

Document d'études

Direction de l'animation de la recherche, des études et des statistiques

Numéro **242**

Octobre 2020


RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

DARES
direction de l'animation de la recherche,
des études et des statistiques

L'enquête Conditions de Travail- Risques Psychosociaux 2016 : apurement et redressement

Amélie MAUROUX,
Selma AMIRA,
Corinne METTE
Dares
Claire BESWICK
DGAFP
Céline DENNEVAULT
Drees

**L'enquête
Conditions de Travail-
Risques Psychosociaux 2016 :
apurement et redressement**

**Volume 1 – le volet « individus »
Volume 2 – le volet « entreprises »**

Résumé

Ce document décrit les méthodes retenues pour le calcul des pondérations du volet « individus » et du volet « employeurs » de l'enquête Conditions de travail et Risques psychosociaux 2016.

Les redressements du volet « individus » présentent deux défis méthodologiques. Pour la première fois, le volet « individus » comporte une dimension panel, ainsi que recommandé par le Collège d'expertise sur le suivi statistique sur les risques psychosociaux au travail. De plus, comme dans l'enquête Conditions de travail 2013, des échantillons supplémentaires de salariés de la fonction publique (trois versants) et du secteur hospitalier privé, ont été financés par la Drees et la DGAFP. Le plan de sondages de la partie « en coupe » est donc particulièrement complexe et un recours intensif à la méthode généralisée du partage des poids a été nécessaire. Le volume 1 consacré au volet « individus » présente les différentes étapes de correction de la non-réponse des différents échantillons, le partage des poids et les calages permettant d'obtenir les jeux de pondérations pour le fichier panel et les jeux de poids pour le fichier en coupe, représentatifs des actifs occupés en 2016.

Comme dans l'enquête Conditions de travail 2013, un volet « employeurs » a complété l'interrogation auprès des personnes en emploi. Les établissements interrogés sont, pour partie, tirés aléatoirement dans la base Sirène et, pour partie, parmi les employeurs des salariés ayant répondu au volet « individus ». Cet échantillonnage indirect rend nécessaire le recours au partage des poids. Le volume 2 consacré au volet « employeurs » dresse le bilan de la collecte et décrit les différentes étapes d'apurement et de redressement (correction de la non-réponse des différents échantillons, partage des poids et calages sur marges). Deux jeux de pondérations sont obtenus, correspondant aux deux types d'exploitation possibles du volet « employeurs » : une pour le fichier couplé salariés-employeurs et une pour le fichier de l'ensemble des employeurs répondants, représentatifs des établissements actifs en 2016.

Mots-clefs : correction de la non- réponse ; partage des poids ; panel ; calage sur marges ; fichiers couplés ; risques psychosociaux ; enquête entreprise ; enquête ménage.

**L'enquête
Conditions de Travail-
Risques Psychosociaux 2016 :
apurement et redressement
Volume 1 – le volet « individus »**

Redressement du volet individu – CT-RPS 2016¹

Ce document décrit la méthode retenue pour le calcul des pondérations en coupe et panel du volet individu de l'enquête conditions de travail et risques psychosociaux 2016 (CT-RPS 2016). Il a également une vocation pratique, en apportant aux utilisateurs de l'enquête les résultats détaillés des différentes étapes de ces redressements.

Depuis 1978 les enquêtes conditions de travail visent à obtenir une description concrète du travail et de son organisation et ce pour l'ensemble des actifs occupés. Seule enquête de cette ampleur sur les conditions de travail, elles permettent également des analyses fines par catégorie socioprofessionnelle et secteur d'activité. Leur reconduction tous les sept ans environ de 1978 à 2013, et tous les trois ans depuis, permet par ailleurs de disposer de la profondeur temporelle nécessaire pour apprécier les évolutions récentes.

L'édition 2016 du dispositif d'enquête Conditions de travail et risques psychosociaux et en œuvre les recommandations du Collège d'expertise sur le suivi statistique sur les risques psychosociaux au travail (2011), réuni en 2009-2010 à la demande du ministre chargé du travail. L'enquête Conditions de Travail 2013 (CT 2013) avait constitué une première étape en ce sens. La problématique des risques psychosociaux avait en effet été approfondie au travers d'un ensemble de questions, notamment dans une nouvelle partie auto-administrée du questionnaire sur les relations avec les autres, les satisfactions et difficultés au travail et dans la vie quotidienne. Autre innovation de l'édition 2013, un volet « employeurs » a été mis en place où les responsables d'établissements sont interrogés sur les politiques managériales en matière de conditions, d'organisation du travail et de prévention des risques professionnels.

En 2016, pour compléter la couverture des six axes identifiés par le collège Gollac, le questionnement a été renforcé sur les facteurs psychosociaux de risques professionnels avec de nouvelles questions sur la conciliation vie professionnelle/vie privée, les conflits de valeurs, la qualité empêchée, et sur la santé et son évolution (douleurs articulaires et musculaires, Outil de diagnostic dépression (MINI), risque suicidaire). Comme pour CT 2013, des échantillons supplémentaires de salariés de la fonction publique (trois versants) et du secteur hospitalier privé, ont été financés par la Drees et la DGAFP et un volet auprès des employeurs a complété l'interrogation auprès des personnes en emploi.

Conformément aux recommandations du Collège d'expertise un suivi panel a également été mis en place. Si l'enquête vise en premier lieu à interroger les actifs occupés, elle réinterroge tous les répondants de l'enquête CT 2013, y compris les personnes sorties de l'emploi depuis cette interrogation (les personnes au chômage ou inactives, retraitées, en arrêt maladie de plus d'un an). Le panel constitué dès CT 2013 permet de mesurer l'évolution des conditions de travail et d'approfondir l'analyse des risques psychosociaux au travail.

L'édition CT-RPS 2016 s'insère ainsi dans la continuité des enquêtes Conditions de Travail menée par la DARES depuis 1978 : tous les trois ans, en alternance, a lieu une enquête CT dédiée aux risques psychosociaux (CT-RPS) avec un nombre plus important de questions sur les risques psychosociaux et une enquête CT décrivant plus finement les conditions et les moyens de travail.

Les principaux thèmes abordés dans le volet « Individu » de l'enquête sont :

- les horaires et l'organisation du temps de travail ;
- l'organisation et les rythmes de travail ;
- les risques, les pénibilités et leur prévention ;
- les contraintes psychosociales, les relations avec le public, la violence au travail ;
- un auto-questionnaire, rempli par l'enquêté regroupe les questions les plus sensibles.

¹ Ce travail est le fruit d'une collaboration entre Amélie Mauroux (Dares), Corinne Mette (Dares au moment de la rédaction), Claire Beswick (DGAFP au moment de la rédaction), Céline Dennevault (Drees au moment de la rédaction). Il a bénéficié des conseils de Thomas Deroyon (Insee).

L'objectif du redressement du volet individus de l'enquête CT-RPS est double. On vise en effet la constitution de deux fichiers pour l'exploitation :

- d'une part, le **fichier de l'enquête CT-RPS 2016 (INDIVIDUS_AO, partie transversale)**, sur le champ des actifs occupés, comparable aux éditions antérieures de l'enquête Conditions de travail. Pour cela on cherche à obtenir un jeu de poids qui :
 - o assure que les répondants sont représentatifs des actifs occupés en 2015,
 - o assure la comparabilité de la partie transversale CT-RPS 2016 avec les autres éditions de l'enquête Conditions de travail,
 - o soit unique mais permette d'étudier les extensions, notamment privé vs public,
- d'autre part, les **fichiers du panel CT 2013-2016 (AO_PANEL, IN_PANEL, partie longitudinale)**. Pour cela on cherche à obtenir un jeu de poids qui :
 - o rende les répondants à CT 2013 et CT-RPS 2016 représentatifs de l'ensemble des personnes en emploi en 2013,
 - o inclue les individus décédés dans le panel.

De plus, comme pour CT 2013, pour chaque jeu de pondérations, on souhaite obtenir des poids pour les répondants au questionnaire principal (QI) et des poids pour les répondants (un peu moins nombreux) au questionnaire auto-administré (QAA).

La démarche générale de redressement de la partie panel est similaire à celle de l'enquête SIP².

Concernant le calcul des poids pour la partie transversale, le plan de sondage étant complexe (panel, extensions fonction publique et secteur hospitalier privé), un recours intensif à la méthode généralisée du partage de poids a été nécessaire. Nous avons notamment mis en œuvre un calage sur marges adapté au partage des poids tel que décrit dans le chapitre VII du livre de référence de Pierre Lavallée sur le sujet (Indirect Sampling, Springer, 2007). Cette méthode permet de caler chaque échantillon sur les marges qui lui sont propres (les entrants sur des marges relatives aux actifs occupés au 31 décembre 2015, les panels sur des marges relatives aux actifs occupés en 2013) tout en garantissant qu'après partage des poids l'échantillon complet soit calé sur les marges relatives aux actifs occupés au 31 décembre 2015.

² Traitement de la non-réponse et calages pour l'enquête santé et itinéraire professionnel de 2010. Document de travail, série sources et méthodes, n° 36, septembre 2012. DREES

**Étapes de calcul des pondérations longitudinales
(AO_PANEL, IN_PANEL)**

**INDIVIDUS PANEL Actifs et inactifs
(ssech 01 à 06)**

*Poids CT2013 corrigés de la NR individuelle
après partage des poids, avant calage
(QJ et QAA)*

1) **Correction de la probabilité
d'inclusion dans le panel
(DOM uniquement)**

2) **Correction de la non-réponse
échantillon principal**
non-réponse individu

Poids individuels corrigés de la NR

3) **Calages sur marges**

*(sur l'information auxiliaire pertinente par rapport aux
populations d'intérêt et aux variables d'intérêt en
2013)*

Marges CT2013

*Marges : âge, sexe, CSP, secteur d'activité, statut et
contrat*

Poids longitudinaux finaux

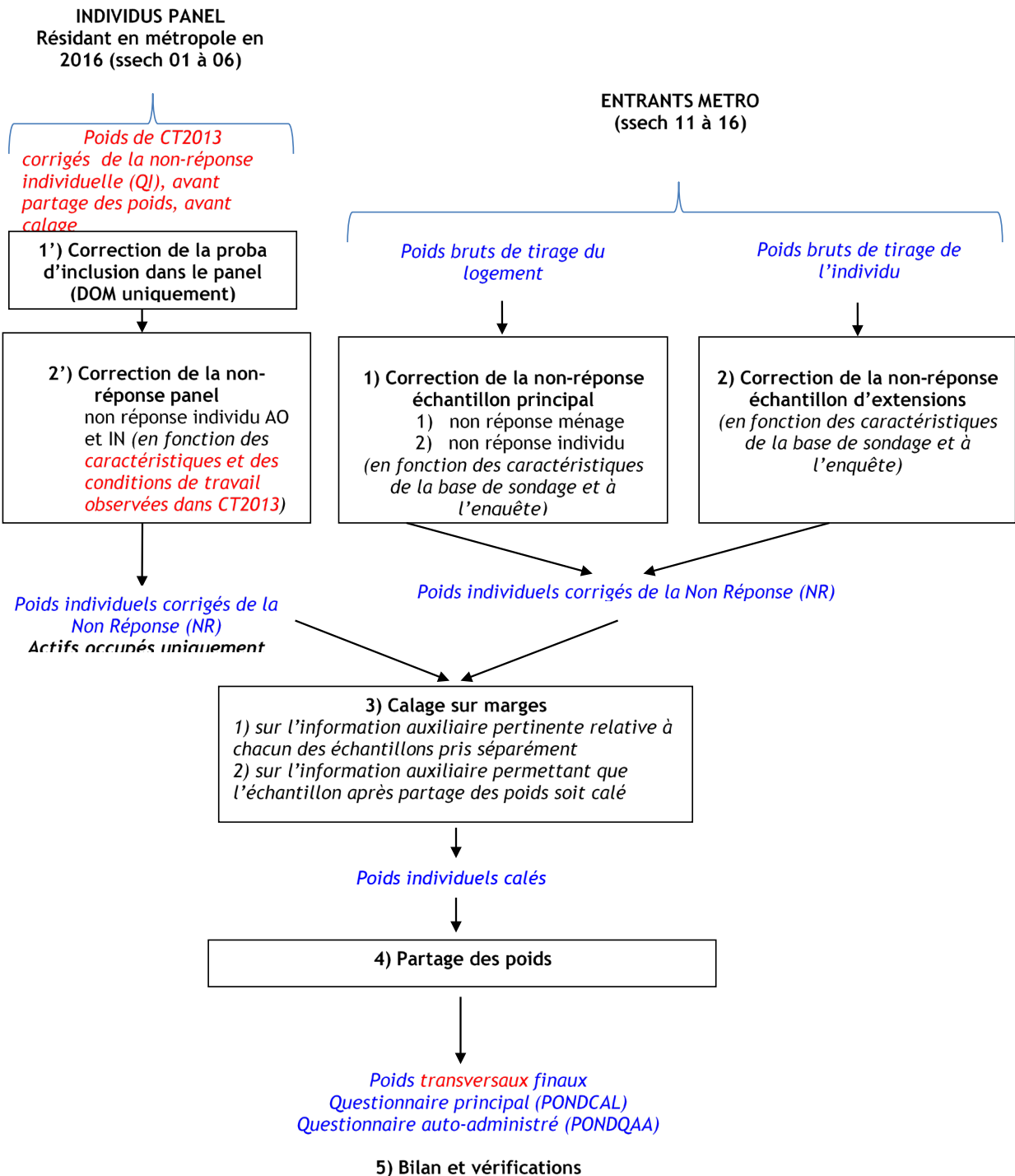
Questionnaire principal (PONDANELCAL)

Questionnaire auto-administré en 2013 ET 2016

(PONDANELQAA)

Étapes de calcul des pondérations transversales (INDIVIDUS_AO)

1- Métropole



Contenu

I. Présentation de l'enquête	11
1. Plan de sondage et échantillonnage	11
a) <i>Echantillon panel</i>	11
b) <i>Echantillon entrant principal (Dares)</i>	11
c) <i>Entrants des extensions DREES et DGAFP</i>	12
2. Protocole de collecte.....	13
3. Bilan de la collecte	13
II. Partie longitudinale	15
1. Données initiales de l'échantillon et population finalement éligible.....	15
a) <i>Les poids de sondage</i>	15
b) <i>Définition de la population (population éligible au Panel) :</i>	16
2. Traitement de la non-réponse.....	17
a) <i>Description de la méthode</i>	17
b) <i>Application au panel de CT-RPS 2016</i>	18
3. Le calage.....	20
a) <i>Le principe du calage</i>	20
b) <i>Application à CT-RPS 2016</i>	20
III. Partie transversale	27
1. Traitement de la non-réponse des panels.....	27
2. Traitement de la non-réponse des entrants.....	28
a) <i>Echantillon principal</i>	28
b) <i>Correction de la non-réponse des échantillons entrants extensions</i>	40
3. Calage sur marges et partage des poids.....	52
4. Calcul des marges.....	61
a) <i>Métropole</i>	61
b) <i>DOM</i>	64
5. Bilan et vérifications.....	69
a) <i>Les pondérations finales</i>	69
b) <i>Effet de la pondération sur quelques variables d'intérêt</i>	71
c) <i>Vérification des effectifs « France entière » du secteur public</i>	76

I. Présentation de l'enquête

1. Plan de sondage et échantillonnage

L'échantillonnage se décompose en 4 sous-échantillons, issus de 4 bases de sondages différentes :

- échantillon panel : les **30 900 individus ayant répondu** à CT 2013,
- échantillon entrants principal : **5 518 logements** tirés au sort dans Octopusse 2014,
- échantillon entrants extensions FP : **912 individus** tirés au sort dans SIASP 2014,
- échantillon entrants extensions secteur hospitalier : **520 individus** tirés au sort dans les SIASP 2014, **365 individus** dans les DADS.

L'échantillon des entrants principaux est un échantillon de logements. Les autres échantillons sont des échantillons d'individus³.

a) *Échantillon panel*

L'échantillon panel est constitué de l'ensemble des individus ayant répondu à l'enquête « Conditions de Travail 2013 » en France métropolitaine et d'un peu plus d'un répondant sur quatre dans les départements des DOM⁴, soit **30 900 individus** répartis dans 22 205 logements ou fiche-adresses (15 471 issus de l'échantillon principal de CT 2013 et 6 734 des extensions Drees et DGAFP).

Les individus du panel ont été réinterrogés quelle que soit leur nouvelle situation vis-à-vis du marché du travail (actifs occupés, retraités, chômeurs...). Les individus éloignés de l'emploi lors de l'enquête CT-RPS 2016 remplissent un questionnaire spécifique.

b) *Échantillon entrant principal (Dares)*

Du fait de l'attrition (non-réponse, sortie de champ, décès, etc.) la taille de l'échantillon panel a diminué entre 2013 et 2016. Cette attrition peut être plus importante pour certains profils de répondants (voir supra) de sorte que l'échantillon panel s'est déformé. De plus, les individus d'un panel vieillissent par définition d'une vague d'interrogation à l'autre, de sorte que les individus les plus jeunes ont 18 ans en 2015 (au lieu de 15 ans en 2012)

Pour garantir compenser l'attrition, couvrir le renouvellement de la population de référence (arrivées sur le territoire national, jeunes actifs de 15 à 18 ans) et ainsi garantir la représentativité de l'enquête en coupe, il est indispensable de tirer des échantillons d'entrants.

Un échantillon de **5 518 logements** a été tiré dans le nouvel échantillon maître (Octopusse) sur la base de tirage constitué à partir de la vague 2014 du recensement. L'échantillon de CT 2013 ayant également été tiré dans le recensement de la population, cela facilite le partage des poids et la correction de la non-réponse.

Dans Octopusse, ne sont retenus que les ménages qui au moment du recensement comprenaient soit au moins un individu en emploi, soit au moins une personne de moins de 60 ans. En effet, il s'agit d'un tirage en deux phases (unités primaires puis logements dans ces unités).

Cette sélection est réalisée afin de concilier un double objectif :

- pour des motifs de coûts, réduire dans l'échantillon la part de ménages sans aucun actif occupé ;

³ Dans l'échantillon panel, plusieurs individus panel peuvent appartenir au même ménage et donc au même logement, le nombre de fiches adresses (FA) panel mises en collecte est donc inférieur au nombre d'individus panel.

⁴ Les quatre DOM suivants (La Réunion, Guadeloupe, Martinique, Guyane) avaient été sur-échantillonnés grâce à un financement de la DGéOM, de sorte que 2 000 FA avaient été mises en collecte dans chacun des départements. Le financement n'ayant pas été reconduit, la taille de l'échantillon a été réduite à 500 FA dans chaque département, réparties en 400 FA panel et 100 FA entrantes. Dans chaque département, 400 FA panel ont été tirées au sort dans la liste des FA de ménages ayant répondu à CT 2013, soit environ un peu plus d'un répondant sur quatre.

- interroger aussi des personnes qui ont pu prendre un emploi depuis le recensement : les ménages ne comprenant que des chômeurs, des étudiants ou des inactifs au moment du recensement ont été inclus dans l'enquête dès lors qu'une personne de moins de 60 ans était présente.

Cet échantillon est stratifié en 4 strates :

- a) logements dont le plus âgé des occupants avait moins de 30 ans au moment de l'enquête,
- b) logements avec au moins un adulte de 30 ans ou plus et un jeune de 18 à 30 ans,
- c) autres logements (uniquement des personnes de 30 ans ou plus),
- d) surreprésentation de l'Île-de-France.

Pour favoriser les entrées de champ, la strate 1 a été surreprésentée d'un coefficient 3, la strate 2 conserve l'allocation initiale et la strate 3 a été sous-représentée.

Plusieurs ménages peuvent résider dans un même logement (budgets séparés). Si plus de deux actifs en emploi vivent dans le ménage, un tirage au sort (dit tirage « kish ») sélectionne les deux individus à interroger. Il s'agit donc d'un tirage en 3 phases.

c) *Entrants des extensions DREES et DGAFP*

Les entrants extensions sur les champs « secteur hospitalier » et « fonction publique » permettent de rafraîchir le panel et, du fait du nombre important de répondants, permettent des comparaisons publics privés avec une plus grande précision.

Le tirage des entrants sur les champs « secteur hospitalier » et « fonction publique » repose sur l'utilisation des bases DADS et SIASP 2014. Les DADS ont servi de base de sondage pour le secteur hospitalier privé. Le fichier SIASP a servi de base de sondage pour le secteur hospitalier public et la fonction publique.

Échantillon entrants secteur hospitalier (champ DREES)

Un échantillon de **520 individus** salariés du secteur hospitalier public a été tiré au sort dans SIASP 2014 et un échantillon de **365 individus** salariés du secteur hospitalier privé a été tiré dans les DADS 2013.

Sur la base du champ décrit plus haut, l'échantillon a été stratifié de la façon suivante :

- a) Salariés résidant en Ile-de-France vs salariés des autres régions.
- b) Pour les salariés du public : une indicatrice pour les établissements de l'AP-HP et les Hospices lyonnais.
- c) Pour les salariés du secteur hospitalier privé : secteur lucratif vs secteur non lucratif
- d) Taille de l'établissement : ETP de personnels non médicaux inférieurs à 200 vs 200 et plus (privé lucratif) ; inférieurs à 500 vs 500 et plus (privé non lucratif) ; inférieurs à 1 000 vs 1 000 et plus (hôpitaux publics).
- e) L'âge de l'enquêté : individus âgés de moins de 30 ans à l'enquête (strate 1), de 30 à 34 ans (strate 2) et de 35 ans ou plus (strate 3).

Pour favoriser les entrées de champ, on pondère différemment les strates relatives à l'âge : les moins de 30 ans sont ainsi surreprésentés d'un coefficient 6 et les 30-34 ans sont surreprésentés d'un coefficient 3.

Échantillon entrants fonction publique – hors hôpitaux (champ DGAFP)

Un échantillon de **912 individus** entrants a été tiré aléatoirement dans SIASP. Il a été stratifié selon :

- a) Le sous-échantillon : FPE, FPT ou FPH (hors hôpitaux publics échantillonnés par la Drees).
- b) Pour les agents de la FPE : distinction ministères (dont Éducation nationale) / établissements publics administratifs (EPA).
- c) Pour les agents de la FPT : distinction région, département d'une part et communes d'autre part (et selon la taille des communes) / établissements publics administratifs (EPA).

- d) L'âge de l'enquêté : individus âgés de moins de 30 ans à l'enquête (strate 1), de 30 à 34 ans (strate 2) et de 35 ans ou plus (strate 3).

Pour favoriser les entrées de champ et pallier le plus faible taux de réponse des jeunes, la strate 1 a été surreprésentée d'un coefficient 6, la strate 2 d'un coefficient 3 et la strate 3 conserve l'allocation initiale.

Au total, 885 individus salariés du secteur public (hors hôpitaux publics) ont été tirés au sort dans SIASP 2014.

Les échantillons d'entrants des suppléments DREES et DGAFP n'ont pas pris en compte les salariés qui ont déménagé entre le tirage et la collecte, ni ceux qui sont entrés dans la fonction publique ou le secteur hospitalier pendant cette période. En revanche, comme pour CT 2013, si un individu échantillonné change d'employeur entre le tirage et la collecte, mais que son employeur reste dans le champ des extensions DGAFP-Drees, il a été convenu de le conserver et de l'enquêter.

2. Protocole de collecte

La Collecte s'est déroulée du 1^{er} octobre 2015 au 30 juin 2016⁵. Le protocole de collecte est similaire à celui de Conditions de travail 2013 et assure donc la continuité :

- entretien en face-à-face au domicile avec un enquêteur Insee,
- auto-questionnaire,
- extensions dans la fonction publique et le secteur hospitalier (partenariat avec la Drees et la DGAFP),
- suivi en panel.

Le temps d'entretien était de 60 minutes en moyenne. Dans l'échantillon principal entrant ou panel, deux personnes peuvent être interrogées par ménage. La personne interrogée répond pour elle-même. La réponse par proxy n'est pas autorisée.

L'ensemble des individus répondant à CT 2013 (échantillon principal métropole, DOM et extensions) ont été suivis dans l'application ASAR de l'INSEE depuis le test papier de CT 2013. L'Insee assure ce suivi en panel selon les procédures habituelles de suivi et de fidélisation : remplissage d'une fiche suivi panel en fin d'enquête, saisie des informations dans l'application ASAR, envoi annuel d'une carte de vœux et mise à jour des informations en fonction des retours.

3. Bilan de la collecte

Au total, 31 000 fiches adresses (FA) ont été confiées aux enquêteurs Insee (29 000 en métropole et 2 000 dans les DOM), dont 23 805 FA panel et 7 295 FA entrantes. Le taux de collecte moyen (défini comme le rapport entre le nombre de FA réussies et le nombre total de FA) est de 69,6 %. Il varie de 48,1 % (Île-de-France) à 82,1 % (Alsace). Le taux de réussite moyen (défini comme le rapport entre le nombre de FA réussies et le nombre total de FA appartenant au champ) est de 80,9 %. Il varie de 58,1 % à 90,8 % selon les régions.

La collecte a permis de recueillir 27 642 questionnaires individuels (25 979 en métropole et 1 663 dans les DOM). Parmi les personnes interrogées, on compte 22 852 individus panels : 17 546 issus de l'échantillon principal et 5 306 issus des échantillons DREES et DGAFP (tableau I.1). On compte également 4 790 individus entrants : 3 853 issus de l'échantillon principal et 935 des échantillons extensions. Parmi les répondants, 24 672 sont actifs occupés, 19 882 panel (15 136 issu de l'échantillon principal, 4 746 des extensions) et, par définition, les 4 790 entrants. L'enquête a également permis d'interroger 2 970 personnes du panel éloignées de l'emploi : 2 410 issues de l'échantillon principal et 560 issues des échantillons d'extensions.

⁵ La collecte a été étendue jusqu'au 9 juillet en Île-de-France.

Tableau I.1 : Répartition des individus répondants par sous-échantillon

	Échantillons		Total répondants
	Principal	Extensions	
Individus entrants	3 855	935	4 790
Individus panels	17 546	5 306	22 852
dont			
- <i>actifs occupés</i>	15 136	4 746	19 882
- <i>inactifs</i>	2 410	560	2 970

Au total, le taux d'attrition France entière a été de 26,5 % (tableau I.2). Il est inférieur à celui initialement attendu au regard de l'enquête SIP (30 %). Il est plus élevé pour les individus « panels » de l'échantillon principal que pour ceux des extensions (27,6 % contre 21,2 %). Les individus du panel issus des extensions étant majoritairement fonctionnaires, cela a pu favoriser une certaine stabilité de l'emploi et géographique.

Au contraire, pour les entrants, les taux de réponse ont été plus élevés pour l'échantillon principal (79,6 %) que pour les échantillons des extensions (64,5 %).

Tableau I.2 : Taux de réponse par sous-échantillon

	Échantillon		Moyenne
	Principal	Extensions	
Échantillons panel	72,4 %	78,8 %	73,5 %
Échantillons entrants	79,6 %	64,5 %	76,1 %

II. Partie longitudinale

Le fichier longitudinal (panel) permet l'étude des trajectoires individuelles, notamment des évolutions des conditions de travail en lien avec le parcours professionnel et les évolutions de l'état de santé entre 2013 et 2016.

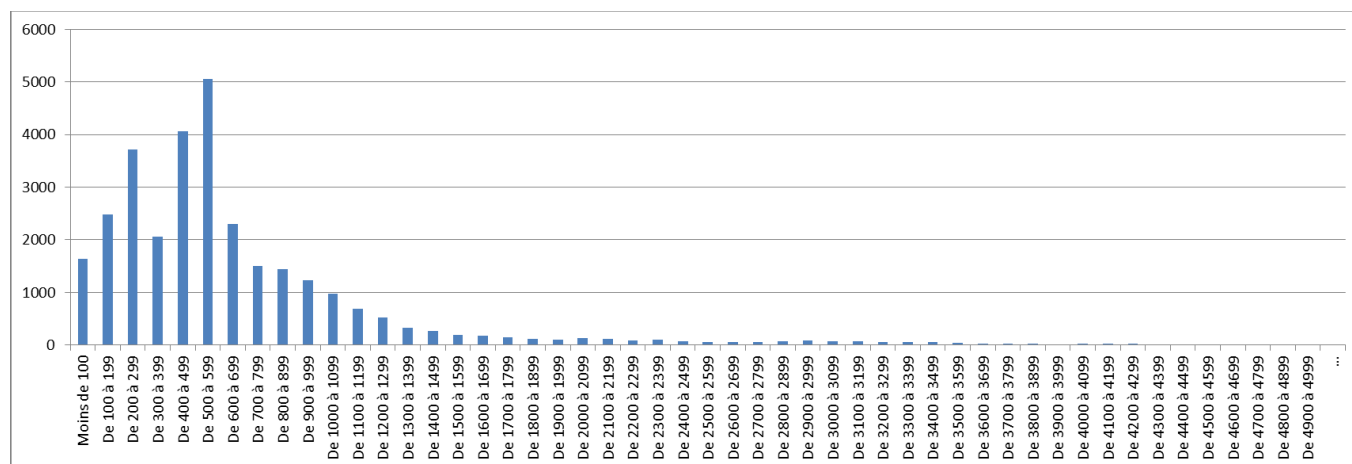
L'objectif ici est d'obtenir une pondération qui permettra de rendre les individus qui ont répondu à CT 2013 et à CT-RPS 2016 (y compris ceux issus des extensions) représentatifs de l'ensemble des personnes en emploi en 2013. La création du jeu de poids consiste, dans un premier temps, dans le traitement de la non-réponse et, dans un second, au calage sur marges. Avant cela, il est toutefois nécessaire de présenter les données initiales de l'échantillon ainsi que de bien déterminer la population éligible au Panel.

1. Données initiales de l'échantillon et population finalement éligible a) Les poids de sondage

Le calcul des pondérations pour la partie longitudinale s'appuie sur les fichiers de répondants à l'enquête CT 2013. Les poids initiaux sont les poids corrigés de la non-réponse, après partage des poids mais avant calage de l'enquête CT 2013, autant pour les résidents en métropole que pour les résidents dans les DOM.

La dispersion des poids initiaux sur l'ensemble de l'échantillon éligible au panel (individus résidents en métropole et en DOM, des secteurs privé et public) est importante puisqu'elle oscille entre 5 et 18 175 (Graphique II.1). Les poids se concentrent autour de deux modes, le premier entre 200 et 299, le second entre 400 et 599.

Graphique II.1 : Distribution des poids initiaux des individus de l'échantillon panel



La courbe de Lorenz fait apparaître que les ménages du dernier décile de poids concentrent 40 % de la somme totale de poids (graphique II.2). La dispersion des poids est donc très élevée.

Graphique II.2 : Courbe de Lorenz des poids initiaux des individus de l'échantillon panel

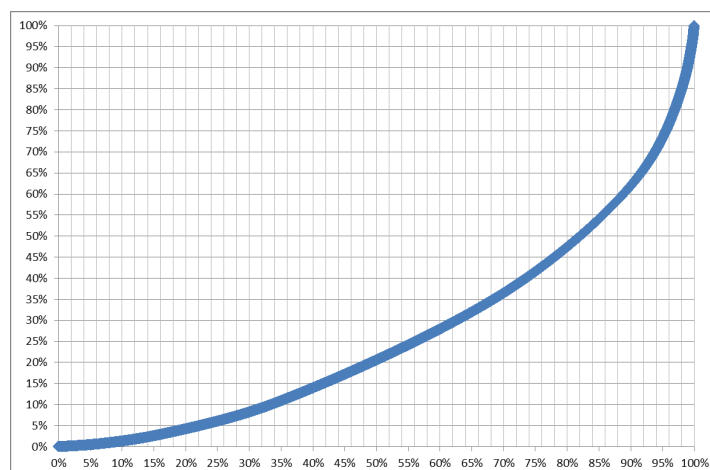


Tableau II.1 : Indicateurs de distribution des poids initiaux des individus de l'échantillon panel

	Poids initiaux avant correction de la non-réponse 2016
Moyenne	753
Médiane	525
Écart type	968
Q1-Q3	296-823
rapport interquartile	2,78
D1-D9	155-1 330
Rapport inter-décile	8,58
Minimum-Maximum	5-18 175

b) Définition de la population (population éligible au Panel) :

La partie longitudinale ou autrement appelée « Panel » se compose des individus interrogés et répondants à l'enquête CT 2013 qui ont également répondu à l'enquête CT-RPS 2016 ou qui sont décédés sur la période. Conserver les personnes décédées dans le panel permet de renseigner sur les liens entre conditions de travail et santé, le décès est, en quelque sorte, une indication de l'état de santé⁶.

Pour les autres, les individus sortis du champ des enquêtes ménages Insee (« hors champ »), c'est-à-dire les individus partis en institution ou à l'étranger entre 2013 et 2016, et ceux qui n'ont pas répondu pour d'autres raisons, ils se voient attribuer un poids longitudinal nul.

Sur les 30 904 individus de CT 2013 ayant un questionnaire exploitable (Tableau II.2), 22 853 (74 %) restent dans le champ, ont répondu à CT-RPS 2016 ; au moins 57 individus sont décédés entre les deux vagues⁷.

Tableau II.2- Répartition des individus du panel CT 2013 – CT-RPS 2016

⁶ Hors cas de mort accidentelle.

⁷ Ne sont considérés ici que les décès dont l'enquêteur Insee a été informé (retour courrier, autre membre du ménage, voisin, etc.), il s'agit donc d'une sous-estimation du nombre de décès entre les 2 vagues.

Individus de l'échantillon panel issus de CT 2013		30904
« Hors-champ » de CT-RPS 2016 (parti en institution, à l'étranger...)		51
Non-répondants à CT-RPS 2016		7 943
Panel	Répondants à CT-RPS 2016	22853
	Décédés	57

Pour rappel, l'unité enquêtée ici est l'individu actif occupé interrogé en 2013 en métropole et dans un DOM, que ce soit dans le cadre de l'échantillon principal de CT 2013 ou dans un des échantillons supplémentaires de la fonction publique et du secteur hospitalier. L'unité d'analyse quant à elle est l'unité enquêtée qui répond à l'enquête en 2016 quelle que soit sa situation à cette date vis-à-vis de l'emploi (jeunes retraités, chômeurs, autres inactifs...) et son échantillon de provenance. Si, en 2013 toutes les personnes enquêtées étaient actives occupées, ce n'est plus nécessairement le cas en 2016.

2. Traitement de la non-réponse

a) Description de la méthode⁸

La correction de la non-réponse⁹ vise à prendre en compte, par un jeu de pondération, les biais de représentativité d'un échantillon dont une partie seulement a répondu. Le poids initial des non-répondants devient nul et celui des répondants est dilaté. Dans le cadre d'un panel, le poids obtenu tient compte de la probabilité d'être répondant à CT-RPS 2016 conditionnellement au fait d'être répondant à CT 2013.

La probabilité de répondre ou pas à CT-RPS 2016 est estimée à partir d'informations disponibles par ailleurs, récoltées ici lors de CT 2013.

Ces informations vont permettre de définir des groupes de réponses homogènes (GRH), qui divisent la population en sous-populations supposées homogènes au sens de la non-réponse. La probabilité de réponse est supposée constante au sein d'un GRH, et le comportement de réponse indépendant d'un GRH à l'autre. Au sein de chaque GRH, le mécanisme de réponse suit donc une loi de Bernoulli. Pour chaque GRH h , cette probabilité (h Panel P) est estimée par le rapport entre le nombre non pondéré de répondants en 2013 et le nombre non pondéré d'individus de cette sous-population répondante en 2016. On remarquera que les taux de réponse non pondérés et pondérés sont très proches.

Différentes techniques de construction de profils de répondant existent, deux sont testées pour ce travail et une seule est retenue. L'une vise à estimer par régression logistique, pour les répondants de CT 2013, la probabilité de répondre à CT-RPS 2016. Cette probabilité permet ensuite de regrouper la population par groupes de probabilité de réponse proches. Cette méthode a été testée ici, mais ne fournit pas des groupes satisfaisants.

La seconde, dont les résultats sont présentés, fait appel à l'algorithme de CHAID basé sur le test du χ^2 , qui permet d'identifier les caractéristiques divisant le mieux l'échantillon en groupes selon leurs propensions à répondre (encadré II.1). L'algorithme regroupe pour chaque variable explicative ses modalités les moins liées à la variable à expliquer (réponse en 2013) ; il le fait de proche en proche par paire ou regroupement de modalités. Cette méthode offre une grande lisibilité des GRH constitués.

⁸ Ce paragraphe est en grande partie issu du paragraphe 2.2.3 du document de travail sur le traitement de la non-réponse et le calage de l'enquête Santé et itinéraire Professionnel de 2010 (http://drees.solidarites-sante.gouv.fr/IMG/pdf/seriesource_method36.pdf).

⁹ Il s'agit de la non-réponse totale (le questionnaire de l'individu reste vierge ou n'est que très partiellement rempli) et non de la non-réponse partielle (une seule partie du questionnaire n'est pas renseignée) qui est en général traitée par imputation.

Encadré II.1 – L’algorithme CHAID

L’algorithme Chi-square Automatic Interaction Detector (CHAID) est une technique d’arbre de décision, dont la méthode de partitionnement est récursive et basée sur des tests du Chi-2 de Pearson.

