises).

4.4. Extraction des données passives

En parallèle du recueil de données par questionnaire, le protocole prévoyait le recueil de données passives par extraction dans les systèmes d'information (SI) de la CCMSA et interrégimes. Ces données seront utilisées afin de documenter passivement l'état de santé et la situation professionnelle des cohortistes mais elles permettront également de comparer les répondants et les non-répondants afin d'estimer le biais potentiellement associé à la non-réponse et corriger les poids de sondage pour la non-participation.

L'accès aux systèmes d'informations suivants a été demandé :

- les données du Sniiram y compris les données médicalisées du PMSI ;
- les données de carrières du Système national de gestion des carrières (SNGC) de la Cnav ;
- les données du régime agricole relatives aux cotisations, aux contrats de travail, à l'assurance vieillesse, et aux accidents de travail et maladies professionnelles (données CCMSA).

Cependant, le recueil des données de la Cnav n'a pas été réalisé dès la phase pilote et le sera ultérieurement.

Le champ de l'extraction concernait tous les actifs tirés au sort dans le cadre de l'enquête pilote, c'est-à-dire à la fois les participants et les non-participants, en excluant les personnes considérées comme non informées (voir 4.2.4. Routage et gestion des plis non distribuables) et celles qui avaient signalé leur refus de l'accès à ces informations lors de l'inclusion pilote.

Le fichier de suivi passif ayant été consolidé, avant la recherche des PND (voir 4.2.4) les extractions réalisées excluaient de fait les personnes dont l'adresse n'avait pas été retrouvée, celles dont on savait que l'adresse était erronée (adresse employeur) et celles qui avaient signalé leur refus de l'accès à ces informations avant cette date. Les données de carrière n'ont donc pas été extraites pour 226 personnes et les données de santé pour 232 personnes.

Après cette date, un refus tardif d'accès aux bases a été reçu et 406 personnes ont été finalement considérées comme PND (voir 4.2.4). Les données individuelles extraites des SI concernant ces individus ont donc été supprimées dans un second temps.

En prenant en compte les PND, le pourcentage de personnes pour lesquelles les données des SI ne seront pas utilisées s'élève à 6 % ; ce pourcentage est élevé et pénalise les stratégies envisagées de prise en compte de la non-réponse qui nécessitent que des informations soient disponibles à la fois pour les répondants et les non-répondants. Dans le cadre de la phase pilote, ces données ayant été extraites, elles seront utilisées sous forme agrégée pour réaliser un calage sur marges. Il est donc recommandé, lors de l'extension nationale, de disposer des données agrégées des SI pour toutes les personnes tirées au sort. La phase pilote permettra de déterminer quelles seront les données agrégées les plus pertinentes (voir prochain rapport bilan du pilote).

4.4.1. Données du Sniiram

Le circuit initial prévoyait que la CCMSA transmette le numéro AnoSniir1 (issu du premier niveau d'anonymisation de la fonction Foin) et le numéro de confidentialité spécifique aux données de la CnamTS, ce qui demandait une adaptation du logiciel Camel pour pouvoir transmettre un fichier avec ce format. Cependant, en raison des difficultés d'adapter le logiciel de transmission à la CCMSA, la CnamTS a préféré que la CCMSA transmette directement les éléments nécessaires pour générer AnoSniir1 (NIR, sexe et date de naissance), qu'elle-même génère AnoSniir1 et AnoSniir2 et conserve la correspondance entre AnoSniir2 et le numéro de confidentialité propre à la CnamTS. Concernant le champ des données, il était plus simple pour la CnamTS et beaucoup plus rapide que l'extraction concerne l'ensemble des données du Sniiram (format Echantillon généraliste des bénéficiaires et données du PMSI) plutôt que de faire une sélection de certaines variables. La sélection des données d'intérêt dans le cadre de Coset sera faite suite à la phase pilote afin de limiter le champ des données pour les extractions ultérieures. Ce filtre sera appliqué sur les données transmises par la CnamTS.

En termes de calendrier, l'extraction des données du Sniiram initialement prévu en septembre a dû être retardée en raison du délai pour obtenir l'accord Cnil et du retard dans les échanges préalables entre les organismes. Le fichier a donc été transmis par la CCMSA à la CnamTS en décembre 2010 et l'extraction a pu être réalisée en janvier 2011 pour les données de 2008 et 2009 et en mai pour les données de 2010 qui n'étaient pas encore disponibles en janvier.

Ces premières extractions ont permis de constater que l'ensemble des données présentes dans le Sniiram n'avaient pas été extraites, notamment pour le PMSI, puisque seules les données concernant les hospitalisations en médecine, chirurgie, obstétrique ou odontologie avaient été extraites alors les données concernant les consultations externes, et les hospitalisations en soins de suite et de réadaptation, hospitalisation à domicile et psychiatrie sont théoriquement disponibles. Une demande complémentaire a été adressée à la CnamTS dans ce sens.

4.4.2. Données de la CCMSA

L'extraction des données de la CCMSA a été retardée en raison du délai pour obtenir l'accord Cnil et du retard dans les échanges préalables entre les organismes. Afin d'extraire dans le même temps les données de l'année 2010, l'extraction a été réalisée en avril pour les données sur les accidents du travail et maladies professionnelles (bases Simpat et Oréade). Pour les données de cotisations, contrats et carrières (bases Siera, Sisal, et Sara), une première extraction a été réalisée en mai et une seconde mise à jour en septembre.

Les premières extractions ont permis de repérer quelques défauts (variables absentes, problème de doublons ou données extraites pour le chef d'exploitation au lieu du conjoint ou de l'aide familial). Des corrections (champ, période à extraire) ont été apportées par la CCMSA pour les données professionnelles et le seront pour les données des AT-MP. Le flux Pathos qui centralise les données de mise en ALD est en cours de qualification à la CCMSA avec la remontée des données 2010. Elles devraient être disponibles au cours de l'année 2011.

4.5. Enquête complémentaire

L'organisation et la réalisation du recueil des données de cette enquête complémentaire ont été confiées à l'Institut Ipsos, prestataire sélectionné à l'issue d'un appel d'offre. La rédaction du cahier des charges de cette enquête, attribuée à l'équipe Coset, a été extrêmement lourde non seulement du

fait d'une méthodologie très exigeante, mais également du fait des éléments administratifs et légaux qu'il a été nécessaire d'intégrer malgré des marges d'incertitude importantes quant à la capacité future du prestataire à contacter et convaincre les personnes de participer. Ceci a demandé d'élaborer des scénarii. Par ailleurs, le coût proposé par le prestataire étant directement fonction du nombre d'interviews réalisées, le budget alloué a dû être réévalué au cours du déroulement de l'étude ce qui a demandé des négociations avec Ipsos et le service financier de l'InVS.

D'un point de vue technique, cette étude a débuté par un tirage au sort de 500 non-participants à enquêter parmi les non-répondants au questionnaire. Ces personnes ont été réparties par tirage au sort dans deux groupes d'enquête : enquête téléphonique puis en face-à-face si pas de contact téléphonique possible / enquête en face-à-face puis au téléphone en cas d'impossibilité (éloignement géographique). Le tirage au sort des modes d'enquête avait pour but de vérifier la comparabilité des réponses en fonction du mode d'enquête. Quinze personnes ont été retirées du fichier de contacts, parce qu'elles faisaient partie des 114 personnes pour lesquelles le questionnaire n'avait pas été envoyé (n=7) ou parce qu'elles avaient précédemment manifesté au DST leur refus de participer ou qu'on accède à leurs données passives (n=8).

Le terrain d'enquête a été préparé de septembre à octobre 2010. La phase de téléphonisation des adresses a permis de retrouver un numéro pour 73 % des personnes ; 93 % après recherche manuelle complémentaire, justifiée par la variabilité des libellés d'adresses postales en zone principalement rurale, avec une fiabilité du numéro de téléphone incertaine.

Le terrain d'enquête s'est déroulé de novembre 2010 à février 2011. Au final, parmi les 485 personnes à enquêter, 64,3 % ont répondu au questionnaire (tableau 5).

I Tableau 5 I

Nombre et mode d'interviews réalisées selon le groupe de recueil initial

Groupe initial	Effectif à enquêter N	Interviews réalisées			
		Par téléphone n	En face-à-face n	Total	
				n	%
EPTel	339	169	60	229	67,6 %
EPFaF	146	8	75	83	56,8 %
Total	485	177	135	312	64,3 %

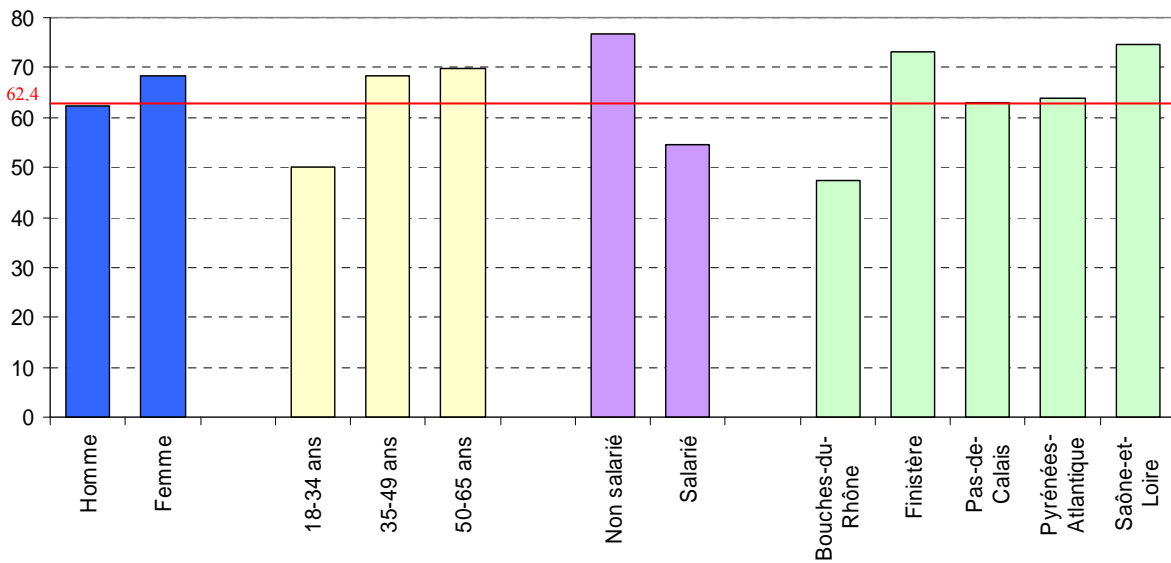
Ipsos a livré un fichier de parodonnées (données inhérentes au suivi de terrain des 485 personnes à enquêter) et un fichier de données de questionnaire. Cette étape s'est avérée longue, du fait notamment du contrôle insuffisant, avant livraison, sur la cohérence des données recueillies, qui a imposé au DST de nombreuses vérifications, et a rallongé la durée de l'étude.

