

## REDRESSEMENT DES DONNÉES DE L'ENQUÊTE SUMER 2010

Décembre 2011

Le redressement vise à corriger l'échantillon enquêté de ses éventuelles déformations par rapport à la population cible de l'enquête. Chaque questionnaire est affecté d'un poids (coefficient multiplicateur) afin que les expositions professionnelles des 48 000 salariés répondants représentent correctement celles des 21,7 millions de salariés du champ enquêté (équivalent à 90 % de l'ensemble des salariés en 2010).

### SOMMAIRE

I - Poids de tirage.....	p. 3
II - Correction de la non-réponse totale.....	p. 13
III - Calage sur marges.....	p. 17
Annexes .....	p. 23
Bases d'informations disponibles pour la population de référence de médecins du travail et comparaisons avec les informations de l'échantillon des médecins-enquêteurs.....	p. 23
Variables explicatives de la non-réponse.....	p. 27
Taux appliqués pour la correction de la non-réponse totale pour chaque sous-population définies précédemment.....	p. 29

L'enquête Sumer 2010 s'est déroulée dans un contexte de changements forts. La réforme de la médecine du travail de 2004 a espacé les visites périodiques obligatoires (l'intervalle passant d'un à deux ans) et incité les entreprises qui souhaitent conserver une visite annuelle à déclarer leurs salariés comme devant bénéficier d'une Surveillance Médicale Renforcée (SMR). Alors qu'il existe une réelle pénurie des effectifs au sein de la profession, les médecins du travail tendent à concentrer leurs convocations sur les salariés en SMR ou sur ceux dont les expositions leur semblent le justifier particulièrement. Alors que ce n'était pas le cas en 2003, il devient plausible que la probabilité pour un salarié d'être convoqué en visite médicale (et donc d'être enquêté par Sumer) soit fortement corrélée aux risques liés à son poste de travail. Les médecins-enquêteurs verraient dans leur cabinet des salariés en moyenne plus exposés que les autres. En l'absence de correction spécifique de ce biais, l'enquête risquerait de surestimer les fréquences d'expositions.

D'autre part, le volontariat des médecins-enquêteurs a donné lieu à des critiques en termes de méthodologie d'enquête : il y aurait dans le relevé des expositions un biais potentiel lié au volontariat des médecins-enquêteurs. Aussi, alors que l'enquête Sumer 2010 est entrée dans le cadre des grandes enquêtes statistiques reconnues par le Conseil national de l'information statistique (Cnis) (avis d'opportunité et label de qualité statistique accordés respectivement en 2007 et 2008), il devient indispensable de renforcer la méthodologie de l'enquête et de son redressement.

C'est pourquoi la méthode de redressement de l'enquête Sumer 2010 a été complexifiée par rapport à celle de Sumer 2003.

Du protocole de collecte défini par le comité scientifique découle la procédure de pondération : le médecin se déclare enquêteur et c'est lui qui tire au sort les salariés enquêtés. La procédure de pondération sera donc assimilée à celle d'un sondage à deux degrés.

Ainsi, les quatre grandes étapes de ce redressement seront les suivantes.

- 1- Redressement en fonction des caractéristiques du médecin-enquêteur : pour corriger des éventuelles déformations liées au volontariat du médecin, on calcule des poids initiaux en fonction des caractéristiques du médecin-enquêteur comparées à celles de la population des médecins des différents secteurs professionnels.
- 2- Correction du biais induit par la corrélation entre fréquence des visites et fréquences des expositions.
- 3- Correction de la non-réponse totale par la méthode des groupes homogènes
- 4- Calage sur marges

Par la suite, nous allons préciser ces étapes du redressement et étudier au fur et à mesure l'impact de ces corrections sur quelques fréquences d'expositions :

- vibrations transmises aux membres supérieurs 10h ou plus par semaine
- bruit supérieur à 85db plus de 20h par semaine
- travail de nuit (99 nuits ou plus par an)
- poussières de bois
- manutention manuelle de charge
- travail le dimanche et jours fériés
- conduite sur la voie publique 20h ou plus par semaine
- travail sur écran 20h ou plus par semaine
- contraintes posturales 2h ou plus par semaine (genou, bras en l'air et autres contraintes posturales).

## I- Poids de tirage

La probabilité initiale d'être tiré au sort pour un salarié est la probabilité d'inclusion d'ordre 1 notée P :

$$P = P_a \times P_b \times P_c$$

Avec :

$P_a$  : Probabilité que le médecin soit volontaire

$P_b$  : Probabilité que le salarié ait une visite prévue avec son médecin pendant la période d'enquête

$P_c$  : Probabilité que ce salarié soit tiré au sort par son médecin

La pondération initiale est l'inverse de la probabilité d'inclusion P.

### 1) $P_a$ : Probabilité que le médecin soit volontaire

Pour calculer cette probabilité, on suppose que le volontariat du médecin est l'équivalent d'un tirage aléatoire : tous les médecins ont a priori la même probabilité de se porter volontaire pour réaliser l'enquête. Cependant, pour vérifier que cette hypothèse reflète correctement la réalité, le Comité scientifique de Sumer 2010 a recommandé la collecte d'informations permettant de comparer les médecins-enquêteurs avec l'ensemble des médecins du travail, afin de pouvoir redresser d'éventuels biais liés au volontariat. En cas de distorsion, ce redressement consiste donc à affecter à chaque médecin (et à chaque salarié interrogé par ce médecin) un coefficient  $1/P_a$  corrigeant en partie cet éventuel biais.

C'est donc par un **calage sur les marges de la population de référence des médecins par secteur d'exercice** qu'on calcule la probabilité  $P_a$  d'inclusion du salarié via son médecin du travail.

Pour rassembler l'information nécessaire à ce calage, les moyens mis en place en parallèle à la collecte sont :

- fiches médecins volontaires (informations sur les caractéristiques sociodémographiques des médecins-enquêteurs, leur charge de travail et les moyens à disposition) renseignées par les médecins-enquêteurs eux-mêmes.

- bases « informations de la population de référence médecins du travail » : bases de données au niveau régional pour le régime général et par secteur d'exercice plus spécifiques comme la MSA, Air France, la Poste, EDF-GDF, la SNCF, les Fonctions Publiques d'Etat, Territoriale et Hospitalière (respectivement FPE, FPT, FPH), les gens de mer. Ces bases d'informations auxiliaires ont été constituées par la Dares à partir des informations disponibles de la base régionale IMT2 de l'Inspection médicale du Travail pour le régime général et des sources des secteurs spécifiques. (Cf. ces informations disponibles en annexe 1)

Ces informations concernent l'âge et le sexe du médecin, son type de diplôme, le type de service (autonome ou interentreprises), sa charge de travail (nombre de salariés suivis, nombre de visites réalisées en 2007). On pouvait par exemple faire l'hypothèse que les médecins les plus surchargés (ceux qui ont le plus grand nombre de salariés à suivre) ont moins de disponibilité pour se porter volontaire pour Sumer, ce qui justifierait de redresser sur ce critère<sup>1</sup>.

### Étapes et mode de calcul

1- 71 % des médecins-enquêteurs ont renseigné leur « fiche médecin ». Or, il est nécessaire d'avoir l'ensemble des informations pour tous les médecins-enquêteurs pour pouvoir réaliser le calage sur marges. Un profil a donc été imputé à chaque médecin non-répondants des fiches médecins par la **méthode d'imputation de hot-deck hiérarchisé (HDH)**. Le fichier de l'ensemble des profils médecins-enquêteurs est ainsi complété.

**Hot-deck hiérarchisé** : la donnée manquante est remplacée par la valeur observée pour un individu répondant ayant les mêmes caractéristiques pour les variables  $X_1, X_2, \dots, X_k$ . S'il n'en

<sup>1</sup> On verra ci-dessous que ce n'est pas le cas.

existe pas, elle est remplacée par une unité ayant les mêmes valeurs pour X1, X2, ..., Xk-1 etc.