L’algorithme se décompose en deux étapes : la première étape consiste dans le regroupement deux-à-deux des modalités des prédicteurs. On regroupe deux modalités quand elles ne sont pas statistiquement différentes dans le test du Chi-2 de Pearson entre le prédicteur et la variable binaire de comportement de réponse.

La deuxième étape de l’algorithme consiste dans la sélection des variables utilisées dans le découpage du groupe des individus en groupes de réponse homogène. Les variables de découpage sont déterminées de façon récursive, la variable retenue à chaque balayage est celle ayant la p-value la plus faible. L’algorithme repose sur deux paramètres à choisir, le seuil de significativité en dessous duquel une corrélation entre comportement de réponse et variable auxiliaire est jugée pertinente, ainsi qu’un critère avec la taille minimale de chacun des groupes.

Soit $W_{Panel,i}^{TNR}$ le poids longitudinal traité de la non-réponse de l’individu i pour la partie *Panel*. En définissant l’indicatrice $I_{Panel,i}^h$ (= 1 si l’individu i de *Panel* appartient au GRH h ; 0 sinon et H le nombre de GRH), on a :

$$W_{Panel,i}^{TNR} = \left(\sum_{h=1}^H \frac{1}{P_{Panel}^h} I_{Panel,i}^h \right) W_{v1,i}^{TNR,C}$$

Le processus de non-réponse est ainsi assimilé à la réalisation d’une enquête en deux phases, avec post-stratification. En effet, tout se passe comme si l’on tirait, au sein de l’échantillon initial post-stratifié par la mise en place des GRH, un second échantillon avec un taux de sondage égal, pour chaque GRH, à la probabilité qu’a l’individu de répondre.

b) Application au panel de CT-RPS 2016

L’enquête CT 2013 a permis de récolter des informations qui peuvent s’avérer pertinentes dans l’explication de la non-réponse à l’enquête CT-RPS 2016. Elles ont trait aux caractéristiques sociodémographiques des personnes (sexe, âge, niveau de diplôme, lieu de résidence, être immigré ou non), à leur état de santé avec en particulier la perception qu’elles en ont, et à leur situation vis-à-vis de l’emploi (nature du contrat de travail et conditions de travail) (

Tableau II.3). Enfin, une variable de gestion d’enquête a été ajoutée puisque les réunions de bilan de la collecte révélaient que dans de nombreux cas de ménages Panel avec deux répondants, la 2e personne, découragée par la longueur du questionnaire, ne répondait pas. Le nombre d’individus panel à enquêter au sein du ménage semble donc impacter le taux de réponse. On tient compte de cela en ajoutant ainsi au modèle de correction de la non-réponse individuelle le nombre d’individus panel dans le ménage.

L’algorithme CHAID est mis en œuvre avec la macro SAS TreeDisc. La taille minimale d’un GRH a été fixée à 250 individus, afin de ne pas avoir d’effectifs trop faibles. L’algorithme CHAID permet ainsi de déterminer 88 GRH (qui ne sont pas présentés ici en raison de leur nombre important).

Parmi les 164 variables choisies pour construire des groupes de réponse homogène, 48 sont retenues par l’algorithme pour diviser la population :

Variables sociodémographiques :

- Sexe (homme/femme)
- Age (Moins de 30 ans ; De 30 à 39 ans ; De 40 à 49 ans ; De 50 à 59 ans ; 60 ans et plus)

- Lieu de résidence : Paris, Lyon, Marseille et 52 modalités obtenues en croisant la région (département dans les DOM) de résidence avec le caractère rural ou urbain de la commune de résidence (voir annexe I pour la liste exhaustive)
- Salarié du secteur privé (oui/non)

Variable de gestion d'enquête :

- Nombre d'individus panel à enquêter dans le logement (Une/deux/trois et plus¹⁰)

Variables de conditions de travail (voir annexe I pour la liste exhaustive)

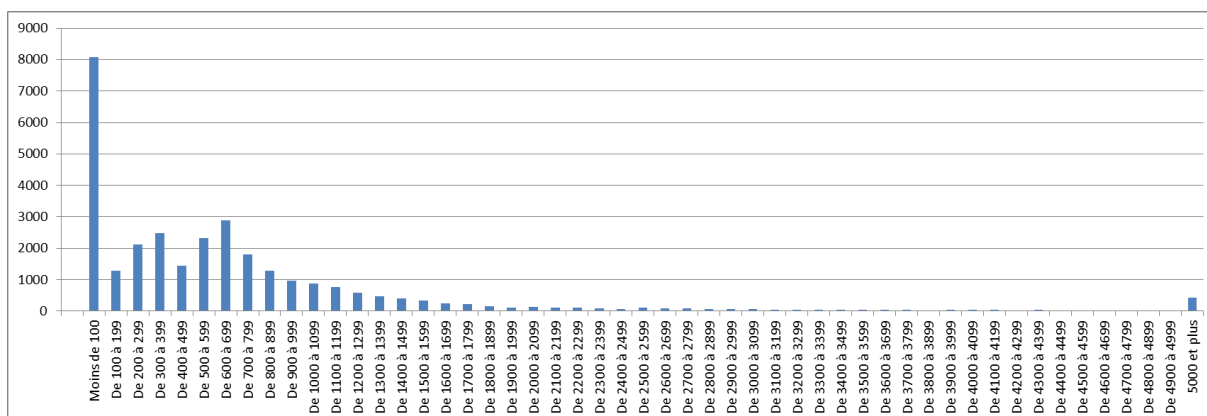
Après traitement de la non-réponse, le rapport des poids se situe dans l'intervalle [1,19 ; 1,34] ; 98 % des poids corrigés de la non-réponse se situent entre 125 et 6 721, ils se situaient initialement entre 82 et 5 130.

Tableau II.3 : Quantiles des pondérations avant et après traitement de la non-réponse à CT-RPS 2016

	Poids traités de la non-réponse en 2013, après partage des poids mais avant calage de l'enquête CT 2013 pour la population éligible au panel	Poids panel corrigé de la non-réponse à CT-RPS 2016	Rapport des poids
100Max 100 %	15530	20853	1,34
99%	5 130	6 731	1,31
95%	2 254	2 977	1,32
90%	1 330	1 844	1,39
75% Q3	823	1 061	1,29
50 % Médiane	525	654	1,25
25 % Q1	310	390	1,26
10 %	193	250	1,29
5%	145	190	1,31
1%	82	125	1,53
0% Min	5	6	1,19

¹⁰ Lors de la mise en collecte de CT-RPS 2016, les ménages d'un même logement qui avaient déclaré faire budget séparé ont été refusionnés sur la même FA. Il peut donc y avoir plus de 2 individus panel dans le même ménage, même si lors de l'enquête CT 2013 il y avait au maximum deux répondants par ménage.

Graphique II. 3 : Distribution des poids corrigés de la non-réponse à CT-RPS 2016 des répondants du panel



3. Le calage

La correction de la non-réponse permet de limiter les biais de représentativité dus au fait que toutes les personnes de l'échantillon n'ont pas nécessairement répondu au questionnaire, elle ne permet pour autant pas d'assurer la représentativité des répondants par rapport à la population cible. À cette fin, on a recours au calage sur marges qui va permettre de rendre extrapolable les résultats de CT-RPS 2016 à la population cible donc, mais également de calculer un second jeu de poids permettant de corriger la moindre participation au questionnaire auto-administré.

a) Le principe du calage

Le calage permet d'égaliser les effectifs ainsi que les proportions pondérées issus de l'échantillon avec les effectifs et les proportions connus de la population, par l'intermédiaire d'informations externes, tout en assurant une déformation minimale des pondérations initiales.

Les pondérations corrigées de la non-réponse sont corrigées par un coefficient qui permet ainsi de réduire la variance d'échantillonnage.

On utilise ici la macro « Calmar 2 » de SAS qui permet de minimiser la distance entre les poids corrigés de la non-réponse à CT-RPS 2016 $W_{pan}^{T, NR, el, i}$ et les pondérations recherchées $W_{v, 1, i}^{T, NR, C}$.

b) Application à CT-RPS 2016

L'objectif du calage sur marges longitudinal est d'assurer la représentativité des répondants panel de CT-RPS 2016, relativement à la population cible au moment de la sélection du panel : les actifs occupés de 15 ans et plus résidant en métropole ou en Dom en 2013. Le calage est donc fait en fonction des caractéristiques des répondants lors de la 1^e interrogation en 2013. L'information provient ainsi de CT 2013.

Sept variables sont retenues, le sexe, l'âge, le niveau de diplôme, la tranche d'unité urbaine, le statut et la classification d'emploi, ainsi que la fonction exercée (Tableau II.). Pour les besoins du calage, les modalités ont été regroupées comme indiqué dans le tableau suivant.

Tableau II.4 : Variables utilisées pour le calage de la partie panel

Nature des variables	Variables	Modalités
Socio-démographique	Sexe	Homme/femme
	Age	-Moins de 30 ans ; -De 30 à 44 ans ; -45 ans et plus ;
	Niveau de diplôme	-Au maximum un CAP, BEP ou l'équivalent -Baccalauréat -Diplôme de niveau Bac+2 et plus
	Tranche d'unité urbaine 2012	-Commune rurale -Commune appartenant à une unité urbaine de 2 000 à 4 999 habitants -Commune appartenant à une unité urbaine de 5 000 à 9 999 habitants -Commune appartenant à une unité urbaine de 10 000 à 19 999 habitants -Commune appartenant à une unité urbaine de 20 000 à 49 999 habitants -Commune appartenant à une unité urbaine de 50 000 à 99 999 habitants -Commune appartenant à une unité urbaine de 100 000 à 199 999 habitants -Commune appartenant à une unité urbaine de 200 000 à 1 999 999 habitants -Commune appartenant à l'unité urbaine de Paris
Caractéristiques de l'emploi exercé en 2013	STATUT2	-Salarié de l'État -Salarié d'une collectivité territoriale -Salarié d'un hôpital public ou du secteur public social et médico-social -Salarié d'un établissement de santé privé -Salarié d'une entreprise, d'un artisan, d'une association -Salarié d'un ou plusieurs particuliers -Le reste (Aide un membre de votre famille dans son travail sans être rémunéré, Chef d'entreprise salarié et indépendant ou à son compte)
	Fonction principale de l'emploi exercée	-Production, chantier, exploitation -Installation, réparation, maintenance, gardiennage, nettoyage, entretien ménager -Manutention, magasinage, logistique -Secrétariat, saisie, accueil -Gestion, comptabilité

		<ul style="list-style-type: none"> -Commercial, technico-commercial -Études, recherche et développement, méthodes -Enseignement -Soin des personnes -Autre fonction
	Classification de l'emploi	<ul style="list-style-type: none"> -Manœuvre ou ouvrier spécialisé, ouvrier qualifié ou hautement qualifié -Technicien -Personnel de catégorie B ou assimilé -Agent de maîtrise, maîtrise administrative ou commerciale, vrp (non cadre) -Personnel de catégorie A ou assimilé -Ingénieur, cadre (à l'exception des directeurs généraux ou de ses adjoints directs) -Personnel de catégorie C ou D ou assimilé -Employé de bureau, employé de commerce, personnel de services Directeur général, adjoint direct

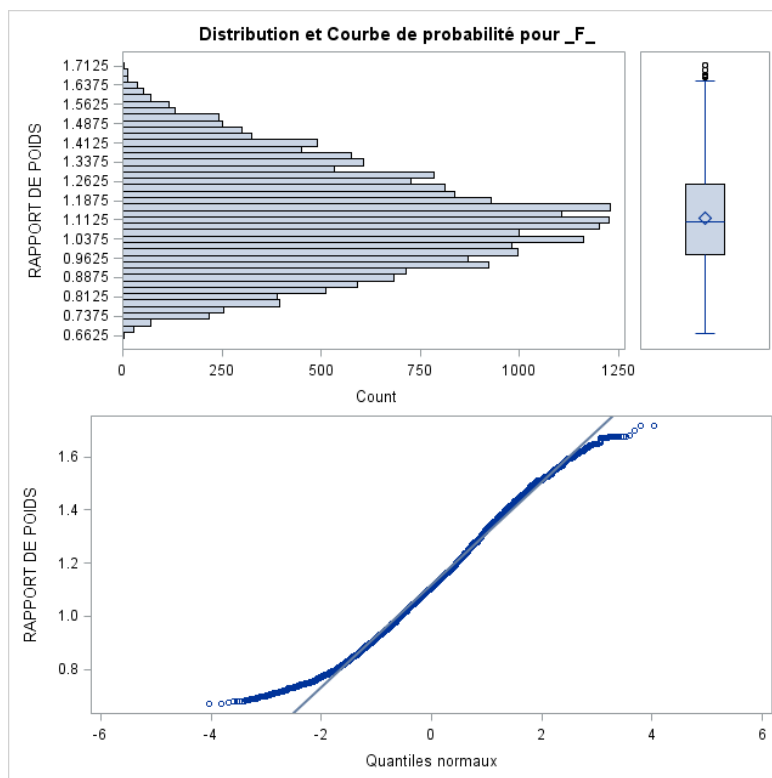
Tableau II.5 : Les marges et la répartition avant calage

		Marges issues de l'échantillon	Marges de la population	Pourcentage échantillon	Pourcentage échantillon QAA
SEXE	Homme	11 344 448	13 935 819	49,8	49,0
	Femme	11 460 574	12 684 736	50,3	51,0
Age	Moins de 30 ans	3 492 679	4 738 607	15,3	15,7
	De 30 à 44 ans	8 823 053	10 243 911	38,7	39,4
	45 ans et plus	10 489 290	11 638 037	46,0	44,9
Niveau de diplôme	Au maximum un CAP, BEP ou l'équivalent	9 081 499	11 378 353	39,8	38,6
	Baccalauréat	4 055 594	4 713 198	17,8	18,2
	Diplôme de niveau Bac +2 et plus	9 667 929	10 529 004	42,4	43,3
Tranche d'unité urbaine 2 012	Commune rurale	5 896 151	6 720 397	25,9	25,9
	Commune appartenant à une unité urbaine de 2 000 à 4 999 habitants	722 656	834 051	3,2	3,2
	Commune appartenant à une unité urbaine de 5 000 à 9 999 habitants	1 165 119	1 321 137	5,1	4,9
	Commune appartenant à une unité urbaine de 10 000 à 19 999 habitants	1 914 780	2 200 765	8,4	8,5
	Commune appartenant à une unité urbaine de 20 000 à 49 999 habitants	1 459 517	1 763 694	6,4	6,4
	Commune appartenant à une unité urbaine de 50 000 à 99 999 habitants	1 583 431	1 859 665	6,9	7,1
	Commune appartenant à une unité urbaine de 100 000 à 199 999 habitants	1 637 079	1 893 148	7,2	7,5
	Commune appartenant à une unité urbaine de 200 000 à 1 999 999 habitants	4 696 835	5 672 692	20,6	21,0
	Commune appartenant à l'unité urbaine de Paris	3 729 455	4 355 006	16,4	15,6
Statut d'emploi	Salarié de l'État	2 400 379	2 396 851	10,5	10,8
	Salarié d'une collectivité territoriale	1 887 835	1 841 364	8,3	8,4
	Salarié d'un hôpital public ou du secteur public social et médico-social	1 042 068	1 121 685	4,6	4,7
	Salarié d'un établissement de santé privé	337 072	293 248	1,5	1,5
	Salarié d'une entreprise, d'un artisan, d'une association	13 564 459	16 818 583	59,5	59,5
	Salarié d'un ou plusieurs particuliers	807 028	983 792	3,5	3,5
	Le reste (Aide un membre de votre famille dans son travail sans être rémunéré, Chef d'entreprise salarié et indépendant ou à son compte)	2 766 181	3 165 032	12,1	11,6
Fonction principale de l'emploi exercée	Production, chantier, exploitation	3 307 302	4 142 072	14,5	14,1
	Installation, réparation, maintenance, gardiennage, nettoyage, entretien ménager	2 443 012	3 100 419	10,7	10,4
	Manutention, magasinage, logistique	950 842	1 255 354	4,2	4,2
	Secrétariat, saisie, accueil	1 620 412	1 676 169	7,1	7,2
	Gestion, comptabilité	1 782 560	1 932 126	7,8	8,0
	Commercial, technico-commercial	1 935 135	2 338 250	8,5	8,5
	Études, recherche et développement, méthodes	1 758 020	1 820 098	7,7	8,0
	Enseignement	1 308 434	1 371 789	5,7	5,9
	Soin des personnes	2 077 316	2 468 246	9,1	9,4
	Autre fonction	2 849 541	3 344 151	12,5	12,7
	Non renseigné	2 772 449	3 171 881	12,2	11,7
Classification de l'emploi	Manœuvre ou ouvrier spécialisé, ouvrier qualifié ou hautement qualifié	3 833 302	5 230 902	16,8	16,1
	Technicien	1 468 036	1 670 289	6,4	6,6
	Personnel de catégorie B ou assimilé	1 143 317	997 982	5,0	5,2
	Agent de maîtrise, maîtrise administrative ou commerciale, VRP (non-cadre)	1 360 244	1 500 547	6,0	6,1

Personnel de catégorie A ou assimilé	1 754 703	1 787 284	7,7	7,9
Ingénieur, cadre (à l'exception des directeurs généraux ou de ses adjoints directs)	2 894 485	3 157 750	12,7	12,7
Personnel de catégorie C ou D ou assimilé	2 077 023	2 178 900	9,1	9,3
Employé de bureau, employé de commerce, personnel de services	5 331 241	6 731 142	23,4	23,7
Directeur général, adjoint direct	171 018	194 539	0,8	0,8
Non renseigné	2 771 654	3 171 220	12,2	11,6

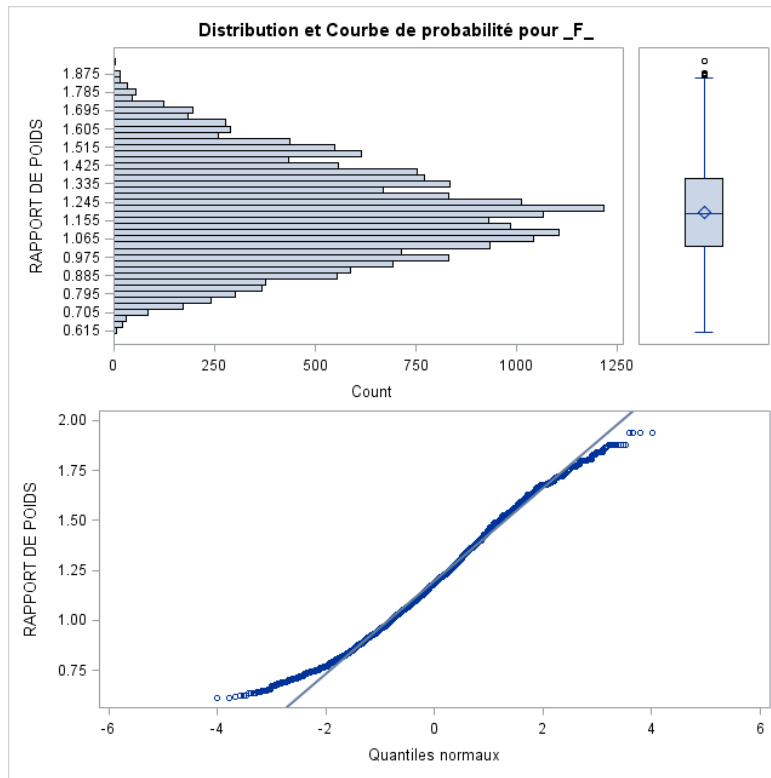
La qualité du nouveau jeu de pondérations se mesure principalement à la distribution du rapport des poids avant et après calage, c'est-à-dire du coefficient de correction des poids. Une distribution concentrée autour des valeurs centrales du rapport des poids est gage d'une bonne qualité. À l'inverse, une distribution concentrée autour des valeurs extrêmes signifierait une qualité moins bonne. Le rapport des poids associés aux répondants à l'ensemble du questionnaire oscille entre 0,66 et 1,71 et se concentre principalement autour des valeurs centrales, entre 0,9 et 1,18.

Graphique II.4 : Distribution du rapport des poids issus du calage sur marges de l'ensemble du questionnaire



La distribution du rapport des poids calculé pour la partie du questionnaire auto-administré se concentre quant à elle entre 0,6 et 2,16 (Graphique II.5).

Graphique II.5 : Distribution du rapport des poids issus du calage sur marges pour le questionnaire auto-administré



Graphique II.6 : Histogramme des poids calés des répondants au panel

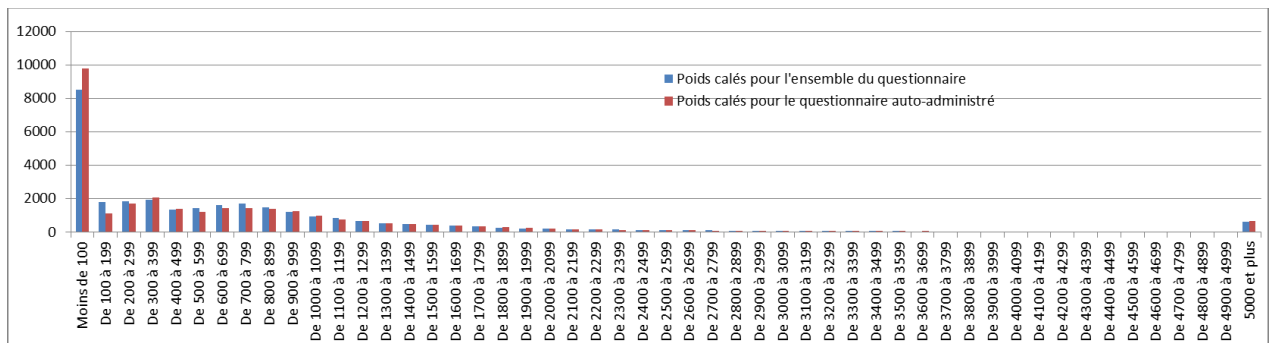


Tableau II.6 : Dispersion des poids des répondants au panel

	Poids panel corrigé de la non-réponse	Poids panel après calage pour l'ensemble des répondants	Poids panel après calage pour les répondants au questionnaire administré
Effectifs	22834	22834	21211
Moyenne	998,7	1 165,8	1 255,0
Premier quartile	390,4	396,8	420,0
Médiane	654,3	750,6	804,9
Troisième quartile	1 060,6	1 252,8	1 349,1
Minimum	5,9	5,8	6,1
Maximum	20 853,3	28 242,0	30 353,6
Somme	22 805 022	26 620 555	26 620 555
Rapport interquartile	2,7	3,2	3,2
Rapport inter-décile	7,4	9,1	9,5

III. Partie transversale

Le fichier transversal est celui qui correspond à l'édition 2016 du volet individus de l'enquête Conditions de Travail. La pondération transversale sera celle utilisée pour les analyses en coupe. L'objectif est donc d'obtenir une pondération qui permettra de rendre les répondants à l'enquête CT-RPS 2016 représentatifs des actifs en emploi au 31 décembre 2015¹¹. L'unité de référence est l'individu.

Pour rappel, l'objectif est également de calculer deux jeux de poids, un pour le QI et un pour le QAA. Cela implique de corriger de la non-réponse (et donc de faire toute la procédure) pour le QI, mais aussi pour le QAA.

L'observation transversale concerne uniquement les personnes en emploi au moment de l'enquête. Il s'agit donc :

- des individus des échantillons entrants : ils peuvent appartenir à l'échantillon principal ou aux extensions ;
- des actifs occupés de l'échantillon panel (y compris ceux issus des échantillons extensions de CT 2013).

Les individus panel inactifs en 2016 sont hors champ pour le volet transversal.

La correction de la non-réponse a été effectuée échantillon par échantillon, selon les bases de sondages (panel, entrants principaux tirés dans Octopusse, entrants extensions tirés dans SIASP, entrants extensions tirés dans les DADS) car les variables auxiliaires sont différentes.

Les poids initiaux sont, pour les échantillons entrants, les poids de tirages et, pour le panel, les poids de CT 2013 corrigés de la non-réponse en 2013 mais avant partage des poids et avant calage.

1. Traitement de la non-réponse des panels

Par rapport au calcul des poids transversaux, les poids initiaux sont ici les poids de tirage en 2013 corrigés de la non-réponse en 2013, non calés mais non partagés (tableaux III.1 et III.2). Les sources de recouvrement entre échantillons sont plus nombreuses que dans l'enquête CT 2013 du fait de la présence de l'échantillon entrant et des extensions. La méthode de partage de poids retenue prendra en compte tous les recouvrements (entre 2013 et 2016, entre échantillons principaux et extensions) lors d'une étape unique. Les recouvrements entre échantillon principal et extensions des individus panel en 2013 seront donc à nouveau pris en compte.

La non-réponse en 2016 a été traitée dans la partie précédente : chaque panel répondant actif occupé a été affecté dans un GRH pour lequel on a calculé un taux de réponse dans le GRH.

Les poids corrigés de la non-réponse en 2016 sachant que l'individu était répondant en 2013 sont calculés en multipliant les poids initiaux par l'inverse du taux de réponse dans le GRH.

¹¹ Au 31 décembre 2015.

Tableau III.1 : Indicateurs de distribution des poids initiaux (corrigés de la non-réponse en 2013, non partagés, non calés), ensemble des individus de l'échantillon panel

Moyenne	Poids initiaux
Max	18 175
Min	9
25 % Q1	481
50 % Médiane	613
75 % Q3	976
Rapport inter quartile	2,0
Rapport inter-décile	5,6
Somme des observations	28 075 873

Tableau III.2 : Indicateurs de distribution des poids corrigés de la non-réponse individuelle à CT-RPS 2016, individus actifs occupés en 2016 de l'échantillon panel

Moyenne	Poids corrigés de la non-réponse à CT 2016
Max	20 853
Min	11
25 % Q1	598
50 % Médiane	811
75 % Q3	1 277
Rapport inter quartile	2,1
Rapport inter-décile	6,0
Somme des observations	24 230 872

2. Traitement de la non-réponse des entrants
 a) *Échantillon principal*

L'objectif est de réduire le biais causé par la non-réponse totale.

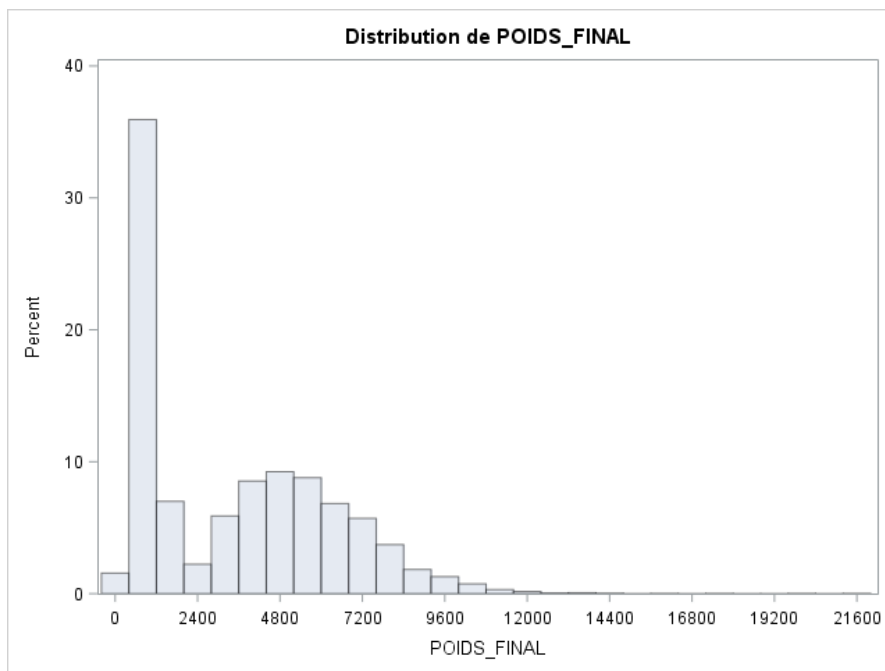
Pour rappel, l'échantillon des entrants principaux est constitué de logements au sein desquels résident les ménages composés des répondants. La correction de la non-réponse de l'échantillon principal se fait donc en deux étapes :

- Correction de la non-réponse ménage (CNRM)
- Correction de la non-réponse individuelle (CNRI)

Poids initiaux

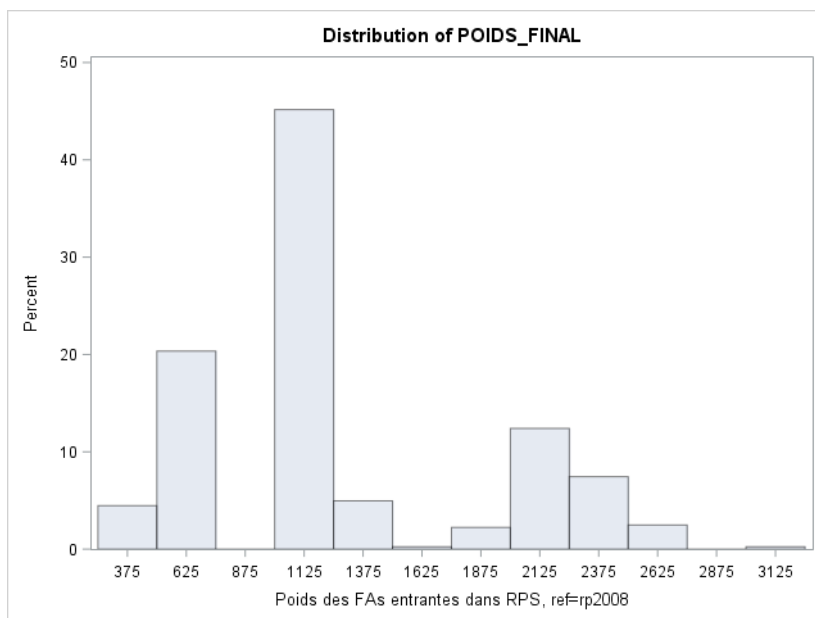
Les poids initiaux des logements sont les poids de tirage dans Octopusse (cf. note de tirage de l'échantillon) pour le détail des règles d'échantillonnage ; graphique III.1 et III.2).

Graphique III.1 : Distribution des poids de tirage des logements de l'échantillon entrant principal, France métropolitaine



Source : Insee - Octopusse

Graphique III.2 : Distribution des poids de tirage des logements de l'échantillon entrant principal, DOM



Source : Insee - Octopusse

Tableau III.3 : Indicateurs de distribution des poids de tirage des logements de l'échantillon entrant principal

	Metro	DOM	Tout
Unités	logements	logements	logements
Moyenne	3 516	1 282	3 364
Max	21916	3 134	21916
Min	32	456	32
25 % Q1	960	1 046	960
50 % Médiane	3 289	1 174	2 689
75 % Q3	5 637	1 743	5 478
Rapport inter quartile	5,9	1,7	5,7
Rapport inter-décile	11,8	4,5	12,4
Somme des observations	19 399 879	516 668	19 916 547

i. Traitement des hors-champ

On identifie ici les ménages hors-champ ici au sens des enquêtes ménage Insee, c'est-à-dire les ménages dont le logement n'est pas un logement ordinaire utilisé au titre de résidence principale en France métropolitaine ou dans un des quatre DOM couverts par l'enquête. Il ne s'agit pas à ce stade des ménages hors du champ thématique de l'enquête CT-RPS 2016. Une condition supplémentaire a été appliquée au moment du sondage au champ « logements » de l'enquête : le ménage devait comprendre au moins un actif occupé ou une personne de moins de 60 ans. Les poids de sondage ont été calculés en conséquence et on ne vérifie pas la permanence de cette condition à l'enquête.

Un « hors-champ » est donc un ménage dont le logement n'est plus habité en résidence principale. On se base sur les codes résultats des enquêteurs INSEE pour les fiches-adresses (FA) correspondantes aux logements mis en collecte. Pour le code résultat « impossible à joindre », on considère comme hors champ les 260 ménages pour lesquels l'enquêteur ne sait pas si la résidence est habituelle ou non, les 1 323 autres sont considérés non-répondants.

Il s'agit des situations suivantes :

Code	Intitulé	Conditions	Nombre de ménages
AHC	Autres Hors Champ	Logement détruit, absorbé, partagé...	33
INC	INConnu	Logement inconnu, impossible à identifier	107
NT1	Non Traitée	FA non traitée avant la fin de la collecte, repérage non fait	108
VAC	VACant	Logement vacant ou occupé exceptionnellement	593
ALD	Absent de Longue Durée	Personne présent, absence habituelle ou non mais certaine	36
BIS	Déjà interrogé	Ménage déjà interrogé pour cette enquête	1
IAJ	Impossible A Joindre	NSP si le logement est habité en résidence principale	94

Lors de la collecte de CT-RPS 2016, la part des FA classées hors-champ a été particulièrement importante dans le sous-échantillon principal des entrants (16,3 % des FA). Pour mémoire, lors de la collecte de CT 2013, seules 9,8 % des FA du sous-échantillon principal avaient été classées hors-champ¹². En particulier, 10 % des FA de l'échantillon entrant principal ont été signalées comme vacantes, alors qu'en 2013 il n'y en avait que 6,6 % sur l'échantillon principal. 108 FA ont été classées NT1 (1,8 % contre 0,5 % en 2013 sur l'échantillon principal).

Un quart de ces hors-champ sont situés dans la région Île-de-France (361). Les NT1 y sont particulièrement nombreux (84 sur 108 NT1 au total). Il n'est pas possible de savoir si les logements concernés auraient été non-répondants, répondants ou effectivement hors champ. Par conséquent, on peut anticiper un déficit de ménages répondants, notamment en Île-de-France.

ii. Correction de la non-réponse totale ménage

Est non-répondant total un ménage pour lequel le questionnaire ménage est vide ou incomplet.

Les codes résultats correspondants sont les suivants :

Code	Intitulé	Conditions	Nombre de ménages
EVT	EViTement	Pas de contact, quelqu'un était présent : refus déguisé	185
IAJ	Impossible A Joindre	Pas de contact, absence inhabituelle et incertaine	368
IMP	IMPossible	Personne présente inapte ou pas habilitée à répondre	108
DEJ	DEJà enquêté	Ménage interrogé pour une autre enquête	30
NT2	Non Traitée	FA n'ayant pu être traitée, le repérage a pu être fait	68
NT3	Non Traitée	FA non traitée, un habitant du logement a été contacté	29
REF	REFus	Ménage refusant l'enquête (ouvertement ou non)	368
THP	THL Partiel	Tableau des Habitants partiellement rempli, abandon	
Total			1 161

La part de ménages non-répondants du sous-échantillon principal est également en hausse, par rapport aux résultats du sous-échantillon principal de l'enquête CT 2013 (19,5 % des ménages en 2016 contre 17,2 % en 2013). Les refus sont en baisse, mais la part des évitements (EVT), des INC et des IAJ sont en hausse.

Les ménages répondants sont tous les autres, c'est-à-dire ceux qui ont un des codes RES suivants (les codes commençant par U signalent qu'il s'agit d'une unité de vie dans un logement qui en comprend plusieurs).

Code	Intitulé	Nombre de ménages
THV	Tableau des Habitants du Logement, Validé	4
ERP	Enquête Réalisée Partiellement	47
ERV	Enquête Réalisée Validée	2 588
ERW	Enquête Réalisée Validée pour un seul actif au lieu de 2	194
URP	Unité Réalisée Partiellement	2
URV	Unité Réalisée Validée	20

¹² Le taux de hors-champ est particulièrement faible pour les FA panel et les FA sous-échantillons extensions (moins de 3 %) mais cela tient au protocole de collecte qui prévoit pour ces FA une recherche nominative des répondants, même s'ils ont déménagé. Ainsi, de 15 à 23 % des FA entrants extensions ont été classées « sans adresse connue ».

URW	Unité Réalisée Validée pour un seul actif au lieu de 2	0
SAO	Sans Actif Occupé	970
Total		3 821

Les ménages sans actifs occupés (code SAO) sont considérés comme répondants car ils ont accepté de recevoir l'enquêteur Insee et de remplir la première partie du questionnaire, le tableau des habitants du logement (THL). Néanmoins, le champ de l'échantillon entrant principal étant restreint aux actifs occupés, ils seront considérés hors du champ du volet transversal de l'enquête CT-RPS 2016.

Dans CT-RPS 2016, 64,2 % des ménages du sous-échantillon principal sont répondants, soit 9 points de moins qu'en 2013 (73 %). Le taux de réussite (ménages répondants sur total des ménages dans le champ) est également en léger recul pour les ménages de l'échantillon principal, 76,7 % contre 81 % en 2013. Les consignes données aux enquêteurs de privilégier le panel pour limiter l'attrition, ont vraisemblablement pénalisé la collecte de l'échantillon entrant principal.

La correction de la non-réponse a été réalisée France entière, l'échantillon des entrants dans les DOM (400 FA dans chaque département) étant trop restreint pour être traité séparément.

La méthode de correction de la non-réponse retenue est la constitution de groupes homogènes de réponse (GRH) en utilisant de l'algorithme CHAID (voir encadré 1).

10 variantes ont été testées, selon la taille minimale des groupes (100, 200 ou 500), le seuil de significativité en dessous duquel une corrélation entre le comportement de réponse et la variable auxiliaire jugée pertinente (0,1 ou 0,2) et l'estimateur du taux de réponse dans les groupes (moyenne empirique non pondérée dans la classe, ou pondérées par les poids de tirage).

