

# Enquête complémentaire auprès d'un échantillon de non-participants lors de la phase pilote de la cohorte Coset-MSA

---

## Protocole d'étude

---

Version du 3 février 2010

### **Rédaction du protocole**

Gaëlle SANTIN

Laetitia BENEZET

Béatrice GEOFFROY-PEREZ

### **Ont également contribué à la mise au point de ce protocole**

Nicolas VIAROUGE (Direction des études répertoires et statistiques - CCMSA)

Florient BREMAUD (Direction de la maîtrise d'ouvrage institutionnelle - CCMSA)

Ellen IMBERNON (DST-InVS)

## SOMMAIRE

1	Introduction .....	5
2	Objectifs .....	6
3	Méthode .....	6
3.1	Type d'enquête .....	6
3.2	Échantillonnage .....	6
3.3	Modalités de recueil des données .....	7
3.3.1	Données recueillies .....	7
3.3.1.1	Questionnaire d'enquête .....	7
3.3.1.2	Paradonnées .....	7
3.3.1.3	Numéros de téléphone .....	7
3.3.2	Recueil des données .....	8
3.3.2.1	Recueil des données par téléphone .....	10
3.3.2.2	Recueil des données en face-à-face .....	10
3.3.3	Formation des enquêteurs .....	10
3.4	Circuit de transmission des données et confidentialité .....	10
3.5	Information des personnes .....	12
3.6	Analyses statistiques .....	12
3.6.1	Principe général .....	12
3.6.2	Analyse des données .....	13
3.6.2.1	Description des données relatives au déroulement de la collecte .....	13
3.6.2.2	Étude des biais liés au recueil des données .....	13
3.6.2.3	Étude des biais de non-participation .....	14
4	Confidentialité .....	14
5	Durée et modalité d'organisation de l'étude .....	15
5.1	Groupe de travail .....	15
5.2	Gestion de l'étude .....	15
5.3	Calendrier .....	15
6	Bibliographie .....	16

## Résumé

### État de la question et objectif de la recherche

Ce protocole entre dans le cadre du programme Coset (Cohortes pour la surveillance épidémiologique en lien avec le travail). Un premier protocole décrivant le dispositif pilote d'inclusion d'actifs couverts par le régime de la mutualité sociale agricole a précédemment été soumis CCTIRS et autorisé par la CNIL. Un deuxième protocole, soumis actuellement au CCTIRS, décrit les modalités de recueil et d'analyse de données individuelles (dites données « auxiliaires »), relatives à la santé et à la situation socioprofessionnelle, disponibles dans les systèmes d'informations existants. Le présent protocole est en lien direct avec les deux précédents ; il a pour objet la réalisation d'une enquête complémentaire auprès d'un sous-échantillon de non-participants à l'enquête initiale.

Les objectifs de cette enquête sont avant tout descriptifs, puisqu'elle a pour objet de permettre d'établir au mieux des estimations sans biais des variables d'intérêt majeur pour la cohorte Coset-MSA ; pour cela, le taux de réponse à cette enquête complémentaire devra être maximal, d'où l'accent particulier mis sur le recueil de données.

Les estimations obtenues en combinant les réponses des participants à l'enquête initiale de la phase pilote et de cette enquête complémentaire seront confrontées à celles obtenues après correction des biais de non-participation en utilisant les données issues des systèmes d'information. Cette double approche permettra 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.

### Population concernée

Afin de répondre aux objectifs précédemment énoncés, l'enquête concernera un échantillon aléatoire de 500 éligibles non-participants à l'enquête initiale (100 personnes dans chacun des 5 départements participant à l'enquête pilote). Pour rappel, cette dernière concerne 10 000 actifs de 18 à 65 ans affiliés à la MSA en 2008, dépendant d'une des cinq caisses pilotes (Bouches-du-Rhône, Finistère, Pas-de-Calais, Pyrénées-Atlantiques, Saône-et-Loire).

### Méthode d'observation ou d'investigation retenue

L'enquête complémentaire est ponctuelle, transversale par sondage aléatoire. Contrairement aux personnes ayant répondu à l'enquête initiale, les participants à cette enquête complémentaire ne seront pas suivis par questionnaire. Chaque personne recevra au préalable une lettre d'annonce présentant l'enquête et les droits de l'enquêté. Le recueil des données sera réalisé par un prestataire. Deux modes de recueil seront attribués aléatoirement par le Département santé travail de l'InVS (DST) et comparés en termes de taux de réponse et de qualité des données recueillies : (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 qui reprend une partie des questions posées lors de l'enquête postale initiale, portera sur l'état de santé et la situation professionnelle de la personne. La durée de l'interview n'excédera pas 15 minutes. La saisie des données sera effectuée au moment de l'interview par l'enquêteur. Par ailleurs, des données sur déroulement de l'enquête (nombre de tentatives avant contact, refus, etc.) seront également recueillies.

### Origine et nature des données nominatives recueillies, Justification du recours à celles-ci

L'enquête nécessitant un contact en face-à-face ou par téléphone des personnes sélectionnées par tirage au sort, leurs nom, prénom, adresse et numéros de téléphone seront conservés durant toute la durée du recueil des données. Les nom, prénom et adresse seront issus des fichiers administratifs de la MSA, les numéros de téléphone seront recherchés par le prestataire à partir des nom, prénom et adresse postale. Ces informations seront cryptées et rattachées à un numéro de confidentialité NC-DST.

## **Mode de circulation des données**

Le dispositif envisagé pour le circuit des données fait intervenir (1) le DST qui tire au sort et transmet à la CCMSA le numéro de confidentialité NC-DST des personnes à enquêter, ainsi que leur caisse d'affiliation, leur groupe (recueil prioritairement par téléphone ou en face-à-face) et le statut NPAI de leur adresse postale, (2) la CCMSA qui transmet un fichier crypté au prestataire, contenant les informations transmises par le DST ainsi que des informations directement nominatives telles que les civilité, nom, prénom et adresse postale des personnes à enquêter (3) un prestataire chargé de l'organisation et de la réalisation du recueil des données. Le prestataire constituera deux fichiers de données. Le premier fichier comprendra les réponses au questionnaire ; les personnes seront indexées par l'intermédiaire d'un numéro de confidentialité unique NC-DST et ce premier fichier ne contiendra aucune donnée nominative. Ce fichier sera transmis par le prestataire au DST à l'issue du recueil. Le second fichier contiendra les données identifiantes (nom, prénom, adresse, numéro de téléphone) durant le déroulement de l'enquête, afin que les enquêteurs puissent joindre les personnes ; il contiendra par ailleurs le numéro de confidentialité NC-DST ainsi que toutes les informations relatives au déroulement de la collecte. Ce fichier de données sera conservé avec un système de sécurité et de cryptage. A la fin du recueil, les informations relatives au déroulement de la collecte seront transmises au DST après suppression des informations identifiantes.

## **Durée et modalités d'organisation de la recherche**

L'enquête complémentaire est supervisée par le DST, l'organisation et la réalisation du recueil des données étant confiées à un prestataire. Le DST est en charge de la procédure d'échantillonnage, de la rédaction du protocole, des questionnaires, du dossier pour la Cnil et du cahier des charges pour l'appel d'offre en collaboration avec les services concernés de l'InVS, ainsi que de l'analyse statistique des données. Les propositions de participation à l'enquête complémentaire ainsi que le recueil des données auront lieu de septembre à octobre 2010. L'analyse des données recueillies se déroulera jusqu'à fin juin 2011. Les données de l'étude seront conservées par l'InVS au minimum pendant toute la durée d'activité du programme.

## **Méthode d'analyse des données**

Les analyses des données de l'enquête consisteront en une description des données relatives au déroulement de l'enquête (paradonnées). L'accord de participation attendu étant d'environ 70%, une description du statut « répondant/non-répondant » sera effectuée en fonction des paradonnées. Puis une étude des biais liés aux modes de recueil des données sera réalisée. Enfin, les estimations issues des données de l'enquête postale initiale combinées avec les données provenant de la présente enquête complémentaire seront comparées aux estimations obtenues à partir des données de l'enquête postale initiale, corrigées ou non pour la non-participation par l'intermédiaire des informations auxiliaires. Ces analyses permettront de discuter sous plusieurs aspects les biais liés à la non-participation dans l'étude pilote Coset-MSA.

## **Justification du nombre de sujets ou analyse de puissance**

La taille de l'échantillon de l'enquête complémentaire a été déterminée selon un principe de parcimonie entre les coûts de collecte, qui seront plus élevés que ceux de l'enquête initiale, compte tenu de l'objectif de maximisation du taux de réponse, et les effectifs, qui doivent être suffisamment grands pour obtenir des estimations fiables. Un échantillon de 100 non-participants par caisse départementale sera constitué par tirage au sort ; la taille totale de l'échantillon sera donc de 500 personnes.

# 1 Introduction

Ce protocole entre dans le cadre du programme Coset (Cohortes pour la surveillance épidémiologique en lien avec le travail) et plus particulièrement dans le cadre de la phase pilote d'inclusion d'actifs affiliés au régime de la Mutualité sociale agricole (MSA). Un premier protocole [1] a été précédemment soumis et accepté par la Cnil (réf. dossier n° 909091, voir annexe 1) ; il décrit le dispositif pilote d'inclusion d'actifs couverts par le régime de la Mutualité sociale agricole et les données recueillies, c'est-à-dire les données du questionnaire et les données nécessaires pour suivre la personne. Pour rappel, le recueil des données est effectué par autoquestionnaire postal. Dans la suite du document, l'étude pilote d'inclusion Coset-MSA sera dénommée « enquête initiale ». Un autre protocole est actuellement en cours de soumission à la Cnil, pour le recueil complémentaire d'informations auxiliaires issues des bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav pour l'ensemble des personnes incluses lors de la phase pilote de la cohorte Coset-MSA [2].