Le bilan de cette enquête est positif. Ipsos a, au final, livré les fichiers dans le format souhaité et le terrain s'est très bien déroulé. Le taux de réponse à l'enquête (figure 2) est particulièrement satisfaisant compte tenu de la population ciblée (non-participants à l'enquête initiale).

Par ailleurs, parmi les personnes ayant effectivement été contactées (n=377), le taux de refus s'est révélé très faible puisqu'il était de 15,4 % contre 82,8 % d'interviews réalisées et 1,8 % de personnes en incapacité de répondre. Cette enquête ne sera pas reproduite lors de l'extension nationale. En effet, elle est destinée à appréhender les biais potentiels de non-participation à l'enquête pilote d'une façon complémentaire à celle envisagée au travers des informations passives. Lors de l'extension nationale, seules les données passives seront utilisées pour étudier ce biais et corriger les estimations pour la non-participation.

I Figure 2 I

Taux de participation à l'enquête complémentaire (%)

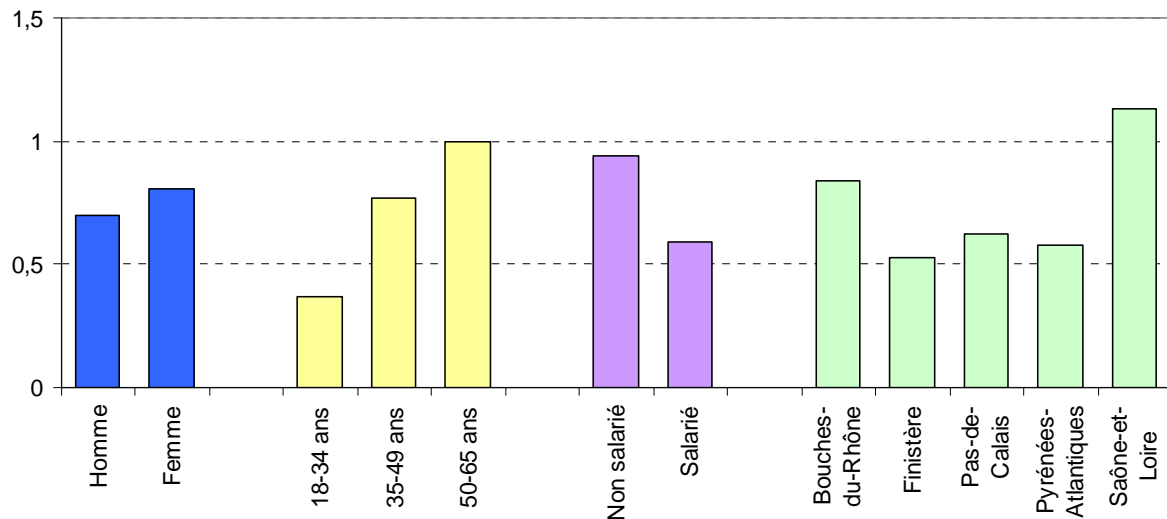


5. Premiers résultats

5.1. Refus de participation à l'enquête postale

I Figure 3 I

Refus exprimés de participation active (%)



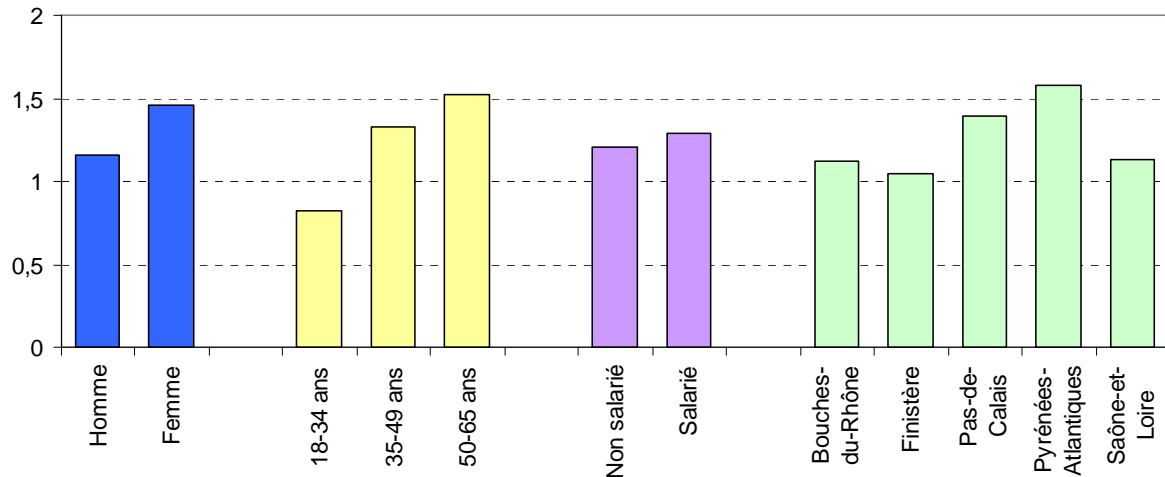
Le pourcentage de personnes ayant manifesté leur refus de participer à l'enquête postale (figure 3) est faible (moins de 1 %) ; ce refus a été exprimé principalement *via* la hotline ou le renvoi d'un questionnaire vide ou d'un courrier.

5.2. Refus d'accès aux données issues des systèmes d'information existants

Le pourcentage de personnes ayant refusé l'accès aux bases de données SI est faible (1,3 %) (figure 4) (113 refus d'accès aux données de carrière et 119 refus d'accès aux données de santé) ; les personnes ayant refusé un accès aux données de carrière ont également refusé un accès aux données de santé alors que certaines ont refusé l'accès aux données de santé seules.

I Figure 4 I

Refus d'accès aux bases de données passives de santé (%)

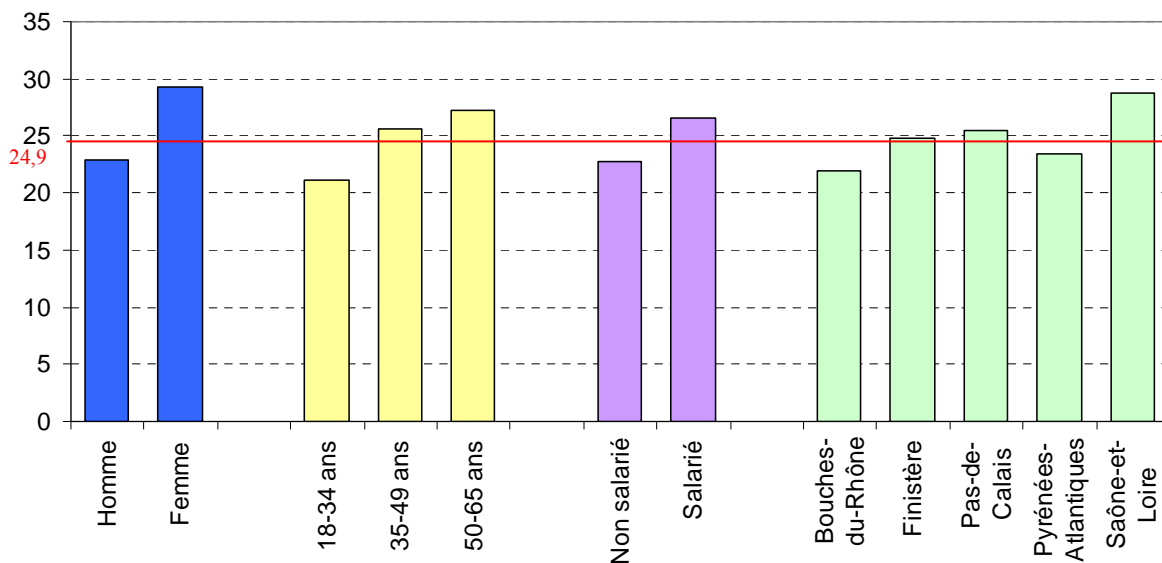


Le pourcentage de personnes ayant refusé d'accéder à leurs données SI est de 1,8 % parmi les personnes ayant renvoyé un questionnaire contre 1,1 % parmi les personnes n'ayant pas renvoyé de questionnaire.

5.3. Taux de participation

I Figure 5 I

Taux de participation à l'enquête postale (N=9 477) (%)

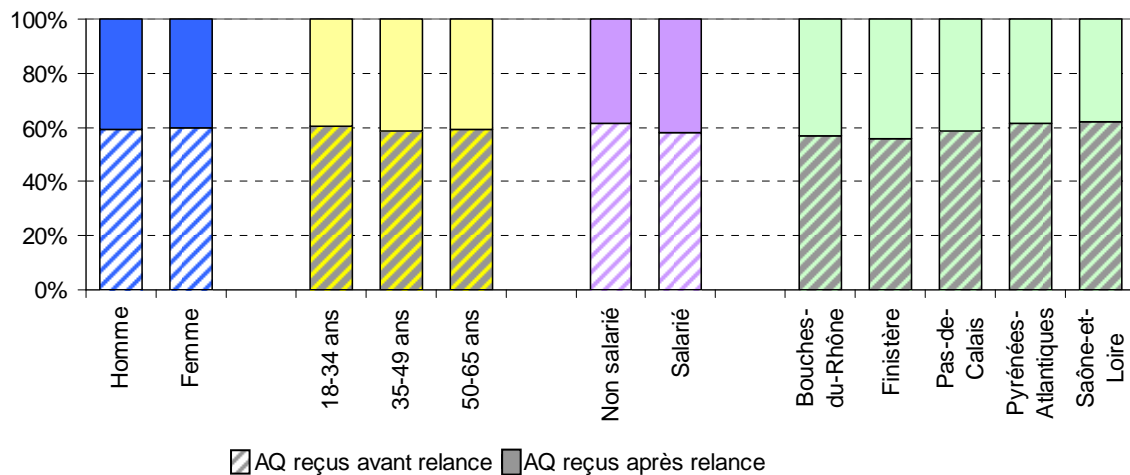


Lorsque l'on s'intéresse aux personnes ayant *a priori* reçu un questionnaire (N=9 477), le taux de réponse est de 24,9 % (contre 23,6 % en considérant l'ensemble des 10 000 personnes tirées au sort) (figure 5).