Les caractéristiques sélectives utilisées dans notre cas ont été : la région, le type de service (service autonome, interentreprises ou de prévention) et le secteur d'exercice (régime général, MSA, Air France, etc.).

2- Il est possible ensuite de réaliser le **calage sur marges des médecins (macro CALMAR de SAS) par secteur** en utilisant les bases d'informations auxiliaires : « informations population de référence médecins du travail » constituées par la Dares.

Il a été difficile de rassembler les informations pertinentes pour calculer les marges de calage « médecins du travail » : la base IMT2, incomplètement renseignée, a dû demander beaucoup d'extrapolations ; la plupart des secteurs d'exercice spécifiques comme la FPE, FPT, ou autres régies de transports urbains n'ont pas de base de données spécifique pour leur population de médecins du travail. Il a été nécessaire d'utiliser et de recroiser différentes sources pour chaque secteur.

Finalement, les distorsions liées au volontariat concernent le sexe et l'âge des médecins. Ainsi, par exemple, pour les médecins du régime général : les 1 341 médecins-enquêteurs volontaires sont plus souvent des femmes (76 %) que l'ensemble des 6 866 médecins du travail (64 %) du régime général (Annexe 1). Ils sont également plus jeunes : 55 % ont moins de 55 ans contre 45 % de l'ensemble des médecins. En revanche ils se distinguent peu sur les autres caractéristiques (type de service, diplôme, charge de travail) qui ne sont donc pas intégrées dans les variables servant au calage sur marges et ce, quel que soit le secteur. Ce calage sur marge est donc réalisé par secteur sur la structure par sexe et âge.

Cf. en annexe 1 : mise en parallèle des informations auxiliaires issues de la population de référence des médecins et des informations issues de l'échantillon des médecins-enquêteurs après HDH.

### Impact de Pa sur les fréquences d'expositions

	Fréquence d'exposition (en %)		Evolution après intégration de Pa (en points)
	Brut (sans pondération aucune)	Avec pondération 1/Pa	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	2,1	2,0	-0,1
Bruit supérieur à 85db plus de 20h par semaine	5,9	5,9	0
Travail de nuit (99 nuits ou plus par an)	2,9	2,9	0
Poussières de bois	1,6	1,6	0
Manutention manuelle de charge	37,3	37,4	+0,1
Travail le dimanche et jours fériés	34,9	34,9	0
Conduite sur la voie publique 20h ou plus par semaine	3,6	3,4	-0,2

Travail sur écran 20h ou plus par semaine	25,0	24,7	-0,3
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	20,1	19,6	-0,5

Les médecins-enquêteurs volontaires ont un peu plus de salariés exposés au travail sur écran 20h ou plus ou exposés aux contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes).

2) **Pb : Probabilité que le salarié ait une visite prévue avec son médecin pendant la période d'enquête**

Intuitivement :

$Pb = (\text{durée de la collecte}) / (\text{intervalle de temps moyen entre deux visites pour le salarié})$

Les pondérations  $1/Pb$  permettront de corriger le biais induit par la corrélation entre fréquence des visites et importance des expositions.

Supposons par exemple qu'un médecin réalise la collecte de l'enquête Sumer 2010 en 3 mois. Si l'un des salariés suivis vient une fois tous les deux ans (huit trimestres) pour une visite périodique, ce salarié a une chance sur huit d'avoir une visite durant la période d'enquête. Si un second salarié, plus fortement exposé, est déclaré SMR et est convoqué tous les ans, alors il aura une chance sur quatre d'avoir une visite durant la période d'enquête. Le salarié en SMR a donc une probabilité deux fois plus élevée d'être vu au cours du trimestre d'enquête : si l'autre salarié a un poids égal à 1, il faut affecter au salarié en SMR un poids de 0,5 pour que sa fréquence élevée d'exposition ne biaise pas vers le haut la moyenne de l'échantillon.

Pour calculer  $Pb$ , il faut donc connaître la fréquence des visites et la durée de collecte. Ces informations sont disponibles dans le questionnaire principal mais ont nécessité des corrections importantes.

**La fréquence des visites**

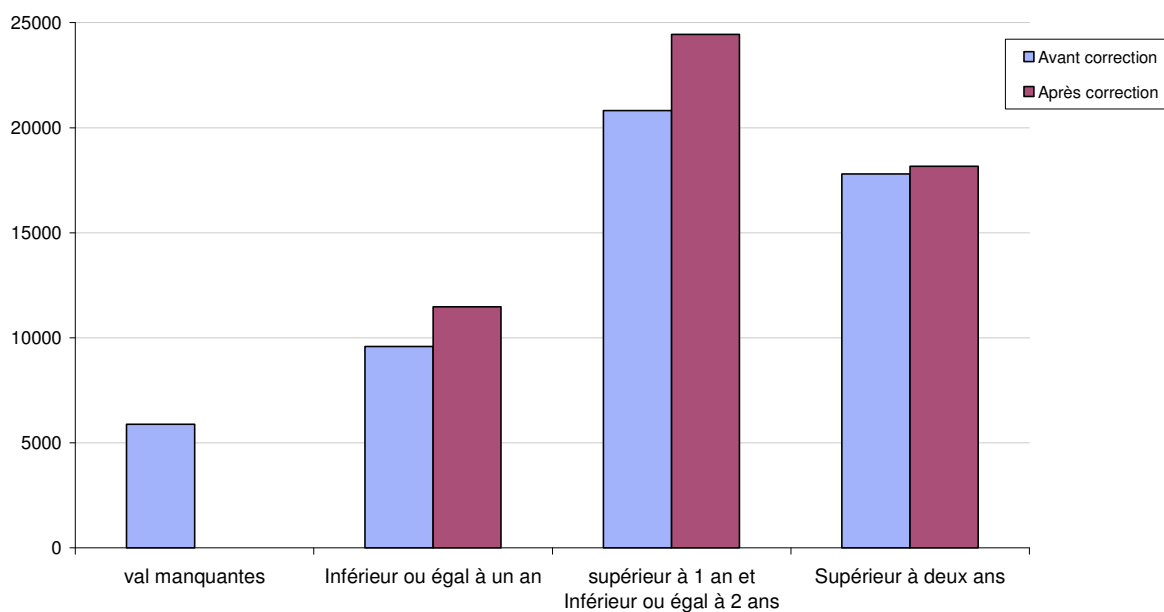
Deux informations sont disponibles :

- \* la fréquence réelle : on demande la date de la dernière visite. La fréquence réelle est donc la durée écoulée entre la visite précédente et la visite durant laquelle le salarié est enquêté (date de dernière visite DDVIS - date enquête DATENQ)
- \* la fréquence théorique (variable FVIS) : on demande à quelle fréquence le salarié doit effectuer une visite médicale.

Si le salarié a donné une réponse exploitable pour la fréquence réelle, on utilise cette réponse. Sinon, on prend la fréquence théorique de visite si elle est également renseignée. Enfin, si le salarié n'a pas répondu à ces deux questions, par convention on lui impute un intervalle d'un an s'il est déclaré en SMR, de deux ans sinon.

## Répartition des fréquences de visite avant et après apurement

Répartition des salariés selon la fréquence des visites, avant et après correction des valeurs manquantes.