Le présent protocole est en lien direct avec les deux précédents ; il a pour objet la réalisation d'une enquête complémentaire auprès d'un échantillon de non-répondants à la phase d'inclusion du pilote afin d'appréhender les biais de non-participation à l'inclusion. Cette enquête ponctuelle sera réalisée au décours de l'inclusion pilote et permettra de prendre en compte la non-participation d'une façon complémentaire à celle envisagée au travers des informations auxiliaires. Pour rappel, il est prévu, lors de la phase pilote de la cohorte Coset-MSA, mais également lors de son extension nationale, d'utiliser les informations auxiliaires contenues dans les bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav pour corriger les biais de non-participation liés au volontariat. Le recours à cette méthode pour minimiser ce type de biais peut être cependant discuté ; en effet, rien ne garantit que les informations auxiliaires dont on disposera soient suffisantes pour corriger efficacement le biais de non-participation [3]. Étant donné que le taux de réponse attendu ne devrait pas excéder 30% et que les objectifs du programme Coset sont avant tout descriptifs, il est apparu primordial d'appréhender au mieux les biais de non-participation liés au volontariat.

C'est pourquoi il a été décidé de mettre en œuvre une enquête complémentaire auprès d'un échantillon de non-participants à l'étude pilote. Le recours à ce type d'enquête est fréquent outre-Atlantique [4-7]. Les moyens envisagés pour cette étude ont pour objectif de maximiser autant que possible le taux de réponse à cette enquête complémentaire. Les réponses des participants à cette enquête combinées à celles des participants à l'enquête initiale devraient permettre d'obtenir des estimations représentatives de la population d'intérêt, si le taux de participation à l'enquête complémentaire est proche de 100%.

Cette double approche permettra de discuter sous plusieurs angles les biais liés à la non-participation à l'inclusion dans 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.

<p>Le présent protocole décrit l'échantillonnage et les modalités de recueil de données de l'enquête complémentaire auprès de non-participants. Il aborde ensuite les analyses statistiques envisagées à partir des données issues de cette enquête complémentaire dans le cadre de l'étude d'inclusion pilote des travailleurs affiliés à la MSA.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## 2 Objectifs

Les objectifs de l'enquête complémentaire sont avant tout descriptifs, puisqu'elle doit permettre d'établir au mieux des estimations sans biais des variables d'intérêt majeur pour la cohorte Coset-MSA ; pour cela, le taux de réponse à cette enquête complémentaire devra être maximal, d'où l'accent particulier mis sur le recueil de données.

Les estimations obtenues en combinant les réponses des participants à l'enquête initiale de la phase pilote et à cette deuxième enquête seront ensuite confrontées à celles obtenues à partir de la première enquête corrigées ou non pour la non-participation à partir des données issues des systèmes d'information.

## 3 Méthode

### 3.1 Type d'enquête

L'enquête complémentaire auprès d'un échantillon de non-participants est une enquête téléphonique ou en face-à-face, transversale par sondage aléatoire. Contrairement aux personnes ayant répondu à l'enquête initiale et donc volontaires pour un suivi de cohorte, les personnes participant à cette enquête complémentaire ne seront pas sollicitées ultérieurement.

### 3.2 Échantillonnage

La population cible correspond à tous les individus non-participants à l'enquête initiale et qui ne se seront pas opposés à l'accès aux données auxiliaires. Pour rappel, la population étudiée lors de l'enquête initiale correspondait aux actifs, salariés ou à leur compte, affiliés à la MSA, âgés de 18 à 65 ans ; dix mille personnes affiliées aux caisses MSA des Bouches-du-Rhône, du Finistère, du Pas-de-Calais, des Pyrénées-Atlantiques et de la Saône-et-Loire avaient été tirées au sort pour participer à l'enquête initiale, soit 2 000 personnes par département.

Les sujets interviewés dans le cadre de l'enquête complémentaire seront sélectionnés par un tirage au sort stratifié selon la caisse départementale d'appartenance du non-participant.

La taille de l'échantillon de l'enquête complémentaire auprès de non-participants a été déterminée selon un principe de parcimonie entre les coûts de collecte, et les effectifs, qui doivent être suffisamment grands pour obtenir des estimations fiables.

Un échantillon de 100 non-participants par caisse départementale sera constitué par tirage au sort. Ils seront répartis aléatoirement en deux groupes avec des modes de collecte de données différents.

Parmi les 100 non-participants tirés au sort dans une caisse :

- 70 seront sélectionnés aléatoirement pour être affectés au groupe 1 : enquête prioritairement par téléphone, et enquête en face-à-face pour les personnes non jointes par téléphone ;
- 30 seront sélectionnés aléatoirement pour être affectés au groupe 2 : enquête prioritairement en face-à-face, et enquête par téléphone pour les personnes hors de la zone géographique de l'enquête en face-à-face (ie pour les personnes habitant plus de 20 kilomètres au-delà des limites du département concerné par l'enquête).

Au total, 500 personnes seront donc échantillonnées ; parmi elles, 350 seront affectées au groupe 1 et 150 seront affectées au groupe 2.

### 3.3 Modalités de recueil des données

Le recueil des données sera réalisé par un prestataire choisi par l'InVS à l'issue d'un appel d'offre. Les données nominatives (nom, prénom, adresse et numéro de téléphone) seront conservées par le prestataire et détruites à l'issue de l'enquête. Les données d'enquête seront liées uniquement à un numéro de confidentialité.

Seules les personnes sélectionnées par tirage au sort devront être enquêtées ; un proche ne pourra pas se substituer pour répondre à sa place.

#### 3.3.1 Données recueillies

##### 3.3.1.1 Questionnaire d'enquête

La longueur du questionnaire initial étant probablement à l'origine de certaines non-réponses et afin de limiter les abandons en cours d'interview, la taille du questionnaire a été conçue afin que sa durée de collecte n'excède pas un quart d'heure. Il se limite à quelques questions sélectionnées parmi celles du questionnaire initial, selon les critères suivants :

- **Variables *a priori* associées à la non-participation [8]** : sexe, âge (mois et année de naissance), catégorie sociale, statut marital, niveau d'étude, comportements (alcool, tabac), état de santé général ;
- **Principales variables d'intérêts** dans le programme Coset : état de santé général, statut vis-à-vis de l'emploi, principales expositions professionnelles, nombre d'épisodes professionnels.

Par ailleurs, les motifs de non-participation à l'enquête postale initiale seront également recueillis. Ceci permettra le cas échéant, d'adapter ou de modifier le questionnaire pour l'extension nationale de la cohorte Coset-MSA.

Le questionnaire de cette enquête figure en annexe 2. La façon de poser les questions sera adaptée au mode de recueil (téléphonique ou en face-à-face) et certaines questions pourront éventuellement être supprimées afin que l'interview n'excède pas le temps prévu. Néanmoins, l'ordre des questions ne sera pas modifié et aucune question ne sera ajoutée.

##### 3.3.1.2 Paradonnées

Un fichier de paradonnées, c'est-à-dire de données relatives à la collecte sera constitué. Ce type d'information est utile pour décrire la non-participation à l'enquête complémentaire. Ce fichier contiendra notamment le nombre et les dates des contacts, les durées d'interview et toute autre information relative au déroulement de l'interview (refus d'interview avec son motif, échec pour contacter la personne, etc.).

##### 3.3.1.3 Numéros de téléphone

Les fichiers de la MSA ne contenant pas les numéros de téléphone de leurs adhérents, une recherche des coordonnées téléphoniques des personnes tirées au sort pour participer à l'enquête complémentaire sera réalisée dans les annuaires téléphoniques disponibles à partir des adresses postales transmises par la CCMSA, que la personne soit enquêtée en première intention par téléphone ou en face-à-face.

### 3.3.2 Recueil des données

Le schéma d'étude (cf. figure 1) de l'enquête complémentaire a pour objectif principal de maximiser le taux de réponse. C'est pourquoi il a été choisi de réaliser une enquête par téléphone plutôt qu'une enquête postale. La MSA ne disposant pas des numéros de téléphone de ses bénéficiaires, il est cependant probable qu'un pourcentage non négligeable de numéros de téléphone ne sera pas retrouvé, car il n'existe pas à ce jour d'annuaire téléphonique exhaustif (cf. annexe 4). La MSA possédant les adresses postales de ses affiliés, il a donc été décidé d'enquêter en face-à-face les personnes non-jointes par téléphone. Cependant, afin d'étudier les biais potentiels liés aux différents modes de collecte des données (par téléphone et en face-à-face) [9], il a été décidé qu'un groupe soit enquêté en première intention en face-à-face que son numéro de téléphone ait été retrouvé ou non.

Sur les 500 personnes échantillonnées, 350 seront affectées au groupe 1 (enquête par téléphone en première intention) et 150 seront affectées au groupe 2 (enquête en face-à-face en première intention).

Un courrier d'annonce (annexe 3) sera envoyé à toutes les personnes échantillonnées (cf. 3.5.).

#### - Enquête auprès des personnes du groupe 1

Une recherche des coordonnées téléphoniques sera réalisée, à partir des adresses transmises par la CCMSA et mises-à-jour par le prestataire (cf.3.4.).

Cas 1 : les coordonnées téléphoniques sont retrouvées. L'enquêteur tente donc de contacter la personne avec deux issues possibles :

- L'enquêteur réussit à contacter la personne et réalise donc l'enquête par téléphone ;
- L'enquêteur ne réussit pas à contacter la personne. Il se rendra alors à son domicile pour tenter de réaliser l'enquête en face-à-face.

Cas 2 : les coordonnées téléphoniques ne sont pas retrouvées : la personne sera alors enquêtée directement en face-à-face.