Les variables auxiliaires que l'on peut mobiliser proviennent de deux sources :

- La base de sondage, à savoir le recensement (feuille de logement et bulletin individuels des occupants) : situation géographique du logement, composition au moment du recensement (âge, sexe, type de ménage, nombre d'enfants, statut d'emploi des occupants, profession, diplôme, etc.), situation d'occupation du logement (propriétaire, locataire, etc.), lieu de naissance (en France métropolitaine, dans un DOM, à l'étranger), taille d'unité urbaine de la commune.
- Le repérage effectué par les enquêteurs : type d'habitat au voisinage du logement (maisons dispersées, immeubles en ville, etc.), type de logement (ferme, maison de ville, appartement dans un immeuble de deux logements, etc.), vague de collecte.

Elles sont toutes incluses en entrée de l'algorithme CHAID. Pour choisir les paramètres de l'algorithme CHAID, on compare l'écart entre la distribution des variables auxiliaires dans l'échantillon (poids de tirage) et après correction de la non-réponse ménage en calculant la somme de la valeur absolue des écarts.

Les paramètres retenus sont les suivants : alpha = 0,2 et leaf = 100. 19 GRH sont créés à partir des variables auxiliaires suivantes (tableau III.4) :

- taille d'unité urbaine de la commune,
- statut d'occupation du logement,
- type d'habitat au voisinage du logement,
- âge de la personne de référence (en 6 classes),
- nombre de personnes mineures dans le ménage,
- nombre de personnes entre 25 et 29 ans dans le ménage,
- nombre de personnes entre 40 et 49 ans dans le ménage,
- nombre de salariés du secteur public dans le ménage,
- nombre de salariés en CDI dans le ménage,
- nombre de non-salariés dans le ménage,

Une fois ces classes constituées, la probabilité de réponse commune à toutes les unités appartenant à la même classe peut être estimée de deux manières :

- l'estimateur standard, égal au nombre de répondants dans le groupe divisé par le nombre de répondants et de non-répondants dans le groupe ;
- l'estimateur pondéré, égal à la somme des poids de sondage des répondants dans le groupe, divisée par la somme des poids de sondage des répondants et des non-répondants du groupe.

Le premier estimateur est sans biais, le second est approximativement sans biais mais il est plus robuste aux écarts au modèle. Ils ont en général des propriétés très proches.

L'estimateur pondéré peut être sensible dans les cas où il y a une grande dispersion des poids. Dans le traitement de la non-réponse ménage, le rapport entre le Q3 et le Q1 des poids corrigés de la non-réponse ménage varie peu selon qu'on utilise l'estimateur pondéré ou non pondéré de la probabilité de réponse à l'intérieur des classes. Néanmoins, au sein de chaque classe il y a une grande dispersion de poids (jusqu'à des rapports de 1 à 22 entre le poids de tirage minimum et le maximum). L'estimateur standard est donc préférable car l'estimateur pondéré est très sensible au comportement de réponse des quelques unités ayant des poids très élevés et a une variance d'estimation trop importante (tableau III.5 et graphique III.3).

Tableau III.4 : Détails des 19 GRH issus de l'algorithme CHAID pour la correction de la non-réponse ménage de l'échantillon entrant principal

GRH	Variable	Obs.	Tx. de réponse	moyenne	Q1	Q2	Q3	Minimum	Maximum	Somme
1	Poids de sondage	257	77%	4 352	1 202	4 756	6 264	32	21 916	1 118 570
	Poids corrigés NR Men	198		5 660	1 545	6 244	8 199	41	28 447	1 120 727
2	Poids de sondage	133	74%	5 302	4 123	5 067	6 862	388	11 891	705 158
	Poids corrigés NR Men	98		7 543	5 917	7 130	9 826	526	16 138	739 183
3	Poids de sondage	165	95%	5 137	3 751	5 232	6 745	471	11 062	847 648
	Poids corrigés NR Men	157		5 445	4 070	5 499	7 051	495	11 626	854 822
4	Poids de sondage	519	87%	4 566	1 215	4 956	6 745	69	13 287	2 369 897
	Poids corrigés NR Men	450		5 275	1 402	5 716	7 884	80	15 324	2 373 890
5	Poids de sondage	121	74%	1 716	906	1 062	1 294	379	9 904	207 678
	Poids corrigés NR Men	90		2 302	1 218	1 419	1 674	509	13 315	207 135
6	Poids de sondage	135	90%	3 630	1 138	2 018	5 922	379	12 314	490 067
	Poids corrigés NR Men	122		3 832	1 259	2 221	6 309	419	12 936	467 455
7	Poids de sondage	196	81%	3 450	1 031	1 753	6 145	503	10 145	676 207
	Poids corrigés NR Men	159		4 219	1 271	2 022	7 652	620	12 506	670 888
8	Poids de sondage	672	81%	4 821	2 755	5 119	6 698	140	19 701	3 239 928
	Poids corrigés NR Men	542		5 955	3 415	6 261	8 276	174	24 426	3 227 829
9	Poids de sondage	253	92%	4 711	2 185	5 083	6 581	365	12 546	1 191 823
	Poids corrigés NR Men	232		5 149	2 712	5 542	7 060	398	13 681	1 194 531
10	Poids de sondage	125	81%	4 598	2 043	5 083	6 675	109	10 260	574 729
	Poids corrigés NR Men	101		5 800	3 134	6 333	8 246	134	12 698	585 759
11	Poids de sondage	125	94%	3 141	821	1 285	5 885	109	11 985	392 584
	Poids corrigés NR Men	117		3 377	877	1 373	6 287	116	12 804	395 112
12	Poids de sondage	122	75%	3 580	973	1 852	6 106	155	15 959	436 775
	Poids corrigés									

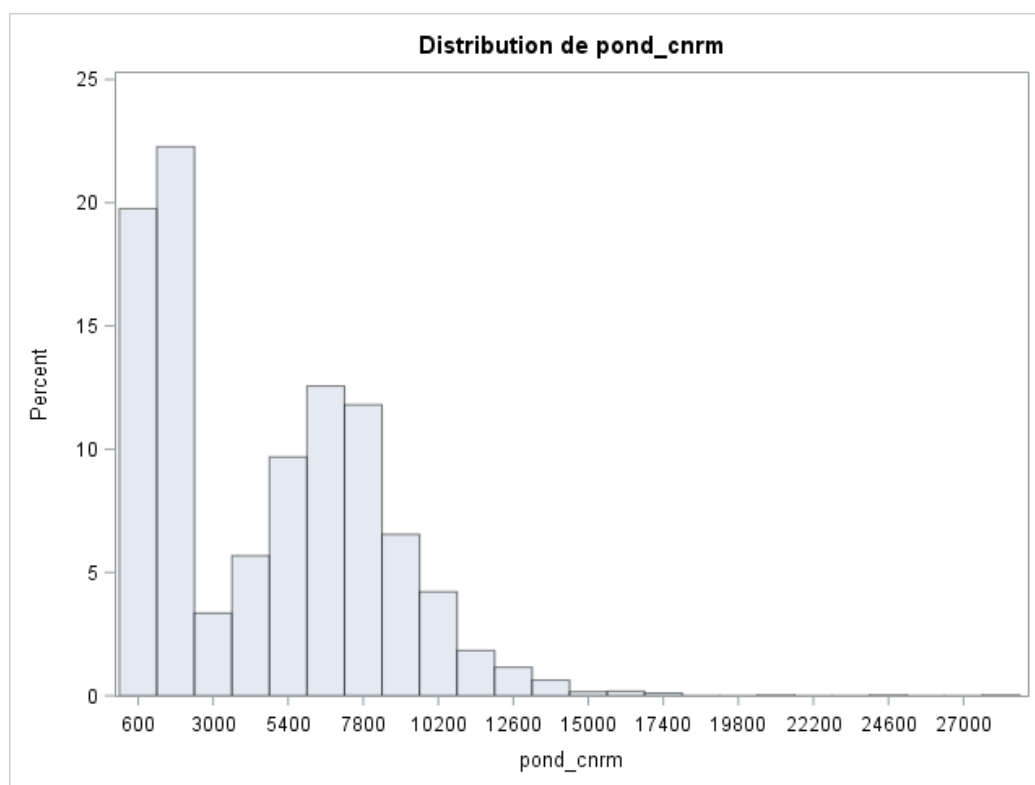
GRH	Variable	Obs.	Tx. de réponse	moyenne	Q1	Q2	Q3	Minimum	Maximum	Somme
13	Poids de sondage	123	67%	4 711	2 491	4 744	6 591	503	12 271	579 394
	Poids corrigés NR Men	83		6 883	3 165	7 030	9 997	751	15 742	571 269
14	Poids de sondage	676	79%	2 401	910	1 180	3 746	199	13 294	1 623 123
	Poids corrigés NR Men	535		3 083	1 171	1 491	4 855	252	16 797	1 649 585
15	Poids de sondage	348	72%	1 399	818	1 017	1 204	292	11 690	486 752
	Poids corrigés NR Men	249		1 870	1 162	1 410	1 683	510	16 338	465 573
16	Poids de sondage	129	70%	4 356	861	4 646	6 492	541	11 047	561 905
	Poids corrigés NR Men	90		6 643	1 411	6 992	9 809	775	15 834	597 867
17	Poids de sondage	189	63%	2 805	663	3 289	4 064	158	9 393	530 134
	Poids corrigés NR Men	120		4 502	1 145	5 380	6 296	249	9 683	540 275
18	Poids de sondage	260	60%	1 944	560	646	3 739	401	5 308	505 309
	Poids corrigés NR Men	156		3 466	939	1 169	6 519	668	8 846	540 620
19	Poids de sondage	434	53%	2 670	684	3 024	3 876	158	8 909	1 158 971
	Poids corrigés NR Men	230		5 193	1 374	5 846	7 361	298	16 812	1 194 443

Lecture : dans le GHR 1, il y a au total 257 ménages, parmi lesquels 198 sont répondants. Le taux de réponse individuel empirique dans ce GRH est de 77 %. Le poids de tirage moyen dans le GRH 1 est de 4 352. Après correction de la non réponse ménage, la moyenne du poids des ménages répondants dans le GRH 1 est de 5 660. La somme des poids de tirage dans GRH 1 avant correction de la non-réponse est égal à 1 118 570.

Tableau III.5 : Indicateurs de distribution des poids avant et après correction de la non-réponse ménage pour les ménages de l'échantillon entrant principal

	Avant CNRM	Avant CNRM	Après CNRM	rapport des poids
Unités	ménages dans le champ	ménages répondants	ménages répondants	ménages répondants
Nombres d'unités	4 982	3 821	3 821	
Moyenne	3 552	3 667	4 667	1,27
Max	21916	21916	28447	1,30
Min	32	32	41	1,30
25 % Q1	1 008	1 031	1 294	1,26
50 % Médiane	3 289	3 492	4 616	1,32
75 % Q3	5 678	5 866	7 347	1,25
Rapport inter quartile	5,6	5,7	5,7	
Rapport inter-décile	11,5	11,1	9,9	
Somme des observations	17 696 650	14 012 790	17 832 890	

Graphique III.3 : Distribution des poids des ménages de l'échantillon entrants principal corrigés de la non-réponse ménage pour les ménages de l'échantillon entrant principal



iii. Correction de la non-réponse totale individuelle

Une fois entré dans le ménage, l'enquêteur doit sélectionner au plus deux actifs occupés pour les interroger (tirage « kish »). Dans les ménages avec plus de deux actifs occupés, il faut donc tenir compte de cette probabilité d'inclusion.

Les poids initiaux individuels sont calculés ainsi :

- les poids bruts de tirage du logement corrigés de la non-réponse totale au niveau ménage ;
- multipliés par l'inverse de la probabilité d'inclusion individuelle :
 - 1 s'il y a un ou deux actifs occupés dans le ménage,
 - $y/2$, s'il y en a y avec $y > 2$, la probabilité d'inclusion étant de $2/y$.

À ce stade, il n'y a plus de problème de définition des hors-champ : tous les individus sélectionnés sont dans le champ, à savoir avoir 15 ou plus et être actif occupé au moment de l'enquête.

Tableau III.6 : Nombres d'individus kish sélectionnés selon le nombre d'actifs par ménage

Nombre d'actifs occupés dans le ménage	Nombre de ménages	Nombre d'individus sélectionnés	Probabilité d'inclusion individuelle
Aucun	974	-	-
1	1 626	1 626	1
2	1 126	2 252	1
3	88	176	2/3
4	11	22	2/4
Total	3 825	4 076	

Sur les 4 076 répondants sélectionnés, 3 878 conserveront le poids du ménage corrigé (tableau III.6). Pour les autres, ce poids est multiplié par 3/2 ou 4/2.

Comme pour CT 2013, on choisit de considérer comme non-répondants parmi les 4 076 individus sélectionnés ceux pour lesquels les variables « statut » en début de questionnaire et/ou « diplôme », en fin de questionnaire, ne sont pas renseignées. Soit ils n'ont pas répondu du tout, soit ils ont abandonné en cours de route.

Sur les 4 076 individus « kishs » tirés au sort :

- 221 n'ont pas répondu du tout ;
- 12 ont abandonné en cours d'entretien et sont considérés comme non-répondant complets.

Au final, il y a 3 843 répondants et 233 non-répondants, soit un taux de non-réponse individuelle de 5,7 %. A titre de comparaison, le taux de non-réponse individuelle des individus kish CT 2013 pour la métropole était de 5,5 % et pour les DOM de 6,9 %.

En général, c'est le second répondant sélectionné dans le logement qui a refusé de répondre. En effet, il n'y a eu qu'un seul cas d'abandon en cours de questionnaire dans un ménage à un seul individu « kish ». On restreint donc l'étape de CNRI aux ménages avec deux actifs sélectionnés¹³. Cela correspond à 2 450 individus « kishs » dans un ménage avec au moins deux actifs occupés : 2 245 individus ont accepté de répondre, 205 ont refusé.

En principe, pour chaque ménage répondant, le questionnaire ménage (THL) est complet, au moins jusqu'à la sélection des répondants. Pour expliquer cette non-réponse individuelle, on dispose donc de plus de variables auxiliaires : celles du questionnaire ménage qui caractérisent l'individu et les variables de la phase précédente.

¹³ Pour les ménages à un seul kish, la correction de la non réponse au niveau individuelle est donc supposée être égale à celle au niveau ménage.

Les variables utilisées pour expliquer la non-réponse individuelle d'un des deux individus kish au sein du ménage sont les suivantes :

- sexe,
- âge (en tranches),
- situation matrimoniale (marié, pacsé, divorcé, veuf, célibataire),
- nombre d'occupants du logement,
- situation principale vis-à-vis de l'emploi (en emploi, autres actifs, pas d'emploi la semaine de référence, aide familial, travail informel, apprenti),
- nombre d'actifs occupés,
- nombre d'enfants,
- kish croisé avec statut d'enfant du ménage,
- type de ménage,
- occupant principal du logement,
- présence dans le logement (toute l'année, souvent, rarement),
- région,
- lieu de naissance (à l'étranger, dans un DOM pour un résident de métropole, dans un DOM pour un résident d'un DOM, en métropole pour un résident de métropole),
- type d'aire urbain,
- vague de collecte,
- classement de la commune en ZUS,
- présence dans le ménage d'un employé, d'un indépendant, un intérimaire, etc.

Comme pour la correction de la non-réponse ménage, l'algorithme de CHAID a été utilisé pour déterminer des GRH. Plusieurs variantes sont testées et le modèle suivant est retenu : $\alpha = 0,01$ et $\text{leaf} = 100$. Il permet d'obtenir 8 GRH à partir des variables suivantes : sexe, âge (en 6 classes), être l'occupant principal du logement, nombre de non-salariés dans le ménage, nombre d'employés, vague de collecte (tableau III.6 et 7, graphique III.4).

Tableau III.7 : Détails des 8 GRH issus de l’algorithme CHAID pour la correction de la non-réponse individuelle dans les ménages de 2 kish de l’échantillon entrant principal

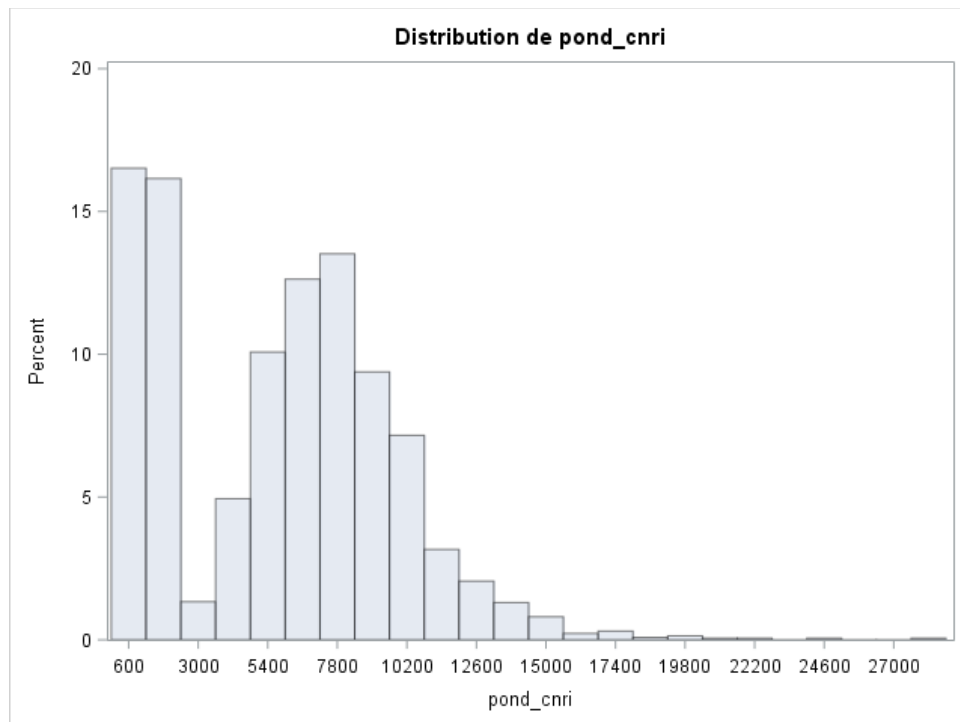
GRH	Variable	Obs.	Tx. de réponse	moyenne	Q1	Q2	Q3	Minimum	Maximum	Somme
1	Poids corrigés NR Men	936	96%	5 503	1 402	5 891	7 977	41	19 437	5 151 189
	Poids corrigés NR ind	895		5 764	1 453	6 164	8 399	43	20 327	5 159 046
2	Poids corrigés NR Men	103	90%	5 942	4 228	6 200	7 920	537	13 313	612 062
	Poids corrigés NR ind	93		6 610	4 639	6 867	8 927	595	14 744	614 693
3	Poids corrigés NR Men	225	96%	5 607	1 374	6 050	8 056	80	17 000	1 261 674
	Poids corrigés NR ind	216		5 848	1 391	6 276	8 402	83	17 708	1 263 129
4	Poids corrigés NR Men	110	96%	6 749	5 488	6 795	8 240	585	19 065	742 440
	Poids corrigés NR ind	106		7 107	5 706	7 052	8 601	609	19 785	753 314
5	Poids corrigés NR Men	445	91%	4 805	1 231	4 826	7 780	41	16 812	2 138 202
	Poids corrigés NR ind	404		5 344	1 355	5 388	8 699	45	18 518	2 158 780
6	Poids corrigés NR Men	258	86%	5 941	4 114	6 105	7 888	576	14 955	1 532 823
	Poids corrigés NR ind	222		7 082	4 914	7 228	9 410	670	17 380	1 572 236
7	Poids corrigés NR Men	175	89%	5 554	1 737	5 573	8 066	530	21 986	971 926
	Poids corrigés NR ind	156		6 190	1 931	6 238	9 087	594	24 664	965 585
8	Poids corrigés NR Men	198	77%	6 210	2 389	6 314	8 376	353	21 986	1 229 487
	Poids corrigés NR ind	153		8 172	3 091	8 535	10 986	457	28 453	1 250 277

Lecture : dans le GHR 1, il y a au total 936 individus sélectionnés (pour lesquels on dispose d’un poids corrigé de la non-réponse ménage), parmi lesquels 895 sont répondants. Le taux de réponse individuel empirique dans ce GRH est de 96 %. Le poids moyen des individus dans le GRH 1 est de 5 503 après correction de la non-réponse ménage et avant la correction de la non-réponse individuelle, celui des individus répondants après correction de la non-réponse individuelle est de 5 764. La somme des poids de l’ensemble individus du GRH 1 (poids corrigés de la non réponse ménage) est égal à 5 151 189.

Tableau III.8 : Indicateurs de distribution des poids après prise en compte du tirage kish avant et après correction de la non-réponse individuelle pour les individus de l'échantillon entrant principal

	après tirage kish, avant CNRI	Après tirage kish, avant CNRI	après CNRI	rapport des poids
Unités	individus sélectionnés (actifs occupés)	individus répondants	individus répondants	individus répondants
Nombres d'unités	4 706	3 843	3 843	
Moyenne	5 211	5 204	5 508	1,06
Max	28447	28447	28453	1,00
Min	41	41	43	1,05
25 % Q1	1 366	1 361	1 424	1,05
50 % Médiane	5 544	5 538	5 792	1,05
75 % Q3	7 777	7 807	8 208	1,05
Rapport inter quartile	5,7	5,7	5,8	
Rapport inter-décile	10,3	10,4	10,8	
Somme des observations	21 238 098	19 998 243	21 169 094	

Graphique III.4 : distribution des poids corrigés de la non-réponse individuelle, pour les individus répondants de l'échantillon entrant principal



b) Correction de la non-réponse des échantillons entrants extensions

Le redressement des échantillons d'extensions de l'enquête CT-RPS 2016 s'inscrit dans la méthode globale de redressement de l'enquête. La correction de la non-réponse totale des sous-échantillons fonction publique a été mise en œuvre par la DGAFP et celle de l'échantillon des cliniques privées de l'extension a été mise en œuvre par la Drees. La méthode de correction des hors-champ et de la non-réponse se fonde sur la méthode utilisée pour l'échantillon entrants principal et le panel. Cette partie décrit les différentes étapes de redressement des échantillons d'extensions avant le partage des poids et le calage sur marges.

Les extensions Drees et DGAFP correspondent aux sous-échantillons 12 à 16. Parmi eux on distingue deux échantillons :

- L'échantillon fonction publique (1 441 individus, tirés dans SIASP 2014), constitué des sous-échantillons :
 - 12 – « Fonction publique d'État » - 465 individus,
 - 13 – « Fonction publique territoriale » - 427 individus,
 - 14 – « Fonction publique hospitalière – Hôpitaux » - 523 individus,
 - 16 – « Fonction publique hospitalière – Autres » - 26 individus.

Cet échantillon inclut l'ensemble des échantillons d'extension DGAFP (FPE et FPT) et une partie de l'échantillon Drees (FPH).

- L'échantillon hôpitaux privés (365 individus, tirés dans les DADS 2013) est le suivant :
 - 15 – « Établissements de santé privé (à but lucratif ou non lucratif) » - 365 salariés d'établissements de santé privés.

La correction de la non-réponse est effectuée sur chaque échantillon séparément car les bases de sondage sont distinctes (SIASP 2014 pour la fonction publique, DADS 2013 pour les établissements de santé privés).

Les étapes de correction de la non-réponse sont les suivantes :

- définition des individus hors-champs et des non-répondants ;
- identification des facteurs corrélés à la non-réponse grâce à une méthode logit ;

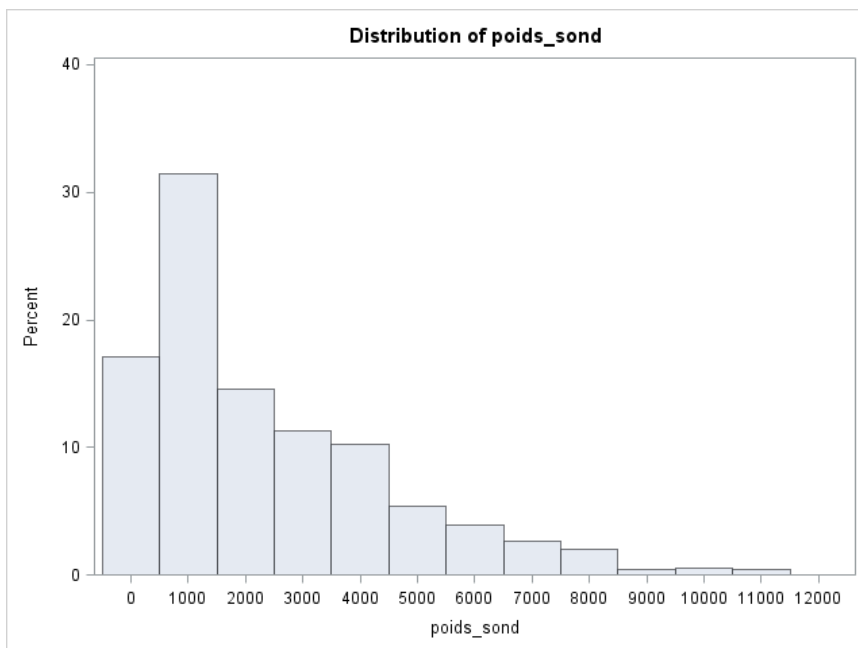
- correction de la non-réponse à l'aide de l'algorithme CHAID.

i. Poids de sondage

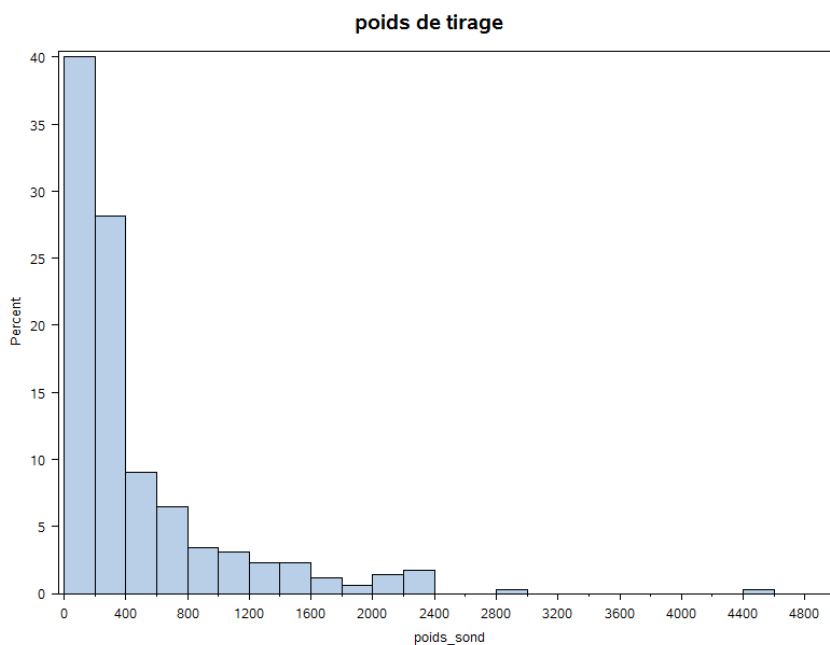
La variable de poids de sondage à considérer est la variable poids_sond.

Les poids sont relativement peu dispersés pour les sous-échantillons fonction publique (graphique III.5, tableau III.9). Pour le sous-échantillon des cliniques privées, tirées dans les DADS 2013, les poids de sondages bruts sont plus dispersés et prennent des valeurs comprises entre 12 et 4 404, avec la moitié des valeurs supérieures à 252 (graphique III.6, tableau III.9). Néanmoins, le neuvième décile des poids bruts s'élève à 1 185, 2 individus ayant un poids plus élevé.

Graphique III.5 : Distribution des poids de sondage dans l'échantillon d'extension fonction publique (12, 13, 14, 16)¹⁴



Graphique III.6 : Distribution des poids de sondage dans l'échantillon extension hôpitaux privés (15)



¹⁴ Un individu de l'échantillon d'extension fonction publique avait un poids de sondage très élevé et très éloigné du reste de la distribution. Il a été retiré pour présenter la distribution des poids de sondage car il la rendait peu visible. Il a cependant bien été pris en compte lors du redressement de l'extension.

Tableau III.9 : Indicateurs de distribution des poids de sondage des individus de l'échantillon d'extension fonction publique (12, 13, 14, 16) et hôpitaux privés (15)

	Extension fonction publique	Extension hôpitaux privés
Unités	Individus	Individus
Moyenne	2 423	458
Max	61 828	4 404
Min	16	12
50 % Médiane	1 605	242
Rapport inter quartile	5,2	4
Rapport inter-décile	16,8	19
Somme des observations	1 843 869	162 517

ii. Identification des hors-champs

Les tables mobilisées sont :

- La table GESTION où figurent notamment les codes résultats de la collecte (RES2) qui permettront de définir les individus hors-champ.
- Les bases de tirage : d'un côté SIASP 2014 pour l'échantillon fonction publique, de l'autre DADS 2013 pour l'échantillon établissements de santé privés. Ces tables contiennent pour chaque individu des informations démographiques (sexe, âge...) et professionnelles (corps, grade, ministère, département de fonction...). Une table contenant deux identifiants (ident_s dans DADS ou id_nir dans SIASP et l'identifiant de la fiche adresse dans l'enquête ident_fa) ainsi que les poids de sondage nous a été transmise par l'Insee et nous a permis de fusionner la base de tirage avec les bases de l'enquête.
- La table TCM_MEN qui contient les réponses au tronc commun des ménages des enquêtés.
- La table INDIVIDUS_AO qui contient l'ensemble des réponses à l'enquête.

Lors de la fusion de ces tables il est à noter que :

- 1 individu de l'échantillon fonction publique n'était pas présent dans la table GESTION et n'a donc jamais été enquêté bien qu'il ait été tiré.
- 3 cas de doublons dans l'échantillon FP se sont présentés avec plusieurs individus référencés pour une seule fiche adresse (rappelons que pour les extensions, contrairement au tronc commun, le tirage est effectué au sein d'une base d'individus et non de logements, en principe il y a donc un seul individu enquêté par logement). Un seul individu a été gardé par fiche adresse. Nous avons identifié l'individu à conserver pour chaque fiche adresse (celui qui avait été tiré dans la base de sondage) grâce son prénom qui figurait à la fois dans la base de tirage et dans la table TCM.
- 1 cas de doublon où un individu était présent dans deux sous-échantillons. Cet individu était également un doublon fiche adresse et n'aurait pas dû être enquêté ; nous ne l'avons donc pas conservé.

La définition des individus hors-champ

Le champ de l'enquête est constitué de l'ensemble des logements ordinaires utilisés au titre de résidence principale. Comme pour l'échantillon entrant principal, les logements détruits, vacants ou durablement inoccupés sont donc exclus du champ. Sont également exclus du champ les individus qui sont devenus inactifs et ceux qui sont passés dans le secteur privé, hors cliniques privées, (Statutext > 5) et qui sont donc sortis du champ des extensions.

Sont considérés comme hors-champ les cas suivants :

- les codes résultats RES2 = ("AHC", "INC", "ALD", "NT1", "BIS", "ENA", "DCD"),
- les individus devenus inactifs (RES2 = "SAO", Actoccup = 0) ou ayant quitté la fonction publique ou le secteur hospitalier privé (Statutext > 5).

Tableau III.10 : Répartition des individus hors-champ selon le code RES2 dans les échantillons d'extension

RES2		Effectifs échantillon FP	Répartition FP	Effectifs échantillon hôpitaux privés	Répartition hôpitaux privés
ALD	Absent longue durée	7	5 %	2	5 %
BIS	Déjà interrogé sur cette enquête	8	6 %	8	19 %
DCD	Décédé	4	3 %	0	0 %
ENA	Enquête non attribuée	7	5 %	2	5 %
INC	Inconnu	3	2 %	0	0 %
Actifs sortis du champ de extensions, passage dans le secteur privé hors cliniques (statutext > 5)		30	23 %	23%	36 %
Devenu inactif (actoccup = 0)		69	54 %	54 %	36 %
Total		128	100 %	43	100 %

Au total, 128 individus sont hors-champ dans l'échantillon d'extension fonction publique (secteur hospitalier public compris) et 43 dans le secteur hospitalier privé (tableau III.10). A noter que 49 % d'entre eux sont des individus devenus inactifs. Les résultats sont sensiblement similaires dans l'échantillon hôpitaux privés.

Nous corrigeons simplement la non-réponse totale des individus appartenant au champ de l'enquête (1 314 individus pour les extensions fonction publique et 322 pour l'échantillon des cliniques privées).

iii. Correction de la non-réponse

Définition des individus non-répondants

Contrairement à l'échantillon principal où l'on considère séparément non-réponse ménage et non-réponse individuelle, la correction de la non-réponse dans les échantillons d'extension se fait en une seule étape car chaque fiche adresse correspond à un individu.

Sont donc considérés comme non-répondants :

- les individus dont le code RES2 indique un évitement, un refus, un individu impossible à joindre, etc.... ("IAJ", "EVT", "IMP", "DEJ", "NT2", "NT3", "REF", "THV", "SAC"),
- les individus pour lesquels la variable « statut » ou la variable « diplôme » sont vides (au moins l'une des deux).

Tableau III.11 : Répartition des individus non-répondants selon le code RES2 et leurs réponses aux variables « statut » et « diplôme » dans les échantillons d'extension

RES2		Effectifs échantillon FP	Répartition FP	Effectifs échantillon hôpitaux privés	Répartition hôpitaux privés
DEJ	Déjà enquêté pour une autre enquête	13	2 %	2	1 %
EVT	Évitement	45	8 %	10	7 %
IAJ	Impossible à joindre	91	16 %	16	11 %
IMP	Impossible (personne inapte ou non habilitée à répondre)	19	3 %	3	2 %
NT2	Non traitée (mais le repérage a pu être fait)	36	7 %	8	5 %
NT3	Non traitée (mais le contact a été pris)	4	1 %	0	0 %
REF	Refus	71	13 %	26	17 %
SAC	Sans adresse connue	271	49 %	83	56 %
Statut vide et diplôme vide		0	0 %	0	0 %
Diplôme vide (et statut renseigné)		3	1 %	1	1 %
Total		553	100 %	149	100 %

553 individus sont considérés comme non-répondants et 761 sont répondants dans l'échantillon d'extension fonction publique (tableau III.11). 149 individus sont considérés comme non-répondants et 173 sont répondants dans l'échantillon d'extension hôpitaux privés.

12 % des individus échantillonnés dans l'échantillon d'extension hôpitaux privés sont considérés comme hors champ, ce qui est plus élevé que l'échantillon fonction publique (9 %). Étant donné que l'on considère comme hors champ les individus devenus inactifs ainsi que ceux qui sont passés dans le privé, les taux de hors-champ des extensions sont mécaniquement plus élevés que pour l'échantillon principal. L'échantillon entrants principal compte 16,3 % de logements hors-champ. Le taux de non-réponse s'élève à 38 % dans l'échantillon FP et 41 % dans l'échantillon hôpitaux privés, ce qui est moins bon que pour l'échantillon principal, pour lequel on considère la non-réponse ménage (taux de non-réponse ménage : 19,5 %) et individu (taux de non-réponse individu : 5,7 %).

Tableau III.12 : Répartition des individus des deux échantillons d'extension selon leur type de réponse

	Effectifs échantillon FP	Répartition FP (en %)	Effectifs échantillon hôpitaux privés	Répartition hôpitaux privés (en %)
Hors-champ	128	8 %	43	12 %
Non-réponse	553	39 %	149	41 %
Réponse	762	53 %	173	47 %
Total	1 429	100 %	365	100 %

Correction de la non-réponse : les variables qui expliquent la non-réponse

Nous avons réalisé des régressions logistiques sur le fait d'être non-répondant à partir des données disponibles dans les bases de tirage (SIASP : sexe, tranche d'âge, région, corps de la FP, statut ; DADS : tranche d'âge, catégorie professionnelle, zone urbaine ou rurale,). Les variables issues du repérage des enquêteurs (type de logement, etc.) sont trop souvent manquantes sur les sous-échantillons 12 à 16 pour être utilisées.

Les deux régressions sont réalisées sur les individus considérés comme étant dans le champ, répondants et non-répondants, soit 1 314 individus pour l'échantillon fonction publique et 322 dans l'échantillon cliniques privées.

Pour l'échantillon fonction publique, les variables discriminantes sont presque les mêmes que pour CT 2013 : l'âge, la région, le statut et le sexe (tableau III.13). Le type de logement n'a pas pu être exploité comme cela avait été fait pour CT 2013 et pour l'échantillon principal car ces données n'existaient pas pour les extensions.

Pour l'échantillon salariés des cliniques privées, différentes régressions logistiques sur le fait d'être non-répondant sur l'ensemble des individus dans le champ (c'est-à-dire les répondants et les non-répondants) ont été mises en place mais les résultats n'étaient pas significatifs. Il n'a donc pas été possible de recourir à des régressions logistiques pour construire des groupes d'individus au comportement de réponse homogène. Ceci peut s'expliquer par la taille plutôt petite de l'échantillon des extensions cliniques privées et par le fait que le comportement de réponse s'explique par des variables inobservées.