Il varie significativement selon le sexe (les femmes ont répondu plus fréquemment que les hommes), l'âge (les 18-34 ans ayant moins fréquemment répondu que leurs aînés), le statut (les salariés ayant plus fréquemment répondu que les non-salariés) et la caisse départementale (les affiliés à la caisse de la Saône et Loire ayant plus fréquemment répondu que les affiliés des autres caisses).

I Figure 6 I

Pourcentage de questionnaires reçus avant/après relance postale



L'effet de la relance postale sur le taux de réponse est incontestable puisque 40,8 % des questionnaires ont été reçus après la relance (figure 6). L'effet a été particulièrement marqué chez les salariés et chez les affiliés des caisses des Bouches-du-Rhône et du Finistère (caisses majoritairement salariées). A noter cependant que 118 personnes tirées au sort dans la caisse du Finistère n'ont reçu un questionnaire que lors de la relance postale en raison de l'absence d'adresse postale au premier envoi (voir 4.4.2. recherche des adresses postales).

5.4. Remplissage des questionnaires (annexe 2)

La médiane de la durée de remplissage est de 50 minutes. Cette durée varie peu en fonction du statut professionnel et de l'âge du participant.

Le niveau de remplissage du questionnaire est très variable d'une partie à l'autre. La distribution des réponses brutes est présentée en annexe 2 de ce document.

Pour la phase pilote, les problèmes de remplissage ont été repérés lors de la saisie, notamment les annotations marginales. Un travail de pré-validation de la base a consisté à vérifier les pages litigieuses.

De même, tous les libellés en clair ont été repris et recodés.

Des analyses ont ensuite été entreprises au sein des programmes thématiques concernés du DST qui permettront d'identifier les défauts du questionnaire et de vérifier l'exploitabilité des données.

Cette validation a déjà permis de repérer certains défauts le plus souvent liés à la présentation, à la compréhension des questions ou à l'absence de prise en compte de certaines situations particulières. Une synthèse sera réalisée à l'issue de la validation et les résultats permettront d'améliorer le questionnaire en vue de l'extension nationale.

5.5. Codage des épisodes professionnels

Afin de pouvoir croiser les épisodes professionnels avec les matrices emplois-expositions (MEE) du programme Matgéné, le codage des épisodes professionnels de Coset-MSA doit être réalisé en nomenclatures NAF 2000 et PCS 1994.

Le tableau 6 présente le nombre moyen d'épisodes déclarés. Parmi les personnes ayant déclaré au moins un épisode professionnel, on compte en moyenne 2,4 épisodes par personne ; 6,4 % des personnes n'ont déclaré aucun épisode. Ceci est en faveur d'un défaut de remplissage puisque la population interrogée est issue des personnes qui étaient actifs affiliés à la MSA en 2008.

I Tableau 6 I

Distribution du nombre d'épisodes par questionnaire

Nombre d'épisodes professionnels	Nombre d'historiques de carrière	%
0	152*	6,4
1	856	36,2
2	577	24,4
3	351	14,9
4	177	7,5
5	109	4,6
6	49	2,1
7	40	1,7
8	52	2,2

Nombre moyen d'épisodes (écart-type) : 2,4 (1,7)

* Dans la base codée par l'hygiéniste industriel, ce nombre est de 149 car 3 épisodes vides ont été supprimés après consolidation de la base.

Pour chaque épisode professionnel, la personne remplissant le questionnaire devait rechercher dans la notice d'aide au remplissage le code NAF correspondant à son activité, à 4 caractères pour les épisodes en tant qu'exploitant et à 2 caractères pour les emplois salariés. En effet, pour les salariés, la NAF sur 2 caractères est suffisante pour coder automatiquement la profession par le logiciel Sicore® [10]. Il était également précisé aux personnes qu'elles pouvaient rechercher le code NAF directement sur leur bulletin de salaire. Cette dernière indication s'est finalement avérée peu judicieuse car les versions des nomenclatures ont évolué au fil du temps et on obtient ainsi des codes NAF de différentes versions.

Quoiqu'il en soit, dans tous les cas, pour croiser les épisodes professionnels avec les MEE, il est indispensable d'avoir la NAF sur 4 caractères. Pour ces raisons, il a été nécessaire de recoder les NAF incorrectes (erronées ou qui n'étaient pas en version 2000) et, dans la mesure du possible, d'affiner le codage des NAF à 2 caractères, à l'aide du libellé en clair du secteur d'activité. Ensuite, la profession a été codée automatiquement par Sicore®.

Ce travail de codage a d'abord été réalisé sur un échantillon de 59 épisodes afin d'analyser finement les divergences entre codage automatique et codage manuel. Puis le codage a été étendu à l'ensemble du fichier.

5.5.1. Test de codage de la profession sur échantillon

Sur les 59 épisodes du fichier test, 3 types de codage ont été réalisés :

- codage manuel de la profession en PCS 1994 par un hygiéniste industriel ;
- codage automatique de la profession par Sicore® sans recodage préalable des NAF ;
- codage automatique de la profession par Sicore® après recodage manuel des NAF.

Le tableau 7 présente le résultat du codage par Sicore® sans/avec reprise manuelle préalable de la NAF et pour chacun, le niveau de concordance avec le code proposé par l'hygiéniste.

Sur les 59 épisodes professionnels testés, Sicore® réussit à en coder 50. Sur les 9 échecs de codage, 5 ont pu être codés sur 4 caractères par l'hygiéniste industriel. L'analyse des divergences au cas par cas montre que la reprise manuelle des NAF permet de rectifier le codage par Sicore® dans 3 cas mais détériore le codage dans 2 cas.

I Tableau 7 I

Comparaison du codage automatique/manuel sur le fichier test

Concordance codage Sicore®/ codage hygiéniste	NAF recodée		NAF non recodée	
Echec Sicore®	9	-	9	-
0 caractère	7	14 %	7	14 %
1 caractère	9	18 %	9	18 %
2 caractères	5	10 %	4	8 %
3 caractères	3	6 %	5	10 %
4 caractères	26	52 %	25	50 %
Total	50	100 %	50	100 %

Sur les 50 épisodes pour lesquels une comparaison entre le codage Sicore® et le codage manuel est possible, dans plus de 50 % des cas, le codage réalisé par Sicore® est identique au codage manuel.

Parmi les 24 cas litigieux (Sicore® avec NAF recodée), l'analyse fine des discordances montre que :

- dans 9 cas, Sicore® donne un code de meilleure qualité du point de vue de l'hygiéniste ;
- dans 10 cas, le codage par Sicore® est jugé douteux par l'hygiéniste ;
- dans 5 cas, aucun des codages (Sicore® ou manuel) n'est jugé satisfaisant par l'hygiéniste, faute d'information suffisante pour coder.

Sicore® permet ainsi de rectifier quelques erreurs de codage manuel. En l'absence des précisions nécessaires pour un codage fin, Sicore® code généralement par défaut et fournit alors systématiquement un code sur 4 caractères. Cela peut entraîner des codages douteux ou litigieux d'après le codeur manuel.

En cas d'échec de codage de Sicore®, la reprise manuelle est justifiée puisqu'elle permet de fournir un code sur 4 caractères dans plus de la moitié des cas, le libellé de la profession étant inexploitable pour le logiciel mais suffisamment explicite (ex : "conjoint collaborateur" ou "conjoint d'exploitant"). Une reprise automatique des libellés les plus fréquents en amont du codage pourra être envisagée.

5.5.2. Codage du fichier complet

Pour 42 % des derniers emplois et seulement 24 % des autres emplois, les codes NAF renseignés sur les questionnaires étaient exploitables. Les activités renseignées pour les autres épisodes (1 285 derniers emplois et 2 345 autres emplois) ont été repris par un hygiéniste industriel. Lorsque le code NAF est absent ou non valide, la reprise manuelle permet le plus souvent (79 à 97 % des cas) de recoder l'activité au moins au niveau de la section de la NAF. Lorsque le code activité est renseigné seulement au niveau division (2 caractères), la reprise manuelle est nécessaire pour corriger ou affiner le code. Au final, faute d'information suffisante dans le questionnaire, 100 et 174 NAF sont manquants à l'issue du recodage respectivement pour le dernier emploi et pour les autres emplois.

Le tableau 8 présente le résultat de la reprise manuelle des codes NAF.

I Tableau 8 I

Résultat de la reprise des NAF sur le fichier complet

	Dernier emploi		Autres emplois	
	n	%	n	%
Code NAF exploitable (NAF 2000)	929	42,0 %	753	24,3 %
Code NAF absent	572	25,9 %	709	22,9 %
<i>recodage NAF</i>	486 (85 %)		558 (79 %)	
<i>non recodables</i>	86 (15 %)		151 (21 %)	
Code inexistant en NAF 2000	121	5,5 %	159	5,2 %
<i>recodage NAF</i>	112 (93 %)		154 (97 %)	
<i>non recodables</i>	9 (7 %)		5 (3 %)	
Code NAF valide à 2 caractères	592	26,7 %	1 477	47,7 %
<i>recodage NAF</i>	318 (54 %)		941 (64 %)	
<i>non recodables</i>	269 (45 %)		518 (35 %)	
<i>incorrect mis à manquant</i>	5 (1 %)		18 (1 %)	
Total	2 214	100,0 %	3 098	100,0 %

Le tableau 9 présente le résultat de codage automatique de la profession avec ou sans recodage préalable de la NAF. Le taux de codage varie de 75 à 80 %. Il est un peu meilleur sur le dernier épisode professionnel (gain de 4 %) et est amélioré d'un demi-point lorsque l'on utilise la NAF recodée par l'hygiéniste industriel.