#### - Enquête auprès des personnes du groupe 2

Une zone géographique jusqu'à 20 kilomètres autour de chaque département pilote sera définie et l'enquête en face-à-face se limitera à cette zone. Comme pour le groupe 1, une recherche des coordonnées téléphoniques sera réalisée.

Cas 1 : la personne réside dans cette zone et ses coordonnées téléphoniques sont retrouvées. Un rendez-vous sera alors pris pour réaliser l'enquête en face-à-face. Si l'enquêteur ne réussit pas à joindre la personne, il se rendra directement à son domicile pour l'interviewer.

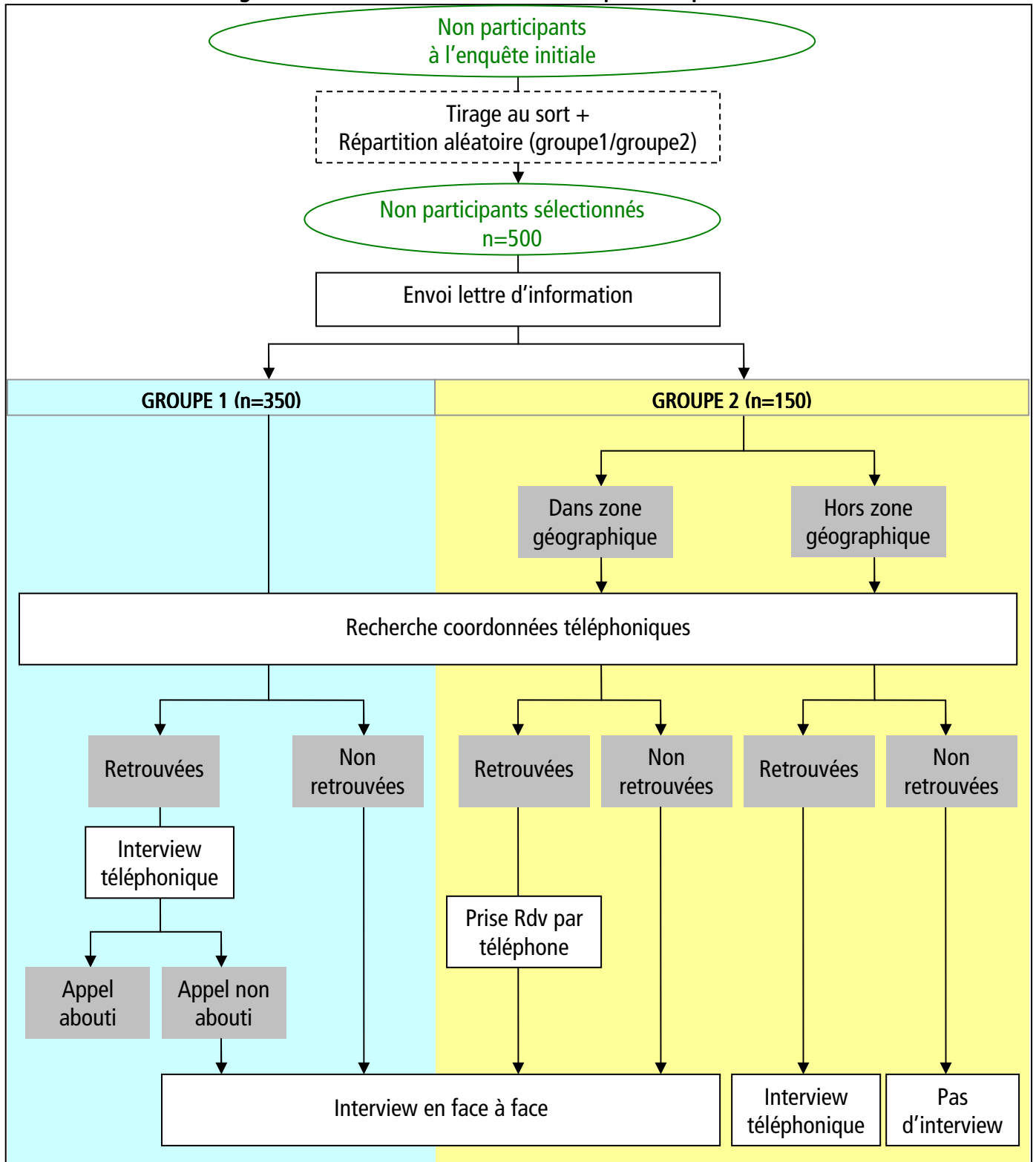
Cas 2 : la personne réside dans cette zone et ses coordonnées téléphoniques ne sont pas retrouvées. L'enquêteur se rendra directement au domicile de la personne pour l'interviewer.

Cas 3 : la personne ne réside pas dans cette zone et ses coordonnées téléphoniques sont retrouvées. L'enquêteur tentera alors de réaliser l'enquête par téléphone.

Cas 4 : la personne ne réside pas dans cette zone et ses coordonnées téléphoniques ne sont pas retrouvées. Elle ne sera alors pas interviewée.

Même si de nombreuses options ont été envisagées, les personnes ne seront effectivement sollicitées qu'une seule fois. En effet, dès lors que la personne est contactée, si elle refuse de répondre à l'enquête, elle ne sera pas sollicitée de nouveau.

**Figure 1 : modalités de réalisation de l'enquête complémentaire**



### 3.3.2.1 Recueil des données par téléphone

Le recueil des données par téléphone concernera :

- Les personnes du groupe 1 pour lesquelles le numéro de téléphone a été retrouvé ;
- Les personnes du groupe 2 qui résident hors de la zone géographique de l'enquête en face-à-face et pour lesquelles le numéro de téléphone a été retrouvé.

On aura recours à la méthode de Collecte Assistée par Téléphone et Informatique (CATI) où l'enquêteur remplit directement les réponses au questionnaire via un ordinateur. Etant donné qu'il s'agit d'une population d'actifs, les créneaux horaires privilégiés pour les appels seront de 12h à 14h et de 17h à 21h en semaine et de 10h à 18h le samedi. En cas de non-réponse ou d'absence (pas de réponse, occupé, répondeur) jusqu'à 15 tentatives seront réalisées à des jours et heures différents (dont au moins un samedi).

Si la personne est indisponible au moment de l'appel, un rendez-vous téléphonique sera proposé.

### 3.3.2.2 Recueil des données en face-à-face

Le recueil des données en face-à-face concernera :

- Les personnes du groupe 2 qui résident dans la zone géographique de l'enquête en face-à-face ;
- Les personnes du groupe 1 pour lesquelles le numéro de téléphone n'a pas été retrouvé ;
- Les personnes du groupe 1 pour lesquelles un contact n'aura pu être établi par téléphone, bien que leur numéro ait été retrouvé.

Comme pour les interviews téléphoniques, on aura recours à une saisie directe des réponses par l'enquêteur via la méthode CAPI (Computer Assisted Personal Interview).

Les personnes dont le numéro de téléphone a pu être retrouvé seront préalablement contactées par téléphone pour fixer une date de rendez-vous.

Les visites seront effectuées de préférence à des créneaux horaires où les personnes ont plus de chance de se trouver à leur domicile, c'est-à-dire les soirées en semaine ou la journée du samedi. En cas d'absence, l'enquêteur laissera un avis de passage et un numéro vert auquel le joindre. Jusqu'à trois passages seront effectués.

### 3.3.3 Formation des enquêteurs

Les enquêteurs seront préalablement formés. Cette formation portera sur la manière de présenter l'étude aux sujets (référence au courrier d'annonce, rappel des dispositifs de confidentialité, durée approximative de l'interview) et sur l'étude elle-même. Elle aura par ailleurs comme objectif de standardiser au mieux le recueil des données. La formation sera dispensée conjointement par le prestataire et l'InVS.

## 3.4 Circuit de transmission des données et confidentialité

La réalisation de l'enquête complémentaire auprès de non-participants sera confiée à un prestataire. Les modalités de transmission des données, dont certaines sont directement identifiantes, entre le prestataire, le Département santé travail de l'Institut de veille sanitaire (DST-InVS) et la Caisse centrale de la mutualité sociale agricole (CCMSA) seront sécurisées (*cf.* 4.).

Les différentes étapes mises en place pour la réalisation de cette enquête sont les suivantes : (*cf.* figure 2).

**Etape 1 :** A l'issue du tirage au sort pour le pilote d'inclusion dans la cohorte, le DST de l'InVS dispose d'un fichier contenant, pour chaque individu, son numéro de confidentialité (NC-DST) et sa caisse départementale d'affiliation. A l'issue de l'enquête initiale, étant donné que le numéro de confidentialité figure sur le questionnaire, le DST est en mesure de repérer les participants et les non-participants à cette première enquête par leur numéro de confidentialité (NC-DST). Il peut ainsi réaliser le tirage au sort stratifié sur la caisse des non-participants à enquêter et constituer les groupes qui bénéficieront de différents modes de recueil des données. Par ailleurs, à l'issue de l'enquête initiale, le prestataire chargé du routage aura transmis à l'InVS la liste des NC-DST qui n'habitent plus à l'adresse indiquée (NPAI) qui pourront ainsi être identifiés parmi les personnes tirées au sort afin de tenter une mise à jour de leur adresse.

**Etape 2 :** Le DST constitue alors un fichier « **Non-participants à enquêter** » qui contient pour chaque personne sélectionnée pour participer à l'enquête complémentaire son NC-DST, sa caisse départementale d'affiliation, son groupe d'appartenance qui détermine le mode de recueil des données (1-téléphone prioritairement / 2-face-à-face prioritairement) et la notion de NPAI. Il transmet alors ce fichier à la CCMSA.