Tableau III.13 : Variables explicatives de la non-réponse dans l'échantillon d'extension fonction publique

Variables explicatives	Modalités	Significativité	Odd-ratios	Taux de non-réponse (en %)
SEXE	Hommes	**	0.640	48,2
	Femmes	Réf.	Réf.	39,1
Classe d'âge	50 ans et plus	**	1.361	29,7
	moins 30 ans	***	0.552	53,4
	30-49 ans	Réf.	Réf.	36,5
Région	Ile-de-France	***	0.389	66,5
	Nord, Est et Ouest	***	1.917	28,5
	Autres	Réf.	Réf.	42,8
Corps de la FP	FPE	NS	0.992	36,9
	FPH	**	0.678	46,3
	FPT			42,1
Statut	Non titulaires	**	0.702	46,6
	Titulaires	Réf.	Réf.	51,8

*** Significatif à 1 % ** Significatif à 5 % * Significatif à 10 %

Construction de groupes homogène de réponse à partir de l'algorithme CHAID et la macro SAS Treedisc (Insee)

Pour *l'échantillon d'extension fonction publique* nous avons comparé plusieurs modèles différents en faisant varier les paramètres de la macro Treedisc entre une taille minimum de classe de 50 à 150 et un seuil de significativité de 0.1 à 0.2. Le modèle que nous avons choisi était celui présentant la plus faible déformation des poids de sondage initiaux (la taille minimum des classes est de 100 et le seuil de significativité de 0.1). Il conduit à la constitution de 10 groupes homogènes de réponse (tableau III.14 et graphique III.7).

Pour *l'échantillon des salariés des cliniques privées* différentes valeurs pour chacun des paramètres ont été testées mais la petite taille de l'échantillon réduit fortement les marges de manœuvre. Le nombre de classes et leur composition étant robustes – presque invariantes - au choix de ces paramètres, nous retenons une position intermédiaire avec une taille minimale des feuilles à 45 et un alpha à 0,10. Comme le nombre de variables et leur corrélation potentielle n'ont pas d'impact sur les sorties de l'algorithme, nous utilisons toutes les variables à disposition (utilisées dans les régressions logistiques précédentes).

Nous obtenons avec l'algorithme CHAID deux classes homogènes d'estimation du comportement de réponses. Cette estimation conduit à dilater les poids d'un facteur 1,4 pour 46 répondants, d'un facteur de 1,8 pour les 161 autres. La forme de la distribution des poids après correction de la non-réponse est donc très proche de celle de la distribution des poids de sondage (tableau III.16 et graphique III.8).

Tableau III.14 : Poids avant et après correction de la non-réponse dans l'échantillon FP par GRH

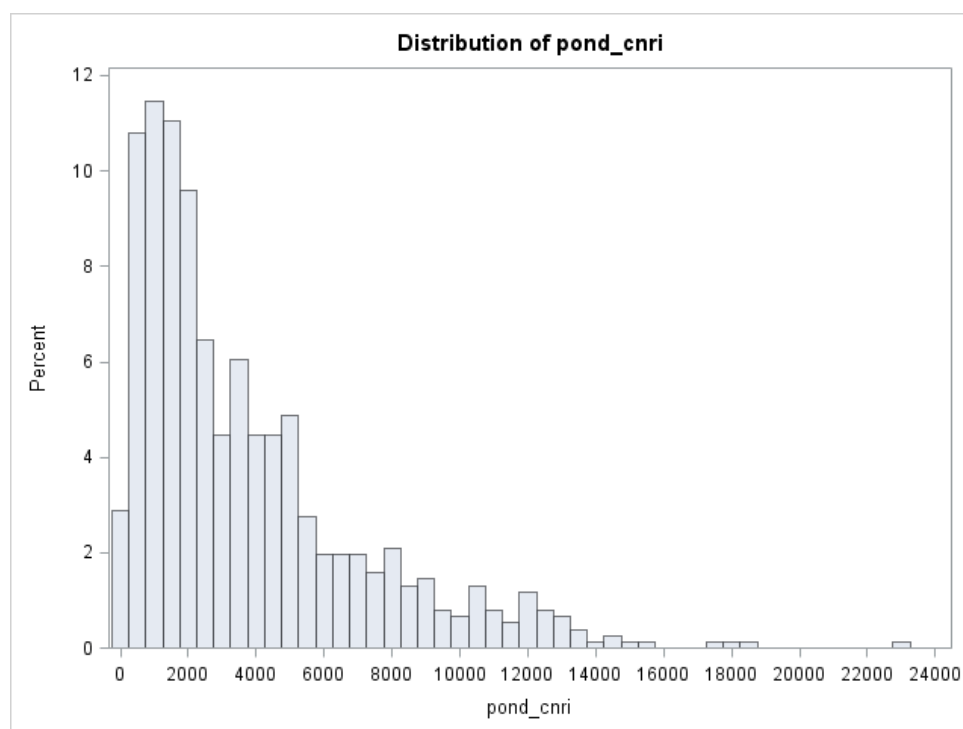
GRH	Population	Effectifs	Taux de réponse (%)	Moy	Q1	P50	Q3	Min	Max	Somme
1	Sondés (poids_sond)	89	69,7	3 372	1 581	2 866	4 398	93	10 870	300 146
	Répondants (pond_cnri)	62		4 859	2 389	4 037	6 313	256	14 906	301 256
2	Sondés (poids_sond)	49	40,8	627	309	460	832	89	2 793	30 726
	Répondants (pond_cnri)	20		1 737	768	1 173	1 837	219	6 843	34 737
3	Sondés (poids_sond)	171	85,4	2 968	1 363	2 545	4 174	16	11 454	507 545
	Répondants (pond_cnri)	146		3 398	1 447	2 898	4 887	19	13 415	496 132
4	Sondés (poids_sond)	63	73,0	2 752	1 211	2 735	3 668	18	7 589	173 402
	Répondants (pond_cnri)	46		3 957	2 247	3 918	5 148	24	10 393	182 029
5	Sondés (poids_sond)	176	46,0	2 996	1 441	2 435	4 012	94	10 694	527 280
	Répondants (pond_cnri)	81		6 950	3 230	5 622	10 218	793	23 236	562 937
6	Sondés (poids_sond)	193	65,8	3 423	1 232	2 906	4 112	27	61 828	660 651
	Répondants (pond_cnri)	127		5 641	1 939	4 448	6 731	41	93 958	716 373
7	Sondés (poids_sond)	94	57,4	3 307	1 579	2 753	4 817	76	8 404	310 860
	Répondants (pond_cnri)	54		5 723	2 857	5 329	8 265	132	13 104	309 058
8	Sondés (poids_sond)	108	72,2	800	366	614	1 086	25	4 655	86 397
	Répondants (pond_cnri)	78		1 019	466	848	1 500	34	3 286	79 478
9	Sondés (poids_sond)	167	48,5	695	337	587	888	14	3 455	116 014
	Répondants (pond_cnri)	81		1 439	647	1 179	1 946	191	6 271	116 548
10	Sondés (poids_sond)	202	32,7	674	369	625	926	6	2 232	136 218
	Répondants (pond_cnri)	66		1 918	1 014	1 941	2 458	184	4 941	126 616

Tableau III.15 : Poids des répondants avant et après correction de la non-réponse dans l'échantillon FP par sous-échantillon

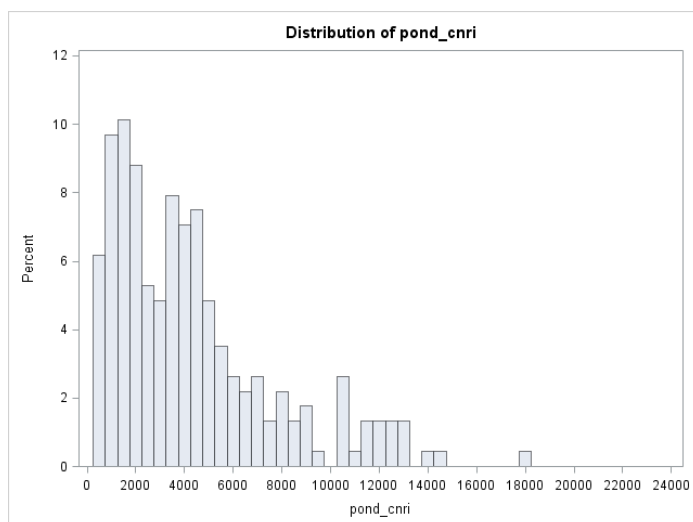
Sous-échantillon	N	Variable	Somme	Moyenne	Minimum	Maximum	Ecart-type	Q3/Q1	D9/D1
FPE	269	Poids de sondage	830 365	3 086,9	159,1	11 462,0	2 371,2	3,7	11,0
		Poids corrigé NR	1 314 059	4 885,0	255,1	23 236,4	3 777,8	3,4	8,8
FPT	228	Poids de sondage	675 474	2 962,6	189,4	61 827,6	4 433,1	4,0	11,8
		Poids corrigé NR	1 077 458	4 725,7	262,3	93 958,4	6 832,5	3,4	11,5
FPH Hôpitaux	249	Poids de sondage	333 380	1 338,9	16,0	11 454,2	1 456,6	4,3	20,5
		Poids corrigé NR	527 080	2 116,8	18,8	13 415,5	2 138,7	3,4	13,3
FPH Autres	15	Poids de sondage	4 650	310,0	43,5	898,0	277,5	3,1	11,7
		Poids corrigé NR	6 568	437,9	60,2	1 212,2	358,5	2,8	8,0
Ensemble de la FP	761	Poids de sondage	1 843 869	2 423,0	16,0	61 827,6	3 043,5	5,2	16,8
		Poids corrigé NR	2 925 164	3 843,8	18,8	93 958,4	4 721,7	4,0	14,3

Graphique III.7 : Distribution des poids corrigés de la non-réponse par sous-échantillon de l'extension fonction publique (hors poids maximum)

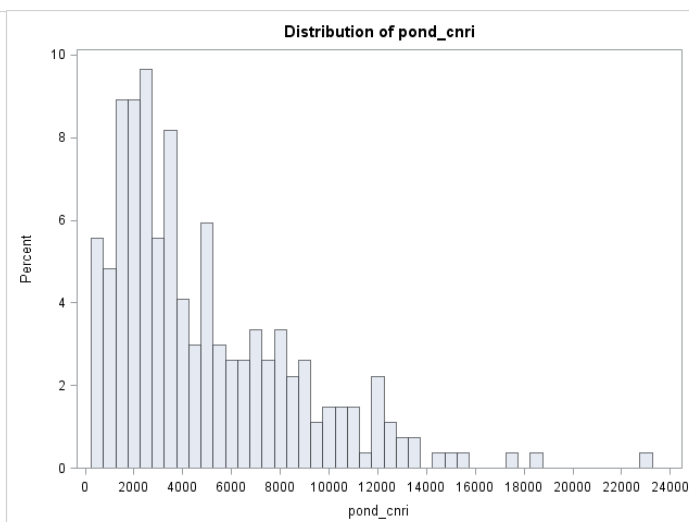
Ensemble de la fonction publique



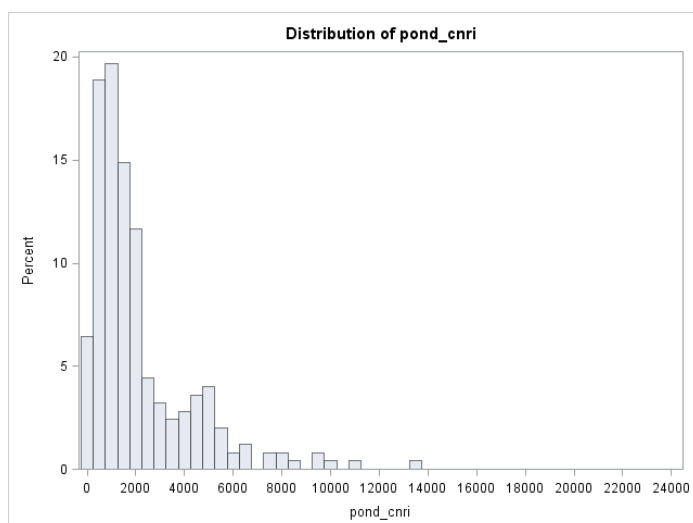
Fonction publique d'Etat



Fonction publique territoriale



Fonction publique hospitalière – Hôpitaux



Fonction publique hospitalière - Autres

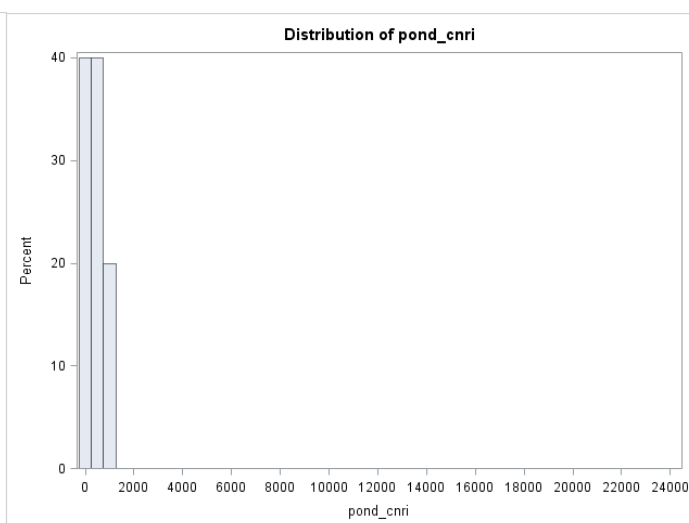
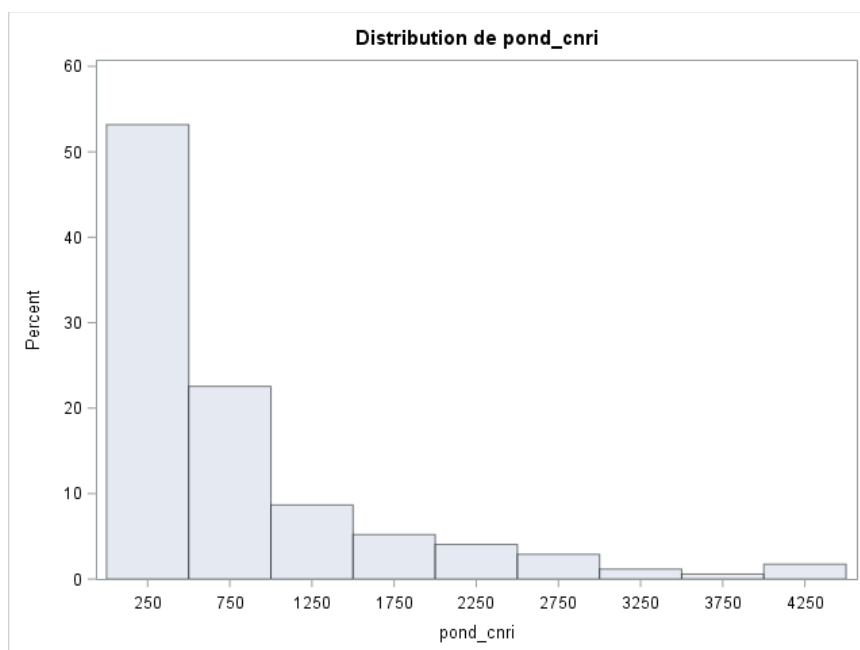


Tableau III.16 : Description des poids de sondage après correction de la non-réponse totale dans l'échantillon des cliniques privées

Sous-échantillon	N	Variable	Somme	Moyenne	Minimum	Maximum	Ecart-type	Q3/Q1	D9/D1
Cliniques privées	173	Poids de sondage	79326	459	12	2354	517	4,6	19,3
		Poids corrigé NR	136029	786	21	4270	881	4,5	22,7

Graphique III.8 : Distribution des poids de sondage après correction de la non-réponse totale dans l'échantillon des cliniques privées



4- Effet de la correction de la non-réponse sur quelques indicateurs des conditions de travail

Nous testons la robustesse de notre repondération en construisant les tableaux de fréquence de certaines variables de l'enquête (tableau III.17).

Tableau III.17 : Effet correction de la non-réponse sur quelques indicateurs des conditions de travail dans l'échantillon dans la fonction publique et dans les cliniques privées (15)

	Proportion en (%)	
	Poids bruts	Poids corrigés de la non-réponse totale, avant partage des poids, avant calage
Fonction publique (ssech 12, 13, 14, 16)		
Porter ou déplacer des charges lourdes	33,0	32,7
Rester longtemps debout	51,1	49,7
Au moins 3 contraintes de rythme	23,7	24,2
Craindre pour son emploi	13,7	13,3
Travailler le dimanche	18,6	17,7
Au moins un accident du travail	8,4	9,0
Penser à trop de choses à la fois	49,6	49,4
Cliniques privées (ssech 15)		
Porter ou déplacer des charges lourdes	65,2	65,5
Rester longtemps debout	70,3	70,7

Au moins 3 contraintes de rythme	38,2	37,4
Craindre pour son emploi	23,9	23,8
Travailler le dimanche	35,0	34,2
Au moins un accident du travail	6,5	6,3
Penser à trop de choses à la fois	64,4	65,8

3. Calage sur marges et partage des poids

i. Partage des poids

Le partage des poids est nécessaire dans les enquêtes avec plusieurs échantillons lorsque ceux-ci ne sont pas disjoints et que donc des individus sont présents dans différents échantillons constitutifs de la même enquête. Sans que cela se soit effectivement produit, ces individus auraient pu être tirés avec une probabilité non nulle dans plusieurs échantillons et sont donc « comptés » plusieurs fois dans les poids de tirage. Pour eux, il faut procéder au partage des poids et corriger du nombre de fois où un individu aurait pu être sélectionné (voir Favre-Martinoz et Gros 2017 pour une présentation de la méthode du partage des poids). Le cas des doublons physiques, effectivement échantillonné au titre de deux sous-échantillons, bien que rare, doit également être traité.

Il y a dans le plan de sondage de l'enquête CT-RPS 2016 plusieurs sources de recouvrement de champs :

- Les individus de l'échantillon principal (panel ou entrants) éligibles aux extensions en 2013 et/ou en 2016.
- Les individus entrants qui étaient éligibles à l'enquête CT en 2013 et auraient pu être dans le panel.
- Les salariés du périmètre DREES qui cumulent un emploi dans le privé et dans le public (et auraient pu être tirés dans SIASP et dans les DADS).

Etant donné le nombre de sources de recouvrement de champs, la méthode la plus simple du décompte des liens (ici le nombre de bases de sondages dans lesquelles les individus auraient pu être tirés) a été retenue. Elle est sans biais (Deville, Lavallée 2006).

Calage

L'objectif est d'obtenir à l'issue du calage une pondération qui permette d'estimer certaines structures connues sur la population complète des actifs occupés de 15 ans ou plus au 31 décembre 2015 en métropole. De plus, comme en 2013, on souhaite que l'échantillon final décrive deux populations distinctes :

- Les actifs du secteur privé (données 2015 de l'Enquête Emploi en Continu, EEC2015)
- Les salariés du secteur public (SIASP2015 : répondants ayant un statut 1, 2, 3, 5).

Articulation des deux

Il est d'usage de réaliser le partage des poids avant le calage sur marges. Pour le redressement de CT 2013, le calage a en effet été réalisé sur les poids corrigés de la non-réponse et après partage des poids. Pour le redressement de CT-RPS 2016, cette méthode a été néanmoins écartée au profit d'un calage sur marges adapté au partage des poids (chapitre VII du livre de référence de Pierre Lavallée sur le sujet (Indirect Sampling, Springer, 2007).

En effet, après correction de la non-réponse individuelle, il persiste d'importantes différences de profil entre les répondants du panel et les entrants qui étaient actifs occupés en 2013 (dans le champ de CT 2013), alors que, une fois la non-réponse prise en compte, ces deux sous-populations sont en théorie comparables (cf. tableau III.18).

Tableau III.18 : caractéristiques des répondants de l'échantillon panel principal (01) et de ceux de l'échantillon entrant principal (11) dans le champ de CT 2013, après correction de la non-réponse individuelle

		Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*
Diplôme	Aucun diplôme	6,7	8,2
	CEP ou diplôme étranger de même niveau	1,8	2,5
	Brevet des collèges	4,6	5,5
	CAP, BEP	24,5	26,8
	Baccalauréat technologique ou professionnel	11,8	11,9
	Baccalauréat général	5,8	6,5
	Diplôme de niveau Bac+2	14,9	13,9
	Diplôme de niveau bac +3 ou bac +4	14,6	12,8
	Diplôme de niveau supérieur à bac + 4	15,3	11,9
	NA	0,1	0,1
Catégorie socio-professionnelle	Agriculteurs	2,3	2,9
	Artisans, commerçants	6,4	6,5
	Cadres	22,0	18,1
	Professions intermédiaires	25,5	26,1
	Employés	26,4	28,1
	Ouvriers	17,4	18,4
Statut	FPE	11,6	10,4
	FPT	7,9	7,6
	FPH	4,1	5,2
	Privés	63,1	64,2
	Non-salariés	13,4	12,7
Région	Île-de-France	13,1	10,6
	Métropole hors IDF	86,9	89,5

* actifs occupés au 31 décembre 2012, nés avant 1998 (15 ans au moment de l'enquête CT 2013) ou arrivés sur le territoire français avant 2013.

Lors du redressement de CT 2013, le calage sur marges semble avoir joué un rôle important, la correction de la non-réponse n'ayant pas permis de contrôler la non-réponse liée à la catégorie sociale, au secteur, au sexe dans la fonction publique (voir Algava et al., 2015). Ces variables sont en effet relativement peu précises dans Octopusse car connues uniquement au niveau du logement. Or, les poids initiaux pour les individus panel sont les poids de tirage de CT 2013 après correction de la non-réponse en 2013, avant partage des poids et avant calage. Un calage sur marges de CT-RPS 2016 uniquement sur les marges de la population des actifs occupés au 31 décembre 2015 et, pour les

individus du panel, à partir de poids imparfaitement corrigés de la non-réponse de 2013 ne serait donc pas suffisant pour ajuster les biais résiduels de la non-réponse en 2013.

Dans le cadre d'une enquête panel, l'approche « traditionnelle » présente en effet comme limite de ne pas mobiliser toute l'information auxiliaire disponible pour chacun des sous-échantillons. Au contraire, la méthode du partage du calage sur marges adapté au partage des poids (Lavallée, 2007) permet de mobiliser l'ensemble de l'information auxiliaire disponible sur les échantillons pris séparément à savoir :

- la structure des actifs occupés au 31 décembre 2012 pour l'échantillon panel,
- la structure des actifs occupés au 31 décembre 2015 pour l'échantillon des entrants principal .

Cette méthode permet de caler chaque sous-échantillon sur des marges qui lui sont propres¹⁵, en plus du calage de l'ensemble des répondants sur les marges de la population cible. Le calage est réalisé sur les poids corrigés de la non-réponse et, pour le calage de l'ensemble des répondants actifs occupés sur des **variables transformées** (voir encadré III.1). Le partage des poids est appliqué en dernière étape, sur les poids calés.

Encadré III.1 : Principe du calage sur marges adapté au partage des poids

Soit Y une variable d'intérêt, le nombre total de femmes par exemple.

Lors du calage sur marges, on cherche les poids finaux W_i tels que, après calage, on ait :

$$W_i \text{ tels que } \sum_i W_i 1_{\text{femme}} = \text{Nombre de femmes}$$

Le partage des poids est effectué après le calage sur marges. Pour les marges pour les actifs occupés au 31 décembre 2015, il faut donc en tenir compte lors du calage pour s'assurer qu'à l'issue du redressement, pour chaque marge on ait bien :

$$\sum_{i \in S} W_i X_i = N_F$$

La solution tirée de Lavallée (2007) est, dans un premier temps, de réaliser le calage sur une variable transformée, puis d'appliquer le partage des poids sur le poids obtenu à l'issue du calage.

Soit $X_i = 1$ si femme, 0 sinon

Soit W_i le poids final après calage et partage des poids.

On cherche les poids W_i tels que $\sum_{i \in S} W_i X_i = N_F$ (1)

où W_i est obtenu par la formule

$$W_i = \frac{\sum_{k=1}^K 1_{(i \in S_k)} \pi_{ik}}{\sum_{k=1}^K 1_{(i \in U_k)}}$$

Le poids final est égal à la somme des poids de sondage π_{ik} de l'individu i dans chacun des échantillons où il apparaît divisé par le nombre d'échantillons dans lesquels l'individu i aurait pu figurer.

A l'issue du redressement, on veut :

- que l'équation de calage (1) soit vérifiée ;
- caler chaque échantillon S_k sur des marges qui lui soient spécifiques.

On va donc réaliser le calage sur le fichier obtenu en empilant les échantillons S_k .

¹⁵ Les sous-échantillons des extensions n'ont pas été inclus, ce qui a permis de ne pas avoir à tenir compte du partage des poids inhérent à l'enquête CT2013.

Pour chacune de ces lignes, on définit la variable \tilde{X}_{ik}

$$\tilde{X}_{ik} = \frac{X_i 1(i \in S_k)}{L_i}$$

On cherche les poids calés \tilde{d}_{ik} tels que

$$\sum_{k=1}^K \sum_{i \in S} \tilde{d}_{ik} \tilde{X}_{ik} = N_F \Leftrightarrow \sum_{k=1}^K \sum_{i \in S} \tilde{d}_{ik} \frac{X_i 1(i \in S_k)}{L_i} = N_F$$

Le poids final partagé calé s'écrit :

$$\tilde{W}_i = \frac{\sum_{k=1}^K \tilde{d}_{ik} 1(i \in S_k)}{L_i}$$

On peut vérifier que les poids finaux calés et partagés vérifient bien les marges. On calcule l'estimateur du nombre de femmes :

$$\sum_{i \in S} \tilde{W}_i X_i = \sum_{k=1}^K \sum_{i \in S_k} \tilde{d}_{ik} \frac{1(i \in S_k)}{L_i} X_i = \sum_{k=1}^K \sum_{i \in S_k} \tilde{d}_{ik} \tilde{X}_{ik} = N_F$$

Cette méthode garantit, qu'après partage des poids, l'échantillon complet est calé sur les marges relatives aux actifs occupés au 31 décembre 2015.

Le sous-échantillon panel est calé sur les marges de la population active en 2013. On inclut donc les répondants panel inactifs lors de CT-RPS 2016, car ils étaient tous actifs occupés en 2013.

	Variables calées sur les marges de l'ensemble des répondants	Variables calées sur les marges propres aux sous-échantillons	
	Variables communes tous les échantillons AO 2016 secteur privé et FP	Variables spécifiques entrants AO 2016	Variables spécifiques panel AO 2013
Panel			
Actifs occupés	$\tilde{X}_{ik} = \frac{X_{i(k)}}{L_{i(k)}}$	0	$X_{i(k)}$
Inactifs	0	0	$X_{i(k)}$
Entrants	$\tilde{X}_{ik} = \frac{X_{i(k)}}{L_{i(k)}}$	$X_{i(k)}$	0

Au final, cette méthode a permis de réduire les écarts importants entre individus panels (échantillon principal) et entrants actifs occupés en 2013 (échantillon principal) après correction de la non-réponse (voir partie III.5, tableau III.26).

ii. Calcul du nombre de liens pour le partage des poids

On cherche à dénombrer les liens avec les bases de sondages pour tous les répondants à CT-RPS 2016, qu'ils soient panel ou non.

a) Prise en compte de la dimension panel

Comme dans les redressements de l'enquête SRCV (Ardilly, 2010), pour chaque individu i (panel ou entrant, principal ou extension), le nombre de liens avec l'échantillon principal est égal au nombre d'éditions de l'enquête durant lesquelles i se trouve dans le champ¹⁶.

L'enquête CT-RPS 2016 est composée des actifs occupés panel de CT 2013 et des répondants entrants (qui sont par construction des actifs occupés). L'enquête CT-RPS contient des jeunes actifs occupés entre 15 et 18 ans (exclus). Ces jeunes ne peuvent appartenir qu'à l'échantillon des entrants, car les individus panel ont au moins 18 ans en 2016 (car ils avaient 15 ans ou plus en 2013). Cependant, les 18 ans et plus peuvent provenir du panel comme du sous-échantillon des entrants.

Pour les individus du panel, par définition de l'enquête transversale (individu panel actif occupé), le nombre de liens au minimum égal à 2.

Pour les individus de l'échantillon entrant, la question STATUTANT (situation vis-à-vis de l'emploi et statut au 1^{er} décembre 2012) permet de définir s'ils étaient actifs occupés en 2013 ; les questions ANAIS et ANARRIV s'ils avaient l'âge pour participer à l'enquête et s'ils résidaient déjà en France pour les immigrés.

Ont donc un lien supplémentaire les entrants qui vérifient les trois conditions suivantes :

- Actif occupé en 2013 (statuant ne '11','12','13)
706 entrants non actifs occupés au 31 décembre 2012
- 15 ans ou plus au moment de l'enquête CT 2013 (anais < 1998)
8 entrants sont trop jeunes
- Résidant en France au moment de l'enquête (anarriv < 2013)
22 entrants sont écartés

Au final, **4 086** entrants auraient pu être tirés au sort en 2013 et être répondant panel.

b) Prise en compte du recouvrement des différents échantillons extension

Les échantillons d'extensions réalisent une partition, les frontières entre les échantillons d'extensions sont en effet définies à partir des catégories juridiques de l'employeur. Un cas théorique de recouvrement entre FPE ou FPT d'un côté et cliniques privées de l'autre viendrait d'un cas de cumul de deux emplois simultanément, l'un dans le champ FPE ou FPT, l'un dans le champ des cliniques privées. L'appariement entre SIAPS et les DADS pour l'échantillon de ces champs permet néanmoins de conclure il n'y a aucun doublon entre FPE/cliniques privées et FPT/cliniques privées. La probabilité d'être tiré dans deux échantillons d'extensions pour un individu est nulle donc le partage des poids n'est pas nécessaire.

c) Prise en compte du recouvrement de l'échantillon principal et des échantillons extensions

- Prise en compte du recouvrement des échantillons principal et extension en 2013 pour les individus du panel

¹⁶ En l'absence de questions sur le lieu et le type de résidence lors des enquêtes précédentes, on néglige aussi le fait d'être dans le champ des enquêtes ménages de l'Insee (logements ordinaires, résidence principale en France métropolitaine ou Guadeloupe, Martinique, Guyane, Réunion). On néglige ici les déménagements entre la métropole et les DOM.

Pour les individus du panel il est nécessaire de procéder à un partage des poids différent de celui pour CT 2013, puisque le nombre de leurs liens a augmenté (ils auraient tous pu être tirés dans l'échantillon 11 et, pour certains, être tirés dans une extension).

Pour déterminer si le répondant faisait partie du champ d'une extension en 2013 on repart de ses réponses en 2013.

Ont un lien supplémentaire les individus panel de l'échantillon principal 01¹⁷:

- 1) qui étaient dans le champ d'un échantillon d'extension (STATUT=1 à 5)
- 2) en fonction en France métropolitaine,
- 3) qui n'étaient pas militaires de la FPE.

Ce sont des individus pour lesquels soit le libellé de profession contient un des termes clés (« milit », « armée », « caporal », etc.) soit pour lesquels le code profession est "334A", "452B", "532A", "532B" ou "532C"

- 4) qui avaient entre 17 et 62 ans en 2010 (âge révolu au 31/12/2010),

Cela revient à enlever les enquêtés nés avant 1948 ou après 1993.

- 5) qui n'étaient pas en emploi aidé (échantillons DGAFP),

6) qui n'étaient ni des élus, ni des maîtres du privé, ni des élèves stagiaires de la FPH (qui n'ont pas statut de stagiaire de la fonction publique, les internes et résidents sont en revanche inclus), (échantillon DGAFP)

- 7) qui n'étaient ni étudiants, ni stagiaires, ni intérimaires, ni apprentis ni en emploi aidé (échantillons DREES, TYPEEMPLOI et TITPUB),

- 8) en poste au 31 décembre 2010

On enlève tous ceux pour lesquels DATANT>2010 donc 399 en tout.

- 9) qui occupaient un poste principal non annexe au 31/12/2010 (échantillon DGAFP).

On peut supposer que tous les actifs occupés de l'échantillon principal occupent un poste principal. On suppose qu'une arrivée antérieure à 2010 signale en 2012-2013 un emploi non annexe, sauf si le nombre d'heures habituellement travaillées chaque semaine est inférieur à 7,5 (soit au moins 1,5 heures pendant 5 jours). »

- Prise en compte du recouvrement des échantillons entrants en 2016 avec celui des échantillons extension en 2013

Seuls sont considérés ici les individus appartenant au champ de l'enquête de 2013. La question STATUTANT (situation vis-à-vis de l'emploi et statut au 1^{er} décembre 2012) et l'année de naissance permet de définir si les individus entrants étaient dans le champ d'une extension non (STATUTANT = 1 à 5). En l'absence d'informations supplémentaires sur l'emploi occupé en décembre 2012, il n'est pas possible de se placer plus près des caractéristiques de tirage et retirer les personnes hors du tirage (militaires, élus, emplois aidés, département d'outre-mer, etc.)

Ont un lien supplémentaire les individus des sous-échantillons entrants qui étaient dans le champ de l'enquête CT 2013 et qui :

- 1) étaient dans le champ d'un échantillon d'extension (STATUTANT=1 à 5) ;
c'est le cas de 1632 entrants (4 entrants dans le champ des extensions en 2013 ne sont pas dans le champ de l'enquête CT 2013).
- 2) avaient entre 17 et 62 ans en 2010 (âge révolu au 31/12/2010), ce qui revient à enlever les enquêtés nés avant 1948 ou après 1993.
On retire 2 salariés.

Au total, il y a 1 630 entrants enquêtés auraient pu être échantillonnés au titre des extensions de 2013, 846 de l'échantillon entrant principal (11) et 784 des échantillons entrants extensions.

- Prise en compte du recouvrement des échantillons panel et entrant principal avec celui des extensions en 2016

L'échantillon des actifs occupés panel (ssech 01 à 06) et l'échantillon principal 11 contiennent toutes sortes d'individus et donc il faut déterminer pour chaque individu s'il aurait pu être tiré dans un

¹⁷ Voir DT Algava et al. pour le détail du champ des extensions.

échantillon d'extensions de 2016. L'idée est d'être au plus près possible des caractéristiques de tirage. Il est néanmoins impossible de retrouver les individus de l'échantillon principal qui étaient dans le champ des extensions en décembre 2014 au moment du tirage mais n'y sont plus.

Ont un lien supplémentaire les individus qui lors de la collecte de CT-RPS 2016 :

1) étaient dans le champ d'un échantillon d'extension au moment de l'enquête (STATUT=1 à 5)

9524 salariés sont dans ce cas.

2) en fonction en France métropolitaine

37 travaillent à l'étranger et 477 dans un DOM (6 non-réponse)

3) qui n'étaient pas militaires de la FPE¹⁸.

Ce sont des individus qui soit déclarent être militaire (TITPUBR = « 3 – militaires »), soit pour lesquels le code profession est "334A", "452B", "532A", "532B" ou "532C"

Il y en a 116.

4) qui avaient entre 17 et 62 ans en 2015 (âge révolu au 31/12/2014)¹⁹,

Cela revient à enlever 53 enquêtés nés en 1952 ou avant, ou en 1999 ou après.

5) qui n'étaient pas en emploi aidé (échantillons DGAFP),

Il y en a 8

6) qui n'étaient ni des élus, ni des maîtres du privé, ni des élèves stagiaires de la FPH (qui n'ont pas statut de stagiaire de la fonction publique, les internes et résidents sont en revanche inclus), (échantillon DGAFP)

On réussit à identifier 5 élus (code profession « 335A » et libellé ne mentionnant pas d'activité syndicale)

Il est impossible de repérer les autres situations, on considère qu'il n'y en a pas.

7) qui n'étaient ni étudiants, ni stagiaires, ni intérimaires, ni apprentis ni en emploi aidé (échantillons DREES, TYPEMPLOI et TITPUB)

On enlève 23 salariés apprentis, intérimaires, stagiaires, en emploi aidés ou sans contrat.

8) en poste au 31 décembre 2014²⁰

On enlève tous ceux pour lesquels DATANT>2014 donc 207 en tout.

Si en poste au 31 décembre 2013, on enlève tous ceux pour lesquels DATANT>2013 donc 207+162 en tout.

9) qui occupaient un poste principal non annexe au 31/12/2013 (échantillon DGAFP).

On peut supposer que tous les actifs occupés de l'échantillon principal occupent un poste principal²¹. On suppose qu'une arrivée antérieure à 2013 signale en 2015-2016 un emploi non annexe²², sauf si le nombre d'heures habituellement travaillées chaque semaine est inférieur à 7,5 (soit au moins 1,5 heures pendant 5 jours). »

46 salariés sont concernés

Au final, 8 603 enquêtés auraient pu être échantillonnés au titre des extensions de 2016, 3 567 individus du panel principal, 4 290 individus du panel extension et 746 entrants de l'échantillon principal.

¹⁸ En effet, dans l'échantillon DGAFP, seuls des postes provenant des fichiers de paie des agents de l'Etat et des DADS ont été sélectionnés, pour garantir la possibilité de reconstituer l'adresse des salariés sélectionnés. Cela conduit à exclure la quasi-totalité des militaires de la FPE, et 96 % des militaires de la FPT (ainsi que près de 5 % des non-titulaires de la FPE).

¹⁹ Age au 1er janvier 2015 tel que $16 < \text{âge} < 63$.

²⁰ Cette sélection est imparfaite. Pour les salariés des cliniques privées, s'ils sont arrivés dans l'entreprise après 2014 (STATUT=4 et DATANT>2014), il est possible qu'ils aient été auparavant employés dans une autre clinique privée, à ce titre ils auraient pu être échantillonnés. S'ils sont arrivés « dans la fonction publique », sans plus de précisions (STATUT = 1, 2, 3, 5 et DATANT>2014), il est peu vraisemblable qu'ils aient été dans le champ en 2014. Mais c'est possible s'ils travaillaient dans une clinique privée à ce moment-là par exemple.