Pour les professions codées lors des deux passages de Sicore®, environ 90 % des PCS sont identiques sur 4 caractères, et 97 % des PCS identiques sur au moins sur les deux premiers caractères.

I Tableau 9 I

Résultats du codage par Sicore® selon recodage de la NAF

	Dernier emploi		Autres emplois	
	n	%	n	%
PCS codées avec recodage des NAF	1 764	79,7 %	2 340	75,5 %
PCS codées sans recodage des NAF	1 754	79,2 %	2 327	75,1 %
Concordance codage Sicore® (avec/sans recodage)				
0 caractère	9	0,5 %	17	0,7 %
1 caractère	26	1,5 %	60	2,6 %
2 caractères	67	3,8 %	72	3,1 %
3 caractères	65	3,7 %	93	4,0 %
4 caractères	1 587	90,5 %	2 085	89,6 %
Total	1 754	100,0 %	2 327	100,0 %

Au total, le lourd travail de recodage de la NAF préalable au traitement par Sicore® présente peu de bénéfice en vue du codage automatique de la profession. Cependant, ce codage reste indispensable en vue du croisement avec les matrices emplois-expositions existantes. L'exploitation des données passives de carrières des individus, actuellement à l'étude dans le cadre du programme Coset, devrait

permettre d'alléger ce travail de reprise des codes NAF. Quant au codage automatique de la profession par Sicore®, il s'avère performant et allège considérablement le travail de codage manuel dans une optique à très grande échelle.

5.6. Résultats de l'appariement des données

5.6.1. Données du Sniiram

La structure des données extraites du Sniiram est identique à celle de l'échantillon généraliste des bénéficiaires. Elles comprennent :

- les données de remboursement de ville pour l'ensemble des régimes de sécurité sociale. Elles ont été extraites pour les années 2008, 2009 et 2010 ;
- les données du PMSI qui contiennent l'ensemble des résumés de sorties anonymisés (RSA) produits pour les séjours réalisés dans le champ MCO en établissements publics ou privés et des données de facturation pour les établissements privés. Elles ont été extraites pour les années 2008 et 2009, les données de 2010 n'étant pas encore disponibles ;
- les données du budget global couvrant uniquement les séjours réalisés dans les établissements publics pris en charge par le régime général. Il s'agit de données de facturations qui ne sont présentes dans le PMSI que pour les établissements privés. La 1^{ère} extraction contient les données des années 2008 et 2009 et la seconde les données de 2010 mais elles ne concernent qu'un seul individu en raison du délai nécessaire à l'enregistrement des séjours dans le système d'information ;
- les données sur les motifs d'exonération du ticket modérateur (TM) : celles-ci excluent les données du contrôle médical du régime agricole. Les deux extractions transmises, contiennent la totalité des données disponibles jusqu'à la date d'extraction, seule la seconde extraction, plus à jour sera donc utilisée.

Les données extraites ont été appariées aux données du fichier des personnes tirées au sort pour les 9 358 personnes concernées par le suivi passif santé afin de décrire l'effectif retrouvé dans chaque base en fonction des variables de stratification. Ces résultats sont présentés selon la nature des données dans les tableaux 10 à 13.

I Tableau 10 I

Proportion de sujets retrouvés dans les données de remboursement de soins ambulatoires

		2008		2009		2010		Période 2008-2010	
		N*	%	%	%	%			
Sexe	Homme	6 322	86,3	86,8	85,2	95,6			
	Femme	3 036	94,6	95,2	94,4	98,2			
Classe d'âges	18-34	2 424	88,0	88,9	87,8	97,4			
	35-49	4 078	89,4	89,5	87,9	96,5			
	50-65	2 856	89,3	90,2	89,1	95,7			
Statut	Non-salarié	4 010	89,5	89,9	88,0	97,5			
	Salarié	5 348	88,7	89,3	88,4	95,7			
Caisse	RA13	1 768	83,5	83,5	81,8	91,0			
	RA29	1 883	91,6	92,9	91,2	98,0			
	RA62	1 915	90,3	90,5	88,8	97,4			
	RA64	1 874	90,2	91,1	90,1	97,8			
	RA71	1 918	89,1	89,3	88,6	97,8			
Total		9 358	89,0	89,5	88,2	96,5			

* Nombre de personnes n'ayant pas refusé le suivi passif santé.

Environ 96,6 % des personnes sont retrouvées dans les données de remboursement de soins de ville sur la période 2008-2010 (tableau 10), avec un taux de recours aux soins annuel de 88 à 90 % (proportion de personnes ayant effectué au moins une demande de remboursement de soin dans l'année). Ce taux est plus élevé chez les femmes que chez les hommes. Ces chiffres sont similaires à ceux retrouvés dans l'échantillon généraliste des bénéficiaires avec 92 % en 2008 pour les bénéficiaires du régime général [11] et dans la cohorte AZF avec 94 % en 2005.

I Tableau 11 I

Proportion de sujets retrouvés dans les données du PMSI

			2008	2009
		N	%	%
Sexe	Homme	6 322	9,1	9,7
	Femme	3 036	13,9	14,4
Classe d'âges	18-34	2 424	10,9	10,4
	35-49	4 078	9,2	10,3
	50-65	2 856	12,7	13,3
Statut	Non-salarié	4 010	10,9	10,7
	Salarié	5 348	10,5	11,7
Caisse	RA13	1 768	10,2	11,6
	RA29	1 883	9,9	11,3
	RA62	1 915	9,8	11,3
	RA64	1 874	12,2	11,0
	RA71	1 918	11,2	11,9
Total		9 358	10,7	11,3

Le taux d'hospitalisation est respectivement de 10,7 % et 11,3 % en 2008 et 2009, il est plus élevé chez les femmes que chez les hommes (tableau 11). Pour comparaison, les données de cadrage de l'Irdes sur les hospitalisations donnent des taux d'hospitalisations en population générale variant de 15 % à 31 % entre 20 et 60 ans [12].

I Tableau 12 I

Proportion de sujets retrouvés dans les données du budget global

		2008	2009	2010
		N	%	%
Total	9 358	0,4	0,3	0,01

Le pourcentage de personnes retrouvées dans les données budgétaires des établissements de soins sous dotation globale est faible (0,4 % en 2008 et 0,01 % 2010) (tableau 12). Cela s'explique par la remontée tardive de ces données dans le Sniiram. Leur champ est par ailleurs très limité car elles sont utilisées pour calculer des montants ; elles ont peu d'intérêt dans le cadre du programme Coset et ne seront donc pas exploitées au cours du suivi.

Environ 4,5 % des personnes sont retrouvées dans les données des exonérations du TM (tableau 13) ; comme ces données n'incluent pas les données en provenance du contrôle médical du régime agricole, elles correspondent à des exonérations qui ont été accordées alors que les personnes étaient sous la protection d'un autre régime.

I Tableau 13 I

Proportion de sujets retrouvés dans les données des exonérations du TM en fonction du motif d'exonération

		N	Tous motifs %	ALD %	ALD non exo. %	ATMP %	Autre motif %
Sexe	Homme	6 322	4,3	1,7	0,7	2,2	0,3
	Femme	3 036	4,3	2,0	1,4	1,3	0,2
Classe d'âges	18-34	2 424	3,2	1,4	0,4	1,5	0,1
	35-49	4 078	5,1	2,0	1,2	2,3	0,4
	50-65	2 856	4,0	1,8	0,9	1,6	0,3
Statut	Non-salarié	4 010	2,9	1,2	0,8	1,2	0,3
	Salarié	5 348	5,3	2,2	1,0	2,5	0,3
Caisse	RA13	1 768	5,4	2,0	1,1	3,0	0,3
	RA29	1 883	4,9	2,0	1,1	2,2	0,4
	RA62	1 915	3,2	1,2	0,8	1,5	0,3
	RA64	1 874	4,9	2,6	0,9	1,7	0,2
	RA71	1 918	3,0	1,2	0,7	1,1	0,3
Total		9 358	4,3	1,8	0,9	1,9	0,3

5.6.2. Données de la CCMSA

Les données extraites à la CCMSA concernent à la fois des données de santé et professionnelles, elles sont le plus souvent issues de systèmes d'information différents pour les salariés et pour les non-salariés. Les données suivantes ont été extraites :

- les données concernant les accidents du travail et maladies professionnelle pour les salariés et pour les exploitants ;
- les données professionnelles :
 - o de cotisation des actifs non salariés (exploitants agricoles, les conjoints collaborateurs et aides familiaux). Seules les données à partir de 2006 sont disponibles ;
 - o des contrats de travail salariés au régime agricole. Elles comportent une ligne par contrat de travail ;
 - o des retraites du régime qui retracent l'ensemble de la carrière professionnelle à la MSA à la fois pour les salariés et les non-salariés.

Les données extraites ont été appariées aux données du fichier des personnes tirées au sort pour les 9 358 personnes concernées par le suivi santé d'une part et les 9 364 personnes concernées par le suivi carrière d'autre part selon la base concernée afin de décrire l'effectif retrouvé dans chaque base en fonction des variables de stratification. Les résultats de l'appariement sont présentés dans les tableaux 14 et 15 suivants.

Environ 12 % des individus sont retrouvés dans la base des AT-MP des non-salariés, sachant qu'il s'agit des AT-MP survenus entre 2002 et 2010 (tableau 14). Et 21,6 % des individus sont retrouvés dans la base des AT-MP des salariés sachant que l'ancienneté des données disponibles est supérieure à celle des non-salariés et que contrairement à la base des AT-MP des non-salariés on y retrouve les individus qui touchent des prestations pour des accidents survenus dans le passé.