**Etape 3 :** La CCMSA dispose du fichier de routage de l'enquête initiale avec pour chaque individu son NC-DST, ses civilité, nom, prénom et adresse postale. Elle ajoute ces éléments au fichier « Non-participants à enquêter » et le transmet au prestataire en charge de l'enquête complémentaire.

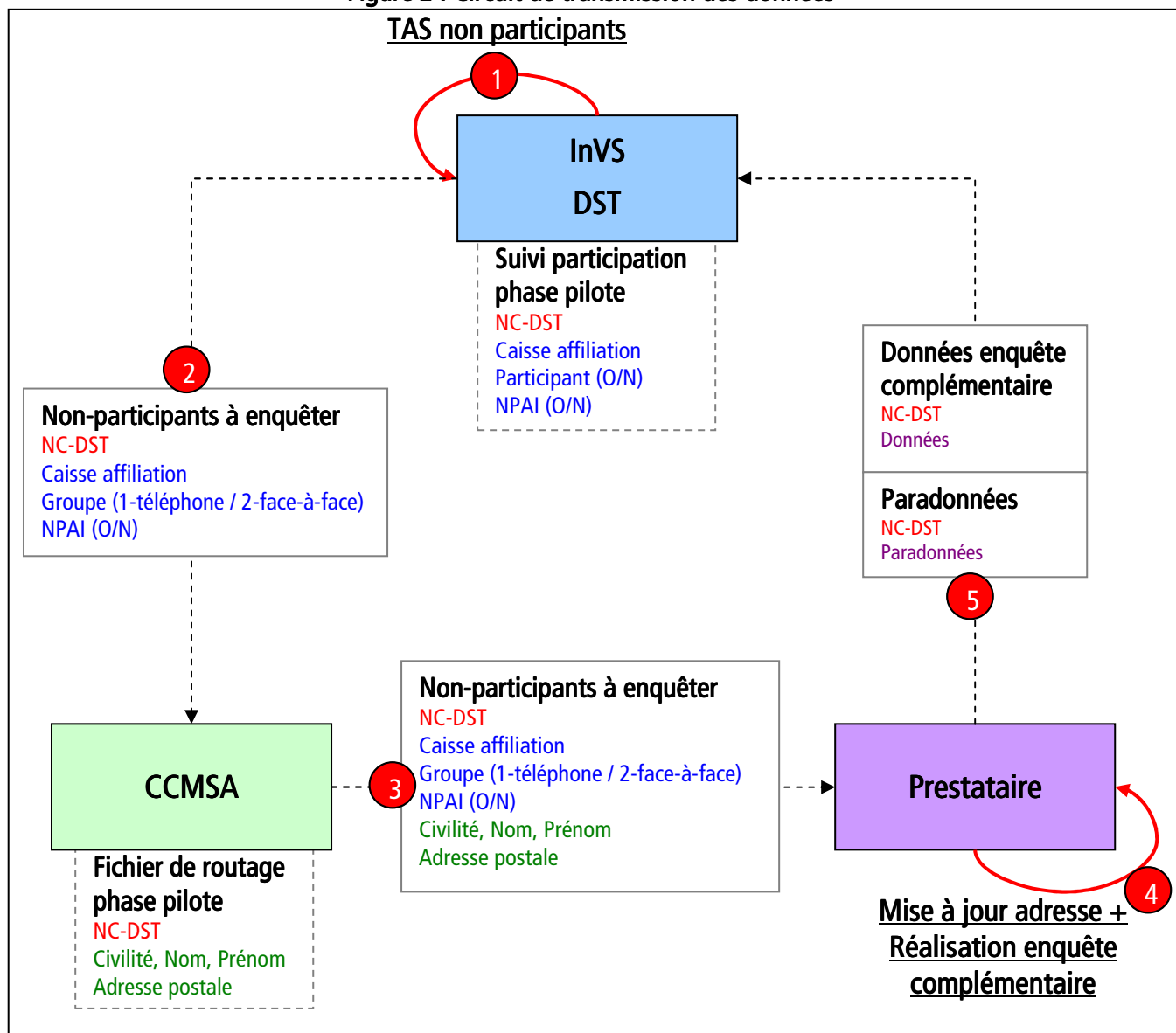
**Etape 4 :** A partir de ce fichier, le prestataire réalise l'enquête complémentaire. Dans un premier temps, il met à jour le fichier des adresses (afin de retrouver éventuellement l'adresse des NPAI) et envoie la lettre d'annonce à l'ensemble des personnes sélectionnées (cf. 3.5). Il effectue ensuite la recherche des coordonnées téléphoniques. La base de données nominatives des personnes contactées, qui inclura par ailleurs une base de gestion des mailings des appels téléphoniques et des visites des enquêteurs restera séparée des bases de données de questionnaires tout au long de la réalisation de l'enquête ; le lien entre les deux bases pourra être réalisé uniquement via le numéro d'identification unique NC-DST, qui figurera dans les deux fichiers. La base des coordonnées (adresse postale et numéro de téléphone) sera conservée uniquement pendant la durée nécessaire aux contacts avec les sujets à enquêter. Après réalisation de l'enquête complémentaire selon les modalités décrites au paragraphe 3.3.2., le prestataire constitue deux fichiers :

- « **Données enquête complémentaire** » contenant les réponses au questionnaire d'enquête complémentaire ;
- « **Paradonnées** », contenant les seules informations individuelles propres au déroulement de l'enquête.

Ces deux fichiers ne contiendront aucune information nominative, les individus y seront repérés uniquement par l'intermédiaire de leur NC-DST fourni au prestataire avec le fichier d'enquête initial.

**Etape 5 :** Un fois l'enquête réalisée, le prestataire transmet au DST le fichier des données d'enquête ainsi que le fichier de paradonnées et détruit le fichier des non-participants à enquêter.

Figure 2 : Circuit de transmission des données



### 3.5 Information des personnes

Un courrier préalable à l'enquête complémentaire, cosigné par l'InVS et la MSA, sera envoyé par le prestataire à toutes les personnes sélectionnées. Il précisera le contexte de l'étude et l'appel téléphonique ou la venue prochaine d'un enquêteur. La confidentialité des données, le traitement des réponses dans le respect de l'anonymat et les droits de la personne seront rappelés. Un numéro vert sera indiqué pour toute demande d'information, prise de rendez-vous et éventuellement pour que la personne puisse s'opposer à l'enquête.

### 3.6 Analyses statistiques

#### 3.6.1 Principe général

La stratégie envisagée de prise en compte des biais liés à la non-réponse dans le programme Coset repose principalement sur l'exploitation des données auxiliaires issues des bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav [2]. Elle consiste à utiliser ces informations,

recueillies à la fois pour les primo-participants et les non-participants, afin de corriger les poids d'échantillonnage des primo-participants pour la non-participation. Cette façon de procéder est justifiée par le fait que ces informations auxiliaires sont liées à la fois à la non-participation et à nos variables d'intérêt.

Néanmoins, on ne peut jamais être certain d'avoir toute l'information nécessaire pour obtenir des estimations sans biais (par définition, on n'observe pas les non-participants, on ne connaît donc leurs caractéristiques qu'au travers des informations auxiliaires).

L'exploitation des données de l'enquête complémentaire permettra avant tout d'appréhender sous un deuxième angle les biais liés à la non-réponse [10].

En effet, si on suppose qu'il n'existe pas d'autres sources de biais (en particulier des erreurs de mesure), les données d'une enquête auprès d'un échantillon de non-participants combinées aux données de l'enquête initiale permettront d'obtenir des estimations approximativement sans biais, sous réserve d'obtenir un taux de participation maximal (idéalement proche de 100%) [3;11].

Cette deuxième méthode est également probablement imparfaite. Tout d'abord, elle est valide pour un taux de réponse proche de 100%, ce qui semble peu réaliste en pratique. Par ailleurs, les personnes pour lesquelles il est plus difficile d'obtenir une réponse lors d'un premier contact sont susceptibles de répondre moins soigneusement aux questions qui leur sont posées, ce qui peut entraîner un biais dans les estimations [12]. En outre, maximiser le taux de réponse est synonyme dans notre enquête de multiplication des modes de recueil des données (par courrier, par téléphone et en face-à-face), ce qui peut entraîner là aussi d'autres formes de biais [9] ; c'est pourquoi pour cette enquête complémentaire, il a été décidé de sous-échantillonner directement des personnes à enquêter par téléphone et d'autre en face-à-face, afin de pouvoir discuter les possibles biais induits par les modes de recueil différents.

Ces deux approches complémentaires (correction pour la non-participation des poids d'échantillonnage des participants initiaux et enquête auprès d'un sous-échantillon de non-participants avec pour objectif de maximiser le taux de participation à cette deuxième enquête) permettront de discuter et documenter au mieux les problèmes liés à la non-participation dans la cohorte Coset-MSA.

## 3.6.2 Analyse des données

### 3.6.2.1 **Description des données relatives au déroulement de la collecte**

A partir de la base de parodonnées, une description des données relatives au déroulement de l'étude sera réalisée. Elle portera notamment sur une description de l'activité de la ligne téléphonique en relation avec le numéro gratuit mis en place (nombre d'appels, synthèse des questions posées) , le nombre d'adresses non retrouvées et mises-à-jour, le nombre de numéro de téléphone non retrouvés, le nombre de courriers envoyés, le nombre d'appels téléphoniques réalisés, et leur durée moyenne, par type d'appel, le nombre de visites et leur durée moyenne. Ces indicateurs seront étudiés selon le groupe d'appartenance du sujet.

Une description des parodonnées en fonction du statut de « répondant » ou « non-répondant » à l'enquête complémentaire sera par ailleurs réalisée.