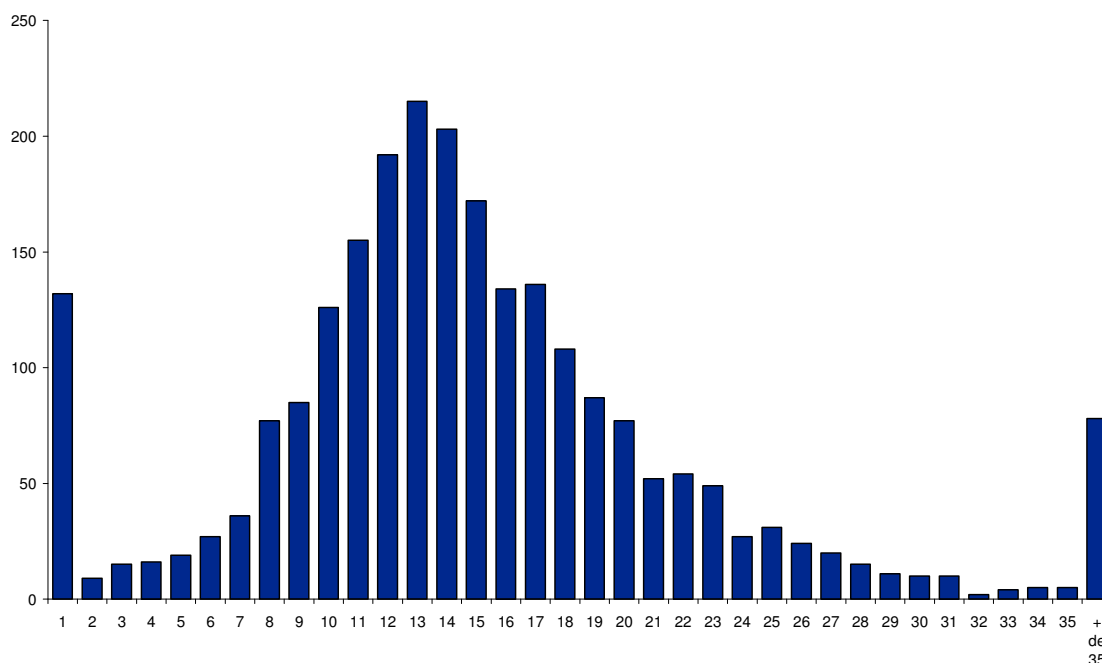


- pour la durée de la collecte : pour un médecin donné, on dispose normalement de la date de la 1ère et la dernière enquête. La durée est la différence entre ces deux dates

### La durée de collecte

La durée de la collecte en semaines calculée simplement par le nombre de jours entre premier et dernier questionnaire, fait apparaître des situations aberrantes. D'après la consigne donnée en formation, les médecins auraient dû réaliser l'enquête pendant 12 semaines. Pour la plupart, les médecins qui ont une durée de collecte d'une semaine sont ceux qui ont réalisé une seule enquête. Mais si la durée la plus fréquente est de 13 semaines, il apparaît aussi que beaucoup de médecins ont des durées très longues : plus de 50 % des médecins du travail ont plus de 3 mois de collecte.

*Répartition des médecins selon la durée de la collecte (en semaines),  
calculée par le nombre de jours entre premier et dernier questionnaire*



Cela est à relier à de fréquentes périodes d'interruption : plusieurs jours, voire plusieurs semaines sans questionnaire. Il s'agit alors de savoir quelle est la durée prise en compte par le médecin dans le calcul du taux de sondage : un médecin qui interrompt la collecte Sumer pendant quinze jours ou un mois compte-t-il les salariés vus en visite durant cette période d'interruption ? Vraisemblablement non dans la plupart des cas.

Les périodes d'interruption (semaines sans questionnaire rempli) sont relativement fréquentes et peuvent donc déformer la répartition des durées de collecte.

60 % des médecins ont au moins une interruption de 14 jours (un quart en ont même deux ou plus).

86 % des médecins ont au moins une interruption d'une semaine.

55 % des médecins du travail ont eu 1 à 3 périodes d'interruption de 7 jours, plus de 40 % ont eu 1 période d'interruption de 14 jours.

*Répartition des médecins selon le nombre d'interruptions de collecte :*

Nombre d'interruptions	Interruption = au moins 7 jours sans enquête		Interruption = au moins 14 jours sans enquête	
	Effectif	%	Effectif	%
0	345	14,27	975	40,32
1	456	18,86	847	35,03
2	529	21,88	385	15,92
3 et plus	1088	44,98	211	8,72

Il est nécessaire de calculer la **durée de collecte réelle** des médecins-enquêteurs, en retirant les périodes d'interruption. Pour savoir comment retirer ces périodes d'interruption, il faut s'assurer avant tout de la cohérence entre la durée de collecte déclarée et le taux de sondage du médecin du travail. C'est cela qui a guidé le choix de la correction effectuée et du nouveau mode de calcul de la durée de collecte (voir encadré).

Finalement, on corrige la durée de la période d'enquête en supposant que le médecin n'a été actif dans sa recherche de salariés à enquêter que durant les semaines où il a effectivement tenté d'enquêter un salarié (avec succès ou non : questionnaire complet ou incomplet).

## Impact de Pb sur les fréquences d'expositions

	Fréquence d'exposition (en %)		Evolution après intégration de Pb (en points)
	Avec pondération 1/Pa	Avec pondération 1/(Pa x Pb)	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	2,0	1,7	-0,3
Bruit supérieur à 85db plus de 20h par semaine	5,9	4,5	-1,4
Travail de nuit (99 nuits ou plus par an)	2,9	2,3	-0,6
Poussières de bois	1,6	1,3	-0,3
Manutention manuelle de charges lourdes	37,4	34,7	-2,7
Travail le dimanche et jours fériés	34,9	33,5	-1,4
Conduite sur la voie publique 20h ou plus par semaine	3,4	3,2	-0,2
Travail sur écran 20h ou plus par semaine	24,7	26,1	+1,4
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	19,6	18,2	-1,4

Les fréquences d'expositions aux :

- bruit supérieur à 85db plus de 20h par semaine,
- travail de nuit (99 nuits ou plus par an),
- manutention manuelle de charges lourdes,
- travail le dimanche et jours fériés,
- contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)

sont révisées à la baisse par l'intégration de cette pondération 1/Pb dans le calcul des pondérations d'inclusion. Ce sont des risques qui appellent en principe le recours à la surveillance médicale renforcée, et poussent donc à la hausse les fréquences observées d'expositions. En revanche l'intégration de cette pondération révisé à la hausse l'exposition au travail sur écran plus de 20h par semaine : les salariés soumis à cette contrainte voient moins souvent un médecin du travail que les autres.

Les pondérations 1/Pb permettent donc de corriger le biais induit par la corrélation entre fréquence des visites et fréquences des expositions. Elles permettent ainsi la correction d'éventuelles surestimations de fréquences d'expositions.

Après l'intégration de la pondération 1/Pb, la fréquence de salariés déclarés en SMR est de 37 % contre 47,3 % avant correction.



### **Encadré - Comment corriger la durée de collecte ?**

Pour tenir compte des interruptions de collecte dans l'intervalle entre premier et dernier questionnaire collecté, de nombreux modes de calcul étaient disponibles.

Par exemple, la correction la plus drastique consiste à ne compter comme jours de collecte que les jours où un questionnaire (complet ou incomplet) a été collecté. A l'opposé, on peut n'enlever de la durée de collecte que les jours d'interruption qui dépassent 14. Par exemple un médecin s'arrête pendant 21 jours, on enlève 7 jours à la durée de collecte pour ramener l'interruption à 14 jours.

Six modes de calcul de la durée de collecte ont été comparés.

**Durée 1** : durée calculée comme nombre de jours entre premier et dernier questionnaire (pas de correction)

**Durée 2** : nombre de jours d'enquête (nombre de questionnaires collectés si 1 par jour)

**Durée 3** : à chaque fois qu'il y a une interruption d'au moins 7 jours, on enlève les jours qui dépassent 7 au nombre de jours entre premier et dernier questionnaire. Par exemple pour une interruption de 10 jours on enlève 3 jours.

**Durée 4** : à chaque fois qu'il y a une interruption d'au moins 7 jours, on enlève les jours au nombre de jours entre premier et dernier questionnaire. Par exemple pour une interruption de 10 jours on enlève 10 jours.

**Durée 5** : à chaque fois qu'il y a une interruption d'au moins 14 jours, on enlève les jours qui dépassent 14 au nombre de jours entre premier et dernier questionnaire. Par exemple pour une interruption de 17 jours on enlève 3 jours.