I Tableau 14 I

Proportion de sujets retrouvés dans les données d'accidents du travail et de maladies professionnelles

		N	ATMP non-salariés %	ATMP salariés %
Sexe	Homme	6 322	14,9	24,2
	Femme	3 036	6,7	16,1
Classe d'âges	18-34	2 424	5,3	30,9
	35-49	4 078	14,3	20,0
	50-65	2 856	15,3	16,0
Statut	Non-salarié	4 010	28,3	2,9
	Salarié	5 348	0,2	35,6
Caisse	RA13	1 768	4,6	28,3
	RA29	1 883	12,6	23,1
	RA62	1 915	13,9	17,0
	RA64	1 874	11,3	22,1
	RA71	1 918	18,3	17,9
Total		9 358	12,3	21,6

Sur les 9 364 personnes concernées par le suivi passif professionnel, les données de 22 personnes ne sont pas retrouvées dans les données de retraites, alors que le tirage au sort des personnes a été réalisé dans cette même base. Ces 22 personnes ont toutes en commun un état civil incomplet à la date du tirage au sort (mois et jour de naissance manquants) probablement complété depuis avec affectation d'un NIR différent. En raison des difficultés de suivi, il sera nécessaire d'exclure du tirage au sort en amont les personnes avec un état civil incomplet. Parmi les 9 358 personnes de l'échantillon actuel qui n'ont pas refusé le suivi passif carrière, cela ne concerne 152 personnes (1,6 %) dont 22 sur ces 152 (14 %) ont connu une mise à jour de leur données d'état-civil.

I Tableau 15 I

Proportion de sujets retrouvés dans les données professionnelles de la CCMSA

		N	Cotisations %	Contrats %	Retraites %
Sexe	Homme	6 325	47,7	62,9	99,7
	Femme	3 039	39,9	66,9	100,0
Classe d'âges	18-34	2 425	24,9	89,3	100,0
	35-49	4 081	48,2	60,4	99,9
	50-65	2 858	58,1	48,7	99,4
Statut	Non-salarié	4 014	99,9	16,5	100,0
	Salarié	5 350	4,10	99,9	99,6
Caisse	RA13	1 769	27,4	77,8	98,8
	RA29	1 883	40,5	68,0	100,0
	RA62	1 915	52,0	57,7	100,0
	RA64	1 878	46,9	65,7	100,0
	RA71	1 919	57,6	52,7	100,0
Total		9 364	45,2	64,2	99,8

La quasi-totalité des non-salariés et des salariés à la date du tirage au sort sont retrouvés respectivement dans les bases de cotisations et dans les bases de contrats de la même année.

Au total, 5 personnes ne sont retrouvées ni dans les données de cotisation, ni dans les données de contrat, ni dans les données de retraites. Elles font partie des 22 personnes dont l'état-civil a été mis à jour entre le tirage au sort et l'extraction des données professionnelles. Le fait que 17 d'entre elles soient quand même retrouvées dans les bases cotisations et/ou contrats peut s'expliquer par une date d'archivage antérieure à la mise à jour de leur état-civil. En effet, contrairement à la base carrière qui est une base "vivante" donc mise à jour en continu, les bases statistiques de cotisations et de contrats sont des bases archives sans mise à jour.

6. Bilan et conclusions

6.1. Ressources humaines

La phase pilote de la cohorte Coset-MSA a mobilisé de nombreux personnels tant à l'InVS qu'à la MSA.

A la **CCMSA**, la charge de travail n'a pu être estimée précisément en termes de ressources humaines. Dix réunions techniques et 4 réunions du comité de pilotage ont été nécessaires pour mettre au point les différents protocoles de la phase pilote. Les tests des questionnaires ont également fait appel aux médecins conseils de plusieurs caisses régionales.

Plusieurs journées ont été consacrées d'une part à la constitution de la base de sondage et au tirage au sort, puis, à l'issue de l'inclusion à la préparation des requêtes pour l'extraction des données passives.

Une charge de travail, difficile à estimer mais déterminante, est représentée par l'effort important de coordination interne entre les trois principales directions impliquées (l'Echelon national santé travail, la Direction des études, répertoires et statistiques et la Direction de la maîtrise d'ouvrage institutionnelle), ainsi qu'avec la Direction de la communication et les Directions des caisses participantes dont la CCMSA a recueilli l'adhésion. Chaque hiérarchie a assuré la mise à disposition effective des ressources nécessaires et l'implication de leurs personnels, indispensables au bon déroulement du projet.

Deux réunions et des échanges par mail entre les services communication de l'InVS et de la CCMSA ont permis d'organiser la communication autour du lancement de l'étude. Les services de communication des caisses départementales concernées par l'enquête ont été impliqués afin de relayer l'information sur leurs sites Internet locaux et pouvoir rediriger éventuellement les questions qui leur seraient posées localement par les participants.

Les caisses MSA qui ont pris part au pilote ont ainsi été sollicitées en amont à plusieurs niveaux - Directeur, médecin conseil, service communication et service informatique pour la recherche des adresses. Par ailleurs, le projet a été présenté aux comités de protection sociale nationaux et locaux ainsi qu'aux conseils d'administration.

Au **DST de l'InVS**, l'essentiel du travail en amont du lancement a été représenté par la rédaction des protocoles et la soumission de ceux-ci aux instances éthiques (Cnil) ; par la rédaction des cahiers des charges principalement pour la partie technique mais également pour les clauses administratives ainsi que par l'analyse des réponses aux appels d'offres et la rédaction des rapports afférents.

Les étapes de validation de la phase de prépresse du questionnaire ont été réalisées en relation avec le Scm de l'InVS. Cette phase a été particulièrement lourde et chronophage en raison du volume du questionnaire d'une part, mais surtout du manque de contrôle qualité de la part du prestataire. Beaucoup de temps a été perdu en raison notamment du délai de 9 mois d'obtention de l'accord Cnil pour l'envoi des questionnaires, et des contrôles itératifs de la qualité des documents produits en prépresse. En concertation avec les services communication de l'InVS et de la CCMSA, l'équipe Coset a également suivi et coordonné les activités de communication, rédigé les documents envoyés aux personnes, le communiqué de presse et les éléments diffusés sur le site Coset,

Après l'envoi des plis, l'équipe Coset a assuré la réponse aux appels sur le numéro vert et aux mails adressés sur la boîte mail dédiée. La saisie des codes-barres des questionnaires reçus a été assurée principalement par le secrétariat du DST. Un temps considérable a été perdu du fait de l'absence de

mise en œuvre réelle du suivi du courrier et de la nécessité de recourir au service de mise à jour des adresses postales.

L'équipe Coset s'est également chargée de la coordination des échanges de données entre CCMSA, tiers de confiance, CnamTS et DST avec l'appui de la direction scientifique de l'InVS. Enfin, un travail important de validation du travail des prestataires a été réalisé par l'équipe Coset, aussi bien au niveau de l'édition que de la saisie ou de la livraison des données de l'enquête complémentaire. L'analyse des différents fichiers de données d'enquête se poursuit et fera l'objet d'un document de synthèse ultérieur.

Le Scm de l'InVS a été fortement sollicité pour mettre au point et suivre le marché d'édition des questionnaires, pour le développement d'une charte graphique spécifique au programme en vue de l'édition des documents, et pour la mise au point du dossier thématique puis du site Coset. Ces étapes ont été réalisées avec l'équipe Coset du DST, en concertation avec la CCMSA. Le SFLE a quant à lui été sollicité pour l'élaboration de conventions, dans les échanges avec la Cnil afin d'obtenir les autorisations nécessaires à la mise en œuvre du projet et pour gérer la partie administrative et juridique des appels d'offres.

6.2. Adaptations du protocole

La phase pilote s'est déroulée de manière satisfaisante et a permis de pointer d'ores et déjà certaines améliorations possibles sur le plan organisationnel et sur les difficultés à attendre lors de l'extension nationale qui vise à inclure tous les départements métropolitains (notamment l'adhésion de l'ensemble des 35 directions de caisse sera probablement difficile à obtenir).

La base de sondage sera constituée selon ses critères définitifs, intégrant le critère des 90 jours travaillés qui portera sur l'ensemble de l'activité de l'individu indépendamment de la caisse d'affiliation de l'employeur. Le processus de recherche des adresses postales au niveau des centres informatiques sera modifié afin de retrouver les adresses des personnes même lorsque les bases départementales ne sont pas fusionnées.

L'adaptation des questionnaires sera précisée à l'issue de la phase de validation des données recueillies lors de la phase pilote.

La lettre d'information explicitera le champ des personnes concernées pour éviter en particulier la non-participation à tort des personnes qui ne sont plus en activité ou qui ne travaillent pas dans le secteur primaire.

Concernant les échanges de données entre les différents acteurs, il est apparu nécessaire d'avoir une bonne visibilité des échanges de fichiers pour et de la part de tous les différents acteurs impliqués dans les circuits avec rétroinformation systématique par mail à chaque transmission.

Si cela s'avérait efficace dans la prise en compte de la non-réponse, il est également envisagé, lors de l'extension nationale, de demander une extraction des données agrégées des SI pour toutes les personnes tirées au sort.

6.3. Adaptation des cahiers des charges

Gestion des appels

La prise en charge et le traitement des appels sur le numéro vert seront confiés à un prestataire, les répondants seront formés par l'équipe Coset qui leur fournira notamment une liste de questions-réponses en fonction des questions les plus fréquemment posées lors de la phase pilote.

Édition, impression routage et saisie

La nécessaire coordination des étapes d'édition, d'impression, de personnalisation et de routage des documents impose un marché unique avec éventuel allotissement. Le cahier des charges pour choisir le prestataire chargé de la saisie des questionnaires sera concomitant de celui pour l'édition, impression et routage. L'interdépendance entre l'impression et la saisie des questionnaires nécessite en effet le lancement simultané des deux marchés.

L'accent sera mis sur la nécessité d'avoir des contacts étroits avec le prestataire en charge de l'édition en prévoyant des réunions systématiques entre celui-ci, le service édition de l'InVS et l'équipe du

projet Coset ; soient au moins deux réunions de suivi en plus des échanges par courrier électronique. Les différents points de contrôle devront être listés précisément dans les cahiers des charges.

Une attention particulière sera portée à la mise en œuvre du suivi du courrier de manière à identifier correctement les personnes qui n'ont pas reçu l'information et ne pas surestimer ainsi le nombre de personnes à exclure du suivi passif.

Ainsi, hormis le questionnaire qui devra être adapté au vu de l'analyse des données de la phase pilote, le protocole d'étude sera peu modifié pour l'extension nationale. Les différents écueils repérés au cours de la phase pilote ont pu être pris en compte et corrigés en vue de la mise en œuvre de la cohorte Coset à plus grande échelle.

Les différentes étapes nécessaires à la réalisation des prestations devront être programmées très en amont par rapport au lancement de l'étude avec, notamment, une marge de temps raisonnable en cas de retard sur une prestation, ceci en raison de l'interdépendance de ces étapes.