### 3.6.2.2 **Étude des biais liés au recueil des données**

A l'issue du recueil des données de l'enquête complémentaire, quatre modes de recueil de données auront été utilisés : les personnes du groupe 1 enquêtées par téléphone, les personnes du groupe 1 enquêtées en face-à-face car non-joignables par téléphone, les personnes du groupe 2 enquêtées

en face-à-face, les personnes du groupe 2 enquêtées par téléphone car hors de la zone géographique de l'enquête en face-à-face. Ces quatre groupes seront comparés selon les informations recueillies dans l'enquête.

### **3.6.2.3 Étude des biais de non-participation**

Les caractéristiques des participants à l'enquête initiale et des non-participants ayant répondu à l'enquête complémentaire seront comparées. Ces caractéristiques pourront être issues de la base de sondage de l'enquête initiale (caisse, sexe, âge, statut d'emploi) mais également des variables communes aux questionnaires de l'enquête initiale et de l'enquête complémentaire.

Si le taux de réponse à l'enquête complémentaire auprès des non-participants est proche des 100%, elle permettra, en combinant les données de l'enquête complémentaire avec les données recueillies lors de l'enquête initiale, d'obtenir des estimations sans biais pour la non-participation. Malgré tous les efforts qui seront mis en œuvre pour maximiser le taux de réponse à l'enquête complémentaire, cette condition s'avère difficilement réalisable. Néanmoins, compte tenu de l'accent mis sur ce point, le taux de réponse attendu à l'enquête complémentaire est d'environ 70%. Les estimations issues de la combinaison des réponses à l'enquête initiale avec celles de l'enquête complémentaire pourront être comparées à celles obtenues lors de l'enquête initiale, corrigées ou non pour la non-participation grâce à l'exploitation des informations auxiliaires issues des bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav.

## **4 Confidentialité**

Le circuit des données décrit dans le présent protocole garantit que la réalisation de l'enquête complémentaire se fasse dans le respect de la confidentialité des personnes.

Les échanges physiques de données dans le cadre du présent protocole seront réalisés sur support physique de type CD-ROM ou DVD, soit envoyé par courrier AR, soit remis en main propre par l'émetteur au destinataire.

Les personnes responsables de ces échanges sont :

- Pour l'InVS : Dr Béatrice GEOFFROY, responsable du programme Coset au Département santé travail ;
- Pour la CCMSA : Mr Alain PELC, directeur des Études Répertoires et Statistiques.

La personne responsable du traitement des données à l'InVS est :

- Dr Béatrice GEOFFROY, responsable du programme Coset au Département santé travail.

Les personnes chargées de la mise en œuvre du traitement des données à l'InVS sont :

- Mme Gaëlle SANTIN, statisticienne au Département santé travail ;
- Mme Laetitia BENEZET, statisticienne au Département santé travail ;
- Mme Juliette CHATELOT, épidémiologiste au Département santé travail.

## 5 Durée et modalité d'organisation de l'étude

### 5.1 Groupe de travail

Un groupe de travail constitué de statisticiens du programme Coset, de la cohorte Constances (Inserm) [13] et de la cohorte Elfe [14]<sup>1</sup> a été mis en place afin de se pencher sur les questions statistiques et méthodologiques relatives à la non-réponse dans les enquêtes longitudinales. Ce groupe de travail se réunira tout au long de l'enquête.

### 5.2 Gestion de l'étude

L'enquête complémentaire auprès de non-participants est supervisée par le Département santé travail, l'organisation et la réalisation du recueil des données étant confiées à un prestataire externe à l'InVS, sélectionné par appel d'offre avec cahier des charges conforme au présent protocole, dans le cadre de la loi sur les marchés publics. Le DST est en charge de la rédaction du protocole, du tirage au sort, de la préparation des questionnaires, de la soumission du dossier au CCTIRS et à la Cnil, et de la rédaction du cahier des charges pour l'appel d'offre, en collaboration avec les services concernés de l'InVS. Le DST collaborera avec le prestataire pour la préparation des documents et des procédures utiles à la gestion du recueil des données. En cours de recueil des données, des points réguliers seront organisés entre l'InVS et le prestataire afin de disposer des éléments qualitatifs et quantitatifs nécessaires à l'appréciation de son bon déroulement. Le prestataire fournira les bases de données consolidées, avec leurs dictionnaires de données. Le DST assurera les analyses statistiques et la rédaction du rapport final.

### 5.3 Calendrier

Février 2010 :	Finalisation du protocole, envoi du dossier au CCTIRS
Mars 2010 :	Envoi du dossier à la Cnil
Début mai 2010 :	Date prévisionnelle de choix d'un prestataire
Mai à août 2010 :	Période de préparation au terrain d'enquête
Septembre à octobre 2010 :	Période de recueil des données
Décembre 2010 :	Envoi par le prestataire à l'InVS des bases de données
Février à mars 2011 :	Validation des données transmises
Mars à juin 2011 :	Analyses statistiques
Novembre 2011 :	Rapport d'étude

---

<sup>1</sup> Groupe de travail statistique sur les enquêtes longitudinales : Gaëlle SANTIN, Laetitia BENEZET (DST-InVS), Alice GUEGUEN, Rémi SITTA (U1018-Inserm), Hélène SARTER (DSE-InVS), Nicolas RAZAFINDRATSIMA (INED).

## 6 Bibliographie

- [1] Équipe Coset. Programme Coset - Cohortes pour la surveillance épidémiologique en lien avec le travail : protocole d'inclusion d'actifs relevant de la Mutualité Sociale Agricole. 2009.
- [2] Équipe Coset. Protocole d'étude des données issues des systèmes d'information existants dans le cadre de la phase pilote du projet Coset-MSA. 2010.
- [3] Särndal C, Swensson B, Wretman J. Model assisted survey sampling. New York: Springer-Verlag, 1992.
- [4] Dimitropoulos L, Campbell L, Iannacchione V. The effect of method of delivery on response to a non-response follow-up. American association for public opinion research 2009;74-7.
- [5] Jenkins P, Scheim C, Wang JT, Reed R, Green A. Assessment of coverage rates and bias using double sampling methodology. J Clin Epidemiol 2004 Feb;57(2):123-30.
- [6] Smith H. Un échantillon double pour minimiser le biais dû à la non-réponse dans un sondage par courrier. In: Guilbert P, Haziza D, Ruiz-Gazen A, Tillé Y. Méthodes de sondage : applications aux enquêtes longitudinales, à la santé, aux enquêtes électorales et aux enquêtes dans les pays en développement. Paris: Dunod, 2008. p. 334-9.
- [7] Zell E, Shepard C, Rosenstein N. Follow-up of persons impacted by terrorism: respondents and nonrespondents. 2003 p. 4717-21.
- [8] Goldberg M, Chastang JF, Leclerc A, Zins M, Bonenfant S, Bugel I, et al. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. Am J Epidemiol 2001 Aug 15;154(4):373-84.
- [9] Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-91.
- [10] Stang A. Nonresponse research--an underdeveloped field in epidemiology. Eur J Epidemiol 2003;18(10):929-31.
- [11] Hansen M, Hurwitz W. The problem of nonresponse in sample surveys. JASA 1946;41:517-29.
- [12] Stang A, Jockel KH. Studies with low response proportions may be less biased than studies with high response proportions. Am J Epidemiol 2004 Jan 15;159(2):204-10.
- [13] Équipe Risques Post Professionnels - Cohortes / Unité mixte 687 Inserm-CNAMTS. La cohorte Constances - Laboratoire épidémiologique ouvert. Protocole scientifique, phase pilote. 2009.
- [14] Vandentorren S, L'Azou M, Sarter H, Pirus C, Bois C, Salines G, et al. Cohorte Elfe (Etude longitudinale depuis l'enfance) : résultats du premier pilote. Journées de veille sanitaire. 2007.
- [15] Accardo J, Bascheri E, Jaluzot J. Enquêtes téléphoniques auprès des ménages : peut-on utiliser l'annuaire ? Actes des journées de méthodologie statistique 2009.
- [16] Bigot R, Croutte P. Enquête "Conditions de vie et aspirations des français : la diffusion des technologies de l'information et de la communication dans la société française". Credoc, 2008.
- [17] Beck F, Legleye P, Peretti-Watel P. Aux abonnés absents : liste rouge et téléphone portable dans les enquêtes en population générale sur les drogues. Bulletin de méthodologie sociologique 2005;5-29.

- [18] Roy G, Vanheuverzwyn A. Le téléphone mobile dans les enquêtes par sondage. Actes des journées de méthodologie statistique 2000.
- [19] Arcep. Tableau de bord au 23 juin 2009. Internet 2009 Available from: URL: <http://www.arcep.fr/index.php?id=8742>
- [20] Haziza D. Treatment of unit and item nonresponse in surveys. 22-23 juin 2009. Université de Neuchâtel.

## Annexe 1 : Notification d'autorisation de la Cnil n°909091



Le Président

Madame Françoise WEBER  
Directrice  
INSTITUT DE VEILLE SANITAIRE  
INVS  
12 RUE DU VAL D'OSNE  
94415 - SAINT MAURICE CEDEX

Paris, le 9 octobre 2009

*A l'attention de Madame Nathalie Ferragne*

N/Réf. : AT/YPA/SV/SN/GDP/FLR/AR091223

Instruction du Dossier :  
Frédérique LESAULNIER

Objet : NOTIFICATION D'AUTORISATION

**DEMANDE D'AUTORISATION N° 909091**

**A rappeler dans toute correspondance**

Madame la directrice,

Vous avez saisi, le 4 mars 2009, notre Commission d'une demande d'autorisation relative à un traitement de données à caractère personnel ayant pour finalité

**LA MISE EN OEUVRE D'UN PILOTE DE LA COHORTE COSET, COHORTE POUR LA SURVEILLANCE EPIDEMIOLOGIQUE EN MILIEU DE TRAVAIL, PORTANT SUR L'INCLUSION D'ACTIFS RELEVANT DE LA MUTUALITE SOCIALE AGRICOLE**

La cohorte Coset a pour objet de suivre de façon prospective un échantillon représentatif d'actifs de 18 à 65 ans en France, relevant des trois principaux régimes de sécurité sociale, à savoir le régime général, le régime agricole et le régime social des indépendants. Cette cohorte, qui devrait inclure 200 000 sujets est destinée à permettre la surveillance épidémiologique de l'état de santé de la population en relation avec les risques professionnels.