²¹ Pour le tirage des échantillons d'extension, le poste principal final est celui qui a la date de fin la plus proche du 31/12 parmi les postes actifs du salarié. Un poste est actif lorsqu'au moins une des périodes qui le constituent est active et sont considérées comme actives, les activités à temps complet, à temps partiel ou au forfait, les cessations progressives d'activité, les congés de maladie ordinaire et les congés formation.

²² Dans la définition de l'emploi, on ne retient pas tous les postes actifs, mais seulement ceux qui contribuent de manière non négligeable à l'appareil productif. Ces postes actifs seront qualifiés de non annexes. En adaptant un peu les définitions, il faut que soit le salarié ait un salaire net depuis son arrivée chez l'employeur supérieur à 3 fois le Smic mensuel net; soit que les trois conditions suivantes soient simultanément remplies: une ancienneté chez l'employeur supérieure à 30 jours et un nombre total d'heures travaillées supérieur à 120 heures depuis l'arrivée chez l'employeur et un nombre habituel d'heures travaillées par jour supérieur à 1,5 heure.

d) Doublons physiques

Il y a 27 cas de doublons physiques, à savoir des individus déjà présents dans le panel et tirés dans un des échantillons entrants ou bien tirés à la fois dans l'échantillon principal et dans les extensions.

D'après la théorie du partage des poids, lorsqu'une unité est présente dans plusieurs échantillons, son poids d'estimation après partage des poids est égal à la somme des poids de sondage dans chacun des échantillons où elle apparaît, divisé par son nombre total de liens.

En reprenant les notations de Favre-Martinoz et Gros (2017), leur poids partagé s'écrit:

$$W_i = \frac{1}{L_i} (w_{i,1} L_{ii}^1 \mathbf{1}_{\{i \in S_1\}} + w_{i,2} L_{ii}^2 \mathbf{1}_{\{i \in S_2\}} + \dots + w_{i,6} L_{ii}^6 \mathbf{1}_{\{i \in S_6\}})$$

Si l'individu est présent dans deux échantillons, son poids partagé est donc égal à :

$$W_i = \frac{1}{L_i} (w_i^1 L_{ii}^1 + w_i^2 L_{ii}^2)$$

avec w_i^k poids de tirage de l'unité i dans la base de sondage S_k .

avec L_i la somme des liens de i avec les 6 bases de sondage de l'enquête, soit le nombre de bases dans lesquelles l'unité i aurait pu être échantillonnée :

$$L_i = L_{ii}^1 + L_{ii}^2 + \dots + L_{ii}^6$$

Ici, le partage des poids sera effectué après le calage. Pour les 27 doublons physiques, le calage sera donc effectué sur leur poids de tirage corrigé de la non-réponse.

On conserve la FA qui a été enquêtée, suivant la règle de priorité données aux enquêteurs lors de la collecte (panel > entrant extension > entrant). On réalise la correction de la non-réponse pour cette FA avec le sous-échantillon de collecte, puis on fait la somme des poids W_1 et W_2 :

$$W_i^{sond} = W_1 + W_2$$

$$W_i^{CNR} = W_i^{sond} \times \frac{1}{pGRHi} = \frac{(W_1 + W_2)}{pGRHi}$$

Avec W_1 le poids de sondage de la FA qui a été enquêtée, W_2 le poids de sondage de la FA en BIS, et $pGRHi$ le taux de réponse dans le GRH de l'individu i .

e) Synthèse

L'algorithme peut s'écrire comme suit :

- Unités de l'échantillon panel : +2
- Unités de l'échantillon entrant, n'appartenant pas au champ de l'enquête en 2013 (= 'actoccup_2013 = 0 ou age_2013 < 15 ans' ou anarriv > '2013') : +1
- Unités de l'échantillon entrant, appartenant au champ de l'enquête en 2013 (= 'actoccup_2013 = 1 et age_2013 >= 15 ans' et anarriv <= '2013') : +2
- Unités de l'échantillon principal entrant appartenant au champ d'une extension de l'enquête en 2016 : +1 (par extension*)
- Unités des échantillons extensions entrants de l'enquête en 2016: +1 (par extension*)
- Unités appartenant au champ de l'enquête de 2013 ('actoccup_2013 = 0 et 'age_2013 >= 15 ans' et anarriv <= '2013') et au champ d'une extension de l'enquête en 2013 : +1 par extension

Une façon équivalente de représenter l'algorithme de partage des poids est de créer des indicatrices qui valent 1 chaque fois que l'individu est dans le champ d'un échantillon 0 sinon. Le nombre de lien est égal à la somme des indicatrices (voir annexe 2).

Tableau III.19 : nombre de liens selon le sous-échantillon d'origine du répondant

	Panels principal	Panels extension	Entrants principal	Entrants extension
1 lien			530	
2 liens	11230		2494	114
3 liens	953	471	196	38
4 liens	2924	4285	623	782
5 liens				1

4. Calcul des marges

Comme pour CT 2013, on souhaite que l'échantillon final décrive deux populations distinctes :

- Les actifs du secteur privé (données 2015 de l'Enquête Emploi en Continu, EEC2015).
- Les salariés du secteur public (SIASP2015 : répondants ayant un statut 1, 2, 3, 5).

Pour les salariés du secteur public, les marges sont calculées sur SIASP pour obtenir des effectifs dans la fonction publique qui soient les plus justes possibles, SIASP étant une information exhaustive sur ce sujet plus fiable que l'EEC. Pour les actifs occupés du secteur privé, les marges sont calculées sur l'EEC.

La réunion de ces deux échantillons permet d'avoir une population de répondants représentative de l'ensemble des actifs en emploi. La pondération PONDCAL, grâce à ce double calage, retrouve aussi des marges connues, par exemple sur les trois versants de la fonction publique.

Les DOM sont calés séparément.

Pour tenir compte de l'information auxiliaire disponible sur chaque sous-échantillon, on va caler :

- l'échantillon d'entrants principal sur des variables spécifiques (caractéristiques des actifs occupés au 31 décembre 2015). Ces variables seront nulles pour l'échantillon panel et pour les entrants extensions.
- l'échantillon panel principal sur des variables spécifiques (caractéristique des actifs occupés au 31 décembre 2012, issues de l'EEC 2012). Cela nécessite de faire participer au calage tous les répondants à CT 2013 répondants au panel, indépendamment de leur statut en 2015 (AO, chômeurs, inactifs, décédés, partis à l'étranger). Ces variables seront nulles pour l'échantillon panel extensions et pour les entrants.
- l'échantillon complet du secteur privé et du secteur public sur des variables transformées (caractéristiques des salariés du secteur public et, respectivement, des actifs du secteur privé). Ces variables seront égales à la variable transformée $\frac{Y_{i(j)}}{L_{i(j)}}$ pour les actifs occupés (valeur de la variable de calage pour l'individu divisée par son nombre total de liens) et nulles pour les panels non occupés en 2015/2016.

Le calage ici est différent de celui pour le panel, les variables retenues ne sont donc pas nécessairement les mêmes.

Sur les 24 670 répondants au questionnaire individuel de l'échantillon principal, il y a 1 692 non-réponses au QAA, soit 6,9 %²³. Cela justifie l'introduction d'une pondération spécifique, PONDQAA.

a) Métropole

i. Marges pour les actifs occupés du secteur privé

La définition du champ des actifs occupés et le traitement des « ajouts »

Le champ de l'enquête CT-RPS 2016 est défini par la variable ACTOCCUP (être actif occupé au moment de l'enquête). Dans l'enquête Emploi la variable définissant les actifs occupés, ACTOP, a une définition proche mais à quelques différences près (de formulation et de règles d'inclusion). Depuis 2015, le champ de l'activité dans l'EEC a été étendu, d'une part, à l'exercice d'une activité en tant

²³ 1 460 n'ont pas du tout répondu au QAA (60 ne comprennent aucune des langues proposées, 1 122 refusent, pour 278 l'entretien a été impossible à réaliser) et 172 ont commencé et abandonné en cours de route. Vu le faible nombre d'abandons, on choisit de les assimiler à une non réponse complète.

qu'aide familial ou conjoint collaborateur, et d'autre part, à l'exercice d'activités informelles ce qui contribue à rapprocher les définitions. Il persiste néanmoins quelques différences :

- 1) Les questions sur l'activité professionnelle commencent dans CT 2013 par une question SITUA (« Quelle est actuellement votre situation principale vis-à-vis du travail ? »), qui n'existe pas dans l'EEC (le questionnaire commence par « Avez-vous actuellement un emploi ». Cette différence est considérée comme négligeable ;

Lorsque la personne n'a pas travaillé la semaine de référence, les règles diffèrent un peu dans CT et dans l'EEC en fonction du motif sur la durée maximale sans avoir travaillé : les personnes en congés parental ou autre type de congés non rémunéré sont considérés actifs dans CT-RPS 2016 en cas d'absence de moins d'un an, alors que si cette absence dépasse contre moins d'un trimestre dans l'EEC.

Tableau III.20 : Comparaison des définitions

	Actifs occupés au sens de l'enquête Emploi	Actifs occupés au sens de l'enquête RPS
Enquête Emploi 2015 pondérée, secteur privé	154 523	+ 738 = 155 200
	20 568 969	+ 107 457 = 20 676 426
Effectifs enquête CT-RPS 2016, secteur privé	14 332	+110 = 14 442
Calage	Calage sur marges détaillé ci-dessous	Calage différentiel sur différentiel EEC, par raison de non travail et sexe ²⁴

Dans l'EEC 2015, 124 individus sont considérés comme actifs au sens du BIT mais pas au sens de l'enquête CT-RPS. Il s'agit de personnes dont l'absence dépasse 1 an (365 jours).

On note un important déficit d'actifs en volume dans le secteur privé. Plusieurs pistes d'explications : la part élevée de logements classés HC pour l'échantillon entrant, en particulier le nombre élevé de FA non traitées lors de la collecte pour l'échantillon entrant en Île-de-France.

La collecte de l'enquête Emploi est également réalisée par le réseau des enquêteurs ménage de l'Insee et subit *a priori* les mêmes aléas et biais liés aux destructions de logements, déménagements, etc.

La définition du statut au sens de CT-RPS 2016 et du secteur privé dans l'enquête Emploi

La variable « statut » de l'enquête CT-RPS 2016 a été construite pour coller à la définition des sous-échantillons supplémentaires et il faut reconstruire au plus près cette variable dans l'enquête Emploi.

On utilise pour cela plusieurs variables :

- **PUB3FP** qui permet de distinguer les trois fonctions publiques. Cette variable qui a fait l'objet d'un travail important conjoint entre l'Insee et de la DGAFP, dont est issu un document de travail

http://www.fonction-publique.gouv.fr/files/files/statistiques/documents_methodes/Agents_3FP_oct2011.pdf

²⁴ Peu d'informations sont disponibles pour ces 110 personnes, les marges sont calculées sur seulement un millier de répondants dans l'enquête Emploi. On cale ainsi sur la répartition par sexe (68,7 % de femmes) et la raison de non travail (45,4 % pour congé parental ; 31,2 % pour morte saison ou période précédant le début d'un emploi ; 11,8 % pour mise à pied ou fin d'emploi ; 11,50 % pour d'autres congés non rémunérés).

- **NAFN**, activité à un niveau fin de l'établissement employeur (intérimaire reclassé dans le secteur utilisateur), qui permet de construire la variable indiquant le secteur hospitalier (NAFN= « 8610Z, activités d'hospitalisation »),
- **CJRF**, catégorie juridique de l'employeur, qui permet d'identifier les salariés du secteur hospitalier public des établissements d'hospitalisation (CJRF=7364) et des syndicats inter hospitalier (CJRF = 7365),
- **STC** qui permet de distinguer les non-salariés ou chefs d'entreprise salariés des autres salariés et des aides familiaux,
- **CHPUB** qui permet de repérer les salariés de particuliers

Le calage sur marges

Le nombre d'actifs occupés dans le secteur privé s'élève à 20 568 969 dans l'EEC. Les marges utilisées pour le calage sont la répartition de ces actifs occupés selon :

- leur sexe ;
- leur âge quinquennal au 31 décembre (en 8 tranches) ;
- le secteur d'activité (16 positions²⁵) ;
- la CS en 15 positions ;
- le statut et le type de contrat en 7 positions (salariés en CDI, salariés autres contrats, salariés de particuliers, non-salariés, pour les salariés des cliniques privées : Médecins et cadres, Prof, inter, Employés et ouvriers).

Les marges sont calculées sur l'enquête Emploi 2015.

ii. Marges pour les salariés du secteur public

Pour la fonction publique, les variables retenues sont les mêmes que pour le calage sur marges de CT 2013 : sexe, tranches d'âge, catégorie hiérarchique et statut de la fonction publique (titulaire ou militaire/autres) croisé avec le versant de la fonction publique. Les marges sont calculées à partir de SIASP 2015.

Les variables utilisées sont :

- sexe,
- statut (titulaire, non-titulaire) x employeur (Fonction publique d'État, FP territoriale, FP hospitalière),
- catégorie (A, B ou C)
- âge (3 tranches).

iii. Marges pour l'échantillon entrants principal

Ce calage sur l'échantillon des entrants principaux a pour objectif de corriger la déformation résiduelle par âge, CS, région et de compenser les limites de la correction de la non-réponse sur les variables sociodémographiques à partir des données issues d'Octopusse.

Les variables utilisées pour le calage spécifique aux entrants de l'échantillon principal sont les suivantes :

- catégorie socioprofessionnelle agrégée (6 catégories),
- âge (6 tranches),
- région de résidence (Île-de-France vs de la France métropolitaine).

²⁵ La non réponse sur la variable NAF est d'un ordre de grandeur très différent dans l'EEC (1,5 %) et dans CT-RPS 2016 (0,32 %) de sorte qu'elle relève vraisemblablement de processus différents. Pour le calcul des marges de CT-RPS 2016, la NAF a fait l'objet d'une correction de la non réponse *ad hoc* dans l'EEC à partir des variables sexe, CS, diplôme et âge quinquennale. 51 GRH ont été constitués par l'algorithme CHAID (alpha= 1%, leaf = 400). Pour le calage uniquement, les 43 NA sur la NAF17 dans RPS sont reclassés dans la modalité la plus fréquente (MN =Administration publique, enseignement, santé humaine et action sociale).

Pour rappel, pour le calage, ces variables sont spécifiques à l'échantillon d'entrants principal et sont donc nulle sur l'échantillon panel (et les autres échantillons d'entrants extensions).

iv. Marges pour l'échantillon panel principal (y compris inactifs)

Ce calage sur l'échantillon des panels principaux a pour objectif de corriger la déformation résiduelle par âge et CS, liée à l'utilisation des poids de CT 2013 avant calage, et donc vraisemblablement biaisés par rapport à la non-réponse en 2013. Le calage se faisant sur la population des actifs occupés en 2013, les inactifs sont inclus dans le calage sur cet échantillon (ils prennent néanmoins des valeurs nulles pour toutes les variables spécifiques aux autres échantillons).

Les variables utilisées pour le calage spécifique aux entrants de l'échantillon principal sont les suivantes :

- sexe,
- catégorie socioprofessionnelle agrégée en 2013 (6 catégories),
- statut en 2013 (salariés du privé, salariés de FPE, salariés de FPT, salariés de FPH, salariés de particuliers, non-salariés),
- région de résidence en 2013 (Île-de-France vs de la France métropolitaine).

Pour rappel, pour le calage, ces variables sont spécifiques à l'échantillon panel principal et sont donc nulle sur tous les échantillons entrants.

b) *DOM*

Le calage pour les départements d'Outre-mer est également fait sur l'Enquête Emploi 2015. En raison du nombre relativement faible de répondants dans chaque DOM, il n'a pas été possible de réaliser un calage distinct pour chaque département.

Le champ

On retrouve la différence de définition de l'activité entre ACTOCCUP (CT-RPS) et ACTOP (EEC). On négligera néanmoins cette différence car les effectifs sont très faibles (4 sur les 1 373 répondants à CT-RPS 2016 dans les 4 DOM enquêtés).

Tableau III.21 : Comparaison des définitions

	Actifs occupés au sens de l'enquête Emploi	Actifs occupés au sens de l'enquête RPS
Enquête Emploi 2015 pondérée, secteur privé	17 159	+ 67 -7 = 17 2016
	2 318 641	+9175-1107 = 2 326 709
Effectifs enquête CT-RPS 2016, DOM	1 369	+ 4 = 1 373
Calage	Calage sur marges détaillé ci-dessous	

NB : il y a également 7 actifs occupés au sens de l'EEC mais pas de CT-RPS, on les néglige.

Tableau III.22 : Actifs occupés dans l'EEDOM et dans l'enquête CT-RPS

	EEC		CT-RPS	
	Non pondéré	Pondéré	Non pondéré	Pondéré*
Guadeloupe	3 698	122 555	350	81 819
Martinique	3 933	127 116	376	95 083
Guyane	3 999	63 512	238	31 599
Réunion	5 529	266 476	409	218 657

* Poids de sondage corrigés de la non-réponse, après partage des poids

On note un déficit d'actifs dans les quatre départements d'outre-mer enquêtés. Cela est vraisemblablement lié à l'étape de correction de la non-réponse, réalisée en France entière en raison d'effectifs trop faibles dans les DOM. Les variables de région de résidence et de lieu de naissance ont été introduites mais il est possible que les éventuels déterminants régionaux n'aient pas été statistiquement significatifs à l'échelle nationale.

Le calage sur marges

Le nombre d'actifs occupés en Guadeloupe, Martinique, Guyane et à la Réunion s'élève à 579 660 dans l'EEC en 2015.

Les marges utilisées pour le calage sont la répartition de ces actifs occupés selon :

- leur sexe x leur âge au 31 décembre (3 tranches d'âge par sexe) ;
- leur lieu de naissance x lieu de résidence (dans le DOM de résidence et il s'agit de la Guyane, dans le DOM de résidence hors Guyane, ailleurs en France, à l'étranger) ;
- leur CSP en 10 positions ;
- leur statut en 3 positions (salariés du privé, salariés du secteur public, non-salariés) ;
- leur département de résidence.

En raison des effectifs faibles de l'EEC, il n'a pas été possible de détailler autant les modalités de certaines variables que pour le calage de l'échantillon en métropole. De plus, les marges sont calculées sur l'Enquête Emploi 2015 pour les quatre départements enquêtés pris dans leur ensemble. Cela assure que les marges sont toutes calculées sur au moins 100 enquêtés de l'EEC.

On remarque avant calage un déficit important de répondants en Guyane, où le taux de collecte a été relativement faible.

Avant calage, il y a également un léger excédent relatif de non-salariés. Il y a parmi les répondants à CT-RPS 2016 un déficit de personnes nées en dehors du DOM, en particulier de répondants nés en métropole. C'est un des déterminants connus de la non-réponse dans les départements ultramarins. Concernant la structure par CSP, il y a un déficit de cadres

Tableau III.23 : Variables de calage

Variable	Modalité ou variable	Marge échantillon	Marge population	ech.	pop.
Sexe et âge					
Hommes	15-34	58 416	76 226	13,7	13,2
	35-49	84 269	118 716	19,7	20,5
	50 et plus	81 314	102 166	19,0	17,6
Femmes	15-34	43 750	70 967	10,2	12,2
	35-49	83 062	118 320	19,5	20,4
	50 et plus	76 346	93 265	17,9	16,1
Type d'emploi					
	Non-salariés	74 087	83 728	17,3	14,4
	État, collec. locales et hôpitaux	124 239	179 244	29,1	30,9
	Autres salariés	228 833	316 688	53,6	54,6
Lieu de naissance					
	Dans le DOM de résidence (hors Guyane)	331 964	399 501	77,7	68,9
	Dans le DOM de résidence (Guyane)	13 497	26 682	3,2	4,6
	Ailleurs en France	54 739	109 526	12,8	18,9
	A l'étranger	26 959	43 951	6,3	7,6
Région de résidence					
	Guadeloupe	81 819	122 555	19,2	21,1
	Martinique	95 083	127 116	22,3	21,9
	Guyane	31 599	63 512	7,4	11,0
	Réunion	218 657	266 477	51,2	46,0
CS					
	Agriculteurs	19 997	14 633	4,7	2,5
	Artisans, comm, chefs d'entreprise	41 535	45 432	9,7	7,8
	Cadres et prof. intellectuelles sup.	36 695	62 493	8,6	10,8
	Professions intermédiaires FP	50 672	83 277	11,9	14,4
	PF entreprises	39 340	54 618	9,2	9,4
	Employés FP	58 939	87 517	13,8	15,1
	Employés ent	54 071	65 633	12,7	11,3
	Employés de particuliers	41 480	50 978	9,7	8,8
	Ouvriers qualifiés	53 790	78 890	12,6	13,6
	Ouvriers non qualifiés	30 641	36 189	7,2	6,2

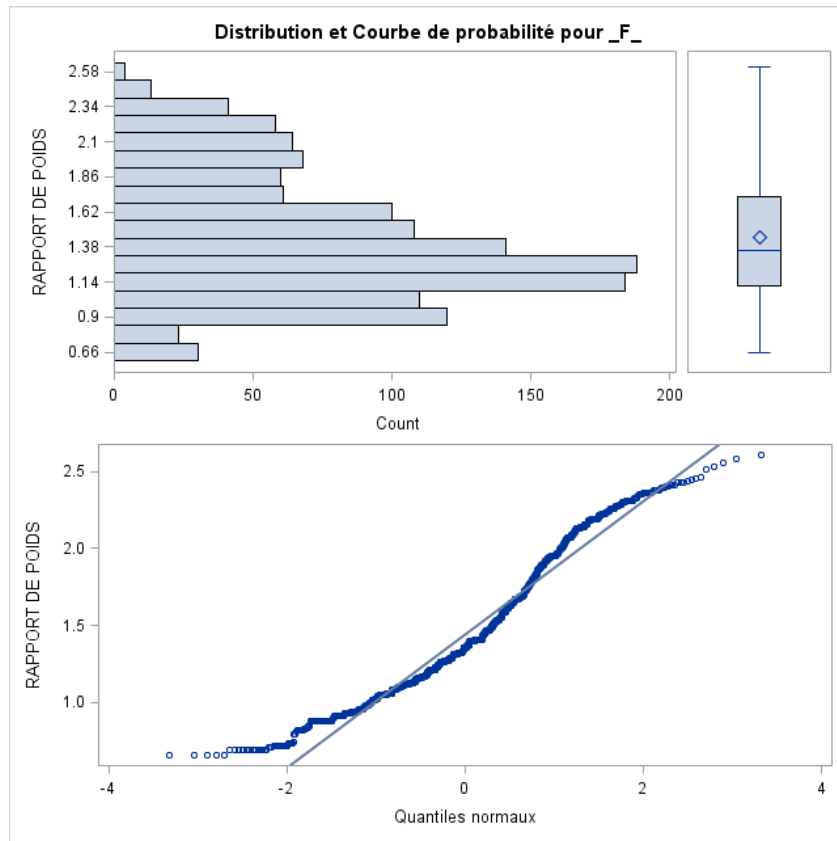
Un calage a été testé sur le jeu alternatif de variables suivant : sexe, variable croisant l'âge et le DOM de résidence, lieu de naissance selon le DOM de résidence, statut, CS. Le calage du QAA ne convergeant pas, la variable croisant âge et DOM de résidence a dû être écartée.

Le calage est réalisé avec la méthode logistique (macro CALMAR2), en prenant comme bornes des rapports de poids 0,5 et 2,9. L'histogramme conserve ainsi une allure relativement gaussienne (Graphique III.9), recommandée par le département des méthodes statistiques de l'Insee.

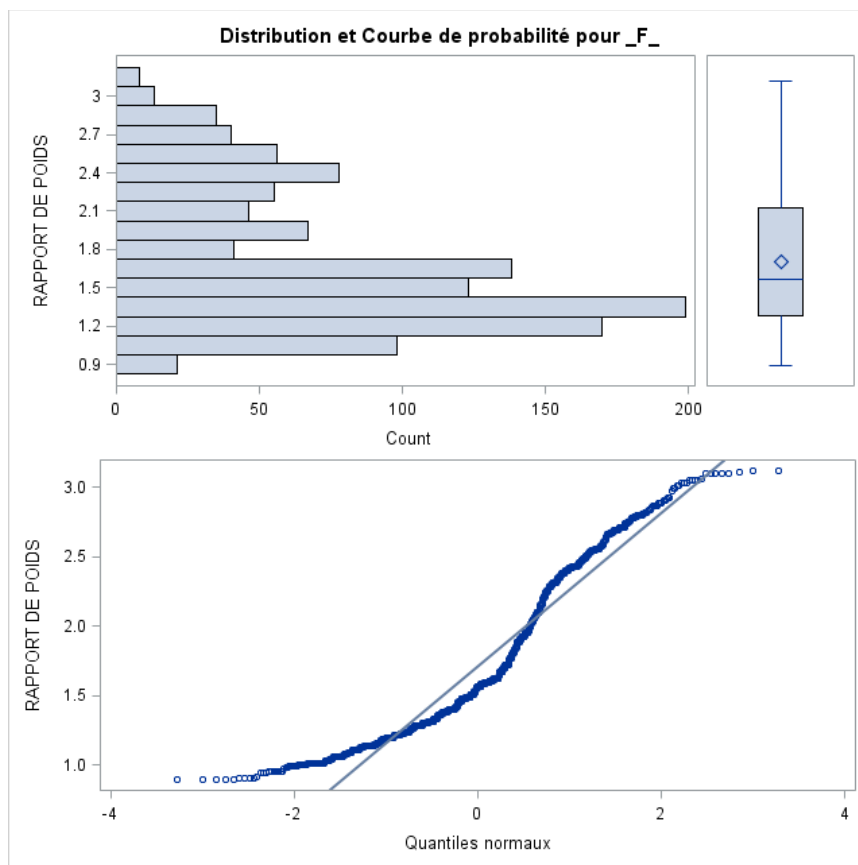
La méthode et les marges retenues sont strictement les mêmes pour les répondants au QAA.

L'intervalle entre les bornes pour le calage sur marges est toutefois de plus grande amplitude (0,6-3,6) et on obtient des poids (PONDQAA) légèrement plus dispersés que les poids pour le questionnaire individuel (Graphique III.10).

Graphique III.9 : rapport des poids DOM avant/après calage (pondcal)



Graphique III.10 : rapport des poids DOM avant/après calage (pondqaa)



5. Bilan et vérifications

a) Les pondérations finales

Le calage est réalisé avec la méthode logistique (macro CALMAR2), en prenant comme bornes des rapports de poids 0,4 et 2,5. La méthode et les marges retenues sont strictement les mêmes pour les répondants au QAA. L'intervalle entre les bornes pour le calage sur marges est toutefois de plus grande amplitude (0,4-2,7) et on obtient des poids (PONDQAA).

Après ce calage sur marges, on regroupe les poids obtenus avec ceux des « ajouts » (cf. tableau III.19).

Les poids de l'échantillon final des actifs occupés sont très dispersés. Cela tient à la complexité du plan de sondage, au panel, et aux taux de sondages très différents selon les départements et les secteurs (public/privé). Pour limiter cette dispersion, on décide de tronquer les 1 % des poids les plus élevés. On ramène les poids extrêmes à la valeur du 99^e percentile puis on refait un calage à l'identique sur ce nouveau jeu de poids (tableaux III.24 et III.25, graphique III.11).

Tableau III.24 : Distribution des poids finaux

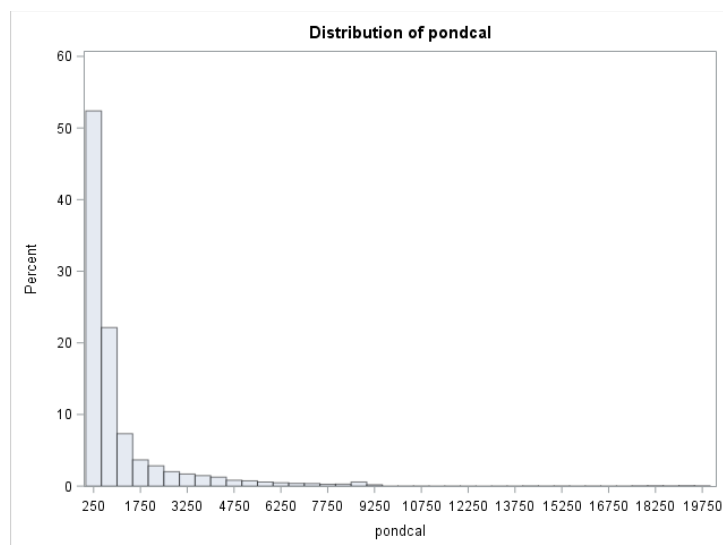
	France métropolitaine		DOM	
	Pondcal	Pondqaa	Pondcal	Pondqaa
Unité	individus	individus	individus	individus
Nb obs.	23269	21792	1 371	1 188
Moyenne	1 123,9	1 199,3	422,8	487,9
25 % Q1	259	266	223	262
50 % Médiane	472	502	293	343
75 % Q3	1 020	1 082	452	512
Min	2	2	95	130
Max	19940	22225	5 446	5 628
Rapport inter quartile	3,9	4,1	2,0	2,0
Somme des observations	26 151 238	26 133 975	579 660	579 660

Tableau III.25 : Distribution des poids finaux par département, DOM

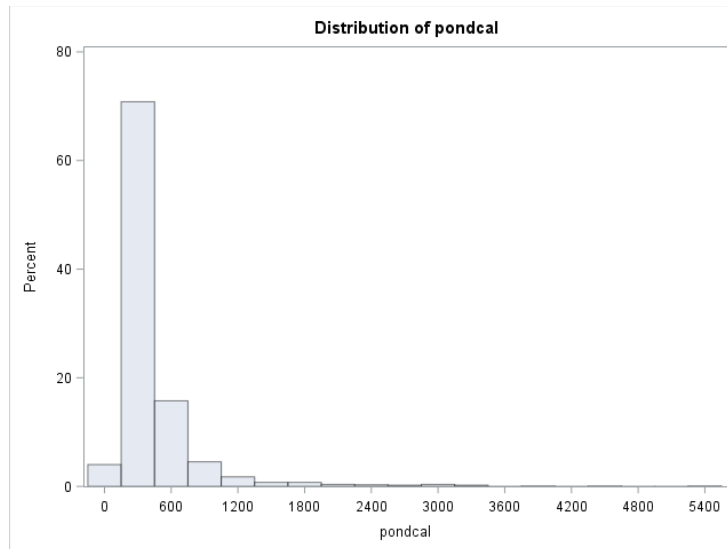
	Guadeloupe		Martinique		Guyane		Réunion	
	POND CAL	PONDQA A	PONDC AL	PONDQA A	PONDCA L	PONDQA A	PONDCA L	PONDQA A
N	350	291	376	312	238	202	409	383
Moyenne	350	421	338	407	267	314	652	696
Médiane	254	302	244	291	229	274	429	462
Minimum	132	176	113	146	95	130	198	237
Maximum	3 111	4 222	2 946	3 214	1 852	1 663	5 446	5 628
Rapport interquartile	1,7	1,7	1,9	2,0	1,7	1,6	1,9	1,9
Rapport inter décile	3,0	3,1	3,7	3,5	3,0	2,9	4,3	3,9

Graphique III.11 : Distribution des poids calés (pondcal)

(a) France métropolitaine



(b) DOM



b) Effet de la pondération sur quelques variables d'intérêt

i. Vérification de la structure du panel et des entrants actifs en 2013

Un des objectifs des redressements est de garantir la comparabilité des résultats de CT-RPS 2016 avec les éditions précédentes de l'enquête CT. Il est donc important de vérifier que la dimension panel et l'ajout des enquêtés issus des extensions ne modifie pas sensiblement les résultats.

Après correction de la non-réponse individuelle, des différences sociodémographiques persistaient entre les répondants du panel et les entrants qui auraient pu être interrogés en 2013 (échantillon principal, tableau III. 18). La méthode de partage de poids généralisé mise en place a permis la réduction notable entre le profil des panels et des entrants dans le champ du panel (tableau III.26).

Tableau III.26 : caractéristiques des répondants de l'échantillon panel principal (01) et de ceux de l'échantillon entrant principal (11) dans le champ de CT 2013, après correction de la non-réponse individuelle et après calage et partage des poids

		Echantillons corrigés de la non-réponse, non calés, non partagés		Echantillons corrigés de la non-réponse, calés, non partagés	
		Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*	Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*
Diplôme	Aucun diplôme	6,7	8,2	7,0	8,5
	CEP ou diplôme étranger de même niveau	1,8	2,5	1,8	2,6
	Brevet des collèges	4,6	5,5	4,8	5,3
	CAP, BEP	24,5	26,8	26,0	26,7
	Baccalauréat technologique ou professionnel	11,8	11,9	12,3	11,8
	Baccalauréat général	5,8	6,5	5,7	6,3
	Diplôme de niveau Bac+2	14,9	13,9	14,7	13,7
	Diplôme de niveau bac +3 ou bac +4	14,6	12,8	13,3	12,6
	Diplôme de niveau supérieur à bac + 4	15,3	11,9	14,3	12,4
NA	0,1	0,1	0,1	0,1	
Catégorie socio-professionnelle	Agriculteurs	2,3	2,9	2,0	2,1
	Artisans, commerçants	6,4	6,5	6,5	6,8
	Cadres	22,0	18,1	19,7	18,8
	Professions intermédiaires	25,5	26,1	23,7	25,5
	Employés	26,4	28,1	28,3	27,2
	Ouvriers	17,4	18,4	19,9	19,8
Statut	FPE	11,6	10,4	9,8	9,7
	FPT	7,9	7,6	7,4	7,5
	FPH	4,1	5,2	4,4	5,3
	Privés	63,1	64,2	66,8	66,4
	Non-salariés	13,4	12,7	11,7	11,2
Région	Ile de France	13,1	10,6	13,5	11,9
	Métropole hors IDF	86,9	89,5	86,5	88,1

* actifs occupés au 31 décembre 2012, nés avant 1998 (15 ans au moment de l'enquête CT 2013) ou arrivés sur le territoire français avant 2013, France métropolitaine.

À l'issue du redressement, leurs réponses devraient être similaires. Elles sont présentées dans le tableau III.27.

Pour la plupart des variables, les différentes étapes de redressement réduisent les écarts entre les panels et les entrants dans le champ de CT 2013 de façon très significative (par ex. penser à trop de choses à la fois, travailler sous pression). Dans certains cas, l'écart se réduit mais persiste néanmoins (par ex. travail répétitif, posture debout, port de charges lourdes). Le nombre de variables pour lesquelles l'écart reste supérieur à 2 % ou s'accroît légèrement (par ex. être soumis à 3 contraintes de rythme ou plus, conciliation vie professionnelle-vie privée très satisfaisante ou satisfaisante) est très limité (moins d'une vingtaine, voir tableau III.28). L'échantillon principal entrant étant relativement petit (3 843 individus), la variance est vraisemblablement grande sur cette sous-population.

On peut en conclure que, à l'issue des étapes de redressement, les enquêtés dans le champ de CT2013 (actifs occupés nés avant 1998 et résidant en France en 2013) répondent de façon similaire qu'ils soient dans le panel ou dans l'échantillon entrant principal (tableau III.27).

Tableau III.27 : Impact des différentes étapes du redressement sur quelques variables d'intérêt, l'échantillon panel principal (01) et de ceux de l'échantillon entrant principal (11) dans le champ de CT 2013

	Echantillons non corrigés de la non-réponse, non calés, non partagés (poids de tirage)		Echantillons corrigés de la non-réponse, non calés, non partagés		Echantillons corrigés de la non-réponse, calés, non partagés	
	Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*	Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*	Panel échantillon principal	Entrants échantillon principal, dans le champ de CT2013*
Travail répétitif	37,0	42,1	37,2	42,6	38,9	42,5
Posture debout	46,5	50,8	46,4	51,0	48,5	51,2
Port de charges lourdes	39,1	42,2	38,8	42,4	40,9	42,5
Vouloir faire son travail jusqu'à 60 ans	61,5	60,2	60,6	60,2	59,4	58,9
Se dépêcher toujours (toujours, souvent)	45,2	46,1	44,9	45,7	45,0	46,2
Craindre pour son emploi	21,5	24,6	21,8	24,7	22,5	25,5
Conciliation vie professionnelle-vie privée satisfaisante	82,0	79,5	82,0	79,5	81,8	78,7
Plus de 3 contraintes de rythme	30,0	32,2	29,9	32,5	31,0	33,9
Travailler sous pression (toujours, souvent)	31,1	28,7	30,6	28,5	32,0	31,5
Penser à trop de choses à la fois (toujours, souvent)	46,0	42,1	43,4	41,7	45,5	45,6
Score de WHO5	16,0	16,1	16,1	16,1	16,2	16,2

* actifs occupés au 31 décembre 2012, nés avant 1998 (15 ans au moment de l'enquête CT 2013) ou arrivés sur le territoire français avant 2013, France métropolitaine.