Par ailleurs, cette étude de faisabilité donne la mesure des moyens humains mobilisés et souligne la variété des compétences nécessaires, à la fois scientifiques mais également informatiques, juridiques, et en matière d'édition et de communication. Elle montre ainsi l'importance de ne pas sous-estimer les ressources humaines à mettre en œuvre pour mener à bien une opération d'une telle envergure.

Seuls les aspects opérationnels et les premiers résultats sur la participation ont été abordés dans ce document. Un prochain rapport sera consacré à l'exploitabilité des données recueillies dans le cadre de cette étude et aux travaux consacrés à la prise en compte de la non-réponse.

Références bibliographiques

1. Imbernon E. Estimation du nombre de cas de certains cancers attribuables à des facteurs professionnels en France. Rapport InVS 2003 :23 pages et annexes
http://opac.invs.sante.fr/doc_num.php?explnum_id=5864
2. Fevotte J; Dananche B, Delabre L, Ducamp S, Garras L, Houot M, Luce D, Orłowski E, Pilorget C, Lacourt A, Brochard P, Goldberg M, Imbernon E. A Program to Develop Job-Exposure Matrices in the General Population in France. *Annals of Occupational Hygiene* 2011, pp. 1–14.
3. Zins M, Bonenfant S, Carton M, Coeuret-Pellicer 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 Aug 12;10:479.
4. Protocole d'inclusion pilote :
http://www.coset.fr/publications/protocoles/protocole_inclusion_CosetMSA_27112009.pdf
5. Site constances : www.constances.fr/
6. Le programme Matphyto : matrices cultures-expositions aux produits phytosanitaires. Exemple de matrices cultures-expositions aux pesticides arsenicaux :
http://opac.invs.sante.fr/doc_num.php?explnum_id=1017
7. Protocole pour l'accès aux données des SI existants :
http://www.coset.fr/publications/protocoles/20100218_ProtoAccesDonneesSi_Passives.pdf
8. Protocole pour enquête complémentaire auprès des non-répondants :
http://www.coset.fr/publications/protocoles/20100203_ProtoCompleet_enquetecomplem.pdf
9. Agrican. <http://www.grean.org/agrican.html>
10. Carton M, Santin G, Geoffroy-Perez B, Chanut A. Contribution des variables annexes au codage des libellés de profession par le logiciel Sicore® :
11. L'Échantillon généraliste de bénéficiaires : représentativité, portée et limites. Points de repère CnamTS (septembre 2009 - n°25) : <http://fulltext.bdsp.ehesp.fr/Cnamts/PR/2009/25/1.pdf>
12. Site internet IRDES – données de cadrage hospitalisation – maj 23/11/2010 :
<http://www.irdes.fr/EspaceEnseignement/ChiffresGraphiques/Cadrage/Hopital/Sejours.htm>

Annexes

Annexe 1 : Communiqué de presse



Communiqué de presse



Lancement d'une grande étude épidémiologique des risques professionnels auprès du monde agricole

Saint-Maurice, le 4 février 2010

L'Institut de veille sanitaire (InVS) lance, en partenariat avec la Mutualité sociale agricole (MSA), une grande étude permettant de décrire et de surveiller l'état de santé de la population au travail dans le monde agricole (la cohorte Coset-MSA).

Pilotée par le Département santé travail de l'InVS, l'étude Coset-MSA consiste à suivre l'état de santé et les conditions de travail d'un groupe de volontaires pendant plusieurs années ; elle a pour objectifs d'identifier les métiers et les conditions de travail à risque et de mesurer leur impact sur la santé de la population dans le but de proposer des recommandations en matière de prévention.

La première phase de cette étude débutera le 8 février 2010 dans 5 départements : les Bouches-du-Rhône, le Finistère, le Pas-de-Calais, les Pyrénées-Atlantiques et la Saône-et-Loire.

10 000 personnes, choisies par tirage au sort parmi les actifs du monde agricole âgés de 18 à 65 ans, seront invitées à participer à cette étude en remplissant un questionnaire postal sur leur santé et leur activité professionnelle.

Cette étape est primordiale avant le lancement de l'étude à l'échelon national en 2012.

Toutes les données recueillies seront analysées, dans le strict respect de l'anonymat des participants, afin de décrire les liens entre les facteurs professionnels, les conditions de travail et les problèmes de santé.

Parallèlement à cette étude dans le monde agricole, des informations similaires concernant des actifs d'autres milieux professionnels seront analysées au sein du programme Coset qui deviendra ainsi le 1er grand programme de surveillance de la population au travail à l'échelle nationale.

Pour plus d'informations :

www.coset.fr / www.invs.sante.fr/coset

www.msa.fr: [Lancement d'une étude épidémiologique des risques professionnels \(Coset\)](#)

Contacts presse :

InVS : Gabrielle Issaverdens : 01 41 79 67 79 (g.issaverdens@invs.sante.fr)

MSA : Géraldine Vieuille : 01 41 63 12 35, vieuille.geraldine@cmsa.msa.fr ;

Annexe 2 : Résultats préliminaires issues des données brutes des questionnaires

1- Temps de remplissage des questionnaires

Les résultats suivants ont été calculés sur les 2 023 répondants ayant renseigné une heure de début et une heure de fin de remplissage valides.

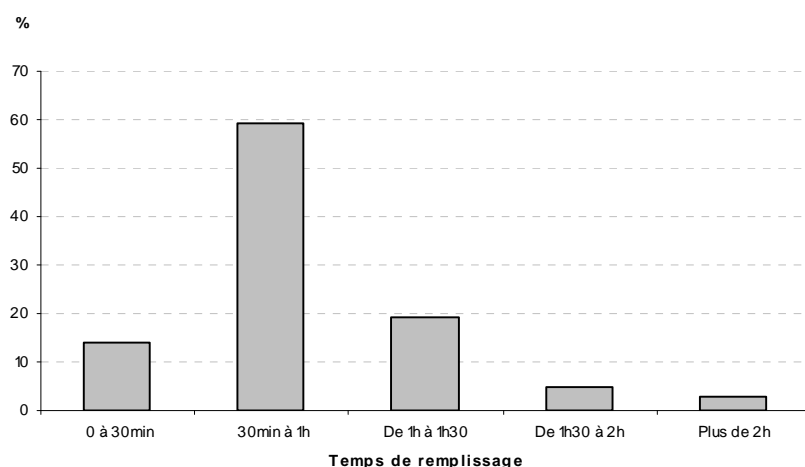
Annexe 2 - Tableau 1

Durée de remplissage totale en minutes

N	Moyenne	Médiane	Ecart-type	Minimum	Maximum
2 023	56,1	50,0	26,5	5,0	203

Annexe 2 -Figure 1

Durée de remplissage des questionnaires



Plus de la moitié des personnes répondantes ont mis 30 à 60 minutes pour remplir le questionnaire Coset. Les durées totales de remplissage du questionnaire sont peu différentes selon le statut professionnel et selon l'âge du participant. De plus, la qualité de remplissage ne semble pas dépendre de la durée de remplissage du questionnaire.

2- Remplissage du questionnaire

Le remplissage du questionnaire a été étudié par partie, suivant le caractère vide ou non des items (items ne comportant que des réponses manquantes). Il a été évalué sur la partie relative à la santé d'une part (10 items) et sur celle liée à la profession actuelle et aux expositions professionnelles d'autre part (11 items).

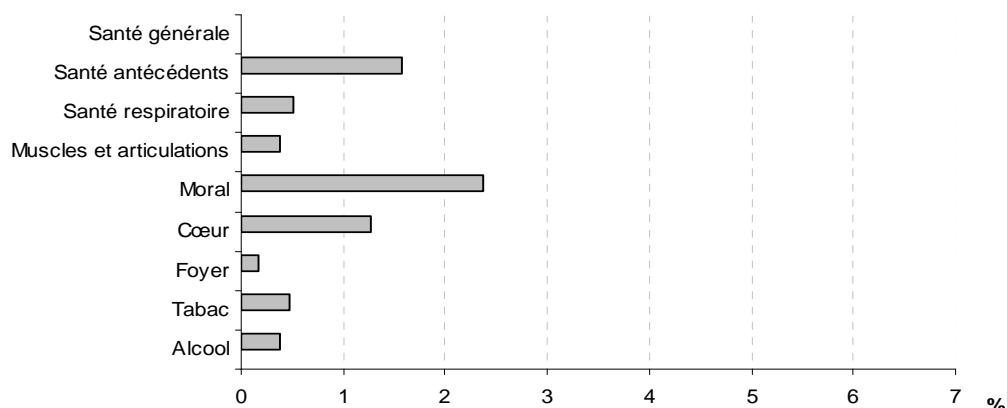
Sur les 2 363 questionnaires récupérés, 1 473 ont été remplis correctement (aucun item manquant), soit environ 60 % des questionnaires recueillis. Environ 97 % des individus ont correctement rempli plus de 80 % de leur autoquestionnaire (moins de 4 items manquants sur 21 au total).

Cependant, la qualité de remplissage du questionnaire est variable selon les différents items des deux parties santé et expositions professionnelles (voir Annexe 2 – Figure 2).

Les questions sur la santé générale du répondant ont été très bien renseignées, aucun item « santé générale » n'est complètement manquant. L'item concernant le moral est le seul item pour lequel on compte plus de 2 % de manquants.