**Durée 6** : à chaque fois qu'il y a une interruption d'au moins 14 jours, on enlève les jours au nombre de jours entre premier et dernier questionnaire. Par exemple pour une interruption de 17 jours on enlève 17 jours.

#### **Exemple**

*Un médecin a collecté 30 questionnaires, entre le 10 avril et le 26 novembre, aux dates suivantes :*

*Avril : 10, 14, 15, 16, 22, 23, 28 (2), 29*

*Mai : 12, 13, 14, 15, 18, 20, 25, 27, 28*

*Juin : 4, 8, 10 (2), 11, 16, 17(2), 23, 24, 29*

*Novembre : 26*

*Si on prend comme limite 7 jours pour compter une interruption, il y en a deux :*

- une interruption de 13 jours entre le 29 avril et le 12 mai*
- une interruption de 150 jours entre le 29 juin et le 26 novembre*

*Si on prend comme limite 14 jours, il y en a une seule, celle de 150 jours.*

*Durée 1 = 231 jours*

*Durée 2 = 27 jours*

*Durée 3 = 231 jours - 150 - 13 + 7 + 7 = 82 jours*

*Durée 4 = 231 - 150 - 13 = 68 jours*

*Durée 5 = 231 - 150 + 14 = 95 jours*

*Durée 6 = 231 - 150 = 81 jours*

Afin de choisir le mode de correction, nous avons recherché le mode de calcul qui semblait le plus cohérent avec les taux de sondage déclarés par les médecins-enquêteurs.

Pour un médecin donné, on compare donc l'information « nombre de visites en 2007 » de la fiche médecin et le nombre de visites estimé en 2010 calculée comme suit à partir du taux de sondage et de la durée de collecte :

Nombre de visites théorique 2010 = Nombre de salariés vus pendant la durée collecte x (1 an/durée collecte)

Où :

nombre de salariés vus pendant durée collecte = pas de tirage x nb salariés enquêtés pendant durée de collecte

(Pas de tirage = 1/taux de sondage)

Pour un médecin-enquêteur, le pas de tirage est « 1 salarié sur 10 » pendant 3 mois. Il a enquêté 30 salariés au total. Alors le nombre de visites théorique 2010 devrait être =  $10 \times 30 \times 4 = 1\ 200$  par an

Ensuite, pour conclure sur la cohérence entre la durée de collecte déclarée et le taux de sondage du médecin du travail, il faut considérer le rapport entre le nombre de visites réelles 2007 et le nombre de visites théoriques (estimé) 2010 en le comparant à 1. Ceci est à considérer en prenant en compte que :

- la définition des visites n'est pas exactement la même
- certains écarts peuvent être liés aux effets de saisonnalité, à l'évolution de l'activité entre 2007 et 2010...

Ainsi le rapport (nombre visites 2007 / nombre de visites théorique 2010) peut être considéré comme :

- crédible entre 0,66 et 1,33
- acceptable entre 0,5 et 2

Sinon il y a une incohérence (erreur de calcul sur le taux de sondage ou la durée de collecte ou incohérence entre les deux).

#### Exemple

*Le médecin de l'exemple précédent a déclaré un taux de sondage de 1/12<sup>ème</sup> et 2 200 visites effectuées en 2007 sur sa fiche médecin.*

*Il dit avoir sélectionné une visite sur 12 et aboutit à 30 questionnaires, donc durant la période de collecte, il a normalement effectué  $12 \times 30 = 360$  visites.*

*Si on considère une durée de collecte de 231 jours, soit presque les deux tiers d'une année, cela ferait 571 visites par an. C'est incohérent par rapport au nombre de visites en 2007 (même si celles-ci incluent les visites non périodiques).*

*Il peut avoir réduit son activité, mais il est plus probable qu'il n'a pas compté les visites sur une aussi longue période.*

*En prenant une durée de collecte de 82 jours, on arrive à 1 565 visites annuelles, ce qui est beaucoup plus cohérent avec sa déclaration pour 2007.*

Il était impossible de corriger manuellement et certains modes de calcul fonctionnent mieux pour un médecin mais pas pour un autre.

Les six scénarios de « correction » de la durée de collecte (interruptions) sont comparés afin de déterminer le meilleur schéma pour minimiser les incohérences du rapport entre nombre estimé de visites en 2010 et nombre déclaré de visites en 2007.

Pour l'enquête Sumer 2010 on obtient :

Mode de calcul de la durée	Rapport entre nombre estimé de visites en 2010 et nombre déclaré de visites en 2007	
	Moyenne	Médiane
Durée 1	0,84	0,68
Durée 2	4,04	2,86
Durée 3	1,15	0,88
Durée 4	2,18	1,16
Durée 5	0,97	0,77
Durée 6	1,32	0,90

Au-delà de la moyenne et de la médiane, on cherche aussi à corriger la durée de collecte pour réduire les incohérences les plus fortes.

Mode de calcul de la durée	Proportion de médecins enquêteurs pour lesquels :	
	Le nombre de visites estimé est <u>vraiment petit</u> (au moins moitié plus petit) par rapport au nombre de visites en 2007	Le nombre de visites estimé est <u>vraiment grand</u> (au moins deux fois plus grand) par rapport au nombre de visites en 2007
Durée 1	29	4
Durée 2	0	79
Durée 3	11	7
Durée 4	4	20
Durée 5	19	5
Durée 6	10	9

Sans correction (mode de calcul 1), 29 % des médecins donnent un taux de sondage qui, rapporté à leur durée de collecte, donne un nombre de visites vraiment petit par rapport à leur activité de 2007 (inférieur de moitié ou plus). Cela s'explique par les interruptions pendant la collecte.

Le mode de calcul de la durée 3 permet de réduire cette proportion à 11 % sans augmenter trop la proportion des médecins qui se trouvent dans la situation inverse (elle passe de 4 % sans correction à 7 % avec). Ce mode de calcul corrigé de la durée de la collecte nous a donc semblé le plus pertinent.

### 3) Pc : Probabilité que ce salarié soit tiré au sort par son médecin

Pc = taux de sondage du médecin-enquêteur

Ce taux de sondage est calculé et déclaré par le médecin du travail dans chacun de ses questionnaires principaux.

#### Impact de Pc sur les fréquences d'expositions

	Fréquence d'exposition (en %)		Evolution après intégration de Pc (en points)
	Avec pondération 1/ (Pa x Pb)	Avec pondération 1/(Pa x Pb x Pc)	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	1,7	1,8	0,1
Bruit supérieur à 85db plus de 20h par semaine	4,5	4,8	0,3
Travail de nuit (99 nuits ou plus par an)	2,3	2,2	-0,1
Poussières de bois	1,3	1,5	0,2
Manutention manuelle de charges lourdes	34,7	35,5	0,8
Travail le dimanche et jours fériés	33,5	32,9	-0,6

Conduite sur la voie publique 20h ou plus par semaine	3,2	3,4	0,2
Travail sur écran 20h ou plus par semaine	26,1	24,6	-1,5
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	18,2	18,8	0,6

Les médecins qui ont utilisé de forts taux de sondage voient des salariés qui manipulent un peu moins souvent que les autres des charges lourdes, et travaillent plutôt plus fréquemment le dimanche ou sur écran.

**Impact des pondérations initiales d'inclusion  $1/(P_a \times P_b \times P_c)$  sur les fréquences d'expositions:**

	Fréquence d'exposition (en %)		Evolution après pondération initiale (en points)
	Brut (sans pondération aucune)	Avec pondération $1/(P_a \times P_b \times P_c)$ (*)	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	2,1	1,8	- 0,3
Bruit supérieur à 85db plus de 20h par semaine	5,9	4,8	- 1,1
Travail de nuit (99 nuits ou plus par an)	2,9	2,2	- 0,7
Poussières de bois	1,6	1,5	- 0,1
Manutention manuelle de charges lourdes	37,3	35,5	- 1,8
Travail le dimanche et jours fériés	34,9	32,9	- 2,0
Conduite sur la voie publique 20h ou plus par semaine	3,6	3,4	- 0,2
Travail sur écran 20h ou plus par semaine	25,0	24,6	- 0,4
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	20,1	18,8	- 1,3

**(\*) :  $1/(P_a \times P_b \times P_c)$  = pondération d'inclusion, initiale**

Les **pondérations d'inclusions** corrigent les biais liés au volontariat et aux fréquences des visites. Les résultats corrigés ne diffèrent pas radicalement des résultats bruts, mais pour la manutention de charges lourdes, le bruit, les contraintes posturales et le travail du dimanche, la pondération diminue de 1 à 2 points les fréquences d'expositions.