Tableau III.28 : variables pour lesquelles l'écart de réponse après correction de la non-réponse individuelle est accru après le calage ou reste élevé

	Panel principal (ssec 01)			Entrants principaux dans le champ de CT 2013 (ssec 11)			Ecart entre les réponses du ssec 01 et du ssec 11		
	poids de tirage	poids corrigés de la non réponse	poids corrigés de la non réponse, calés non partagés	poids de tirage	poids corrigés de la NR individuelle	poids corrigés de la non réponse, calés non partagés	poids de tirage	poids corrigés de la NR individuelle	poids corrigés de la non réponse, calés non partagés
CVFVP	82%	82%	82%	79%	80%	79%	3%	2%	3%
PROCEDUR	46%	46%	46%	48%	48%	49%	2%	2%	3%
HORVAR	34%	34%	34%	37%	37%	37%	3%	3%	3%
STARK "strict"	27%	27%	28%	30%	31%	31%	3%	4%	3%
REPETE	37%	37%	39%	42%	43%	42%	5%	6%	3%
RWNORMJ	18%	18%	18%	21%	21%	21%	3%	3%	3%
RWDEM	83%	83%	82%	85%	84%	85%	2%	1%	3%
CHGTCOLL "non"	43%	43%	43%	45%	45%	46%	2%	2%	3%
PUBLIC	73%	73%	73%	76%	75%	76%	3%	2%	3%
DEBORD	67%	67%	67%	65%	64%	64%	2%	3%	3%
NBCW3*	30%	30%	31%	32%	33%	34%	2%	3%	3%
RPA1G** ('3', '4')	67%	67%	67%	64%	64%	64%	3%	3%	3%
REPOS	80%	80%	80%	77%	77%	77%	3%	3%	3%
RPA1B*** in ('3', '4')	64%	64%	65%	61%	61%	62%	3%	3%	3%
RPA1E**** in ('3', '4')	73%	73%	74%	71%	70%	71%	2%	3%	3%
URGFAM	90%	90%	90%	87%	87%	87%	3%	3%	3%
CWDEBOU	47%	46%	48%	51%	51%	51%	4%	5%	3%
CORRFORM	54%	54%	54%	57%	57%	58%	3%	3%	4%
PUBLIC1tel	62%	62%	62%	65%	65%	66%	3%	3%	4%
CINDUS	46%	46%	47%	49%	50%	51%	3%	4%	4%

France métropolitaine.

* trois contraintes de rythmes parmi 8

** recevoir le respect et l'estime mérités au vu des efforts

*** aide du chef pour mener ses tâches à bien

**** aide des collègues pour mener ses tâches à bien

ii. Effet de la méthode de partage des poids généralisés par rapport à la méthode classique sur les variables d'intérêt

On compare maintenant pour l'ensemble des répondants actifs occupés les résultats obtenus en réalisant différentes stratégies concernant les étapes de partage des poids et de calage :

- la stratégie retenue pour CT 2013²⁶ : application du partage des poids sur les poids corrigés de la non-réponse individuelle, puis calage sur marges sur les poids partagés.

²⁶ Algava, E., Davie E., Loquet J., Mikol F., Schreiber A. (2015), Enquête Conditions de travail 2013 – Métropole : redressement du volet actifs occupés, échantillon principal et extensions, DARES-DREES-DGAFF.

- la stratégie finalement retenue, à savoir calage sur marges de chaque échantillon sur des marges qui lui sont propres (avec prise en compte du nombre de liens quand nécessaire), puis partage des poids.

Au final, les variations sont très modestes (tableau III.29) : pour la quasi-totalité des variables, l'écart est très faible (<1 %) selon que l'on retient l'un ou l'autre des méthodes. Pour la vingtaine de variables pour lesquelles l'écart est le plus élevé il reste inférieur à 2 %. Le choix de la stratégie de partage des poids et de calage a donc relativement peu d'impact sur la valeur moyenne des indicateurs de l'enquête.

La méthode du calage sur marges adaptée au partage des poids reste néanmoins préférable car elle a permis de réduire les écarts sociodémographiques entre le sous-échantillon panel principal et les entrants qui leur sont comparables (sous-échantillon principal, actifs de plus de 15 ans ou résidents en France en 2013) et garantir une plus grande stabilité des analyses par catégories sociodémographiques.

Tableau III.29 : Réponses à des variables d'intérêt de l'enquête, avant redressement, après correction de la non-réponse et selon la méthode de calage retenue

	Poids de tirage	Poids corrigés de la non-réponse, non calés, non partagés	Poids finaux CT-RPS 2016	Poids finaux calculés selon la méthode "classique" (poids corrigés de la non-réponse, partagés puis calés)
Travail répétitif	42,0	39,5	42,1	42,3
Posture debout	51,4	49,4	50,8	50,7
Port de charges lourdes	41,0	39,9	41,0	41,4
Vouloir faire son travail jusqu'à 60 ans	59,0	59,9	56,5	56,5
Se dépêcher toujours (toujours, souvent)	45,7	45,4	45,4	45,5
Craindre pour son emploi	24,2	22,0	25,1	25,2
Conciliation vie professionnelle-vie privée satisfaisante	79,2	81,3	80,4	80,2
Plus de 3 contraintes de rythme	31,6	30,9	33,2	33,2
Travailler sous pression (toujours, souvent)	28,1	29,6	30,7	30,8
Penser à trop de choses à la fois (toujours, souvent)	42,0	43,2	44,3	44,4
Score de WHO5	16,1	16,0	16,3	16,3

Source : enquête CT-RPS2016.

Champ : actifs occupés, France entière.

c) *Vérification des effectifs « France entière » du secteur public*

Le calage sur marges a été effectué de façon séparée pour la fonction publique en métropole (marges issues de SIASP) et pour la fonction publique dans les DOM (marges enquête Emploi). Au final, la prise en compte des deux échantillons pondérés aboutit à des effectifs de fonction publique « France entière » très proches de ceux de Siasp (tableau III.30).

Tableau III.30 : Comparaison de la structure par CS, âge et sexe des répondants

	Conditions de travail et Risques Psychosociaux 2016 (avec variable STATUT)	SIASP 2015
FPE	2 471 451	2 477 457
FPT	1 992 830	1 984 244
FPH	1 187 047	1 186 610
Ensemble	5 651 328	5 648 311

Source : Conditions de travail et Risques Psychosociaux 2016 et SIASP 2015.

Champ : France entière (hors Mayotte), agents de la fonction publique

Références

Algava, E., Davie E., Loquet J., Mikol F., Schreiber A. (2015), Enquête Conditions de travail 2013 – Métropole : redressement du volet actifs occupés, échantillon principal et extensions, DARES-DREES-DGAFP.

Algava, E., Fleuret A. (2015), Enquête Conditions de travail 2013 – DOM : redressement du volet actifs occupés, DARES.

Ardilly, P. (2010), Pondérations longitudinales et transversales dans les échantillons rotatifs. Application à l'enquête SILC, Journée Sondages, Dijon 2010.

Collège d'expertise sur le suivi statistique des risques psychosociaux au travail (2011), Mesurer les facteurs psychosociaux de risques au travail pour les maîtriser, Paris.
www.collegerisquespsychosociaux-travail.fr/index.cfm

de Riccardis, N. (2012), Traitements de non-réponse et calages pour l'enquête santé et itinéraire professionnel de 2010, DREES, Document et de travail, Série sources et Méthodes n° 36, septembre 2012.

Deville, J.-C. et Lavallée, P. (2006). Sondage indirect : les fondements de la méthode généralisée du partage des poids. *Techniques d'enquête*, vol. 32, n° 2, p. 185.

Favre-Martinoz, C. et Gros, E. (2017), La méthode du partage des poids, Fiches méthodologiques, Insee, octobre 2017 (en ligne sur insee.fr)

LaRoche, S. (2003), Pondérations longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu, Statistique Canada, Séries des documents de recherche – Revenu, n° 007.

Lavallée, P. (2007), « GWSM and calibration », in *Indirect Sampling*, Springer, chap. 7, p. 121-150.

Naud, J-F. (2004). Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu, document de recherche Statistique Canada.

Annexe I – Variables lieu de résidence et de conditions de travail en 2013 utilisées pour la correction de la non-réponse des individus panel

Lieu de résidence :

Paris, Lyon, Marseille et 52 modalités obtenues en croisant la région (département dans les DOM) de résidence avec le caractère rural ou urbain de la commune de résidence :

Guadeloupe, Martinique rural, Martinique urbain, Guyane rural, Guyane urbain, La Réunion rural, La Réunion urbain, IDF rural, IDF urbain, Champagne Ardennes rural, Champagne Ardennes urbain, Picardie rural, Picardie urbain, Haute Normandie rural, Haute Normandie urbain, Centre rural, Centre urbain, Basse-Normandie rural, Basse-Normandie urbain, Bourgogne rural, Bourgogne urbain, Nord-Pas-de-Calais rural, Nord-Pas-de-Calais urbain, Lorraine rural, Lorraine urbain, Alsace rural, Alsace urbain, Franche-Comté rural, Franche-Comté urbain, Pays de la Loire rural, Pays de la Loire urbain, Bretagne rural, Bretagne urbain, Poitou-Charentes rural, Poitou-Charentes urbain, Aquitaine rural, Aquitaine urbain, Midi-Pyrénées rural, Midi-Pyrénées urbain, Limousin rural, Limousin urbain, Rhône-Alpes rural, Rhône-Alpes urbain, Auvergne rural, Auvergne urbain, Languedoc-Roussillon rural, Languedoc-Roussillon urbain, PACA rural, PACA urbain, Corse rural, Corse urbain, Non renseigné.

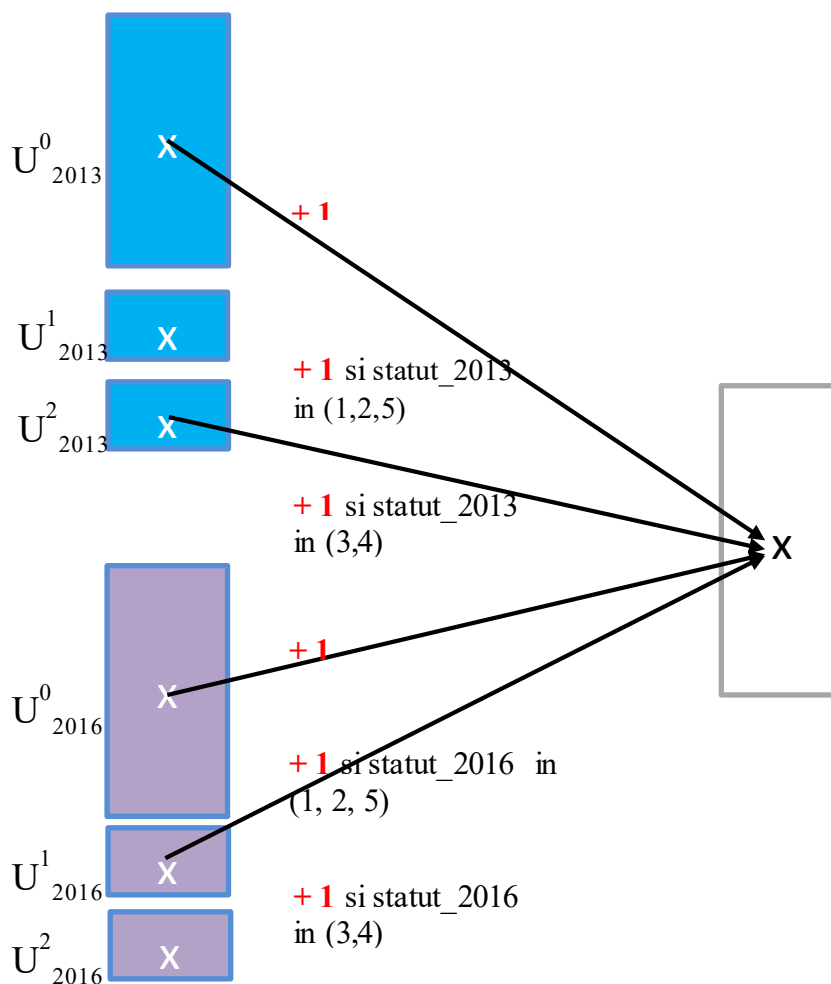
Conditions de travail :

- Pour l'année qui vient, avez-vous des craintes pour votre emploi ? (oui/non)
- Travaillez-vous le samedi ? (Habituellement/ Occasionnellement/ Jamais)
- Êtes-vous obligé de vous dépêcher ? (Toujours/Souvent/ Parfois/ Jamais)
- Votre travail consiste-t-il à répéter continuellement une même série de gestes ou d'opérations ? (oui/non)
- Mon supérieur traite équitablement les personnes qui travaillent sous ses ordres ? (Pas du tout d'accord/ Pas d'accord/ D'accord/ Tout à fait d'accord/ Non concerné, pas de supérieur).
- Les collègues avec qui je travaille m'aident à mener mes tâches à bien ? (Pas du tout d'accord/Pas d'accord/D'accord/ Tout à fait d'accord/ Non concerné, pas de collègue)
- Vu tous mes efforts, mes perspectives de promotion sont satisfaisantes (Pas du tout d'accord/Pas d'accord/ D'accord/ Tout à fait d'accord/ Non concerné)
- Je continue à penser à mon travail même quand je n'y suis pas (Toujours/Souvent/ Parfois/Jamais)
- Il m'arrive de faire trop vite une opération qui demanderait davantage de soin (Toujours/Souvent/ Parfois/Jamais)
- Je sais à l'avance quelles tâches j'aurai le mois suivant (Toujours/Souvent/ Parfois/Jamais)
- J'ai la possibilité de faire des choses qui me plaisent (Toujours/Souvent/ Parfois/Jamais)
- Il m'arrive d'avoir peur pendant mon travail, pour ma sécurité ou celle des autres (Toujours/Souvent/ Parfois/Jamais)
- Dans votre travail, à quelle fréquence vous arrive-t-il d'éprouver les sentiments suivants ? La fierté du travail bien fait (Toujours/Souvent/ Parfois/Jamais)
- Dans votre travail, à quelle fréquence vous arrive-t-il d'éprouver les sentiments suivants ? Le sentiment d'être exploité (Toujours/Souvent/ Parfois/Jamais)
- Travaille à temps complet pour s'occuper de ses enfants, pour s'occuper d'un (autre) membre de votre famille ou pour faire des travaux domestiques
- En cas d'imprévu, pouvez-vous modifier vos horaires en vous arrangeant avec vos collègues ? (Oui/Non)
- Connaissez-vous les horaires que vous devrez effectuer... dans le mois à venir ? (Oui/Non)
- Être soumis à un contrôle horaire (oui/non)
- Dans votre emploi principal, travaillez-vous entre cinq heures et sept heures le matin ? (Habituellement/ Occasionnellement/ Jamais)
- Travaillez-vous entre vingt heures et minuit ? (Habituellement/ Occasionnellement/ Jamais)
- Au cours des 12 derniers mois, avez-vous été joint par votre établissement, vos collègues ou vos supérieurs, en dehors de vos horaires de travail pour les besoins du travail ? (Oui/Non/Sans objet : travaille seul)
- Au cours des douze derniers mois, combien avez-vous eu d'arrêts maladie (hors maternité) ? (Pas d'arrêt/Un arrêt/Deux arrêts/ Trois arrêts ou plus)
- L'exécution de votre travail vous impose-t-elle d'effectuer des déplacements à pied longs ou fréquents ? (oui/non)
- L'exécution de votre travail vous impose-t-elle d'examiner des objets très petits, des détails fins ? (oui/non)

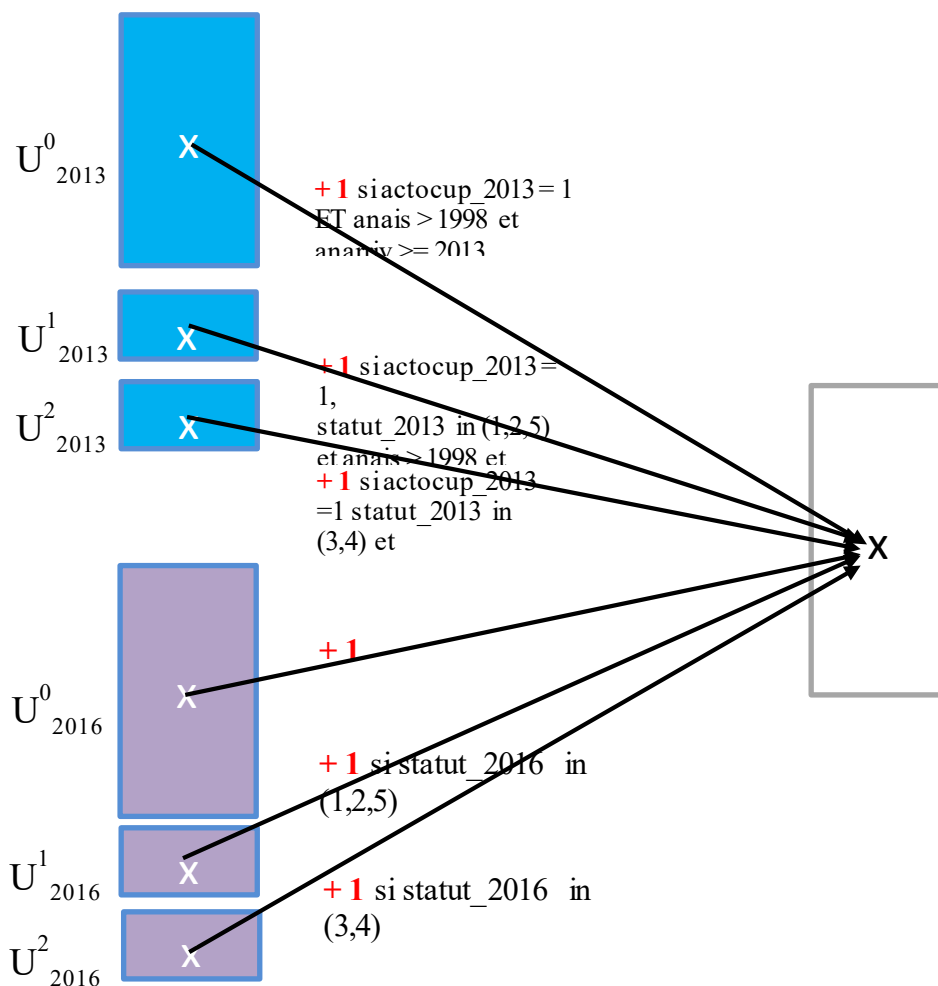
- L'exécution de votre travail vous impose-t-elle de rester longtemps dans une autre posture pénible ou fatigante à la longue ? (oui/non)
- L'exécution de votre travail vous impose-t-elle de faire attention à des signaux visuels ou sonores brefs, imprévisibles ou difficiles à détecter ? (oui/non)
- Votre travail ou votre lieu de travail, présente-t-il les inconvénients suivants absence de vue sur l'extérieur ? (oui/non)
 - Mon supérieur prête attention à ce que je dis (Pas du tout d'accord/Pas d'accord/D'accord/ Tout à fait d'accord/ Non concerné, pas de supérieur)
 - A votre emplacement de travail, êtes-vous amené à respirer des fumées ou des poussières ? (oui/non)
 - Pour préserver votre sécurité ou votre santé dans votre travail, disposez-vous de consignes ou d'instructions écrites, en dehors des consignes d'évacuation en cas d'incendie ? (oui/non)
 - Le rythme de travail est imposé par au moins deux contraintes (machines/collègues/normes de production etc.)
 - Occupez-vous différents postes ? (Oui, vous faites une rotation régulière entre plusieurs postes/ Oui, vous changez de en fonction des besoins de l'entreprise ou de l'administration/ Non.
 - Pour faire votre travail correctement, est-ce que vous appliquez strictement les consignes, ordres et modes d'emploi que vous recevez. (oui/non)
 - Votre travail vous permet-il d'apprendre des choses nouvelles ? (oui/non)
 - Une erreur dans votre travail peut-elle ou pourrait-elle entraîner... des conséquences graves pour la qualité du service ou du produit ? (oui/non)
 - Au cours des douze derniers mois, vous est-il arrivé d'aller (de) travailler tout en pensant que vous vous auriez dû rester à la maison (n'auriez pas dû) parce que vous étiez malade ? (oui/non)
 - Pensez-vous que pour faire correctement votre travail, certaines compétences vous manquent ? (oui/non)
 - Et pensez-vous que certaines de vos compétences ne sont pas utilisées ? (oui/non)
 - Si vous avez du mal à faire un travail délicat, compliqué, est-ce que vous êtes aidé par d'autres personnes de l'établissement ? (oui/non/sans objet)
 - Travaillez-vous seul ? (Toujours/Souvent/Parfois/Jamais)
 - Recevez-vous des ordres ou des indications contradictoires ? (oui/non)
- Êtes-vous adhérent ou sympathisant d'une organisation syndicale de salariés ? (Oui, adhérent/Oui, sympathisant/ Non/Vous ne souhaitez pas répondre)
- Votre établissement est-il couvert par un Comité Hygiène Sécurité et Conditions de Travail (CHSCT) ? (oui/non)
- Au cours de votre travail, êtes-vous amené à être en contact avec des personnes en situation de détresse ? (oui/non)

Annexe II – Dénombrement des liens pour le partage des poids

Graphique 1 : Schéma du comptage de liens pour les individus panel



Graphique 2 : Schéma du comptage de liens pour les individus entrants



Algorithme de calcul du nombre de liens

/ Création des indicatrices sur les cas possibles */*

/ 1. Être dans le champ de CT-RPS 2016*/*

champs_2016 = 1; */*par construction*/*

/ 2. Être dans le champ de CT 2013 */*

champs_2013 = ((ssech <'10' or */* panel par définition dans le champ de CT */*
(statuant not in ('','11','12','13') and anais <=1998 and anarriv not in
(*'2013','2014','2015','2016'*))
or typpanel =1); */* doublons physiques */*

/ 3. Être dans le champ d'une extension de CT 2013 */*

champs_ext2013 = ((ssech =*'01'* and typex_2013=1)
or ssech in (*'02','03','04','05','06'*)
or typex_2013 = 1 or */* doublons physiques */*
(ssech >*'10'* and champs_2013 =1 and 1948 <= anais <= 1993 */* filtre au tirage dans
SIASP et DADS*/*
and statuant in (*'1','2','3','4','5'*)));

/ 4. Être dans le champ d'une extension de CT-RPS 2016 */*

```
if statut in ("1","2","3","4","5")
  and france="1" and rges >'10' /* France métropolitaine */
  and titupbr ne "3" /* Militaires */
  and pe not in ("334A","452B","532A","532B","532C") /* Militaires */
  and 1951 <= anais <= 1998
  and datant not in (2015,2016)
  and (hh ge 8)
then do :
  if statut in ("1","2","5") then do : /* DGAFP */
    if typemploi in ("1","2","4") then champs_ext2016=0; /* emploi aidés, apprentis,  
intérimaires */
    else if typemploi="7" and titupbr in ("4","5","6","7") then champs_ext2016=0; /* sans  
contrat*/
    else champs_ext2016=1 ;
  end;
  else do : /* DREES */
    if typemploi in ("1","2","3","4","7") then champs_ext2016=0; /* apprentis, interim,  
stagiaires, emplois aidés, sans contrats*/
    else if titupbr="9" then champs_ext2016=0; /*Autre*/
    else champs_ext2016=1;
  end;
end ;
else champs_ext2016 = 0;
```

if index(profess,*'REGI'*)>0 or index(profess,*'COMMUNAU'*)>0 and pe = *"335A"* then champs_ext2016 = 0; */* Elus */*

if typex_2016 = 1 then champs_ext2016 = 1; */*doublons physiques */*

if ssech in (*"12"*,*"13"*,*"14"*,*"15"*,*"16"*) then champs_ext2016=1;

/ 5. Doublons physiques tirés dans les 2 extensions */*

champs_2ext2016 = 0;

if typ2ex_2016 = 1 then champs_2ext2016 = 1

/ 6. Nombre de liens */*

nb_liens = champs_2013 + champs_2016 + champs_ext2013 + champs_ext2016 + champs_2ext2016;

**L'enquête
Conditions de Travail-
Risques Psychosociaux 2016 :
apurement et redressement
Volume 2 – le volet « entreprises »**

Table des matières

Introduction.....	3
1. Bilan de la collecte.....	4
a) Constitution de l'échantillon.....	4
b) L'échantillon mis en collecte.....	5
c) Le déroulement de la collecte.....	6
d) Réception des fichiers d'enquête.....	8
2. Délimitation des répondants et apurement.....	9
a) Définition de la non réponse totale.....	9
b) Analyse des taux de réponse.....	10
c) Apurement des réponses.....	11
3. Pondération du fichier couplé.....	14
a) Correction de la non-réponse de l'établissement.....	14
Définition de la non-réponse totale au niveau établissement.....	14
Déterminants de la non-réponse établissement.....	16
Construction des groupes homogènes de réponse.....	17
b) Calage sur marges.....	18
Le calage dans le secteur privé.....	18
Le calage dans la fonction publique.....	21
Le calage dans les DOMs.....	23
c) Bilan et vérifications.....	23
Les pondérations finales.....	23
3. Pondération du fichier des établissements.....	27
a) Poids établissements corrigés de la non réponse - fichier SIR.....	28
b) Partage des poids.....	30
c) Calage sur marges.....	31
L'objectif est d'obtenir pour chaque établissement deux poids :.....	31
Les marges retenues.....	31

L'objectif du volet « employeurs » de l'enquête Conditions de Travail et Risques psychosociaux 2016 (CT-RPS 2016) est, d'une part, de mieux connaître les risques et leur prévention dans les établissements français, et, d'autre part, de confronter les données des deux volets afin d'évaluer autant que possible les convergences ou les divergences des avis des salariés et de leurs employeurs sur leurs conditions de travail. La collecte de l'enquête s'est déroulée de février 2016 à février 2017. La société Ipsos-Observer a été mandatée par le ministère pour mener l'enquête sur le terrain.

Les établissements de tous secteurs, marchand ou non, public ou privé, et toutes tailles sont concernés par l'enquête « employeurs », qui couvre la France métropolitaine et quatre départements d'Outre-mer (Martinique, Guadeloupe, Guyane, Réunion).

Le questionnaire est décliné selon trois modèles afin de poser les questions de façon plus adaptée selon le type d'établissement :

- un [questionnaire « secteur hospitalier »](#), pour les hôpitaux et cliniques ;
- un [questionnaire « fonction publique »](#), destiné aux établissements administratifs de l'État et des collectivités territoriales (hors hospitalière) ;
- un [questionnaire « secteur marchand et associatif »](#), pour tous les autres établissements.

L'échantillon du volet « Employeurs » de l'enquête CT-RPS 2016 est constitué de deux échantillons distincts :

- un échantillon aléatoire : 7 000 établissements d'au moins un salarié tirés aléatoirement dans le répertoire SIRENE.
- un échantillon indirect : les établissements employeurs de 18 150 salariés ayant répondu au volet Actif occupé de l'enquête. Pour des raisons de confidentialité, cet échantillon ne concerne que les établissements de 10 salariés ou plus.

Certains établissements appartiennent aux deux échantillons : le nombre d'établissements distincts finalement contactés est de 19 529.

Ce document dresse le bilan de la collecte et décrit les différentes étapes d'apurement et de redressements après la collecte. Deux bases ont été constituées, correspondant aux deux types d'exploitation de possibles de ce volet employeur.

La première base concerne les **données couplées** salariés-employeur. Chaque table contient **une ligne par salarié dont l'employeur a répondu** et contient les réponses de ce salarié au volet « individu », et celles de son employeur. La pondération à utiliser est PONDCOUPLE (ou PONDCOUPLE_QAA pour les réponses aux questions auto-administrées, Cf. questionnaire « Individu »).

La seconde concerne l'**ensemble des répondants au volet Employeur**. Chaque table contient **une ligne par employeur ayant répondu** et contient les réponses de cet employeur. Les pondérations à utiliser sont POIDS_SALA, qui permet d'obtenir des résultats en nombre de salariés et POIDS_ETAB, pour des résultats en nombre d'établissements.

Chacune de ses bases de données est constituée de 3 tables :

- « SM » pour secteur marchand ;
- « SH » pour secteur hospitalier ;
- « FP » pour fonction publique d'État et collectivités territoriales.

1. Bilan de la collecte

La collecte de l'enquête s'est étalée sur 12 mois, les établissements ont été traités en 5 lots :

- premiers contacts (animation téléphonique) => Lot 1 : fin oct. 2015
- derniers retours questionnaires : => février 2017 (tous lots).

L'enquête s'est effectuée en 5 lots au fur et à mesure des réponses des salariés au volet « Individus ». Les établissements en doublon ont été identifiés au fur et à mesure des lots.

a) Constitution de l'échantillon

Deux échantillons d'établissements sont concernés par l'enquête.

- Un échantillon indirect qui provient des réponses du volet « Individus » de l'enquête. Il s'agit des établissements de 10 salariés ou plus domicilié en France (y compris les DOM, hors Mayotte) dont un salarié a répondu au volet « Individu » de l'enquête et qui a donné des renseignements sur l'établissement qui l'emploie. L'établissement dont plusieurs salariés auront été interrogés n'est contacté qu'une seule fois. Ces établissements sont appelés établissements QAO.
- Un échantillon aléatoire direct de 7 000 établissements tiré dans SIRENE par la division sondage de l'Insee. Il permet d'élargir le champ aux établissements de moins de 10 salariés et de garantir que l'échantillon total soit représentatif de l'ensemble de l'économie (cf. note réf. CT/2015/n° 11 du 11 mars 2015). On parlera par la suite d'établissements SIR.

L'échantillon QAO a fait l'objet de traitements complémentaires et d'enrichissements avant son envoi au prestataire.

➤ Récupérations des informations fournies par les salariés

Pour les salariés répondants, une vingtaine de variables de l'enquête concernant l'employeur principal participent à l'adressage du questionnaire « Employeur ». Il s'agit de variables de la partie « Établissement employeur » du QAO : Désignation, Siret et Adresse ainsi que les variables sur l'activité, le nombre de salariés, le type d'emploi, le statut de l'employeur.

Pour les salariés ayant fourni a minima la raison sociale, le siret ou suffisamment d'éléments d'adresse, ces données ont été transmises par lots aux gestionnaires Insee du répertoire Sirène, pour Sirénisation.

Contrairement à ce qui a été fait pour CT 2013, pour CT-RPS 2016 le traitement des fichiers adresses initiales a été confié au prestataire (Ipsos), aussi bien pour l'échantillon aléatoire SIR que pour les employeurs QAO.

➤ Sirenisation (et traitement des échos)

La Sirénisation renvoie une liste d'échos (Siret dans SIRENE pouvant correspondre à l'établissement présumé¹), jusqu'à 50 pour certains établissements

Le traitement des échos consiste à sélectionner l'écho le plus vraisemblable dans cette liste. En cas d'apport d'information nul du sirénage, ce sont les données « salariés » qui sont retenues pour l'adressage du questionnaire.

➤ Repérage des doublons

¹ La recherche des échos a été par identification automatique (IAU) et par étude de divergence Siret avec notation (SAN), pour plus d'explications, cf. le guide Insee du sirénage).

Une fois l'information sur chaque établissement validée et améliorée, il est nécessaire d'identifier les établissements présents à plusieurs reprises dans l'échantillon parce que :

- ils sont sélectionnés dans l'échantillon SIR et un de leur salarié a répondu (donc ils sont aussi dans l'échantillon QAO).
- plusieurs de leurs salariés ont répondu à l'enquête.

Ces doublons ont été identifiés par le Siret quand celui-ci était disponible, par examen visuel des doublons plausibles sinon (même commune et même secteur).

➤ Préparation des fichiers pour la collecte

Une dernière étape consiste à :

- déterminer le type de questionnaire (secteur marchand, secteur hospitalier ou fonction publique) sur la base du secteur d'activité, de la catégorie juridique, la variable STATUT ;
- afin de pré-remplir le questionnaire, la convention collective majoritaire la plus probable a été imputée (IDCC) grâce aux données ACEMO, lorsque le SIRET ou l'APET étaient disponibles. Cet ajout a été fait à la Dares en aval de la sirenisation par les services de l'Insee.

b) L'échantillon mis en collecte

Les fichiers ont été adressés en 5 lots successifs de novembre 2015 à août 2016, les lots de l'échantillon QAO étant constitués au fur et à mesure de la collecte du volet individus (tableau 1). Le traitement initial des adresses par Ipsos a conduit à un échantillon de 24 047 établissements².

Au fur et à mesure du traitement et de la mise en collecte des lots, des doublons non repérés au départ se sont révélés (4 495 établissements), soit parce que l'établissement était déjà présent dans un lot antérieur, soit parce qu'il s'agit d'un doublon physique non identifié lors des traitements pré-collecte.

Tableau 1 : nombre d'établissements mis en collecte, hors doublons

	Nombre d'établissements	
Secteur Hospitalier public et privé (SH)	1 920	9,7 %
Fonction Publique État et coll. ter. (FPET)	5 490	27,8 %
Secteur Marchand et associatif (SM)	12 338	62,5 %
Lot 1	6 939	35,1 %
Lot 2	3 537	17,9 %
Lot 3	4 770	24,2 %
Lot 4	3 081	15,6 %
Lot 5	1 421	7,2 %
Base établissements	19 748	100,0 %

² Ipsos a omis de traiter les coordonnées des employeurs de 1 195 salariés, les adresses du lot 4 ayant été livrées en deux fichiers, le deuxième fichier (et le plus petit) n'a pas été traité.

c) Le déroulement de la collecte

Pour chacun des cinq lots, une fois les fichiers apurés (contrôle des adresses et recherches des doublons), Ipsos a procédé à :

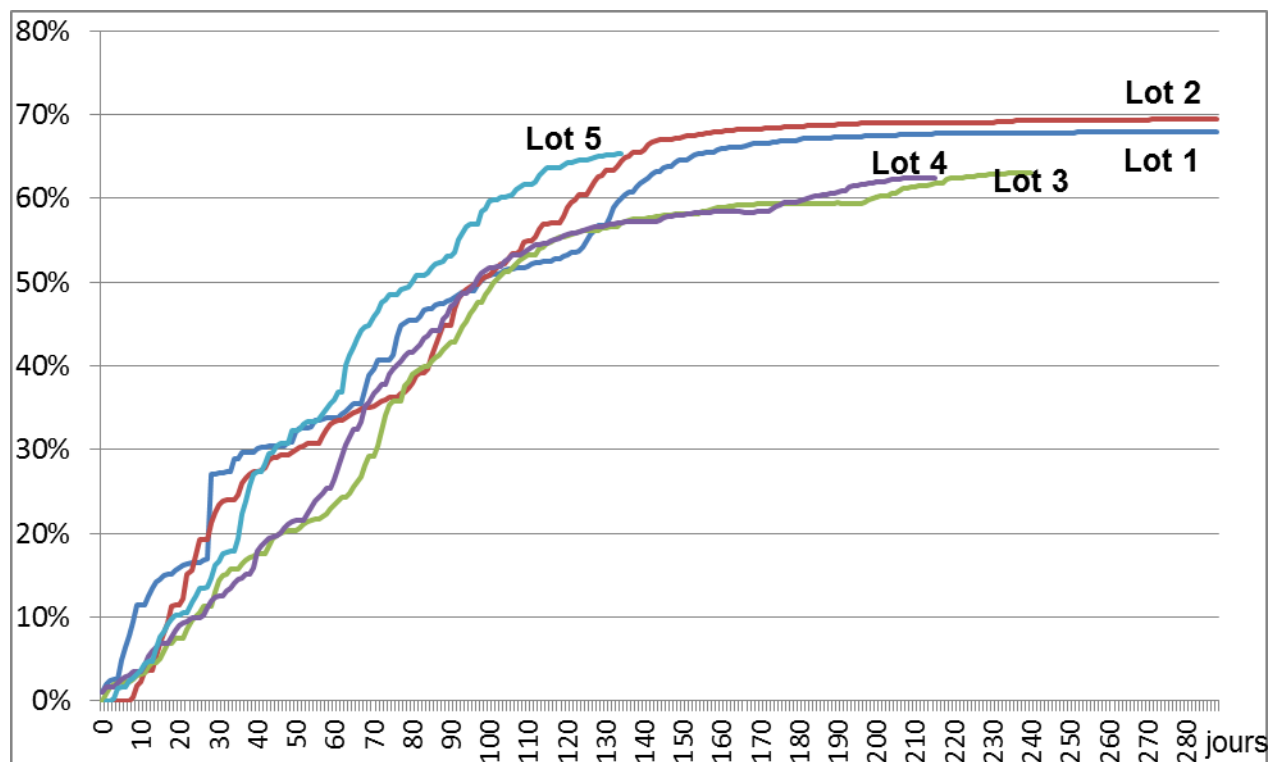
- 1) une phase de recherche téléphonique et de mail,
- 2) une animation téléphonique dans le but de confirmer l'adresse, présenter succinctement l'enquête, de récupérer le nom et éventuellement les coordonnées mail de l'interlocuteur le plus à même de répondre à l'enquête (directeur de l'établissement, DRH, etc.) afin de personnaliser l'envoi du questionnaire, Quelques retours négatifs concernent des établissements qui ont déménagé ou fusionné, ou qui s'avère être hors champ (pas de salarié, en cessation d'activité). Comme pour CT-2013, certains établissements ne se sentent pas concernés par l'enquête, et s'étonnent d'être enquêtés (ex commissariat de police, gendarmerie).
- 3) l'envoi de la documentation (lettre annonce + questionnaire),
- 4) une première relance papier ou par mail des non répondants,
- 5) une seconde relance papier ou par mail des non répondants,
- 6) une dernière relance par téléphone des non répondants.