La constitution de la cohorte sera progressive et impliquera chaque régime d'assurance maladie précité. Le suivi ultérieur sera à la fois actif (auto questionnaire lors de l'inclusion, puis adressé à domicile chaque année) et passif (extraction de données de nature médicale et professionnelle dans les bases de données nationales de l'assurance maladie (le SNIIRAM en particulier) et de la caisse nationale d'assurance vieillesse).

Afin de tester les modalités de recrutement des actifs affiliés au régime agricole et les modalités de recueil des données nécessaires au suivi, vous avez souhaité procéder à une phase pilote.

**Commission Nationale de l'Informatique et des Libertés**  
8 rue Vivienne CS 30223 75083 PARIS Cedex 02 - Tél: 01 53 73 22 22 - Fax: 01 53 73 22 00 - www.cnil.fr  
RÉPUBLIQUE FRANÇAISE

Cette phase pilote se déroulera à l'automne 2009, concernera 5 caisses de MSA et 10 000 personnes qui seront recrutées par tirage au sort, puis envoi postal d'un auto-questionnaire d'inclusion. Une cohorte de non-participants sera constituée à partir de personnes n'ayant pas exprimé un refus de participer à la cohorte. Seul le recueil des informations nécessaires aux suivis actifs et passifs sera testé.

La Commission prend acte que :

- Les personnes non volontaires pour participer activement à la cohorte seront clairement informées que, sauf demande contraire de leur part, elles seront susceptibles d'être tirées au sort pour être suivies de façon anonyme au travers des fichiers de l'Assurance maladie et de la CNAV afin de constituer un échantillon de comparaison.
- Ces personnes, dûment informées, auront la possibilité d'exprimer leur refus de faire l'objet de ce suivi passif auprès du Département santé travail de l'Institut de Veille Sanitaire par téléphone, par l'envoi d'un courrier électronique ou l'envoi postal d'un formulaire de refus.
- Le numéro d'inscription au répertoire (NIR) restera cantonné aux organismes de sécurité sociale déjà autorisés à utiliser ce numéro ;
- Afin de garantir la séparation physique des données nominatives nécessaires à l'envoi des questionnaires et des données relatives à la santé et à la vie professionnelle, les coordonnées postales (nom, prénom, adresse postale) seront conservées par un tiers de confiance (le Centre de ressources informatique (CRI) de l'institut fédératif de recherche n°69 de l'Inserm) dans des conditions de nature à garantir leur confidentialité. Ce tiers de confiance, qui ne détiendra pas le NIR, sera également seul à conserver le table de correspondance entre les deux numéros d'anonymat, propres respectivement à la MSA (NC-MSA) et à l'InVS (N°COSET) attribués à l'inclusion des participants.

J'ajoute que pour la mise en œuvre des phases suivantes, une nouvelle demande d'autorisation devra être adressée aux services de la Commission.

Ce traitement relève de la procédure des articles 54 et suivants de la loi du 6 janvier 1978 modifiée.

En conséquence, conformément aux dispositions de l'article 15 de la loi du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, j'autorise la mise en œuvre de ce traitement.

Je vous prie, Madame la Directrice, d'agréer l'expression de mes salutations distinguées.



Alex TÜRK









- votre niveau de qualification

- 1 Manœuvre ou ouvrier spécialisé
- 2 Ouvrier qualifié ou ouvrier hautement qualifié ou technicien d'atelier
- 3 Agent de maîtrise
- 4 Directeur général ou adjoint direct au directeur
- 5 Technicien, dessinateur, VRP
- 6 Instituteur, assistante sociale, infirmière et autres personnels de catégorie B de la fonction publique
- 7 Ingénieur ou cadre
- 8 Professeur et personnel de catégorie A de la fonction publique
- 9 Employés de bureau, de commerce, agents de service, aides soignantes, gardiennes d'enfants, personnels de catégorie C ou D de la fonction publique
- 10 Autres
- 11 Ne sait pas

## EXPOSITIONS PROFESSIONNELLES

**P15** Êtes-vous en contact physique ou téléphonique avec le public (usagers, patients, clients, voyageurs..) ?

- 0 Non                      1 Oui

**Si OUI** vivez-vous des situations de tension dans vos rapports avec le public ?

- 1 Non ou presque jamais  
2 Rarement  
3 Souvent  
4 Toujours ou presque

**Pour les propositions suivantes, cochez la case correspondant le mieux à votre situation professionnelle actuelle (une case et une seule par question) :**

**S1** Je suis constamment pressé(e) par le temps à cause d'une forte charge de travail

- 1 Pas d'accord            2 D'accord mais ça ne me perturbe pas du tout            3 D'accord et ça me perturbe un peu            4 D'accord et ça me perturbe            5 D'accord et ça me perturbe beaucoup

**S10** Ma sécurité d'emploi est menacée

- 1 Pas d'accord            2 D'accord mais ça ne me perturbe pas du tout            3 D'accord et ça me perturbe un peu            4 D'accord et ça me perturbe            5 D'accord et ça me perturbe beaucoup

**S11** Vu tous mes efforts, je reçois le respect et l'estime que je mérite à mon travail

- 1 D'accord            2 Pas d'accord mais ça ne me perturbe pas du tout            3 Pas d'accord et ça me perturbe un peu            4 Pas d'accord et ça me perturbe            5 Pas d'accord et ça me perturbe beaucoup

Les questions suivantes se rapportent à une **journée de travail typique au cours des 12 derniers mois.**

**J1** Comment évaluez-vous l'intensité des efforts physiques de votre travail au cours d'une journée de travail typique, sur une échelle allant de 6 à 20 ? .....

Cochez la modalité correspondant à votre choix sur l'échelle de 6 à 20 ci-dessous, qui va de « pas d'effort du tout » à « épuisant ».

6	<input type="checkbox"/>	Pas d'effort du tout
7	<input type="checkbox"/>	Extrêmement léger
8	<input type="checkbox"/>	
9	<input type="checkbox"/>	Très léger
10	<input type="checkbox"/>	
11	<input type="checkbox"/>	Léger
12	<input type="checkbox"/>	
13	<input type="checkbox"/>	Un peu dur
14	<input type="checkbox"/>	
15	<input type="checkbox"/>	Dur
16	<input type="checkbox"/>	
17	<input type="checkbox"/>	Très dur
18	<input type="checkbox"/>	
19	<input type="checkbox"/>	Extrêmement dur
20	<input type="checkbox"/>	Épuisant

**J2** Votre travail nécessite-t-il habituellement de répéter les mêmes actions plus de 2 à 4 fois environ par minute ?

- 1 Non ou presque jamais
- 2 Rarement (moins 2 heures par jour)
- 3 Souvent (2 à 4 heures par jour)
- 4 Toujours ou presque

**J8** Au cours d'une journée typique de travail, devez-vous conduire un engin de chantier, un tracteur, un chariot automoteur ou autre machine mobile sur votre lieu de travail ?

- 1 Non ou presque jamais
- 2 Rarement (moins 2 heures par jour)
- 3 Souvent (2 à 4 heures par jour)
- 4 Toujours ou presque

**J11** Au cours d'une journée typique de travail, utilisez-vous des outils vibrants ou devez-vous poser la(les) main(s) sur des machines vibrantes ?

- 1 Non ou presque jamais
- 2 Rarement (moins 2 heures par jour)
- 3 Souvent (2 à 4 heures par jour)
- 4 Toujours ou presque

**E2** EXPOSITIONS AUX BRUITS

12) Etes-vous exposé(e) à des bruits intenses (tels qu'ils vous empêchent d'entendre la voix d'une personne placée à deux ou trois mètres, même si elle crie) ?

- 0 Non
- 1 Oui



## Annexe 3 : Courrier d'annonce

**LOGO InVS**  
**DEPARTEMENT SANTE-TRAVAIL (DST) - InVS**

**Logo MSA**  
**MUTUALITE SOCIALE AGRICOLE - MSA**

Madame, Monsieur,

Vous avez reçu en début d'année un questionnaire pour vous proposer de participer à l'enquête Coset. Nous vous rappelons que cette enquête est mise en place par l'Institut de veille sanitaire (InVS), en partenariat avec la Mutualité sociale agricole (MSA). Il s'agit d'un grand programme national de surveillance de la santé en relation avec le travail qui a pour principal objectif de décrire les liens entre les facteurs professionnels et la survenue de problèmes de santé (allergies, troubles musculaires et articulaires, problèmes cardio-vasculaires et respiratoires, cancers, problèmes de santé mentale).

Depuis février 2010, plus de 2000 personnes dans 5 départements français, ont accepté de répondre au questionnaire et de participer au suivi. Sauf erreur de notre part, vous n'avez pas souhaité répondre au questionnaire postal.

Afin que cette étude prenne en compte toutes les situations, il est essentiel de recueillir un minimum d'information parmi les personnes n'ayant pas participé à l'enquête initiale. C'est pourquoi nous nous permettons de vous contacter une dernière fois pour un bref interview d'une dizaine de minutes. **Un de nos enquêteurs vous contactera prochainement par téléphone ou se rendra à votre domicile pour vous interviewer.**

**Vous êtes libre de vous opposer à ce que l'un de nos enquêteurs vous contacte. Pour cela vous pouvez appeler le 0 800 200 130 de 10h à 18h30 du lundi au vendredi (appel gratuit depuis un poste fixe).**

**Toutes les informations recueillies dans le cadre de cette étude seront traitées de façon confidentielle afin d'être analysées par l'InVS dans le strict respect de l'anonymat des participants.**

Pour cela, les informations concernant votre identité seront conservées séparément selon une procédure contrôlée et agréée par la Commission nationale informatique et libertés (Cnil). Cette procédure garantit que les informations recueillies dans cette étude ne pourront à aucun moment être reliées directement à votre identité, que ce soit par l'InVS ou la MSA, et qu'aucun organisme non autorisé n'ait accès à ces informations.