## II- Correction de la non-réponse totale

Les non-répondants sont ceux qui ont été tirés par le médecin mais n'ont pu être enquêtés. Il est possible que les médecins, lorsque le salarié tiré ne se présentait pas, l'aient remplacé par le suivant, mais nous n'avons aucun moyen de le savoir. On considère donc qu'un non-répondant est une personne sélectionnée pour laquelle le médecin a rempli seulement le début du questionnaire, à savoir les trois premières parties d'identification (identification du questionnaire, caractéristiques de l'établissement employeur, caractéristiques du salarié), et en particulier a renseigné le motif de non-réponse (variable MOTIFNR ayant pour modalités : absence, refus ou impossible de répondre).

11 % des salariés tirés au sort n'ont pas répondu à l'enquête :

- 5,7 % ne se sont pas présentés à la visite,
- 4,9 % ont refusé de répondre,
- 0,4 % n'ont pas pu répondre (difficultés de communication, méconnaissance de la langue française, ...).

Le profil de ces salariés non-répondants est un peu différent de celui des répondants : ils sont plus souvent étrangers hors Union européenne, manœuvres ou ouvriers spécialisés. Ils travaillent beaucoup plus souvent dans les établissements de moins de 10 salariés, dans le secteur de l'agriculture et dans les fonctions de nettoyage, gardiennage et entretien ménager. Ce sont aussi moins souvent des agents à statut (salariés qui travaillent dans une entreprise publique, ou anciennement publique, et bénéficient d'un statut particulier) ou des salariés ayant une ancienneté de 10 ans ou plus dans l'établissement.

La correction de la non-réponse totale (NR) a été réalisée à partir de la **méthode des groupes de réponse homogènes**. Cette méthode de repondération consiste à traiter la non-réponse en modifiant les poids de sondage des individus ayant répondu, afin de compenser l'absence de certaines réponses.

Pour modifier les poids de sondage et ainsi corriger la non-réponse totale, le principe général est d'attribuer à chaque individu  $k$  une probabilité de réponse  $p_k$ . Si celle-ci est connue pour tous les individus, on dispose d'estimateurs sans biais des expositions professionnelles en utilisant comme pondération « intermédiaire » (soit après correction de la non-réponse totale) les poids de sondage divisés par les probabilités de réponse, soit pour tout individu  $k$  :

Poids « intermédiaire » = poids corrigeant de la NR  $\times 1/(P_a \times P_b \times P_c)$

Où :  $1/p_k$  = poids corrigeant de la NR

Soit :

$$\text{Poids intermédiaire} = (1/(P_a \times P_b \times P_c \times p_k))$$

Toutefois, les probabilités de réponse individuelles  $p_k$  sont inconnues et doivent par conséquent être estimées.

La méthode des groupes de réponse homogènes permet d'estimer ces probabilités de réponse individuelles  $p_k$ . Elle consiste à chercher à regrouper les unités ayant certaines caractéristiques communes (connues) conduisant à un comportement de réponse similaire ; on estime alors les probabilités de réponse au sein des groupes ainsi formés, dits groupes de réponse homogènes (GRH).

### **1- Constitution des groupes de réponse homogènes :**

Cette analyse a été réalisée au moyen d'une régression logistique appliquée sur l'échantillon pour ainsi pouvoir choisir les variables auxiliaires les plus explicatives de la non-réponse ainsi que pour effectuer les regroupements en sous-populations.

Descriptif des variables utilisées dans l'étude de la non-réponse

Nom de la variable	Intitulé	Modalités	
NAT <i>(retenue)</i>	Nationalité	1 2 3	Française Etrangère UE Etrangère hors UE
SEXE	Sexe	1 2	Homme Femme
FONCT	Fonction principale exercée	1 2 3 4 5 6 7 8 0	Production Installation Nettoyage Manutention Secrétariat Gestion Commerce Etudes, recherche Autre
NAFC	Secteur d'activité	1 2 3 4 5 6 7 8 9 10	Agriculture, sylviculture, pêche Industries manufacturières, extractives et autres Construction Commerce, transport hébergement, restauration Information, communication Activités financières et assurances Activités immobilières Activités spécialisés, scientifiques et techniques, services administratifs et de soutien Administration publique, enseignement, santé et action sociale Autres activités de service
AGELOGIT	Age	Jeune Adulte Senior	<30 ans >=30 ans et <50 ans >= 50 ans

Nom de la variable	Intitulé	Modalités	
CSP (retenue)	Catégorie socio-professionnelle	1 2 3 4 5 6 7	Agriculteurs Artisans commerçants Cadres Prof. intermédiaires Employés Ouvriers Autres
TAIL1LOGIT (retenue)	Taille de l'établissement	1 2 3 4 5	<10 salariés De 10 à 49 salariés De 50 à 249 salariés De 250 à 499 salariés 500 salariés et plus
SURV	Etre en surveillance médicale renforcée (SMR)	1 0	Oui non

Pour restreindre le nombre de variables explicatives et ne conserver pour le modèle final que celles qui sont les plus discriminantes pour la non-réponse, on utilise l'option « selection=forward ». Cette option introduit en premier lieu la variable explicative qui explique le mieux la variable dépendante, puis en second lieu la variable explicative qui explique le mieux la variance qui reste à expliquer, etc.

Ainsi les trois variables explicatives de la non-réponse les plus discriminantes retenues sont dans l'ordre décroissant d'importance (voir Annexe 2) :

- Taille de l'établissement employeur (les non-répondants travaillent plus souvent dans des petits établissements)
- Nationalité (les ressortissants de pays non européens sont plus souvent non-répondants)
- Catégorie socioprofessionnelle (les ouvriers et les employés sont plus souvent non-répondants)

La variable SURV (« Etre en surveillance médicale renforcée ») ne se classe qu'en 8<sup>ème</sup> position parmi les variables discriminantes de la liste ci-dessus.

Les sous populations homogènes au sens de la non-réponse correspondent aux croisements de toutes les modalités de ces trois variables explicatives les plus discriminantes.

### 2- calcul des taux d'échec (= taux de non-réponse) pour ces sous-populations

Une fois que les sous-populations sont constituées, la probabilité de non-réponse est supposée indépendante d'une sous-population à une autre et constante au sein d'une sous-population.

Le taux de non-réponse (= taux d'échec) est estimé par le rapport  $nr_h / n_h$  où

$nr_h$  = nombre de non-répondants dans la sous-population h

$n_h$  = nombre d'individus de l'échantillon de la sous-population h

En annexe 3, les taux de non-réponse appliqués aux sous-populations pour la correction de la non-réponse totale.

### 3- Calcul des poids corrigeant la non-réponse totale et des poids intermédiaires :

Par chaque sous-population de non-répondants :

Poids corrigeant de la non-réponse = 1/Taux de réponse

Où : Taux de réponse = 1 - Taux d'échec = 1 - ( $nr_h / n_h$ )

Et

Poids intermédiaire = pondérations d'inclusion x Poids corrigeant de la non-réponse

**Impact des pondérations corrigeant la non-réponse totale sur les fréquences d'expositions**

	Fréquence d'exposition (en %)		Evolution après intégration de Pondérations corrigeant de la non-réponse (en points)
	Avec pondération $1/(P_a \times P_b \times P_c)$	Avec Poids intermédiaires	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	1,8	1,9	+0,1
Bruit supérieur à 85db plus de 20h par semaine	4,8	4,8	0
Travail de nuit (99 nuits ou plus par an)	2,2	2,2	0
Poussières de bois	1,5	1,6	+ 0,1
Manutention manuelle de charges lourdes	35,5	35,9	+0,4
Travail le dimanche et jours fériés	32,9	32,9	0
Conduite sur la voie publique 20h ou plus par semaine	3,4	3,5	+0,1
Travail sur écran 20h ou plus par semaine	24,6	24,2	- 0,4
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	18,8	19,1	+0,3

La correction de la non-réponse n'apporte pas de forte correction des fréquences d'expositions aux risques professionnels.