L'envoi des documents papier a eu lieu environ deux semaines après la première phase de contact téléphonique avec une date de réponse attendue à trois semaines environ. Les deux relances téléphoniques devaient se faire, en cas de non réponse, trois semaines après l'échéance prévue pour la réponse, la dernière relance téléphonique de même. Or, on s'est aperçu dès le 2^{ème} lot que les relances étaient un peu trop rapprochées. Les délais ont donc été rallongés. En tenant compte aussi des périodes de vacances scolaires, le calendrier des opérations final est présenté dans le tableau 2.

Tableau 2 : calendrier des opérations réalisées par IPSOS pour chaque lot

	LOT 1	LOT 2	LOT 3	LOT 4	LOT 5
Réception du fichier par IPSOS	<i>18/11/2015 et 27/11/2015</i>	<i>01-déc</i>	<i>27-janv</i>	<i>28-avr</i>	<i>09-août</i>
Nettoyage du fichier	<i>du 18/11/2015 au 23/12/2015</i>	<i>du 05/01/2016 au 20/02/2016</i>	<i>Du 11/02/2016 au 01/04/2016</i>	<i>Du 02/05/2016 au 27/05/2016</i>	<i>Du 16/08/2016 au 16/09/2016</i>
Animation CATI	<i>05/01/2016 au 22/01/2016</i>	<i>23/02/2016 au 08/03/2016</i>	<i>12/05/2016 au 01/06/2016</i>	<i>15/06/2016 au 01/07/2016</i>	<i>19/09/2016 au 30/09/2016</i>
Envoi des questionnaires	<i>10-févr</i>	<i>21-mars</i>	<i>28-juin</i>	<i>21-juil</i>	<i>13-oct</i>
Relance 1 (Mail + Papier)	<i>29-mars</i>	<i>13-mai</i>	<i>08-août</i>	<i>29-août</i>	<i>14-nov</i>
Relance 2 (Mail + Papier)	<i>13-mai</i>	<i>20-juin</i>	<i>05-sept</i>	<i>28-sept</i>	<i>12-déc</i>
Relance Téléphonique	<i>08/06/2016 au 29/06/2016</i>	<i>11/07/2016 au 29/07/2016</i>	<i>26/09/2016 au 12/10/2016</i>	<i>20/10/2016 au 04/11/2016</i>	<i>09/01/2016 au 17/01/2017</i>
Relance mail	<i>11-juil</i>	<i>08-août</i>	<i>07-nov</i>	<i>21-nov</i>	<i>31-janv</i>
Fin du terrain	<i>05-août</i>	<i>02-sept</i>	<i>11-nov</i>	<i>09-déc</i>	<i>17-févr</i>
Saisie	<i>12-août</i>	<i>09-sept</i>	<i>18-nov</i>	<i>16-déc</i>	<i>24-févr</i>

Graphique 1 : Montée en charge des taux de réponse des 5 lots de CT-RPS 2016, en jours depuis l'envoi des questionnaires



Durée de remplissage des questionnaires

89 % des répondants ont répondu à la question sur le temps mis pour le remplissage du questionnaire (tableau 3). 4 % des établissements ont répondu « plus de 4 heures ». Pour rappel, le questionnaire pouvait être rempli par plusieurs personnes (le responsable de l'établissement, le DRH, le responsable santé-sécurité, le comptable...).

Hors durées déclarées supérieures à 4h, la durée moyenne des déclarations est de 60 minutes. Le questionnaire SH étant plus long, le remplissage a nécessité 82 mn contre 60 mn pour la FPET et 57 mn pour le SM.

Dans l'ensemble, 52 % des répondants ont mis moins d'une heure pour remplir le questionnaire.

Tableau 3 : durée de remplissage du questionnaire

	Non réponse	Moins de 1 heure	1 à 3 heures	3 à 4 heures	+ de 4 heures	Total	Nombre d'établissements
SH	11,2	28,0	45,3	5,9	9,7	100	1 349
FPET	10,2	47,7	35,3	3,2	3,7	100	3 700
SM	11,5	48,1	34,4	2,6	3,4	100	8 423
Ensemble	11,1	46,0	35,8	3,1	4,1	100	13 472

Tableau 4 : Répartition des établissements contactés selon l'échantillon d'origine et le type de questionnaire, hors doublons

Échantillon d'origine	Type de questionnaire			Total
	SH	FPET	SM	
SIR	5 %	15 %	80 %	100 %
QAO	12 %	35 %	54 %	100 %
Ensemble	9 %	28 %	63 %	100 %

La surreprésentation des agents de la fonction publique et des salariés des cliniques privées dans le volet individu se traduit par une surreprésentation des établissements SH et de la FP dans l'échantillon QAO par rapport à l'échantillon SIR (à peu près les mêmes chiffres que CT 2013).

d) Réception des fichiers d'enquête

Ipsos a livré la version finale des données d'enquête en mars 2017, à savoir :

- Le fichier des réponses des établissements (13 609 lignes, 3 questionnaires confondus).
- Le fichier de gestion tous secteurs (24 047 lignes) :

contenant les données les variables de gestion des adresses : raison sociale, enseigne..., nouvelle adresse récupérée par Ipsos, numéro de téléphone, date d'appel, d'envoi + dates des relances, date de réception questionnaire, statut du questionnaire (Papier / Cawi-complet ou Cawi-incomplet) + quelques variables (motif de non réponse, commentaires divers, repérage doublons a posteriori...).

Le fichier de gestion permet d'identifier l'échantillon de provenance des établissements et ainsi repérer les doublons entre échantillon QAO et échantillon SIR ainsi que les doublons intra-échantillon.

Le tableau 5 indique le nombre et la nature des doublons identifiés parmi les 24 047 établissements traités par Ipsos :

- dans les 7 000 établissements de l'échantillon SIR, 28 étaient des doublons (code SS),
- 1 388 établissements sont communs aux deux échantillons (code QS),
- au sein de l'échantillon QAO, 3 079 adresses étaient des doublons (mais n'avaient pas été repérées par Ipsos avant l'envoi de questionnaire : code QQ)

Tableau 5 : doublons par type dans l'échantillon d'établissements traités par Ipsos

Code doublon	Nombre	
S	6 899	28,7
SS	28	0,1
QS	1 388	5,8
Q	12 653	52,6
QQ	3 079	12,8
Ensemble	24 047	100

2. Délimitation des répondants et apurement

a) Définition de la non réponse totale

Pour l'ensemble des établissements, on calcule le nombre de questions vides dans les différentes parties des questionnaires (A à G).

5 établissements ont rempli deux questionnaires (de secteurs différents), 133 ont rempli un questionnaire papier et un Cawi. Par ailleurs, certains établissements ont commencé à répondre sur CAWI et ont ensuite renvoyé un questionnaire papier. Dans chacun de ces cas, on ne retient que le questionnaire le mieux rempli (tableau 6 ; colonne C1 à C2).

Ensuite, afin de ne pas gêner l'exploitation, on a choisi comme pour CT 2013 de ne considérer comme répondants que les questionnaires pour lesquels chacune des trois premières parties était, même très partiellement, remplie. Aucun établissement n'est éliminé à cette étape (C3=C2).

À l'issue de ces étapes, il reste 13 471 questionnaires établissements validés sur lesquels porteront les analyses de taux de réponse. Fin juin 2018 une brochure des premiers résultats de l'enquête a été envoyée à tous ces établissements répondants.

Enfin, certains employeurs de l'échantillon QAO ont répondu alors que leur salarié a abandonné en cours de questionnaire individuel ou a été interrogé à tort³. Dans ce cas, le salarié étant non répondant, l'établissement n'aura pas de poids de tirage, ces établissements (38) ont donc été supprimés des fichiers d'exploitation (C4).

Tableau 6 : Premier apurement du fichier livré par le prestataire

	C1	C2	C3	C4
	livraison Ipsos	Suppression des doubles questionnaires et sélection du questionnaire le mieux rempli	Sélection des questionnaires pour lesquels les parties A, B et C sont au moins un peu renseignées	Suppression des établissements QAO avec non réponse partielle du salarié (pas de poids de tirage)
SH	1 363	1 349	1 349	1 345
FPET	3 736	3 699	3 699	3 690
SM	8 510	8 423	8 423	8 398
Fichier de gestion	13 609	13 471	13 471	13 433

³ Une centaine d'individus ont été mis en collecte à tort par l'Insee et ont néanmoins accepté de répondre au volet individus et de communiquer les coordonnées de leur employeur. Ils n'ont été identifiés et supprimés du volet individus qu'en fin de collecte.

b) Analyse des taux de réponse

Le taux de réponse total s'élève à 69 %, soit 9,5 points de plus que pour CT 2013 (tableau 4). C'est dans le secteur hospitalier qu'il est le plus élevé (73,8 %).

Lors de la collecte du volet employeur de CT-RPS 2016, 42 % des établissements ont répondu à la version internet du questionnaire (47 % dans le secteur hospitalier, 39 % dans le secteur marchand et associatif).

Tableau 7 : Comportement de réponse selon le secteur

	SH	FPET	SM	Ensemble
Échantillon hors doublons	1 829	5 414	12 286	19 529
Questionnaires validés	1 349	3 699	8 423	13 471
Taux de réponse (% Quest validés / Échantillon hors doublons)	74 %	68 %	68 %	69 %
Part de la réponse papier	53 %	55 %	61 %	58 %

Les différences de taux de réponse sont également très marquées selon l'échantillon : les établissements du SH et la FPET échantillonnés dans SIR répondent beaucoup mieux que ceux issus de l'échantillon QAO (tableau 9). C'est le contraire pour le secteur SM. Une explication peut être que dans l'échantillon SIR l'adresse est plus précise et donc la cible est plus souvent atteinte (dans QAO, l'employeur a modifié 2 fois plus souvent que dans SIR des éléments d'adresse et/ou l'APET).

Concernant le taux de réponse du secteur marchand et associatif (SM), c'est la présence plus fréquente d'établissements de petite taille que dans le SH ou la FPET qui explique le moindre taux de réponse (tableau 10).

Au total, le taux de réponse de l'échantillon SIR est un peu moins élevé. Les réponses des établissements ont été prises en compte jusqu'au terme de l'enquête, d'où des taux de réponse plus importants pour les premiers lots de QAO (cf. aussi graphique 1).

Tableau 8 : Établissements contactés et répondants par lot

Nombre	SIR	QAO				Total
	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5	
Salariés interrogés dans QAO	-	4 218	6 359	5 358	1 913	17 848
Établissements contactés	6 899	3 505	4 700	3 031	1 394	19 529
Établissements répondants	4 704	2 475	3 285	2 055	953	13 471

Tableau 9 : Taux de réponse par lot et par type de questionnaires

Taux de réponse	SIR (=lot1)	QAO (Lot2 à 5)	Dont QAO...				Ens
			lot2	lot3	lot4	lot5	
SH	77,6	73,3	78,1	72,1	69,4	66,1	74,1
FP	73,2	67,1	66,9	67,2	64,4	74,8	68,2
SM	66,7	69,3	70,3	70,2	69	65,2	68,1
Ensemble	68,2	69	70,4	69,3	67,4	68,2	68,7

Tableau 10. Taux de réponse selon l'échantillon et la taille de l'établissement (EFF_ET dans SIR et NBSALA dans QAO par tranche de taille)

ORIGINE=SIR		ORIGINE=QAO		
Eff et	TR		NBSALA	TR
1 à 9	58,9	1-9	-	-
10-19	68,4	10-19	3	67,5
20-49	70,0	20-49	4	68,6
50-199	71,4	50-199	5	70,9
200-499	72,1	200-499	6	71,2
500 et +	73,1	500 et +	7	70,1
Ensemble 10 salariés et +	71,1	Ensemble		69,0

Pour les établissements de 10 salariés et plus :
Taux de réponse de 71 % pour l'échantillon SIR contre 69 % pour l'échantillon QAO, sans doute un meilleur ciblage de l'établissement dans Sirius

c) Apurement des réponses

a- Apurement des réponses incohérentes :

- pour certaines variables, les modalités « autre » et « non concerné » ont été cochées : « non concerné » est mis à blanc,
- il arrive aussi que la modalité « NSP » et un ou plusieurs autres items soient cochés : « NSP » est mis à blanc,

b- Les batteries de réponses oui/non : lorsque certains « Oui » sont cochés et qu'il n'y a pas de réponse pour les autres questions, la réponse « Non » est attribuée par défaut.

Exemple : les questions sur les outils d'ajustement du volume de travail (C3SM), la liste des changements organisationnels sur les 3 dernières années (C6aSM).

c- Apurement des « fausses » réponses multiples :

Il a été demandé à Ipsos de restituer les réponses telles qu'elles ont été renseignées par les établissements, donc des fausses réponses multiples ont été livrées. Par exemple il était en principe possible de dire que l'établissement avait eu recours à la sous-traitance pour plus de 20 % de son chiffre d'affaires et pour moins de 2 % du chiffre d'affaires. Ces fausses réponses multiples n'ont pas été très fréquentes, mais leur restitution supposait plusieurs variables (une variable indiquant si la case « plus de 20 % » est cochée et une variable indiquant si la case « moins de 2 % » est cochée. Un apurement a été fait sur ces questions pour fournir une seule réponse pour chacune de ces questions en privilégiant l'ordre d'apparition des items dans le questionnaire (hors NSP et non concerné) sauf pour climat social dans l'établissement où c'est inversé (priorité à « très tendu »).

d- Correction des variables numériques

Une correction des variables numériques a été nécessaire :

- les effectifs ont parfois été renseignés avec des précisions en toutes lettres,
- les effectifs et les pourcentages ont été parfois mélangés et le « % » a parfois été saisi,
- calcul du pourcentage à partir des effectifs quand c'est possible et vice versa,

- mise à blanc des variables mal remplies (Ex. pourcentage >100%).

e- Imputation des valeurs manquantes pour les effectifs déclarés

Dans la mesure où la plupart des études distingueront les établissements selon leur taille une variable EFFECTIF_TR a été calculée en 8 tranches (1 à 4, 5 à 9, 10 à 19, 20 à 49, 50 à 199, 200 à 499, 500 à 999, 1 000 et plus).

A la question « Combien votre établissement emploie-t-il de personnes (*effectif rémunéré le mois précédant l'enquête*) » 4,1 % des employeurs n'ont rien répondu (3,5 % du SM, 3,7 % du SH et 5,5 % de la FP).

Pour les SIR, la comparaison entre d'une part les effectifs déclarés par les employeurs et d'autre part ceux de SIRENE montre qu'environ, 77,4 % des valeurs sont dans les mêmes tranches (moins dans la FP) et 18 % sont dans des tranches consécutives.

Pour l'échantillon QAO, confrontées aux effectifs déclarés par le salarié (en tranches), celles déclarées par les employeurs sont moins concordantes (64 % dans la même tranche, 13 % sont dans la tranche au-dessus et 13 % dans la tranche en dessous).

La priorité a été donnée à l'effectif déclaré par l'employeur, à défaut, à l'effectif s'il est retrouvé de Sirene et en dernier, à l'effectif déclaré par le salarié (valeur modale de la tranche déclarée).

Afin de traiter les derniers cas de valeurs manquantes (3 dans FP4 dans SM), l'effectif a été imputé à la valeur modale dans les trois secteurs (tranche 5 : 50 à 199) après avoir vérifié qu'aucun des effectifs déclarés (salariés en SMR, télétravail, etc.) ne contredisait cette affectation.

f- Apurement de la variable APET

La variable d'activité économique selon la NAF rev2 n'est pas toujours remplie ou a été modifiée par l'établissement.

Pour son apurement, 5 variables sont à notre disposition :

- APET provenant de Sirene pour l'échantillon SIR
- XICORRECT4_4 Activité en clair pré-remplie à l'issue du traitement des échos,
- XIMODIFIE4_4 Activité en clair corrigée ou complétée par l'établissement,
- XICORRECT5_5 Code APET pré-rempli à l'issue du traitement des échos pour QAO ; code APET pré-rempli provenant de Sirene pour SIR
- XIMODIFIE5_5 Code APET corrigée ou complétée par l'établissement.

Dans certains cas, on ne dispose que du libellé d'activité en clair (pré-rempli ou rempli par l'établissement) et le code APET de la NAF rev2 est à blanc (pas de pré-remplissage, pas de correction par l'établissement), il faut le retrouver sur la base de l'activité en clair.

Dans d'autres cas, le code APET pré-rempli n'est pas corrigé mais l'établissement a corrigé l'activité en clair. Il faut donc trouver le code APET de la correction et appliquer cette correction.

Dans les 13 471 questionnaires des établissements répondants :

- 11 293 établissements n'ont pas modifié le code APET pré-rempli,
- 1 031 ont modifié le code APET pré-rempli,
- 243 ont confirmé le code Apet pré-rempli (code corrigé = code pré-rempli),
- 904 n'ont pas de code Apet (pas de code pré-rempli, pas de correction par l'établissement).

Les étapes pour apurer l'Apet sont les suivantes :

* On a récupéré la raison sociale dans le fichier de gestion pour vérifier l'activité

* les 7 999 ont un apet corrigé à 5 (ou 4) positions ;
* Reste à vérifier manuellement dans Excel selon le remplissage des 4 variables X1... pré-remplies/modifiées, moyennant la variable ;
remp_apet=compress(code_prerempli|txt_prerempli|code_modifié|txt_modifié).

433 établissements ont remp_apet = '0100' : seul l'activité en clair était pré-remplie,
470 établissements ont remp_apet = '0101' : l'activité était pré-remplie en clair et corrigée en clair,
1 828 établissements ont remp_apet = '1001' : le code était pré-rempli et corrigé seulement en clair,
2 741 établissements ont remp_apet = '1101' : l'activité était pré-rempli en code et en clair, et corrigée en clair seulement.

Au total, 5 472 Apet ont alors été corrigées : on a affecté un code activité corrigé.

Ensuite, on récupère les Apet déclarées en nafrev1 (anciens codes sur 4 positions) ou incohérentes ;

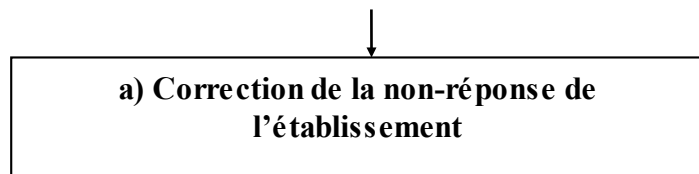
(Ex. dans SM : 70 dont 14 Apet incohérents), on utilise les tables de passage.

On recherche les idempl où il y a eu problème d'affectation en rev2, on les corrige à la main.

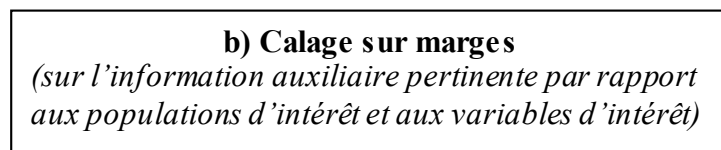
3. Pondération du fichier couplé

Le redressement du fichier couplé a suivi les étapes suivantes :

Poids calés individuels transversaux issus du volet « Actifs occupés »



Poids individuels corrigés de la nr de l'établissement



*Poids finaux
(POND_COUPLE / POND_COUPLE_QAA)*

a) Correction de la non-réponse de l'établissement

Définition de la non-réponse totale au niveau établissement

Définition des actifs en emploi dans le champ du volet Employeur

Le volet Employeur concerne les actifs occupés qui ont répondu :

- qu'ils étaient salariés d'établissements publics ou privés (salariés de particuliers et non-salariés sont exclus du champ)
- que l'établissement où ils travaillent comptait 10 salariés ou plus
- et que l'établissement était domicilié en France.

18 147 salariés sont ainsi dans le champ : leur établissement a été contacté (ou aurait dû l'être).

Non-réponse totale

On considère comme **non-répondant total** tout établissement qui n'a pas rempli au moins une question de la partie A, une de la partie B et une de la partie C (cf. supra).

Cela comprend des établissements qui n'ont pas été traités pour défaut d'informations sur leur adresse, qu'Ipsos n'a pas réussi à contacter, qui n'ont pas répondu ou qui ont renvoyé un questionnaire presque vide.

Réponse

Les établissements répondants sont tous les autres.

Tableau 11 : Répartition des salariés selon la réponse de l'employeur

	Effectif enquête	Taux	Effectif pondéré (PONDCAL)
Ensemble des individus interrogés	24 640		26 730 898
Individus du champ du volet Employeur	18 147		1 8343 790
Individus dont l'établissement a répondu	12 085	66,6 %	
Individus dont l'établissement n'a pas répondu	4 867	26,8 %	
<i>dont : Le salarié n'a pas donné assez d'info</i>	15	0,1 %	
<i>L'établissement a été contacté mais n'a pas répondu</i>	4 852	26,7 %	
Individus dont l'établissement n'a pas été contacté	1 195	6,6 %	
<i>dont : un établissement doublon a été identifié</i>	252	1,4 %	
<i>dont : l'établissement doublon est répondant</i>	187	1 %	
Individus du champ du volet Employeur pour lesquels un établissement a répondu	12 271	67,6 %	
Individus du champ du volet Employeur pour lesquels un établissement a été mis contacté/identifié mais n'a pas répondu	4 932	27,2 %	
Individus du champ du volet Employeur dont l'établissement n'a pas été contacté ou le salarié n'a pas donné assez d'info	943	5,2 %	

Une des craintes initiales générées par la méthodologie retenue pour le volet employeurs - qui est de s'appuyer sur les salariés - était l'introduction d'un biais : on pouvait croire que le salarié refuserait de donner l'adresse de son établissement lorsqu'il s'y sentait en mauvaise posture, de peur des représailles éventuelles.

Mais les établissements qui n'ont pas été traités parce que le salarié n'a pas donné d'information suffisante (classés NTR pour « non traités ») sont assez rares (0,1 % pour l'ensemble des salariés du champ employeur)

Le traitement incomplet du lot 4 par le prestataire a eu pour conséquence que les données fournies par 1 195 salariés n'ont pas été traitées et que donc leur employeur n'a a priori pas été enquêté⁴. Un travail d'identification des doublons dans le fichier complet des établissements a été réalisé post-collecte : 253 de ces établissements étaient le doublon d'un établissement d'un autre lot⁵, 187 d'un établissement répondant. Cela a permis donc de compléter le fichier couplé de 187 individus⁶.

On n'observe pas de difficultés particulières de conditions de travail des salariés travaillant dans un établissement complètement écarté de la collecte (tableau 11.c).

Au final, le fichier couplé est composé de 12 271 couples salarié-établissement employeur, et couvre donc 67,6 % des salariés dans le champ du volet Employeur.

⁴ Ces établissements sont localisés dans 4 régions : Languedoc Roussillon, Auvergne, PACA et Rhône Alpes.

⁵ 27 établissements étaient également doublons mais d'un établissement présent uniquement dans le lot qui n'a pas été mis en collecte. Pour ces établissements, il n'a donc pas été possible de compléter le fichier couplé.

⁶ Ces salariés n'ayant en fin de compte pas participé au sondage indirect permettant d'échantillonner les établissements, leurs poids ne seront toutefois pas utilisés lors du calcul des pondérations du fichier établissements (voir partie IV). En effet, c'est uniquement via un autre salarié (ou le tirage dans SIRENE) que l'établissement a été échantillonné. Un calage spécifique sera réalisé sur le seul champ des salariés d'un établissement traité et ayant répondu.

Tableau 11.b : Établissements répondants et salariés de ces établissements selon le type de questionnaire :

	Ensemble	Secteur marchand	Fonction publique	Secteur hospitalier
Total établissements répondants	13 433	8 398	3 690	1 345
Etablissements répondants QAO	9 216	4 930	3 092	1 194
Nb salariés dont l'étab. a répondu	12 271	5 609	3 800	2 862

Tableau 11.c : Taux de réponse de l'établissement selon les réponses du salarié

	Non traités NTR	Non réponse (mais traités) NR	Réponse REP
Ensemble des individus interrogés	5,2	27,2	67,6
Plan de licenciement	4,6	29,6	65,9
Tension avec supérieurs	4,3	28,3	67,5
Tension avec collègues	4,3	28	67,7
Je suis fier de travailler dans cette entreprise, réponse jamais	5,8	29,1	65,1

Par la suite, on écarte les salariés dont l'établissement n'a pas été traité et/ou pour lesquels aucun établissement doublon n'a été mis en collecte.

Déterminants de la non-réponse établissement

Les variables qu'on peut mobiliser proviennent du volet « actif occupé » et sont de deux natures :

- les variables directement liées à l'établissement : localisation (RGES), nombre de salariés (NBSALA),
- les variables de conditions de travail du salarié (durée, pénibilité, accidents du travail...).

Comme pour le volet Individu, la variable la plus discriminante est la région de gestion : le taux de non réponse est de 38 % en Ile-de-France, 30 % dans les régions délégataires (qui traitent des départements franciliens), 21 % dans les régions de l'Est. Les taux de non réponse sont les plus élevés dans les DOMs (50 % pour Antilles-Guyane et 30 % pour La Réunion).

Afin d'expliquer la non-réponse des établissements, on a utilisé la méthode du Logit. Un grand nombre de variables ont été testées, celles retenues comme les plus pertinentes (avec un effet significatif au seuil de 5 % sur la non réponse) sont présentées dans le tableau suivant.

Tableau 12 : Les facteurs expliquant la réponse de l'établissement

Déclaration du salarié	Variable	Modalité	Taux de réponse	Odds ratio	Significativité
Caractéristiques de l'établissement	RGES (région)	Antilles-Guyane	50,2 %	0.404	***
		Est	78,9 %	1.444	***
		Ile-de-France	62,5 %	0.670	***
		Reste	72,1 %	-	réf.
	ACTIVFIN (activité de l'établissement)	HZ(transport, entreposage)	63,8 %	0.771	**
		IND (Industrie)	76,6 %	1.286	***
		JZ (Information, comm)	58,1 %	0.611	***
		QZ (santé humaine, action sociale ...)	76,2%	1.356	***
	NBSALA (tranche d'effectif salarié de l'établissement)	3 : 10 à 19	68,0%	0.775	***
		4 :	69,2%	0.822	***
		5 :	71,6%	0.944	NS
		Nr	66,7 %	0.699	***
		6 à 8 :	73,3 %	-	réf.
	AUTRETA (autre établissement dans l'entreprise ?)	1 : oui	69,9 %	0.782	***
3 : NSP		71,0 %	0.809	***	
2 : non		74,3 %	-	réf.	
Caractéristiques de la relation du salarié à son établissement	SYNDIC (adhérent ou sympathisant d'une organisation syndicale de salariés)	1 : adhérent	68,0%	0.872	**
		2 : sympathisant	69,7%	0.924	NS
		3 : Non	72,1 %	-	réf.
	EVACRIT (au moins un entretien d'évaluation par an portant sur des critères précis et mesurables)	1 : oui	72,0 %	0.908	**
Conditions de travail	Score de soutien social	0 : inférieur ou égal à 5	70,5 %	1.173	***
		1 : égal à 6	73,3 %		réf.
	FORTMOD3 (restructuration, déménagement dans 12 derniers mois)	1 : oui	68,7 %	0.888	**
		2 : non	71,8 %	-	réf.
	RPA1J (pouvoir organiser son travail de la manière la manière qui convient le mieux)	1 : oui	71,8 %	1.106	**
		2 : non	69,6 %	-	réf.
	RPB4A (agression verbale du public)	1 : oui	70,3 %	0.896	**
		2 : non	71,5 %	-	réf.
	RPB4C (agression verbale des collègues, des supérieurs)	1 : oui	72,8 %	1.151	**
		2 : non	71,1 %	-	réf.

Champ : 16 952 salariés du champ « volet établissement ».

Construction des groupes homogènes de réponse

Pour construire des groupes homogènes de réponse, on a retenu ici la même méthode que pour le redressement de CT2013 : à partir de la variable numérique de probabilité prédite d'être « répondant », on construit des classes (procédures partnum et cahnum des macros SAS d'analyse des données). Ainsi les seuils sont fixés de façon à minimiser la variance intra-groupes et maximiser la variance inter-groupes. On obtient ainsi des groupes de tailles inégales mais adaptés du mieux possible à la distribution des probabilités prédites.

Tableau 13 : Sept groupes de réponse homogène

	N	Probabilités de réponse estimées			Proportion de réponse
		Moyenne	Minimum	Maximum	
1	518	47,3	31,8	52,4	51,1
2	965	57,7	52,5	61,0	55,4
3	1984	64,3	61,0	66,4	63,6
4	3050	68,5	66,4	70,4	69,5
5	2193	81,2	78,8	88,8	80,7
6	4126	72,4	70,4	74,4	72,4

7	4144	76,5	74,4	78,8	76,4
---	------	------	------	------	------

Tableau 14 : Caractéristiques des poids corrigés de la non-réponse

	Poids bruts	Poids redressés de la non réponse
Moyenne	979	1387
Médiane	413	577
Maximum	19940	31377
Minimum	1,9	2,4
Somme	12 008 871	17 024 329
Ecart type	1738	2488
Rapport interquartile	4,0	4,0
Rapport interdécile	18,2	19,5

Champ : 12 272 salariés dont l'établissement a répondu.

b) Calage sur marges

Comme pour le volet « actif occupé », on fait trois calages distincts :

- Un calage pour le secteur privé sur les données 2015 de l'enquête Emploi en continu (EEC2015) : 6 249 répondants ayant STATUT valant 4⁷ ou 6⁸.
- Un calage pour le secteur public sur SIASP 2015 : 5 439 répondants ayant un STATUT à 1, 2, 3, 5. L'objectif est d'avoir des effectifs dans la fonction publique qui soient les plus justes possibles, SIASP étant une information exhaustive sur ce sujet plus fiable que l'EEC.
- Un calage pour les Dom (475 répondants).

Sur les 12 272 salariés dont l'établissement a répondu, 690 n'ont pas répondu au QAA, soit 5,6 %, ce qui justifie l'introduction d'une pondération spécifique, POND_COUPLE_QAA.

Le calage dans le secteur privé

Les variables de calage sont dans le tableau ci-dessous (sexe, âge quinquennal, secteur d'activité en 17 positions, CS en 11 positions, statut en 5 positions).

Les marges sont calculées sur l'EEC 2015.

La structure de la population des répondants du secteur privé (après application des pondérations corrigeant de la non réponse) est assez proche de la population de référence sur les critères retenus.

⁷ Salarié d'un établissement de santé privé ayant une activité hospitalière, à but lucratif (clinique) ou non lucratif.

⁸ Salarié d'une entreprise privée ou publique, d'un artisan, d'une association (hors établissement de santé privé, hors secteur public social et médico-social), des caisses ou organismes de sécurité ou de protection sociale, des organismes consulaires, d'un groupement d'intérêt public (GIP) ou économique (GIE), d'un établissement public à caractère industriel ou commercial (EPIC).

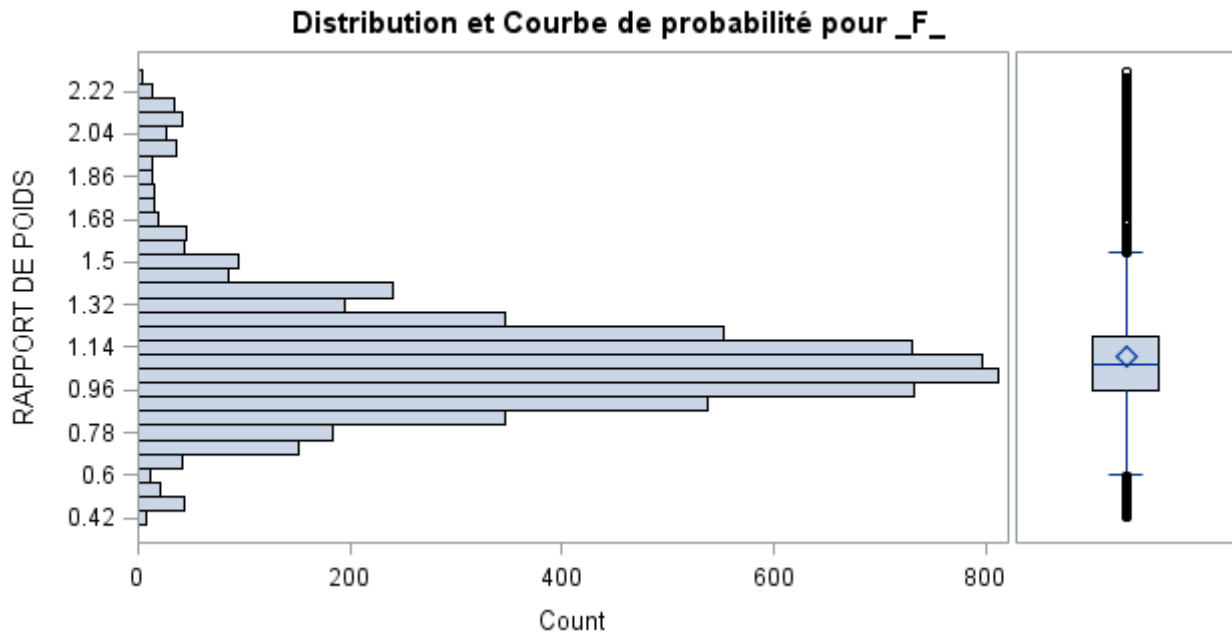
Tableau 15 : Les marges et la répartition dans CT secteur privé avant calage

Variable	Modalité ou variable	Marge échantillon	Marge population	Echantillo n (CT)	Populati on (EE)
Age	15-19	133 736	118 390	1,1	0,9
	20-24	751 445	924 617	6,2	7,1
	25-29	1 380 627	1 504 824	11,5	11,6
	30-34	1 677 733	1 646 155	13,9	12,7
	35-39	1 646 226	1 696 098	13,7	13,1
	40-44	1 647 164	1 812 854	13,7	14,0
	45-49	1 572 115	1 723 045	13,0	13,3
	50-54	1 630 394	1 699 950	13,5	13,1
	55-60	1 210 000	1 325 925	10,0	10,2
	60-64	345 590	454 105	2,9	3,5
	65 et plus	60467	80266	0,5	0,6
CSEI	Non précisée	99467	49422	0,8	0,4
	Chefs d'entreprise de 10 salariés ou plus	62113	88950	0,5	0,7
	Professions libérales	68268	38738	0,6	0,3
	Cadres de la FP, prof. intellectuelles et artistiques	212 696	300 903	1,8	2,3
	Cadres d'entreprise	2 261 264	2 257 336	18,8	17,4
	Prof. interm. enseignement, santé, FP et assimilés	637 586	679 013	5,3	5,2
	Professions intermédiaires administratives et commerciales des entreprises	1 175 802	1 334 944	9,8	10,3
	Techniciens	917 924	885 325	7,6	6,8
	Contremaîtres, agents de maîtrise	478 090	436 182	4,0	3,4
	Employés de la fonction publique	634 450	798 610	5,3	6,2
	Employés administratifs d'entreprise	893 617	1 085 719	7,4	8,4
	Employés de commerce	713 618	746 727	5,9	5,8
	Personnels des services directs aux particuliers	561 206	577 842	4,7	4,5
	Ouvriers qualifiés	2 210 293	2 459 422	18,3	18,9
	Ouvriers non qualifiés	1 069 818	1 173 718	8,9	9,0
	Ouvriers agricoles	59286	73378	0,5	0,6
Secteur d'activité	Non renseigné	55259	102 280	0,5	0,8
	Agriculture, sylviculture et pêche	475 387	477 436	3,9	3,7
	Fab denrées alimentaires, boissons et de pdts de tabac	409 146	442 477	3,4	3,4
	Fab d'équipemts élec, électro, info ; de machines	411 538	395 830	3,4	3,1
	Fabrication de matériels de transport	1 360 200	1 401 191	11,3	10,8
	Fab d'autres produits industriels + Cokéfaction et raffinage	305 574	307 849	2,5	2,4
	Ind extractives, énergie, eau, gest° des déchets et dépollut°	682 536	865 949	5,7	6,7
	Construction	1 706 367	2 007 279	14,2	15,5
	Commerce ; réparation d'automobiles et de motocycles	1 248 565	1 261 612	10,4	9,7
	Transports et entreposage	371 493	451 408	3,1	3,5
	Hébergement et restauration	529 431	570 292	4,4	4,4
	Information et communication	626 592	613 410	5,2	4,7
	Activités financières et d'assurance	151 561	205 641	1,3	1,6
	Activités immobilières	1 414 968	1 548 571	11,7	11,9
	Act scientifiqs et techniqls ; services admet de soutien	2 088 847	1 894 434	17,3	14,6
	Admin publique, enseignement, santé et action sociale	218 033	440 570	1,8	3,4
Autres activités de services	55259	102 280	0,5	0,8	
Sexe	Hommes	7 004 991	7 433 134	58,1	57,2
	Femmes	5 050 506	5 553 095	41,9	42,8
Statut et contrat	Médecins et cadres cliniques privées	19974	19321	0,2	0,2
	Prof. inter. cliniques privées	114 382	133 446	1,0	1,0
	Employés et ouvriers cliniques privées	117 741	121 108	1,0	0,9
	Salariés d'entreprises ou d'associations, en CDI	10 051 420	10 842 149	83,4	83,5
	Salariés d'entreprises ou d'associations, autres contrats	1 751 981	1 870 205	14,5	14,4

Le calage est réalisé avec la méthode logistique, en prenant comme bornes des rapports de poids 0,4 et 2,4. L'histogramme conserve ainsi une allure approximativement gaussienne (Graphique 2), recommandée par le département des méthodes statistiques.