**Nous vous informons que votre participation n'a aucun caractère obligatoire** et ce, conformément à la loi 78-17 du 6 janvier 1978 relative à l'informatique et aux libertés.

**Vous pouvez à tout moment demander l'accès et la rectification des informations vous concernant auprès de l'InVS, à l'adresse indiquée ci-dessous.**

Nous vous remercions par avance pour votre participation à cette étude qui permettra de mieux connaître les liens entre facteurs professionnels et survenue de problèmes de santé et vous prions d'agréer, Madame, Monsieur, l'expression de notre considération distinguée.

Dr Ellen Imbernon  
Directrice du département santé travail – InVS

XXX  
Directeur de la caisse MSA du XXXX

Pour plus d'informations sur cette étude, nous vous invitons à consulter le site Internet de l'InVS ([www.invs.sante.fr/surveillance/coset](http://www.invs.sante.fr/surveillance/coset)) ou de la MSA ([www.msa.fr](http://www.msa.fr)) ou à téléphoner au 0 800 200 130 de 10h à 18h30 du lundi au vendredi (appel gratuit depuis un poste fixe).

Pour nous écrire : Programme Coset – Département santé travail - Institut de veille sanitaire – 12 rue du Val d'Osne 94415 Saint-Maurice cedex

## Annexe 4 : Populations non joignables par téléphone

Les enquêtes téléphoniques permettent généralement d'obtenir de meilleurs taux de réponse que les enquêtes postales, avec des coups de collecte inférieurs à ceux des enquêtes en face-à-face. Cependant, elles sont confrontées à de nouvelles difficultés puisque une part croissante de la population ne figure pas dans l'annuaire téléphonique pour différentes raisons [15;16] :

- Les personnes ne possédant aucun téléphone, environ 1% des français en 2008 ;
- Les personnes détentrices exclusives de mobile, environ 14% des français en 2008 ;
- Les personnes qui ont abandonné leur abonnement téléphonique chez l'opérateur historique (France Télécom) au profit de concurrents (dégrouperage total). Ces derniers ne transmettent pas systématiquement leurs listes aux annuairistes. En 2008, 29% des personnes bénéficiant d'une ligne fixe ont totalement supprimé leur abonnement à l'opérateur historique ;
- Les personnes inscrites sur liste rouge et orange. Ces listes ne sont pas transmises aux annuairistes. On estime qu'environ 12 % des ménages sont inscrits sur liste rouge.

Plusieurs enquêtes ont comparé les individus détenteurs exclusifs d'un téléphone portable avec ceux équipés d'une ligne fixe. Ce sont plus fréquemment des hommes, des jeunes, des personnes vivant seules et selon les études légèrement plus diplômées ou avec des revenus plus faible [16-18].

Les individus issus des ménages inscrits sur liste rouge sont plus fréquemment des femmes, des jeunes, des personnes vivant plus souvent dans de grandes agglomérations, plus diplômées et sur-représentées parmi les hauts revenus [17].

La création d'un annuaire universel devrait atténuer ce défaut de couverture mais sa mise en place reste limitée puisque certains opérateurs participent très faiblement à cet annuaire. D'après l'ARCEP [19] qui gère cet annuaire, en mai 2009, le taux d'abonnés inscrits sur les listes est de 2,8% chez les opérateurs de téléphonie mobile, et de près de 80 % chez les opérateurs de téléphonie fixe.

## Annexe 5 : Annexe méthodologique

### 1/ Enquête complémentaire auprès de non-participants

#### *Principe général*

Hansen et Hurwitz [11] ont été les premiers à réaliser une enquête complémentaire auprès de non-répondants afin d'obtenir des estimations sans biais de moyennes ou de totaux. Ils se sont pour cela replacés dans le contexte d'un échantillonnage double (ou sondage en deux phases).

Le contexte est le suivant. On réalise une enquête auprès d'un échantillon de personnes. Dans la plupart des enquêtes, à l'issue du recueil de données, on observe des répondants et des non-répondants ; or, la présence de non-répondants peut entraîner des estimations biaisées. Hansen et Hurwitz ont proposé de sélectionner par tirage au sort un échantillon de non-répondants que l'on enquête à nouveau. On combine ensuite les données de l'échantillon de non-répondants avec celles de l'échantillon des répondants initiaux pour obtenir des estimations sans biais. La condition essentielle à la validité de cette méthode repose sur la forte participation des non-répondants échantillonnés (aux alentours de 100%).

D'un point de vue plus formel, on peut exprimer le problème comme il suit [3].

On sélectionne par tirage au sort un échantillon ( $s_a$ ) de 1<sup>ère</sup> phase avec comme probabilité d'inclusion  $\pi_{alk}$ , avec  $l$  la strate d'appartenance de la personne  $k$  ; pour la phase pilote du volet «Coset-MSA»,  $l$  correspond à la combinaison des variables suivantes : caisse départementale MSA, âge, sexe, statut d'emploi (non salarié agricole ou salarié agricole). La taille de l'échantillon ( $s_a$ ) est  $n_a$ . Pour la phase pilote du volet «Coset-MSA»,  $n_a$  sera égal à 10 000.

A l'issue du recueil de données, l'échantillon ( $s_a$ ) est constitué d'un échantillon de participants ( $s_{a1}$ ) et d'un échantillon de non-participants ( $s_{a2}$ ). Autrement dit, à l'issue du recueil de données, l'échantillon ( $s_a$ ) peut être divisé en deux strates : une strate de participants et une strate de non-participants.

On considère alors que l'échantillon ( $s_a$ ) sélectionné en 1<sup>ère</sup> phase correspond à une population et que l'on possède comme information auxiliaire pour chaque unité de la population l'information « participant/non participant ».

On sélectionne un échantillon de 2<sup>ème</sup> phase avec les probabilités d'inclusion suivantes :

- si l'individu  $k$  appartient à l'échantillon de participants ( $s_{a1}$ ), il aura une probabilité d'inclusion égale à 1 soit  $\pi_{1k/s_{a1}}=1$ . L'échantillon obtenu ( $s_1$ ) correspond exactement à la strate ( $s_{a1}$ ) et sa taille correspond au nombre de primo-participants.
- si l'individu  $k$  appartient à l'échantillon de non-participants ( $s_{a2}$ ), il aura une probabilité d'inclusion égale à  $\pi_{2l'k/s_{a2}}$  avec  $l'$  la strate d'appartenance (ici la strate sera la caisse départementale) de l'individu  $k$ . On obtient alors un échantillon ( $s_2$ ) de taille  $n_2$  égale à 500 (100 personnes par caisse). L'enquête complémentaire sera donc réalisée auprès de l'échantillon ( $s_2$ ).

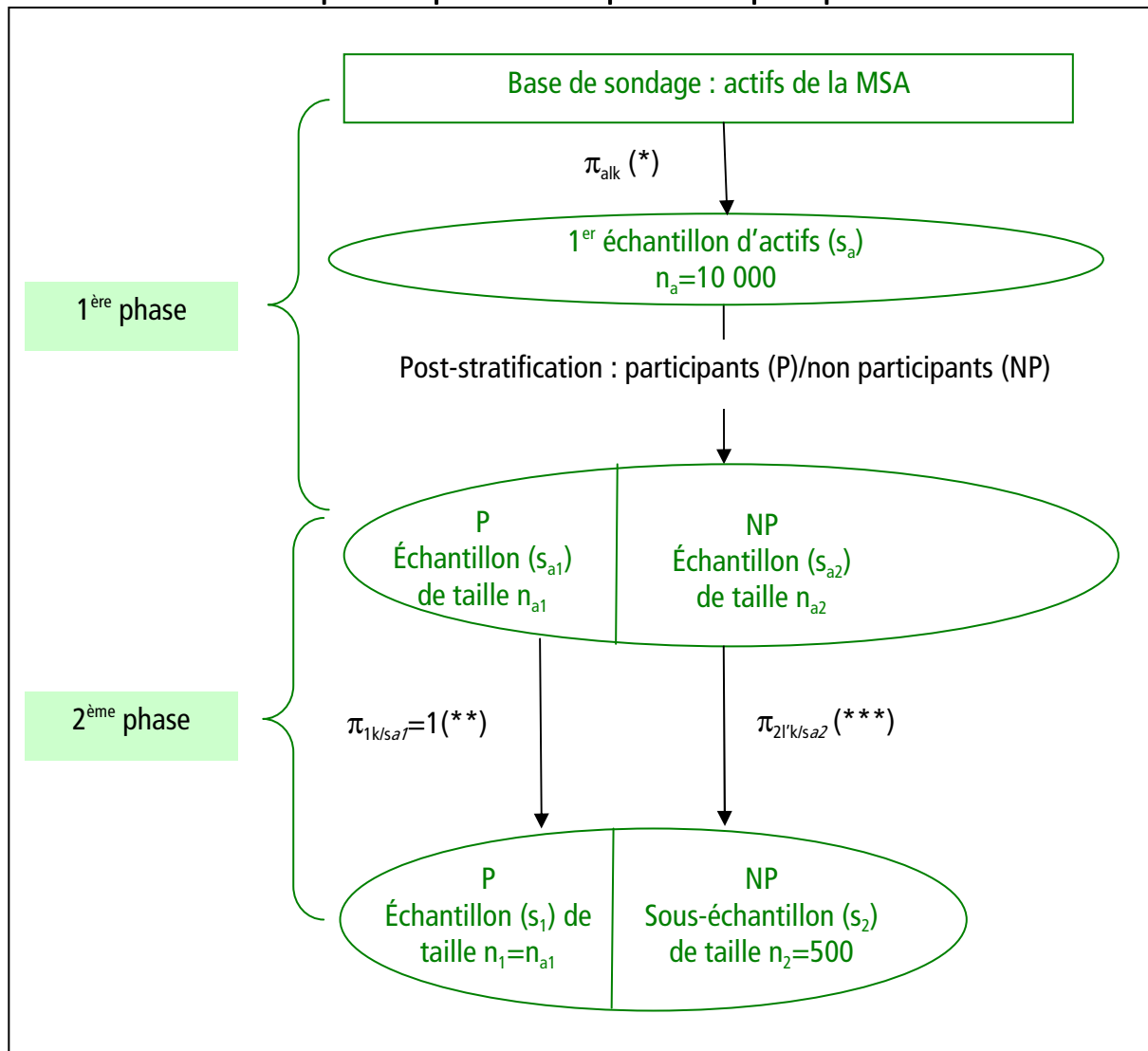
Il existe néanmoins une différence avec les sondages en deux phases classiques ; contrairement aux sondages en 2 phases où la strate obtenue à l'issue de la 1<sup>ère</sup> phase est fixe, la strate « participant/non-participant » est aléatoire (si on enquêtait plusieurs fois l'échantillon de phase 1, la distribution de la strate « participant/non-participant » ne serait pas forcément la même, le statut de « participant » ou « non-participant » d'une personne pouvant varier).