### III- Calage sur marges

Afin d'obtenir les pondérations finales des salariés Sumer 2010, on réalise un calage sur marges des salariés par type de secteur d'exercice des médecins (FPE<sup>(1)</sup>, FPT<sup>(2)</sup>, FPH<sup>(3)</sup>, MSA, Air France, EDF-GDF, la Poste, gens de mer, SNCF, autres régies de transports urbains, et le noyau du régime général hors secteurs déjà cités ci-dessus). Les intérimaires (tous secteurs confondus) sont aussi traités à part, avec un calage sur marges spécifique.

(1) FPE = Fonction Publique d'Etat

(2) FPT = Fonction Publique Territoriale

(3) FPH = Fonction Publique Hospitalière

**Rappel** : la FPE n'a été enquêtée encore qu'à titre expérimental. Seuls quelques ministères (intérieur, équipement, économie, aviation civile, culture, défense, administration centrale de l'éducation, CNRS, Sénat, CDC) ont participé à la collecte Sumer 2010.

Les caractéristiques de ces populations de référence servent de base de calage pour l'échantillon des salariés. Ces caractéristiques sont tirées :

- des **DADS** (Déclarations annuelles de données sociales) pour l'ensemble des secteurs d'exercice sauf les intérimaires et les gens de mer. Contrairement à 2003, en 2010 le fichier « Postes » (unité = salarié) n'était pas disponible. Par contre le fichier « Etablissement » (unité = siret \_ information agrégée par établissement) disponible à cette date a donc été utilisé

- de l'**enquête Emploi** pour les intérimaires

- de **sources spécifiques** pour certains secteurs professionnels :

- \* informations fournies par la CCMSA pour les salariés de la MSA

- \* informations issues du fichier de paie des agents de l'Etat pour la FPE et de l'enquête Colter gérée par l'INSEE pour la FPT. Toutes ces informations étant fournies par la DGAFP

- \* informations transmises pour les gens de mer

Le maximum d'informations a été collecté selon ces différentes sources. Pour les secteurs spécifiques, l'information auxiliaire était parfois commune à deux sources. Par convention, c'est l'information fournie par le secteur lui-même (ex : de la CCMSA pour la MSA, de la DGAFP pour la FPE) qui a été retenue.

Ainsi l'univers de référence est reconstitué à partir de ces variables auxiliaires, de ces marges. Ces critères de pondérations choisis sont quelques peu différents d'un secteur d'exercice à l'autre.

- Modalités des variables utilisées pour le calage sur marges des salariés

Nom variables	Intitulé Variables	Modalités
Sexe	SEXENUM	1 - Homme 2 - Femme
Age	TAGE	1 - moins de 25 ans' 2 - de 25 à 39 ans' 3 - de 40 à 49 ans' 4 - de 50 à 59 ans' 5 - 60 ans ou plus
Nationalité	NAT	1 - Français 2 - étranger UE 3 - étranger hors UE
Type de temps de travail	TPS	1 - temps complet 2 - temps partiel
Catégorie socio-professionnelle	CSPP	3 - Cadres 4- Prof. intermédiaires 5- employés 6 - ouvriers 7- autres
Taille de l'établissement employeur	TAILETAB	1- <10'salariés 2 - De 10 à 49 salariés' 3 - De 50 à 249 salariés' 4 - De 250 à 499 salariés'

		5 - 500 salariés et plus
Secteur d'activité	NAFCC	1- Agriculture, sylviculture, pêche 2 - Industries manufacturières, extractives et autres 3 - Construction 4 - Commerce, transport hébergement, restauration 5 - Information, communication 6 - Activités financières et assurances 7 - Activités immobilières 8 - Activités spécialisés, scientifiques et techniques, services administratifs et de soutien 9 - Administration publique, enseignement, santé et action sociale 10-Autres activités de service
Nomenclature NOSTA regroupée (à partir de la nomenclature NOSTA) :	NOSTAC	1 - Cultures spécialisées : horticulture-Floriculture, Cultures spécialisées: céréales, Champignonnières, Saliculture 2- Elevage gros animaux : bovins lait, Elevage petits animaux : volailles, Chevaux de course, Conchyliculture 3 - Cultures et élevages non spécialisés 4 - Viticulture 5 - Sylviculture, Exploitation du bois, Scieries fixes, Garde-Chasse/Garde-Pêche, Jardinier/Garde-forestier 6 - Entreprises de travaux agricoles, Entreprises paysagistes et métiers de l'environnement 7 - Coopératives de stockage et d'approvisionnement, Coopératives d'approvisionnement, Collecte et distribution de produits laitiers, Traitement de la viande : bovins, Conserverie, Vinification, Insémination artificielle, Distilleries, Meunerie/Panification, Stockage et conditionnement de fruits, Traitement de la viande de volailles, Choucroute 8 - Mutualité Sociale Agricole, Organismes bancaires agricoles 9 - Autres organismes professionnels agricoles 10 - Organismes de remplacement/Travail temporaire, Enseignants / Etablissements techniques, Artisans ruraux,

Selon les secteurs d'exercice, ces variables étaient plus ou moins disponibles.  
Le calage sur marges a donc été réalisé par secteur d'exercice.

- Calage sur marges pour chaque secteur d'exercice : variables de calages utilisées (en fonction de leur disponibilité) et pondérations obtenues

L'échantillon de répondants étant différent entre le questionnaire principal (QP) et l'auto-questionnaire (AQ), deux jeux de pondérations sont calculés pour chacun de ces types de questionnaires. Néanmoins ils restent très proches car le taux de participation à l'auto-questionnaire pour les salariés ayant répondu au questionnaire principal s'élève à 97,9 %.

Les **pondérations finales** du redressement seront nommées :

- **PONDQP** pour le questionnaire principal (QP)
- **PONDAQ** pour l'auto-questionnaire (AQ)

**QP (Questionnaire principal)**

**AQ (Auto questionnaire)**

Secteurs d'exercice	Effectif pop. de référence	Effectif éch. QP	PONDQPmoyenne	PONDQP extrêmes		Effectif éch. AQ	PONDAQ moyenne	PONDAQ extrêmes		Variables de calages
				Min	Max			Min	Max	
MSA	610 922	2 698	226	10	3 991	2 606	234	10	4 085	Sexenum, tage, nat, nostac
Air France-	53 769	124	434	48	2 625	124	434	48	2 625	Sexenum, tage, tps, cspp
La Poste	289 805	1 047	278	2	4 421	998	291	2	4 598	Sexenum, tage, tps, taileétab, cspp
EDF-GDF	138 056	1 408	98	7	1 235	1 391	99	7	1 278	Sexenum, tage, tps, nat, cspp
SNCF	172 028	977	176	10	1 307	969	178	11	1 317	Sexenum, tage, tps, taileétab, cspp
Autres régies transports urb.	91 318	579	158	5	2 206	569	160	6	2 219	Sexenum, tage, tps, taileétab, cspp
Gens de mer	25 201	256	98	6	1 971	255	98	6	1 970	Sexenum, tage, tps, nat, cspp
FPE	1 045 885	2 326	491	8	10 640	2 305	489	8	10 656	Sexenum, tps, cspp
FPT	1 825 031	1 555	746	32	8 171	1 515	764	33	8 274	Sexenum, tage, tps, taileétab, cspp
FPH	1 053 591	1 756	526	15	5 921	1 712	538	15	5 929	Sexenum, tage, tps tps, taileétab, cspp
Noyau	16 015 717	34 303	490	5	19 814	33 595	501	5	20 687	Sexenum, tage, tps, nat, taileétab, cspp, nafcc
Intérimaires	419109	954	439	14	6 243	923	454	14	6 370	Sexenum, tage, nat, cspp
<b>Total</b>	<b>21 740 432</b>	<b>47 983</b>	<b>456</b>	<b>2</b>	<b>19 814</b>	<b>46 962</b>	<b>465</b>	<b>2</b>	<b>20 687</b>	