Graphique 2 : Les rapports de poids lors du calage sur marges

Méthode : logit, inf=0.4, sup=2.4
 Statistiques sur les rapports de poids (= pondérations finales / pondérations
 et sur les pondérations finales



La méthode et les marges retenues sont strictement les mêmes pour les répondants au QAA. L'intervalle entre les bornes pour le calage sur marges est toutefois de plus grande amplitude (0,8-1,9) et on obtient des poids (PONDCOUPLE_QAA) légèrement plus dispersés que les poids pour le questionnaire individuel.

Les poids issus de ces deux calages sont très dispersés, avec un maximum très élevé (50 053 pour pondcouple). Une troncature des valeurs extrêmes supérieures a donc été réalisée : les poids supérieurs au 99^e percentile ont été ramenés à cette valeur (15 803 pour pondcouple), puis les poids modifiés ont à nouveau été calés sur les marges de sorte à garantir sur la somme des poids finaux calés soient égale aux effectifs dans la population.

Tableau 16 : La dispersion des poids

	POND_COUPLE	POND_COUPLE_QAA
N	6 249	5 854
Moyenne	2 078	2 218
Médiane	898	973
Minimum	1,7	1,9
Maximum	20 419	22 225
Rapport interquartile	4,3	4,3
Rapport interdécile	31,2	32,1

Le calage dans la fonction publique

Les variables de calage sont dans le tableau ci-dessous (sexe, tranches d'âge, catégorie hiérarchique et statut de la fonction publique (titulaire ou militaire/autres) croisé avec le versant de la fonction publique).

Les variables de l'enquête mobilisées sont TITPUB, STATUT, CLASSIF, SEXE et ANAIS.

Les marges sont calculées à partir de SIASP-Insee 2015.

La structure de la population des répondants de la fonction publique (après application des pondérations corrigées de la non réponse) est assez proche de la population de référence sur les critères retenus.

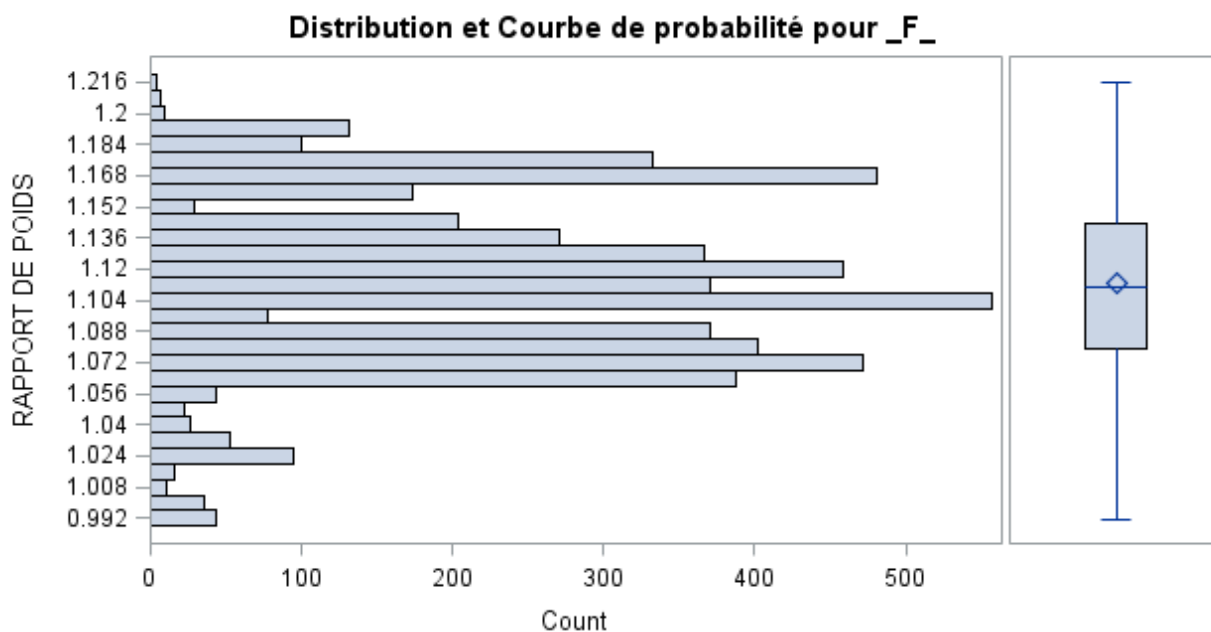
Tableau 17 : Les marges et la répartition dans CT fonction publique avant calage

Comparaison entre les marges tirées de l'échantillon (avec la pondération initiale) et les marges dans la population (marges du calage)

Variable	Modalité ou variable	Marge échantillon	Marge population	Pourcentage échantillon	Pourcentage population
AGCC	Moins de 30 ans	713 217	800 131	15,4	15,6
	30-49 ans	2 418 757	2 672 556	52,3	52,1
	50 ans et plus	1 488 958	1 660 224	32,2	32,3
CAT_HC	A	1 522 631	1 631 220	33,0	31,8
	B	902 241	1 052 778	19,5	20,5
	C	2 196 059	2 448 913	47,5	47,7
SEXE	1 – Homme	1 761 919	1 971 137	38,1	38,4
	2 - Femme	2 859 012	3 161 774	61,9	61,6
STAT_FP	Titulaires ou militaires FPE	1 448 650	1 598 238	31,4	31,1
	Titulaires ou militaires FPT	1 192 245	1 355 958	25,8	26,4
	Titulaires ou militaires FPH	730 311	816 540	15,8	15,9
	Non-titulaires et autres catégories et statuts FPE	537 613	589 530	11,6	11,5
	Non-titulaires et autres catégories et statuts FPT	383 559	431 927	8,3	8,4
	Non-titulaires et autres catégories et statuts FPH	328 554	340 718	7,1	6,6

Le calage est réalisé avec la méthode logistique, en prenant comme bornes des rapports de poids 0,9 et 1,3. L'histogramme conserve ainsi une allure relativement gaussienne (Graphique 3).

Graphique 3 : Histogramme des rapports de poids lors du calage sur marges fonction publique



Comme pour le calage du secteur privé, la méthode retenue pour le calage fonction publique est la même pour les répondants au QAA. Les marges sont rigoureusement les mêmes (bornes du rapport des poids 0,9 et 1,4).

Tableau 18 : Dispersion des poids dans la fonction publique

	POND_COUPLE	POND_COUPLE_QAA
N	5 548	5 296
Moyenne	925	969
Médiane	451	470
Minimum	5,2	5,5
Maximum	31619	34029
Rapport interquartile	3,1	3,1
Rapport interdécile	11,6	11,7

Le calage dans les DOMs

Les effectifs de répondants par département (de 76 à 172), rendaient difficile l'application d'un calage avec de multiples marges. Comme pour l'édition de 2013, il a donc été décidé de caler uniquement l'effectif total de salariés du champ par département avec celui de l'EEC-DOM, pour les répondants au QI et au QAA.

Tableau 19 : Calage dans les Doms

a) Les marges

DOM	Effectifs répondants QI (QAA)	Effectifs avant calage QI (QAA)	Effectifs EEC
Guadeloupe	112 (99)	71 285 (64 126)	80 560
Martinique	115 (98)	72 655 (56 172)	88 968
Guyane	76 (68)	44 144 (37 043)	44 672
La Réunion	172 (167)	159 816 (154 055)	182 781
Ensemble	457 (432)	347 900 (311 397)	396 982

b) La dispersion des poids

	N	Moyenne	Écart type	Minimum	Maximum
Pondcouple	475	836	729	205	7 061
Pondcouple_qaa	432	919	717	244	6 461

RGES	N	Moyenne	Écart type	Minimum	Maximum
Guadeloupe	112	719	469	368	2 940
Martinique	115	774	723	274	7 061
Guyane	76	588	518	205	3 669
La Réunion	172	1 063	879	406	6 228

c) Bilan et vérifications

Les pondérations finales

Il est assez logique compte tenu de la complexité du plan de sondage et des taux de sondage très différents selon les départements (DOM surtout) et les secteurs (public / privé) que la dispersion des poids soit importante.

Tableau 20 : La dispersion des poids

	Métropole		France entière	
	Pondcouple	Pondcouple_qaa	Pondcouple	Pondcouple_qaa
N	11797	11150	12272	11582
Moyenne	1 536	1 625	1 509	1 599
Médiane	637	674	634	675
Minimum	1,7	1,9	1,7	1,9
Maximum	31619	34029	31619	34029
Rapport interquartile	4,4	4,5	4,4	4,5
Rapport interdécile	20,8	21,1	20,8	21,1

Cette forte dispersion des pondérations est en grande partie héritée des pondérations individuelles, et même des pondérations brutes (notamment de l'échantillon principal). Elle a ensuite été augmentée par le tirage kish dans les ménages comprenant beaucoup d'actifs occupés, l'addition d'échantillons de provenances diverses et avec des taux de sondage très différents initialement (par exemple, les actifs occupés de Guyane ou les salariés métropolitains des cliniques privées ont une probabilité d'être interrogés bien plus forte que les salariés métropolitains de l'industrie ou de la banque, et donc des poids bien plus petits). Elle est enfin augmentée par les différentes étapes de redressement.

Tableau 21 : La dispersion des poids par statut du salarié

Variable STATUT (statut employeur, déclaré par le salarié)	N	Moyenne	Médiane	Minimum	Maximum	Rapport interquartile	Rapport interdécile
1 - salariés de l'État	2 248	1 010	480	112,8	31 619,5	3,54	32,91
2 - salariés d'une collectivité territoriale	1 720	1 083	577	83,2	17 535,7	2,12	8,94
3 - salariés d'un hôpital public	1 630	626	5 279	5,2	26 861,0	2,45	6,81
4 - salariés d'un établissement de santé privé ayant une activité hospitalière	820	343	229	1,7	8 193,7	4,18	15,46
5 - Salariés du secteur public social et médico- social	190	869	478	6,9	9 117,9	2,81	14,07
6 - salariés du privé (hors établissements de santé privé ou secteur public social et médico-social)	5 663	2 281	1 007	44,7	20 419,1	3,93	15,49

Effet de la pondération sur quelques variables d'intérêt

Pour vérifier l'effet de cette nouvelle étape de pondération sur les résultats, on compare les résultats sur les variables du questionnaire individuel, du questionnaire auto-administré et du questionnaire employeur (parties SM et SH), (tableau 20). Sur l'ensemble des variables, la différence entre l'application du poids individuel (colonne 2) ou l'application du poids couplé (colonne 3) ne modifie que très peu les résultats. La structure des salariés dont l'établissement a répondu était en effet peu déformée par rapport à celle des salariés du champ. Les écarts entre les résultats pondérés (colonne 3) et non pondérés (colonnes 1) reflètent l'effet de la prise en compte des surreprésentations dans la fonction publique et dans le secteur hospitalier privé. Ainsi les salariés de la FPH, surreprésentés, craignent moins de perdre leur emploi et travaillent plus souvent le dimanche, ce qui explique les écarts importants sur ces lignes. De même, les salariés de la fonction publique étant surreprésentés, la part des salariés en emploi stable parmi les répondants l'est aussi alors que les intérimaires sont sous-représentés. L'interprétation est un peu plus complexe quand on raisonne par type de questionnaire employeur, mais les variations sont assez limitées.

Les 10 % d'établissements ayant les poids les plus élevés contribuent pour 47 % aux résultats pondérés (tableau 21). **Lors de la publication des résultats il faudra donc se montrer prudent en cas d'effectifs limités.** Cette proportion n'est toutefois pas très variable selon les secteurs, ce qui montre qu'il n'y a pas de secteurs particulièrement mal lotis.

Tableau 22 : L'effet de différentes pondérations sur les variables d'intérêt

1- Variable de QAO

	CT-RPS 2016 France entière Salariés dont l'établissement a répondu		
	(1)	(2)	(3)
	Fichier couplé non pondéré	Poids salarié (pondcal pondqaa)	Poids couplé (pondcouple pondcouple_qaa)
n	12 271 (11 582)		
Questionnaire principal			
Intérimaire (TYPEEMPLOI=2)	0,9	2,5	2,3
Se trouve très bien ou bien payé (PAYECOM = 1 ou 2)	16,2	17,3	17,4
Porter des charges lourdes (CWLOURD=1)	41,2	38,1	38,4
Travailler habituellement le dimanche (DIMANCHE=1)	18,0	13,8	13,7
Un ou plusieurs accidents (ACCIDT=1,2,3)	10,0	9,8	9,9
Se dépêche toujours (DEPECH=1)	16,1	15,3	15,7
Craint pour son emploi (CRAINTE=1)	18,7	23,5	23,7
Doit toujours penser à trop de choses à la fois (A2B=1)	15,32	14,0	14,2
Faible bien-être (WHO5<13)	23,12	21,6	21,9

2- Variable du questionnaire « Employeur » SM (n=5 609), (SH n=1 498)

Plus de 50 % du CA en sous-traitance (A4BSM=1,2)	27,7	28,0	26,9
Votre + gros fournisseur appartient au même groupe ou réseau que votre établissement (A9SM=1)	12,5	11,5	11,5
Plus de 50 % des salariés ont été exposés à au moins une des 8 risques... (E1SM_9=1)	20 (52,5)	19,2 (53,6)	19,1 (53,9)
Un ou plusieurs accidents avec AT au cours des 12 derniers mois (E5aSM=1)	69,2 (94,8)	70,3 (92,2)	68,9 (92,2)
Un DUER a été élaboré ou actualisé au cours de 12 derniers mois (E13aSP=1)	85,2 (86,1)	84,6 (84,4)	83,7 (84,3)
Plus de 50 % des salariés sont exposés à la nécessité de travailler dans l'urgence (F1SM_1=1)	7,5 (24,9)	8,2 (27)	8,3 (27,2)
Plus de 50 % des salariés sont exposés à des tensions avec le public, la clientèle (F1SM_5=1)	2,5 (6,7)	2,9 (6,9)	3 (6,9)
Votre établissement est couvert par un CHSCT (G2aSM=1)	68,6 (96,3)	68,1 (95,5)	66,9 (95,6)
Climat social dans l'établissement: très calme (G4aSM = 1)	20,8 (6,2)	22,9 (5,1)	23,3 (4,9)
Climat social dans l'établissement: très tendu (G4aSM=4)	0,6 (1)	0,6 (1,4)	0,7 (1,5)
Sur les 12 derniers mois, l'effectif a diminué (G9SM=1)	20 (19,4)	19,5 (18,7)	19,4 (18,6)
Sur les 12 derniers mois, l'effectif a augmenté (G9SM=3)	51,4 (41,7)	50,4 (41,5)	50,8 (41,4)

Champ : salariés d'un établissement de 10 salariés ou plus, France entière.

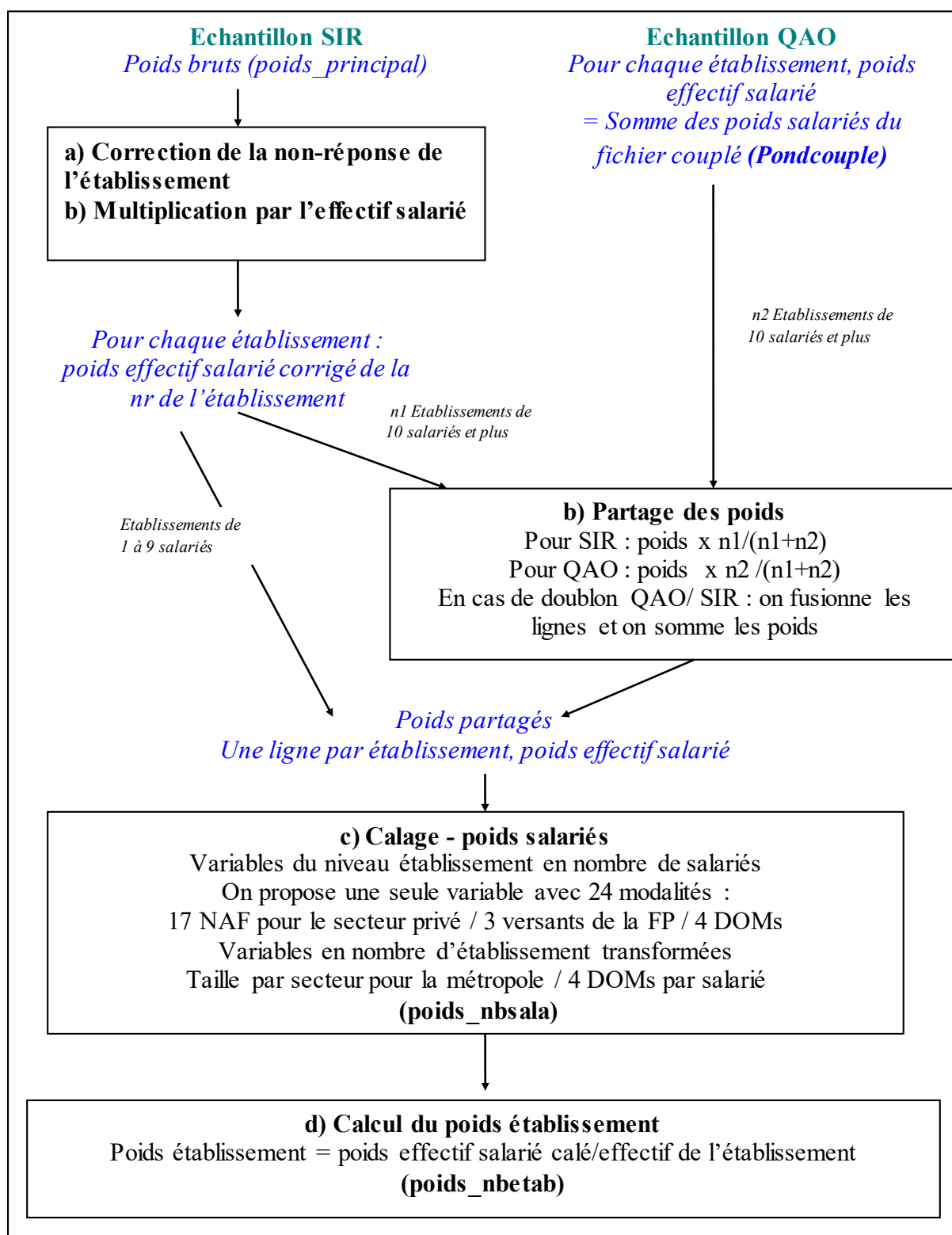
Tableau 23 : Dispersion des poids par secteur (France métro)

France Métro	N	Min	Max	Les 10 % d'établissements ayant les poids les plus élevés représentent ... % de l'ensemble pondéré	Rapport interdécile	Rapport interquartile
NR	12	323	15 656	52 %	22,4	6,4
AZ="Agriculture, sylviculture et pêche"	63	155	11 942	45 %	10,4	3,0
C1="Fabrication de denrées alimentaires, de boissons et de produits à base de tabac"	267	190	20 341	51 %	12,3	2,7
C2="Cokéfaction et raffinage"	16	84	3 306	34 %	6,1	2,2
C3="Fabrication d'équipements électriques, électroniques, informatiques ; fabrication de machines"	146	327	16 303	40 %	10,4	2,7
C4="Fabrication de matériels de transport"	288	86	9 464	41 %	15,5	4,4
C5="Fabrication d'autres produits industriels"	865	84	18 038	45 %	13,3	3,3
DE="Industries extractives, énergie, eau, gestion des déchets et dépollution"	201	217	17 209	41 %	13,2	3,4
FZ="Construction"	389	224	17 558	45 %	12,7	3,0
GZ="Commerce ; réparation d'automobiles et de motocycles"	824	189	20 126	44 %	12,3	3,8
HZ="Transports et entreposage"	498	107	20 251	46 %	17,7	3,5
IZ="Hébergement et restauration"	164	196	17 948	43 %	14,3	3,5
JZ="Information et communication"	248	167	15 978	40 %	14,1	4,6
KZ="Activités financières et d'assurance"	235	117	20 278	43 %	18,4	5,5
LZ="Activités immobilières"	70	180	18 144	44 %	16,1	3,3
MN="Activités scientifiques et techniques ; services administratifs et de soutien"	650	163	20 297	44 %	17,9	4,0
OQ="Administration publique, enseignement, santé humaine et action sociale"	7058	2	31 619	52 %	15,3	3,5
RU="Autres activités de services"	278	171	20 419	45 %	12,7	3,8
Ensemble	12 272	2	31 619	47 %	19,7	4,2

Champ : salariés d'un établissement de 10 salariés ou plus, France entière.

3. Pondération du fichier des établissements

Pour le fichier regroupant l'ensemble des établissements ayant répondu (échantillons QAO et SIR), l'objectif est d'obtenir une pondération assurant la représentativité par rapport à l'ensemble des secteurs économiques et des tailles d'établissements en France. Nous avons choisi de calculer une pondération en nombre de salariés, puis d'en déduire une pondération en nombre d'établissements. Le calcul de ces deux jeux de pondérations suit les étapes suivantes :



a) Poids établissements corrigés de la non réponse - fichier SIR

Définition de la réponse

Comme pour les établissements QAO, pour l'échantillon SIR, les établissements répondants sont ceux qui ont rempli au moins une question de chacune des 3 premières parties du questionnaire :

Tableau 24 : La réponse des établissements selon le type de questionnaire- CT-RPS 2016

	Échantillon SIR	NB étab. pondéré (Poids principal)	Répondants	Taux de réponse
Total établissements contactés	6 972	1 896 000	4 776	68,5
État et collectivités territoriales	1 018	1 711 800	752	73,9
Secteur hospitalier (public et privé)	382	11 400	302	79,1
Secteur marchand et associatif	5 572	173 000	3 722	66,8

Déterminants de la réponse établissement

Les variables mobilisables sont celles de la base de sondage présentes dans le fichier SIR_UMSE. Elles sont relativement peu nombreuses et ont presque toutes été intégrées à la régression logistique. Il s'agit de la région (regroupée en 3 modalités), du secteur d'activité (en 13 modalités), de la taille de l'établissement (en 6 tranches), de la qualité de siège (oui/non) et de l'ancienneté de création (en 3 tranches).

Tableau 25 : Les facteurs expliquant la réponse établissement RPS-2016

Variable	Modalité	Taux de réponse	Odds ratio	Significativité
REGION	DOM	58.2	0.494	***
	IDF	59.7	0.552	***
	Autres	71.6	-	Ref.
NAF	FZ- CST (Construction)	64.2	0.829	*
	IZ (Hébergement, restauration)	60.8	0.805	*
	MN (serv. adm, scientifique, technique)	64.1	0.911	Ns
	QZ (santé humaine, action sociale)	75.6	1.266	***
	Autres	69.0	-	Ref.
EFFECTIF	1. 1 à 5	57.3	0.501	***
	2. 6 à 19	68.3	0.920	Ns
	3. 20 à 49	70.0	-	Ref.
	4. 50 et +	72.7	1.179	*
Multi_SIEGE	0 : siège et mono-etab	69.0	-	Ref.
	1 : siège et multi-étab	77.2	1.035	Ns
	2 : non siège	61.9	0.516	***
ANCIEN	avant 2013	69.0	-	
	2014-2016	62.2	-	

Champ : 6 972 établissements de l'échantillon aléatoire direct (SIR) RPS-2016.

Construction des groupes homogènes de réponse

Pour cet échantillon SIR, la même méthode de redressement de la NR a été appliquée : Logit, suivi de la classification qui fournit 6 classes homogènes de répondants.

Tableau 26 : Six groupes de réponse homogène (RPS-2016)

	N	Probabilités de réponse estimées			Proportion de réponse
		Moyenne	Minimum	Maximum	
1	1 386	64,0	61,0	66,4	61,7
2	1 438	81,9	79,0	84,7	80,2
3	721	47,7	25,7	52,6	49,0
4	1 453	69,0	66,6	72,3	70,5
5	1 228	76,1	72,6	78,8	78,9
6	746	57,8	53,1	60,8	56,6

On multiplie ensuite les poids obtenus par l'effectif salarié (voir partie I-c pour l'apurement de cette variable).

Tableau 27 : Caractéristiques des poids corrigés de la non-réponse

	Poids bruts	Poids redressés de la non réponse	Poids salariés
Moyenne	255	419	5 302
Médiane	51	72	4 220
Maximum	1 244	2 221	1 114 653
Minimum	3	3	843
Écart type	378	657	19271
Rapport interquartile	18,8	20,8	1,7
Rapport interdécile	332	388	3,3
Rapport interpourcentile	367	578	10,5

Champ : 4 768 établissements répondants

b) Partage des poids

Les établissements de 10 salariés ou plus issus de l'échantillon QAO auraient pu être tirés dans l'échantillon SIR, et inversement. Pour tenir compte du recoupement des bases de sondage sur le champ des établissements des 10 salariés ou plus, il est donc nécessaire d'effectuer une table de partage des poids.

Pour vérification et avant partage, les poids corrigés de la non-réponse des établissements donnent en effet des effectifs salariés dans les établissements de plus de 10 salariés similaires à ceux obtenus sur le fichier couplé (tableau 26).

Tableau 28 : Effectifs de salariés dans les échantillons QAO et SIR avant partage

Champ >=10 salariés	MARGES Pour fichier couplé (cf. partie III)	Ech QAO 1 ligne par établissement	Répondants SIR Avec poids cnre (n= 3 796)
Secteur privé (métro)	12 986 229	12 622 076	15 363 551
Fonction publique (métro)	5 132 911	5 498 638	4 756 259
DOMs	396 982	402 924	372 485
Total	18 516 122	18 523 637	20 492 295

Pour les établissements uniquement présents dans l'échantillon SIR, le poids en nombre de salariés partagé est le suivant

$$\text{Pds_cnre} \times n1/(n1+n2)$$

Où n1 et le nombre d'établissements répondants de 10 salariés ou plus issus l'échantillon SIR et n2 le nombre d'établissements répondants de 10 salariés ou plus issus de l'échantillon QAO

Pour les établissements uniquement présents dans l'échantillon QAO, le poids en nombre de salariés partagé est le suivant

$$\text{Pds_cnre} \times n2/(n1+n2)$$

De plus, 522 établissements sont des doublons physiques présents dans les deux échantillons (tableau 29). Le poids est égal à la somme des poids SIR et QAO.

Tableau 29 : Répartition des établissements selon l'échantillon

		Effectif	Prise en compte des doublons entre QAO et SIR
Échantillon SIR	moins de 10 salariés	974	972
	10 salariés ou plus	3 794	3 274
			QAO et SIR : 522
Échantillon QAO (tous ont 10 salariés ou plus)		9 188	8 666
Total		13 956	13 434

Sur fond jaune les établissements pour lesquels les poids sont partagés.

c) Calage sur marges

L'objectif est d'obtenir pour chaque établissement deux poids :

- une pondération en nombre de salariés (**Poids_nbsala**), qui permet d'exprimer les résultats en nombre de salariés concernés par une statistique (par ex. nombres des salariés employés par un établissement ayant élaboré ou mis actualisé un document unique d'évaluation des risques au cours des 12 derniers mois).
- une pondération en nombre d'établissements (**Poids_nbetab**), qui permet d'exprimer les résultats en nombre d'établissements (par ex. nombres d'établissement ayant élaboré ou mis actualisé un document unique d'évaluation des risques au cours des 12 derniers mois).

Le poids établissement est obtenu en divisant, pour chaque établissement, le poids en nombre de salarié par l'effectif de l'établissement.

$$\text{Poids_nbetab} = \text{poids_nbsala} / \text{effectif_etab}$$

L'effectif retenu est la réponse au questionnaire Employeur, redressée par l'effectif de l'établissement dans Sirene (EFF3112_ET) ou à défaut par les données du QAO (taille déclarée par le salarié- NBSALA).

Les marges retenues

(i) En nombre de salariés

Peu de variables étaient disponibles au niveau salarié pour les établissements. Comme pour le calage du volet employeur de CT 2013, il a été choisi de ne construire qu'une seule variable en nombre de salariés : le secteur d'activité et le département en 24 modalités.

Cette variable a été déduite des marges utilisées lors des étapes précédentes (donc calculées sur les enquêtes emploi en métropole et dans les DOM et sur SIASP), en enlevant les salariés de particuliers et les non salariés. Cela permet d'avoir le champ complet des salariés d'établissements d'au moins un salarié.

Tableau 30 : marges en nombre de salariés

	MARGES
Secteur privé (métro)	16 626 876
dont : Agriculture, sylviculture et pêche	259 866
Fabrication de denrées alimentaires, de boissons et de produits à base de tabac	594 870
Cokéfaction et raffinage	9 019
Fabrication d'équipements électriques, électroniques, informatiques ; machines	457 295
Fabrication de matériels de transport	402 284
Fabrication d'autres produits industriels	1 532 139
Industries extractives, énergie, eau, gestion des déchets et dépollution	337 569
Construction	1 256 937
Commerce ; réparation d'automobiles et de motocycles	2 900 802
Transports et entreposage	1 370 218
Hébergement et restauration	775 296
Information et communication	657 599
Activités financières et d'assurance	798 761
Activités immobilières	326 945
Activités scientifiques et techniques ; services administratifs et de soutien	2 032 374

Administration publique, enseignement, santé humaine et action sociale	2 186 722
Autres activités de services	728 179
Fonction publique (métró)	5 452 619
dont : État	2 395 768*
Territoriale	1 898 541*
Hospitalière	1 158 310
DOM	497 550
Guadeloupe	101 359
Martinique	111 632
Guyane	53 584
Réunion	230 975
Total	22 577 045

Source : EEC 2015, SIASP 2015. Il a été in fine impossible de distinguer FPE et FPT : les établissements d'enseignement sont trop difficiles à affecter, locaux et personnels n'étant pas du même ressort.

Tableau 31 : Répartition des établissements par taille et secteur avec poids_nbsala

	Poids partagés, avant calage	poids_nbsala, après calage
Secteur privé (métró)	17 801 048	16 626 876
Fonction publique (métró)	5 580 650	5 452 619
dont : État et territoriale	4 519 722	4 294 309
Hospitalière	1 060 928	1 158 310
DOM	505 319	497 550
Total	22 887 017	22 577 045

(ii) En nombre d'établissements

Un calage uniquement sur une marge en nombre de salariés ne garantit néanmoins pas que les poids établissements seront représentatifs des établissements, notamment par taille.

Pour renforcer le calage de l'échantillon *d'établissements*, on va donc ajouter lors du calage en nombre de salariés des marges sur les effectifs en nombre d'établissements. Les variables retenues sont

- la taille par secteur (en 5 modalités) en métropole,
- le département pour les quatre DOM couverts par l'enquête.

La source CLAP de l'Insee fournit le nombre d'établissements actifs au 31 décembre 2015 d'au moins un salarié, par taille et par secteur⁹ :

⁹ <https://www.insee.fr/fr/statistiques/2021289>

Tableau 32 : marges en nombre d'établissement

		MARGES
Taille par secteur (mètre)		1 850 241
AZ - Agriculture, sylviculture et pêche	1 à 9 salariés	82246
	10 salariés ou plus	3 602
BE- Industrie	1 à 9 salariés	91742
	10 à 19 salariés	19738
	20 à 49 salariés	14606
	50 salariés ou plus	11254
FZ - Construction	1 à 9 salariés	161 063
	10 à 19 salariés	17333
	20 à 49 salariés	9 353
	50 salariés ou plus	3 205
GU - Commerce et services	1 à 9 salariés	989 459
	10 à 19 salariés	99471
	20 à 49 salariés	55648
	50 salariés ou plus	29383
OQ - Administration publique, enseignement, santé humaine et action sociale	1 à 9 salariés	154 676
	10 à 19 salariés	42881
	20 à 49 salariés	35620
	50 salariés ou plus	28961
DOM		45942
	Guadeloupe	11039
	Martinique	10528
	Guyane	4 491
	Réunion	19884
Total		1 896 183

Source : Clap 2015.

Le calage est réalisé en une seule fois sur les marges en nombre de salariés et en nombre d'établissements. Les poids en entrées du calage sont les poids corrigés de la non-réponse et partagés en nombre de salariés. Les poids en sortie du calage sont les poids_nbsala. On calcule ensuite $\text{poids_nbsala}_i = \frac{\text{poids_nbsala}_i}{\text{effectifs}}$.

Pour le calage sur les marges en nombre d'établissement, on a recours à une variable transformée \bar{X}_i qui permet de garantir, qu'après calage et calcul de poids_nbsala, la somme des poids par établissement est bien égale à la marge en nombre d'établissements.

On note \bar{X}_i la variable de calage transformée associée à chaque établissement. Elle est égale à

$$\bar{X}_i = \frac{X_i}{E_i},$$

où X_i la variable de marge exprimée au niveau établissement et E_i le nombre de salariés de l'établissement.

Par exemple, pour caler l'échantillon sur le nombre d'établissements dans le département de la Réunion, X_i est égale à 1 pour tous les établissements se situe à la Réunion, 0 pour les autres. La variable \bar{X}_i associée à l'établissement i est alors égale à $1/E_i$ si est un établissement est situé à la Réunion et 0 sinon.

À l'issue du calage, on obtient les poids en nombre de salariés poids_nbsala. Par définition, pour les variables de calage, ils vérifient l'égalité suivante :

$$\sum_i \text{poids_nbsala}_i * \bar{X}_i = N$$

où N est le nombre d'établissements à la Réunion.

Or, $\text{poids_nbetab}_i = \frac{\text{poids_nbsala}_i}{E_i}$ de sorte que l'égalité suivante est vérifiée pour poids_nbetab :

$$\sum_i \text{poids_nbsala}_i * \frac{X_i}{E_i} = N$$

$$\Leftrightarrow \sum_i \text{poids_nbetab}_i * X_i = N$$

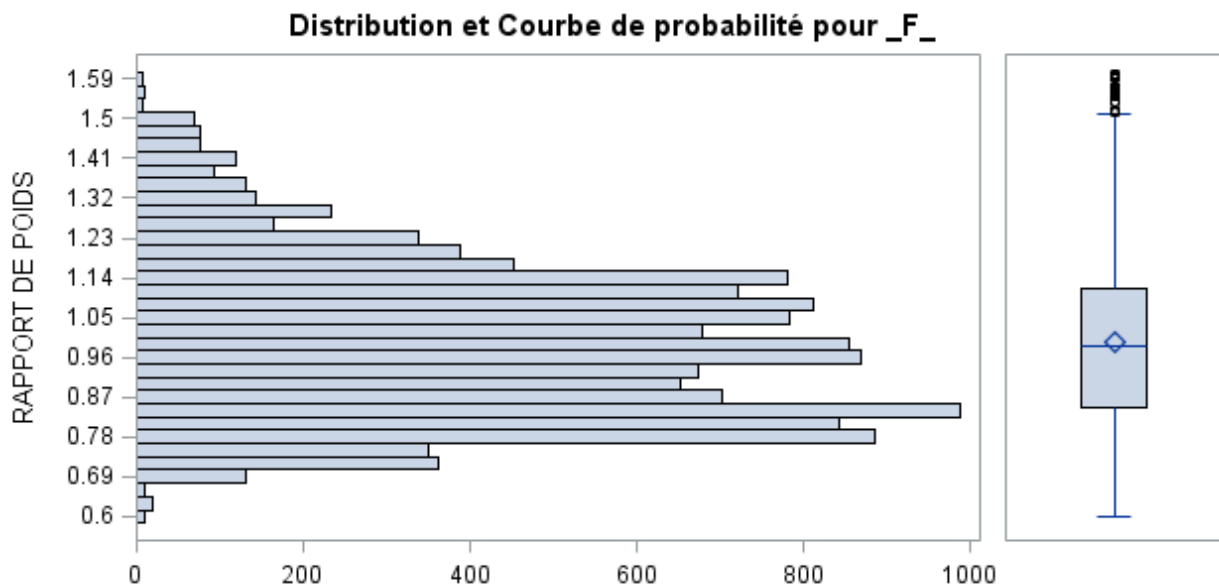
Ce calage sur variables transformés garantit donc bien que les poids_nbetab obtenus respectent les marges en nombre d'établissements.

Pour chaque établissement, on crée 22 variables transformées, soit une par marge en nombre d'établissements.

Le calage est réalisé avec la méthode logistique à partir des poids partagés en nombre de salariés à la fois sur la variable typmarge et sur les 22 variables transformées $\bar{X}_i = \frac{X_i}{E_i}$ pour les marges en nombre d'établissements.

On obtient les poids_nbsala calés. Les bornes des rapports des poids sont 0,6 et 1,6. L'histogramme conserve une allure gaussienne (graphique 4).

Graphique 4 : Les rapports des poids lors du calage sur marges



Finalement, on calcule poids_nbetab à partir de poids_nbetab. Par définition, ces poids sont également calés sur les marges issues de CLAP.

Tableau 33 : Dispersion des poids

	POIDS_NBSALA	POIDS_NBETAB
N	13 434	13 434
Moyenne	1 709	141
Médiane	990	11
Minimum	1,0	0,01
Maximum	233 525	13 052
Rapport interquartile	3,8	17,8
Rapport interdécile	16,3	243,7
Somme	22 960 327	1 896 183

Tableau 34 : Répartition des établissements par taille et secteur avec poids etab

	Secteur marchand et associatif	Secteur hospitalier (public et privé)	État et collectivités territoriales	Total
Établissements de moins de 10 salariés	1 414 958	3 595	98 879	1 517 432
Établissements de 10 salariés et plus	296 808	7 831	74 110	378 750
Total	1 711 766	11 426	172 990	1 896 182