A l'issue des deux phases, la probabilité d'inclusion  $\pi_{k(ED)}$  d'un individu k sera égale à :

- $\pi_{k(ED)} = \pi_{alk}$ , avec l strate *a priori* (ici la combinaison des variables caisse départementale, âge, sexe, statut d'emploi) si k appartient à l'échantillon ( $s_1$ ) des participants de la 1<sup>ère</sup> phase ;
- $\pi_{k(ED)} = \pi_{alk} \times \pi_{2l'k/s_2}$  avec l' strate *a priori* (ici la caisse départementale) si k appartient au sous-échantillon ( $s_2$ ) des non-participants de la 1<sup>ère</sup> phase.

Les poids d'échantillonnage seront donc égaux à  $w_{k(ED)} = 1/\pi_{k(ED)}$ .

**Figure X : Représentation graphique de l'échantillonnage en deux-phases dans le cadre de l'enquête complémentaire auprès de non-participants**



(\*)  $\pi_{alk}$  : probabilité d'inclusion en phase 1 d'une personne k de la strate l

(\*\*)  $\pi_{1k/s_{a1}}$  : probabilité d'inclusion en phase 2 d'une personne k sachant qu'elle appartient à la post-strate des participants

(\*\*\*)  $\pi_{2l'k/s_{a2}}$  : probabilité d'inclusion en phase 2 d'une personne k de la strate l' sachant qu'elle appartient à la post-strate des non-participants

La condition nécessaire pour obtenir des estimations sans biais est d'avoir un taux de réponse proche des 100% pour l'échantillon ( $s_2$ ). C'est pourquoi le protocole opérationnel de l'enquête complémentaire auprès des non-participants prévoit des modes de collecte maximisant autant que possible le taux de réponse (questionnaire court, enquête par téléphone ou en face-à-face).

## Échantillonnage

Afin de conserver la stratification sur la caisse et d'assurer un effectif minimum dans chaque caisse 100 non-participants par caisse seront tirés au sort selon l'algorithme de sondage aléatoire simple.

### Estimations calculées

L'enquête complémentaire permettra de calculer des estimations de moyennes, de pourcentages ou de somme. Considérons le cas où l'on cherche à estimer la moyenne  $\hat{Y}^{ED}$  avec  $\hat{Y}^{ED} = \sum_{k \in (s_{a1} \cup s_2)} \pi_{k(ED)} \sum_{k \in (s_{a1} \cup s_2)} \frac{1}{\pi_{k(ED)}} Y_k$ ,  $Y_k$  valeur observée de  $Y$  chez l'individu  $k$  et  $\pi_{k(ED)}$  probabilité d'inclusion de l'individu  $k$ .

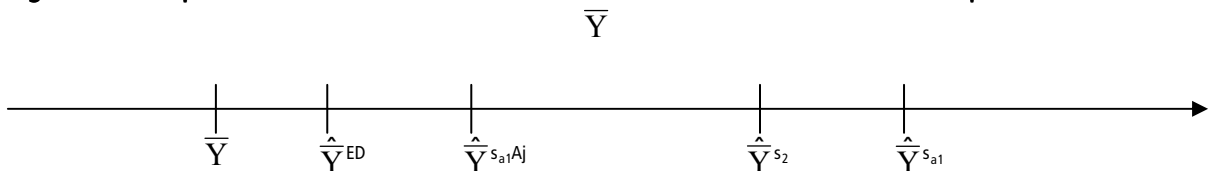
Il faudra par ailleurs calculer la variance de l'estimation  $\hat{Y}^{ED}$ . Cette question sera traitée lorsque les données seront disponibles. Plusieurs pistes seront alors étudiées [20] (l'approche renversée ou le bootstrap avec ou sans remise).

## 2/ Comparaison des estimations

Supposons que l'on cherche à estimer une moyenne  $\bar{Y}$  par  $\hat{Y}$ . Plusieurs estimations pourront être calculées et comparées :

- Soit l'estimation ( $\hat{Y}^{s_{a1}}$ ) obtenue à partir de l'échantillon des participants ( $s_{a1}$ ) et des poids d'échantillonnage  $w_k$  ;
- Soit l'estimation ( $\hat{Y}^{s_2}$ ) obtenue à partir de l'échantillon ( $s_2$ ) et des poids d'échantillonnage  $w_{k(ED)}$ , qui correspond à l'enquête complémentaire auprès des non-participants ;
- Soit l'estimation ( $\hat{Y}^{ED}$ ) obtenue à partir de l'échantillon double ( $s_1 \cup s_2$ ). Si le taux de réponse à l'enquête complémentaire auprès des non-participants est proche de 100%, l'échantillonnage double permettra d'obtenir une estimation  $\hat{Y}^{ED}$  approximativement sans biais, c'est-à-dire que ( $\hat{Y}^{ED} - \bar{Y} \rightarrow 0$ ) ;
- Soit l'estimation ( $\hat{Y}^{s_{a1}Aj}$ ) obtenue à partir de l'échantillon initial de participants ( $s_{a1}$ ) et des poids corrigés pour la non-participation  $w_{k,inclusion_{aj}}$  ; la correction de la non-participation a été réalisée en exploitant les informations auxiliaires contenues dans les bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav [2].

Figure X : Exemple de schématisation des estimations obtenues et de leur distance au paramètre d'intérêt



NB : l'ordre des estimations est fictif.  $\hat{Y}^{s_{a1}Aj}$  aurait pu être représenté plus proche de  $\bar{Y}$  que ne l'ait  $\hat{Y}^{ED}$

Sous l'hypothèse d'absence d'autres sources de biais (par exemple de biais lié aux erreurs de mesure), la comparaison des estimations ci-dessus nous permettra de mieux comprendre les biais liés à la non-participation dans Coset-MSA et de discuter de la qualité des informations auxiliaires

contenues dans les bases de données informationnelles de la MSA, du Sniir-AM et de la Cnav pour corriger ces biais du mieux que l'on peut. On pourra dans un premier temps considérer que deux estimations ne sont pas trop différentes si leurs intervalles de confiance respectifs se chevauchent ; d'autres critères de comparaison plus élaborés pourront également être choisis ultérieurement.

On pourra tout d'abord comparer les estimations  $(\hat{Y}^{s_{a1}})$  et  $(\hat{Y}^{s_2})$  pour avoir une première idée de l'ampleur du biais lié à la non-participation. En effet, si les estimations issues de l'échantillon des participants ( $s_{a1}$ ) se révèlent différentes des estimations issues de l'échantillon ( $s_2$ ) issu de l'enquête complémentaire, on pourra supposer que le recours à des techniques de repondération pour prendre en compte la non-participation peut être envisagé.

On pourra également comparer les estimations  $(\hat{Y}^{s_{a1}})$  et  $(\hat{Y}^{s_{a1Aj}})$ ,  $(\hat{Y}^{s_{a1}})$  et  $(\hat{Y}^{ED})$ ,  $(\hat{Y}^{s_{a1Aj}})$  et  $(\hat{Y}^{ED})$  etc.

Le protocole de l'enquête complémentaire ayant été construit pour obtenir un taux de réponse maximal, on espère atteindre un taux de réponse proche des 70% à l'enquête complémentaire. On pourra alors également corriger pour la non-participation les estimations  $\hat{Y}^{ED}$  à partir des informations auxiliaires contenues dans les bases pour obtenir une estimation  $\hat{Y}^{EDAj}$  et la comparer à  $\hat{Y}^{s_{a1Aj}}$ .

Si  $\hat{Y}^{EDAj}$  et  $\hat{Y}^{s_{a1Aj}}$  sont proches, on pourra considérer que le recours aux informations auxiliaires permet d'obtenir des estimations aussi biaisées que si on avait cherché à maximiser le taux de participation lors de l'enquête initiale. Si elles diffèrent, plusieurs hypothèses devront être discutées : information auxiliaire pour corriger la non-participation insuffisante, biais de mesure dans l'enquête complémentaire (modes de recueil des données multiples, réponses moins précises des personnes enquêtées lors de l'enquête complémentaire), etc.