- Distribution des pondérations finales :  
Pour PONDQP

Poids moyen: 450,30  
Poids modal: 14,91

Quantile	Estimate
100% Max	19816,08
99%	2484,52
95%	1277,13
90%	918,11
75% Q3	545,00
50% Median	308,46
25% Q1	163,72
10%	82,16
5%	50,37
1%	17,72
0% Min	1,71

Pour PONDAQ

Poids moyen: 462,84  
Poids modal: 212,29

Quantile	Estimate
100% Max	20689,36
99%	2484,82
95%	1313,86
90%	946,64
75% Q3	562,52
50% Median	318,63
25% Q1	172,69
10%	90,21
5%	57,75
1%	22,64
0% Min	1,72

- Distribution des rapports de poids (= pondérations finales / pondérations d'inclusion initiales)

Rapport moyen: 1,05  
Rapport modal: 1,16

Quantile	Estimate
100% Max	16,31
99%	3,53
95%	1,96
90%	1,52
75% Q3	1,20
50% Median	0,96
25% Q1	0,78
10%	0,59
5%	0,42
1%	0,15
0% Min	0,02

**Impact des pondérations de calage sur marges des salariés sur les fréquences d'expositions:**

	Fréquence d'exposition (en %)		Evolution après intégration de Pondérations de calage sur marges des salariés (en points)
	Avec poids intermédiaire	Avec pondérations finales (PONDQP)	
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	1,9	1,9	0
Bruit supérieur à 85db plus de 20h par semaine	4,8	4,8	0
Travail de nuit (99 nuits ou plus par an)	2,2	2,2	0
Poussières de bois	1,6	1,7	+0,1
Manutention manuelle de charges lourdes	35,9	37,2	+1,3
Travail le dimanche et jours fériés	32,9	33,3	+0,4
Conduite sur la voie publique 20h ou plus par semaine	3,5	3,3	-0,2
Travail sur écran 20h ou plus par semaine	24,2	22,4	-1,8
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	19,1	20,0	+0,9

Au regard des distributions des rapports des poids (= poids finaux/poids d'inclusion initiaux), le calage sur marge n'impacte pas de façon importante la mesure des fréquences d'expositions.

## Comparaison des impacts des différentes corrections au cours du redressement

En points

	Evolution après intégration de			
	Pondérations corrigeant du volontariat (Pa)	Pondérations corrigeant du biais lié à la fréquence des visites (Pb)	Pondérations corrigeant de la non-réponse	Pondérations de calage sur marges des salariés
Vibrations transmises aux membres supérieurs 10h ou plus par semaine	- 0,1	-0,3	0,1	0
Bruit supérieur à 85db plus de 20h par semaine	0	-1,4	0	0
Travail de nuit (99 nuits ou plus par an)	0	- 0,6	0	0
Poussières de bois	0	- 0,3	0,1	0,1
Manutention manuelle de charge	0,1	- 2,7	0,4	1,3
Travail le dimanche et jours fériés	0	- 1,4	0	0,4
Conduite sur la voie publique 20h ou plus par semaine	- 0,2	- 0,2	0,1	-0,2
Travail sur écran 20h ou plus par semaine	- 0,3	1,4	- 0,4	-1,8
Contraintes posturales 2h ou plus (genou, bras en l'air et autres contraintes posturales)	- 0,5	- 1,4	0,3	0,9

Parmi l'ensemble des redressements effectués, la correction du biais lié à la « fréquence des visites » impacte le plus la correction des poids des salariés. L'effet de la réforme de la médecine du travail sur les pratiques professionnelles des médecins du travail apparaît clairement : ils centrent davantage leur temps de visite sur des salariés particulièrement exposés.

ANNEXES

**Annexe 1 - Bases d'informations disponibles pour la population de référence de médecins du travail et comparaisons avec les informations de l'échantillon des médecins-enquêteurs**

		Informations population de référence médecins du travail		Informations échantillon médecins-enquêteurs (Après imputation par Hot Deck Hiérarchisé)	
Régime général		Source IMT2 (Régime général)		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins HORS médecins de prévention		6 866		1 341	
<b>SEXE</b>	Hommes	2 473	36	322	24
	Femmes	4 393	64	1 019	76
<b>AGE</b>	moins de 39 ans	405	6	94	7
	de 40 à 54 ans	2 690	39	644	48
	Plus de 55 ans	3 771	55	603	45
Nombre de salariés suivis en ETP		2 997		2 756	
Nombre de visites médicales en ETP		2 054		2 149	
Nombre d'entreprises suivies en ETP		261		199	

MSA		Source MSA		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		382		148	
<b>SEXE</b>	Hommes	155	41	57	39
	Femmes	227	59	91	61
<b>AGE</b>	moins de 39 ans	8	2	8	5
	de 40 à 54 ans	149	39	85	57
	plus de 55 ans	225	59	55	37
Nombre de salariés suivis en ETP		2 341		2 585	

		informations population de référence médecins du travail		Informations échantillon médecins-enquêteurs (Après imputation par Hot Deck Hiérarchisé)	

La Poste		Source la Poste		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		149		48	
SEXE	Hommes	34	23	5	10
	Femmes	115	77	43	90
AGE	moins de 39 ans	7	5	1	2
	de 40 à 49 ans	33	22	19	40
	de 50 à 59 ans	87	58	25	52
	Plus de 60 ans	22	15	3	6

EDF-GDF		Source EDF-GDF		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		173		64	
SEXE	Hommes	71	41	33	52
	Femmes	102	59	31	48

SNCF		Source SNCF		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		145		43	
SEXE	Hommes	59	41	8	19
	Femmes	86	59	35	81
AGE	moins de 49 ans	12	12	5	12
	de 50 à 54 ans	36	25	19	44
	de 55 à 59 ans	46	32	11	26
	Plus de 60 ans	46	32	8	19
Nombre de salariés suivis en ETP		1 512		1 310	



		informations population de référence médecins du travail		Informations échantillon médecins-enquêteurs (Après imputation par Hot Deck Hiérarchisé)	
Air France		Source Air France		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		36		4	
SEXE	Hommes	15	42	3	75
	Femmes	21	58	1	25
AGE	moins de 44 ans	7	19	3	75
	plus de 50 ans	29	81	1	25

Fonction Publique Hospitalière (FPH)		Source FPH		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		557		103	
SEXE	Hommes	156	28	13	13
	Femmes	401	72	90	87
AGE	moins de 44 ans	63	26	14	14
	de 45 à 54 ans	156	28	44	43
	de 55 à 59 ans	134	24	37	36
	Plus de 60 ans	122	22	8	8

Fonction Publique d'Etat (FPE)		Source FPE		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		513		126	
Nombre de salariés suivis en ETP		5 097		2 034	

Fonction Publique Territoriale (FPT)		Source FPT		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		739		75	

		informations population de référence médecins du travail		Informations échantillon médecins-enquêteurs (Après imputation par Hot Deck Hiérarchisé)	
Gens de mer		Source gens de mer		Source Fiches médecins du travail	
		effectifs	%	effectifs	%
effectif total des médecins		21		10	
<b>SEXE</b>	Hommes	16	76	7	70
	Femmes	5	24	3	30

Comparaison des profils des médecins-enquêteurs à l'ensemble de la population de référence des médecins du travail par secteur:

- Hors EDF-GDF, Air France, FPE, FPT : plus de femmes et beaucoup moins de médecins âgés (55 ans et plus) que parmi l'ensemble des médecins du travail de chacun de ces services
- EDF-GDF: plus d'hommes

Se référer à l'article « enquête Sumer 2010 : bilan de la collecte » du DMT n°127 de septembre 2011 pour plus de détails concernant le profil comparé des médecins-enquêteurs.