Annexe 2 – Figure 2

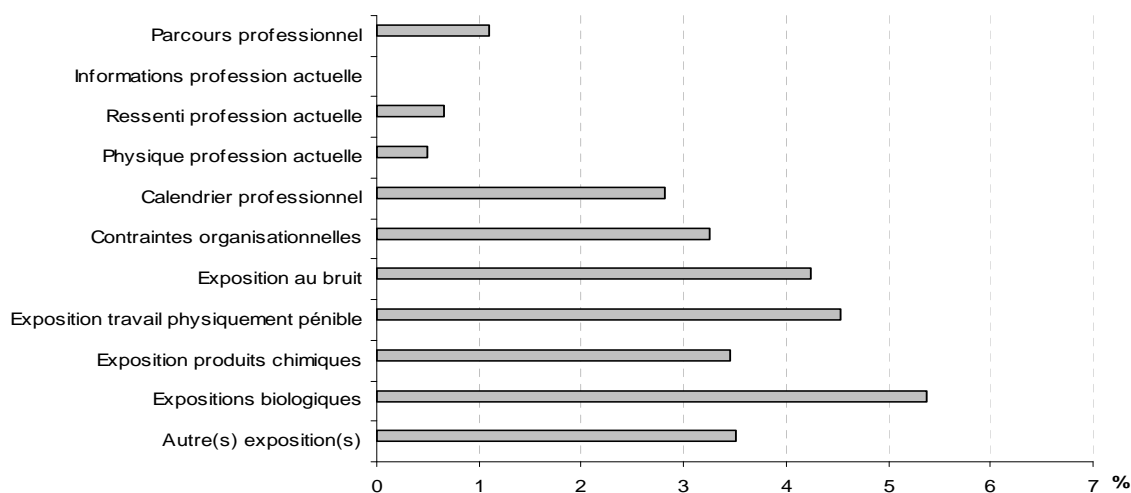
Remplissage de la partie "santé"



Un peu plus de 30 % des items correspondant aux prestations reçues en cas d'incapacité ou d'handicap sont totalement vides (par soucis de clarté, cette item n'apparait pas sur la figure précédente). Ce pourcentage élevé peut s'expliquer par la position de la modalité « aucune (prestation reçue) », cette modalité est en effet la dernière des 13 choix proposés.

Annexe 2 – Figure 3

Remplissage de la partie "profession actuelle, calendrier professionnel et expositions professionnelles"



Les items sur les expositions professionnelles (de contraintes organisationnelles à autre(s) exposition(s)) semblent moins bien renseignés que les items correspondants à la profession actuelle du répondant. Ils comportent de 3 à 5,5 % de manquants selon le type d'exposition.

Toutes les informations sur la profession actuelle des répondants sont renseignées. Les autres items professionnels ont eux aussi un très bon taux de réponse.

3- Sexe, statut et nationalité

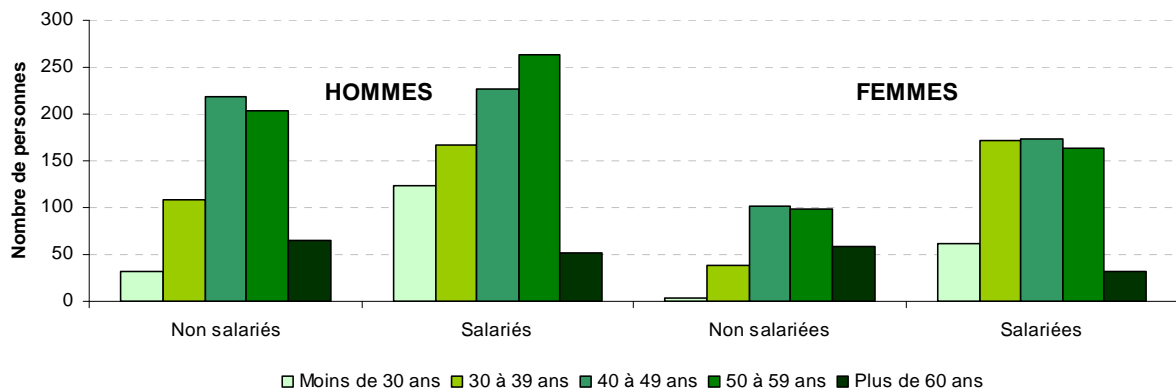
Parmi les 2 363 répondants au questionnaire de la phase pilote Coset, 62 % sont des hommes (1 461). Les salariés représentent 61 % des répondants (1 432 salariés *versus* 922 non-salariés). Environ 6 % des participants sont d'origine étrangère. Ils sont salariés dans presque tous les cas.

4- Age des participants

L'âge moyen des répondants est de 45 ans (minimum 20 ans et maximum 67 ans). Parmi les participants, que ce soit chez les hommes où les femmes, une plus grande proportion de très jeunes actifs est observée dans la population salariée.

Annexe 2 -Figure 4

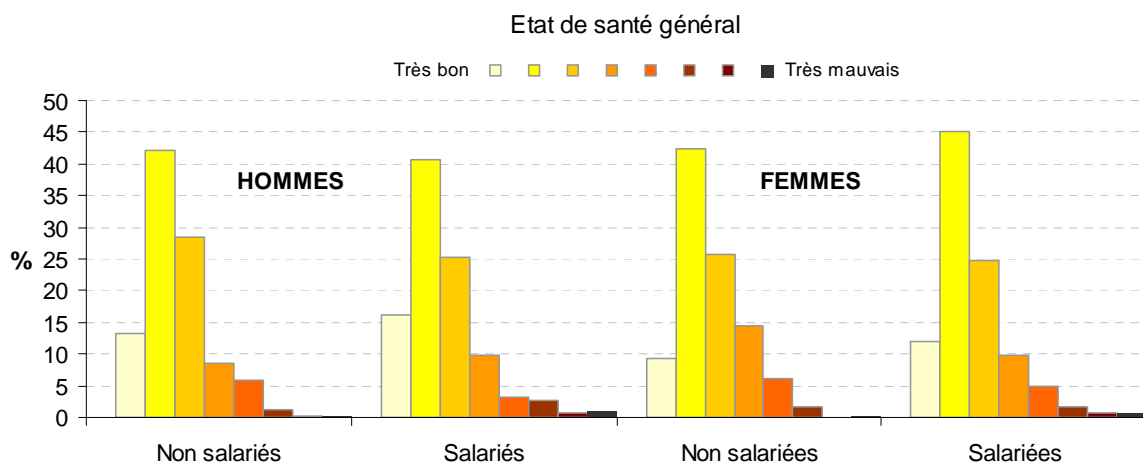
Répartition des répondants par classe d'âge selon le sexe et le statut professionnel, en effectifs



5- Etat de santé général perçu

Annexe 2 -Figure 5

Etat de santé général selon le statut professionnel



L'état de santé général perçu est similaire selon le statut professionnel. Pour les hommes comme pour les femmes, les non-salariés se déclarent cependant un peu moins souvent en très bon état de santé général que les salariés.

6- Antécédents : maladies cardio-vasculaires, métaboliques, endocriniennes, cancer

Annexe 2 - Tableau 2

Maladies cardio-vasculaire et hypercholestérolémie

N (%)	Hypertension artérielle	Angine de poitrine	Infarctus du myocarde	AVC
Oui	275 (11,6)	40 (1,7)	27 (1,1)	38 (1,6)
Non	2 013 (85,2)	2 218 (93,9)	2 225 (94,2)	2 216 (93,8)
Manquant	75 (3,2)	105 (4,4)	111 (4,7)	109 (4,6)
Si oui, âges renseignés	254 (92,4)	19 (47,5)	6 (22,2)	15 (39,5)

N (%)	Artérite des membres inférieurs	Autre affection cardio-vasculaire	Hypercholestérolémie	Diabète
Oui	41 (1,7)	72 (3,0)	133 (5,6)	70 (3,0)
Non	2 211 (93,6)	2 178 (92,2)	1 342 (56,8)	2 219 (93,9)
Manquant	111 (4,7)	113 (4,8)	888 (37,6)	74 (3,1)
Si oui, âges renseignés	18 (43,9)	48 (66,7)	124 (93,2)	51 (72,9)

Un répondant sur dix déclare avoir des antécédents d'hypertension artérielle. C'est le type d'antécédent le plus déclaré, suivi de l'hypercholestérolémie (5,6 %). La question relative à l'hypercholestérolémie n'a pas été renseignée par 37,6 % des cohortistes. L'âge au diagnostic est renseigné de manière très variable selon le type de pathologie (entre 22 et 93 %).

Annexe 2 - Tableau 3

Répartition des traitements pour les personnes se déclarant diabétiques

Traitement diabète	N (%)
Comprimé uniquement	36 (51,4)
Insuline uniquement	5 (7,1)
Deux traitements	6 (8,6)
Aucun traitement	6 (8,6)
Manquant	17 (24,3)
Total	70 (100,0)

Sur les 70 personnes se déclarant diabétiques, 60 % déclarent être traités par comprimés et/ou insulinothérapie et près d'un quart n'ont pas répondu à la question sur le traitement.

Annexe 2 - Tableau 4

Cancers

Cancer	N (%)
Oui	63 (2,7)
Non	2 203 (93,2)
Manquant	97 (4,1)
Total	2 363 (100,0)

Annexe 2 - Tableau 5

Localisations de cancer et âges renseignés aux diagnostics

N (%)	Sein	Col de l'utérus	Utérus sauf col	Ovaires	Poumon, bronches, trachée
Oui	17 (0,7)	8 (0,3)	6 (0,3)	4 (0,2)	4 (0,2)
Si oui, âges renseignés	13 (76,5)	5 (62,5)	3 (50,0)	1 (25,0)	1 (25,0)

N (%)	Prostate	Côlon/rectum	Mélanome	Testicule	Autres localisations
Oui	8 (0,3)	5 (0,2)	5 (0,2)	4 (0,2)	17 (0,7)
Si oui, âges renseignés	5 (62,5)	2 (40,0)	5 (100,0)	4 (100,0)	15 (88,2)

Soixante-trois sujets déclarent des antécédents de cancer. L'âge au diagnostic parmi les personnes ayant déclaré un cancer est très souvent manquant et le pourcentage de valeurs manquantes est très variable selon la localisation.

7- Santé respiratoire

Annexe 2 - Tableau 6

Evolution de l'état des bronches et de la respiration en 10 ans

N (%)	Etat bronches et respiration en 10 ans
S'est amélioré	104 (4,4)
N'a pas changé	1 864 (78,8)
S'est détérioré	312 (13,2)
Manquant	83 (3,5)
Total	2 363 (100,0)

Annexe 2 - Tableau 7

Symptômes respiratoires au cours des 12 derniers mois

N (%)	Sifflement dans la poitrine	Gêne respiratoire	Crise d'essoufflement
Oui	235 (9,9)	198 (8,4)	103 (4,4)
Non	2 041 (86,4)	2 085 (88,2)	2 177 (92,1)
Manquant	87 (3,7)	80 (3,4)	83 (3,5)

N (%)	Crise d'essoufflement, après effort intense	Réveil par une crise d'essoufflement	Réveil par une quinte de toux
Oui	408 (17,3)	59 (2,5)	461 (19,5)
Non	1 888 (79,9)	2 246 (95,1)	1 849 (78,3)
Manquant	67 (2,8)	58 (2,4)	53 (2,2)

Près de 20 % des répondants à la phase pilote Coset ont été réveillés par une quinte de toux sur les douze derniers mois. Sur la même période, 408 personnes ont eu une crise d'essoufflement après un effort intense (17 % des répondants). Le pourcentage de données manquantes est d'environ 3 %.