## Annexe 2 - Variables explicatives de la non-réponse

On a estimé la probabilité de non-réponse par une modélisation Logit. Les estimations des odds ratio ci-dessous, nous permettent d'avancer que :

- un salarié étranger hors UE a 2 fois plus de chance de ne pas répondre à l'enquête Sumer 2010 qu'un salarié français
- un artisan commerçant ou un ouvrier a 1,7 fois plus de chance de ne pas répondre qu'un cadre
- un salarié travaillant dans un établissement de moins de 10 salariés a 2 fois plus de chance de ne pas répondre qu'un salarié travaillant dans un établissement de 500 salariés ou plus

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
NAT étranger UE vs Français	1.348	1.102	1.647
NAT étranger hors UE vs Français	2.103	1.831	2.416
SEXE Femmes vs hommes	0.837	0.781	0.897
FONCT commerce vs 0	1.185	1.064	1.319
FONCT gestion vs 0	0.824	0.717	0.947
FONCT installation vs 0	0.762	0.667	0.871
FONCT manutention vs 0	0.974	0.848	1.118
FONCT nettoyage vs 0	1.268	1.117	1.440
FONCT production vs 0	0.893	0.799	0.998
FONCT secrétariat vs 0	0.959	0.847	1.086
FONCT études, recherche vs 0	0.756	0.650	0.879
nafc Activités financ vs Activités immobi	1.079	0.809	1.440
nafc Activités spécia vs Activités immobi	1.325	1.021	1.719
nafc Administration p vs Activités immobi	1.132	0.871	1.470
nafc Agriculture, sylv vs Activités immobi	1.919	1.434	2.568
nafc Autres activités vs Activités immobi	1.195	0.898	1.592
nafc Commerce, transp vs Activités immobi	1.088	0.841	1.409
nafc Construction vs Activités immobi	1.439	1.091	1.898
nafc Industries manuf vs Activités immobi	0.916	0.703	1.194
nafc Information, com vs Activités immobi	1.475	1.090	1.995
agelogit adult vs jeune	1.116	1.036	1.201
agelogit senio vs jeune	1.322	1.211	1.443
csp Agriculteurs vs Cadres	0.861	0.191	3.871
csp Artisans commerçant vs Cadres	1.688	1.231	2.313
csp Prof. intermédiaire vs Cadres	1.237	1.108	1.380
csp employés vs Cadres	1.532	1.367	1.718

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
csp ouvriers vs Cadres	1.679	1.484	1.901
tail1logit 1 vs 5	1.962	1.779	2.165
tail1logit 2 vs 5	1.445	1.317	1.585
tail1logit 3 vs 5	1.297	1.180	1.426
tail1logit 4 vs 5	1.037	0.913	1.177
surv 0 vs 1	1.168	1.098	1.242

**Annexe 3 - Taux appliqués pour la correction de la non-réponse totale pour chaque sous-population définies précédemment**

Finalement il y a 22 sous-populations. En effet, des regroupements intermédiaires ont dû être réalisés afin de ne pas avoir un effectif nul dans une sous-population.

Les trois variables explicatives de la non-réponse les plus discriminantes					
Taille de l'établissement employeur	Nationalité	CSP	Effectif de non-répondants	Effectif total dans l'échantillon Sumer 2010	Taux de non-réponse
(<10 salariés)	Français	Agriculteurs	4	16	15,6114
		Artisans commerçants	59	208	
		Cadres	117	929	
		Prof. intermédiaires	310	2128	14,5677
		employés	467	3061	15,2565
		ouvriers	643	3386	18,9900
	étranger UE	Agriculteurs	0	0	26,9737
		Artisans commerçants	1	5	
		Cadres	0	7	
		Prof. intermédiaires	5	24	
		employés	17	69	
		ouvriers	27	93	
	étranger hors UE	Agriculteurs	0	0	26,9737
		Artisans commerçants	2	2	
		Cadres	2	14	
		Prof. intermédiaires	5	24	
		employés	18	87	
		ouvriers	46	131	

Taille de l'établissement employeur	Nationalité	CSP	Effectif de non-répondants	Effectif total dans l'échantillon Sumer 2010	Taux de non-réponse
(10 à 49 salariés)	Français	Agriculteurs	0	2	9,2997
		Artisans commerçants	8	84	
		Cadres	158	1699	
		Prof. intermédiaires	284	3048	9,3178
		employés	469	3693	12,6997
		ouvriers	527	4017	13,1192
	étranger UE	Agriculteurs	0	0	22,5443
		Artisans commerçants	0	0	
		Cadres	3	31	
		Prof. intermédiaires	5	38	
		employés	8	46	
		ouvriers	24	131	
	étranger hors UE	Agriculteurs	0	0	
		Artisans commerçants	1	2	
		Cadres	4	15	
		Prof. intermédiaires	7	40	
		employés	24	120	
		ouvriers	64	198	

Taille de l'établissement employeur	Nationalité	CSP	Effectif de non-répondants	Effectif total dans l'échantillon Sumer 2010	Taux de non-réponse
(50 à 249 salariés)	Français	Agriculteurs	0	0	7,2329
		Artisans commerçants	5	29	
		Cadres	127	1796	
		Prof. intermédiaires	231	2710	8,5240
		employés	330	3065	10,7667
		ouvriers	438	3767	11,6273
	étranger UE	Agriculteurs	0	0	19,5382
		Artisans commerçants	0	0	
		Cadres	2	33	
		Prof. intermédiaires	3	35	
		employés	5	28	
		ouvriers	15	98	
	étranger hors UE	Agriculteurs	0	0	
		Artisans commerçants	0	0	
		Cadres	6	24	
		Prof. intermédiaires	5	34	
		employés	15	93	
		ouvriers	59	218	

Taille de l'établissement employeur	Nationalité	CSP	Effectif de non-répondants	Effectif total dans l'échantillon Sumer 2010	Taux de non-réponse
(250 à 499 salariés)	Français	Agriculteurs	0	0	6,7175
		Artisans commerçants	1	7	
		Cadres	62	935	
		Prof. intermédiaires	78	1157	
		employés	113	1091	
		ouvriers	133	1627	
	étranger UE	Agriculteurs	0	0	10,3575 8,8670
		Artisans commerçants	0	1	
		Cadres	0	13	
		Prof. intermédiaires	1	7	
		employés	0	10	
		ouvriers	3	26	
	étranger hors UE	Agriculteurs	0	0	
		Artisans commerçants	0	0	
		Cadres	0	9	
		Prof. intermédiaires	2	17	
		employés	5	31	
		ouvriers	18	86	

On regroupe les salariés étrangers UE et Hors UE avec les ouvrier français car trop peu en effectif



Taille de l'établissement employeur	Nationalité	CSP	Effectif de non-répondants	Effectif total dans l'échantillon Sumer 2010	Taux de non-réponse
(500 salariés et plus)	Français	Agriculteurs	0	0	5,1765
		Artisans commerçants	1	10	
		Cadres	153	2965	
		Prof. intermédiaires	286	4157	6,8800
		employés	329	3472	9,4758
		ouvriers	182	2794	7,3002
	étranger UE	Agriculteurs	0	0	
		Artisans commerçants	0	0	
		Cadres	2	26	
		Prof. intermédiaires	2	25	
		employés	3	25	
		ouvriers	5	22	
	étranger hors UE	Agriculteurs	0	0	
		Artisans commerçants	0	0	
		Cadres	5	45	
		Prof. intermédiaires	2	12	
		employés	7	40	
		ouvriers	14	52	

On regroupe les salariés étrangers UE et Hors UE avec les ouvrier français car trop peu en effectif