Annexe 2 - Tableau 8

Toux

N (%)	Toux au lever, en hiver	Toux journée/nuite, en hiver
Oui	267 (11,3)	298 (12,6)
Non	2 036 (86,2)	1 991 (84,3)
Manquant	60 (2,5)	74 (3,1)

Quatorze pour cent des personnes déclarant tousser au lever, la journée ou la nuit en hiver n'ont pas répondu à la question filtrée « Toussiez-vous comme cela presque tous les jours pendant trois mois de suite chaque année ? ».

Annexe 2 - Tableau 9

Crachats provenant de la poitrine

N (%)	Au lever, en hiver	La journée ou la nuit en hiver
Oui	203 (8,6)	192 (8,1)
Non	2 095 (88,7)	2 095 (88,7)
Manquant	65 (2,7)	76 (3,2)

Seize pour cent des personnes ayant déclaré cracher au lever, la journée ou la nuit en hiver, n'ont pas répondu à la question filtrée suivante « Crachez-vous comme cela presque tous les jours pendant trois mois de suite chaque année ? ».

Annexe 2 - Tableau 10

Essoufflement par une marche rapide sur terrain plat ou par une marche normal sur côte légère

Essoufflements sur terrain plat/côte légère	N (%)
Oui	430 (18,2)
Non	1 887 (79,9)
Manquant	46 (1,9)
Total	2 363 (100)

Les personnes essoufflées par une marche rapide sur terrain plat ou une marche normale sur côte légère sont 6,5 % à ne pas avoir répondu à la question filtrée suivante « Etes-vous essoufflé(e) quand vous marchez avec d'autres gens de votre âge en terrain plat ? ».

Annexe 2 - Tableau 11

Asthme

Asthme	N (%)
Oui	217 (9,2)
Non	2 072 (87,7)
Manquant	74 (3,1)
Total	2 363 (100)

Annexe 2 - Tableau 12

Informations complémentaires sur 217 personnes se déclarant asthmatiques

N (%)	Oui	Non	Manquant
Confirmation médecin	187 (86,2)	14 (6,4)	16 (7,4)
Age première crise	190 (87,6)	27 (12,4)	NA*
Age dernière crise	166 (76,5)	51 (23,5)	NA
Crise les 12 derniers mois	55 (25,3)	146 (67,3)	16 (7,4)
Médicaments	70 (32,3)	135 (62,2)	12 (5,5)

*Non applicable

Annexe 2 - Tableau 13

Allergies nasales

Allergies nasales	N (%)
Oui	552 (23,4)
Non	1 687 (71,4)
Manquant	124 (5,2)
Total	2 363 (100)

8- Muscles et articulations

Annexe 2 - Tableau 14

Problèmes musculaires ou articulaires au cours des douze derniers mois des 2 363 participants

N (%)	Nuque/cou	Epaule	Coude/avant-bras
Oui	952 (40,3)	829 (35,1)	518 (21,9)
Non	1 025 (43,4)	1 107 (46,8)	1 301 (55,1)
Manquant	386 (16,3)	427 (18,1)	544 (23,0)

N (%)	Main/poignet/ doigts	Bas du dos	Genou/jambe
Oui	710 (30,0)	1 345 (56,9)	874 (37,0)
Non	1 172 (49,6)	775 (32,8)	1 075 (45,5)
Manquant	481 (20,4)	243 (10,3)	414 (17,5)

Annexe 2 - Tableau 15

Problèmes musculaires ou articulaires au cours des 7 derniers jours des 2 363 participants

N (%)	Nuque/cou	Epaule	Coude/avant-bras
Oui	510 (21,6)	456 (19,3)	292 (12,4)
Non	1 544 (65,3)	1 581 (66,9)	1 692 (71,6)
Manquant	309 (13,1)	326 (19,3)	379 (16,0)

N (%)	Main/poignet/ doigts	Bas du dos	Genou/jambe
Oui	418 (17,7)	759 (32,1)	454 (19,2)
Non	1 605 (67,9)	1 366 (57,8)	1 596 (67,5)
Manquant	340 (14,4)	238 (10,1)	313 (13,3)

Seuls 1,2 % des répondants n'ont répondu à aucune des 6 questions sur les problèmes musculaires ou articulaires au cours des douze derniers mois et 2,4 % à aucune des questions sur les sept derniers jours. Quatre-vingts pour cent des participants ont déclaré au moins un problème musculaire ou articulaire au cours des 12 derniers mois et 57 % des participants un problème au cours des sept derniers jours. Ces problèmes concernent en premier lieu le bas du dos. On note un pourcentage élevé de valeurs manquantes lorsque chaque question est prise isolément.

9- CESD

56 personnes soit 2,4 % des participants n'ont répondu à aucune des questions concernant le moral. En moyenne, le taux de non-réponses par question est de 6,3 %. 1 963 personnes (83 %) ont répondu à tous les items sur le questionnaire CESD. Plus de 93 % des personnes avaient plus de 16 items renseignés.

10- Foyer

Annexe 2 - Tableau 16

Situation familiale des personnes interrogées

	N (%)
Célibataire	593 (25,1)
Marié(e)/pacsé(e)	1 557 (65,9)
Veuf(ve)	34 (1,4)
Divorcé(e)/Séparé(e)	162 (6,9)
Manquant	17 (0,7)
Total	2 363 (100,0)

11- Comportement et santé

Annexe 2 - Tableau 17

Consommation de tabac

Consommation tabac	N (%)
Fumeur	526 (22,2)
Ex-fumeur	369 (15,6)
Non-fumeur	1 337 (56,6)
Manquant	132 (5,6)
Total	2 363 (100,0)

Annexe 2 - Tableau 18

Nombre de cigarettes journalières et statut tabagique

Nombre cigarettes/j	Fumeur	Ex-fumeur
	N (%)	N (%)
<20 cigarettes	335 (63,7)	153 (41,5)
≥20 cigarettes	119 (22,6)	129 (35,0)
Manquant	72 (13,7)	87 (23,6)
Total	526 (100)	369 (100)

Annexe 2 - Tableau 19

Fréquence de la consommation d'alcool chez les 2 363 participants

Fréquence consommation alcool	N (%)
1- Jamais	366 (15,5)
2- 1 fois par mois ou moins	521 (22,1)
3- 2 à 4 fois par mois	733 (31,0)
4- 2 à 3 fois par semaine	384 (16,3)
5- Au moins 4 fois par semaine	337 (14,2)
Manquant	22 (0,9)
Total	2 363 (100)

Programme Coset : Cohortes pour la surveillance épidémiologique en lien avec le travail

Premier bilan de la phase pilote pour la mise en place de la cohorte d'actifs relevant du régime agricole au moment de l'inclusion - cohorte Coset-MSA

Le Département santé travail (DST) de l'Institut de veille sanitaire (InVS) a souhaité se doter d'un dispositif d'observation longitudinal destiné à améliorer la surveillance épidémiologique des risques professionnels, le programme Coset (Cohortes pour la surveillance épidémiologique en lien avec le travail). Ce programme a pour objectif de décrire la morbidité et son évolution en relation avec les facteurs professionnels à l'échelle de la population active en France. Il s'appuie sur des données de cohortes d'actifs et anciens actifs affiliés à l'un des trois principaux régimes de Sécurité sociale qui couvrent 95 % des actifs en France. C'est dans ce cadre qu'en collaboration étroite avec la Caisse centrale de la Mutualité sociale agricole (CCMSA), le DST de l'InVS a programmé le recrutement et le suivi d'une cohorte de personnes tirées au sort parmi les actifs affiliés au Régime agricole à l'inclusion. Après une phase préparatoire destinée à mettre au point le dispositif d'inclusion et les outils de recueil, une étude pilote a été mise en œuvre à partir d'octobre 2009, afin de tester sur une population restreinte le dispositif en vue de son déploiement à l'échelle nationale. Le présent document fait la synthèse des étapes préparatoires à l'étude et des différents enseignements tirés de la phase pilote.

Mots clés : cohorte prospective, surveillance épidémiologique, risques professionnels, MSA, faisabilité, France

COSET Programme: Cohorts for the Epidemiologic surveillance of occupational risks

Preliminary results of the pilot phase for the implementation of a cohort of workers relevant to the Agricultural Health Insurance Fund - COSET-MSA Cohort

The Department of Occupational Health of the French Institute for Public Health Surveillance (InVS) is setting up a longitudinal epidemiological surveillance programme to monitor occupational risks. This programme, named COSET (Cohortes pour la surveillance épidémiologique en lien avec le travail) aims to describe the morbidity among working people and its evolution over time in relation to occupational factors at the national working population scale. It is based on working population cohorts' data collected from workers covered by the three main French insurance fund systems, representing 95% of the working population in France.

In this context and in partnership with the Agricultural Health Insurance Fund, the occupational health Department of the InVS has planned the recruitment and monitoring of a cohort of agricultural workers (randomly selected employed and self employed workers). After a preparatory phase intended to define inclusion criteria and data collection tools, a pilot study was conducted from October 2009 to December 2010 among a restricted population to consider the national implementation of the programme. The present report summarises the different preparatory stages and the practical lessons learned from this pilot study.

Citation suggérée :

Geoffroy-Perez B, Bénétet L, Santin G, Delabre L, Delézure P, Chatelot J. Programme Coset : Cohortes pour la surveillance épidémiologique en lien avec le travail. Premier bilan de la phase pilote pour la mise en place de la cohorte d'actifs relevant du régime agricole au moment de l'inclusion - cohorte Coset-MSA. Saint-Maurice: Institut de veille sanitaire; 2012. 45 p. Disponible à partir de l'URL : <http://www.invs.sante.fr>