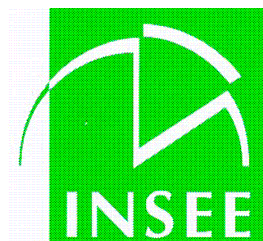


**Direction des Statistiques Démographiques
et Sociales**

**F1101
Le volet Ménages de l'enquête
Handicap-Santé
Présentation, calcul des poids**

Gérard Bouvier
Insee, division Enquêtes et Etudes Démographiques

Document de travail



Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des Documents de Travail
de la

DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES
UNITES DES ÉTUDES DÉMOGRAPHIQUES ET SOCIALES
DIVISION ENQUÊTES ET ÉTUDES DÉMOGRAPHIQUES

NF1101

**Le volet Ménages de l'enquête Handicap-Santé :
Présentation, calcul des poids**

Gérard Bouvier
Insee, division Enquêtes et Etudes Démographiques

Février 2011

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.
Working-papers do not reflect the position of INSEE but only their authors' views.

Résumé

(Redressements et calages pour l'enquête Handicap-Santé, volet ménage)

L'enquête Handicap-Santé, volet ménage a été collectée auprès de personnes vivant en ménages ordinaires en 2008. 29 931 personnes, vivant en métropole et dans les départements d'outre mer ont répondu, soit un taux de réponse de 76,6 %. Cette grande enquête structurale comprend de nombreuses questions sur l'état fonctionnel (maladies, déficiences, limitations), les restrictions de participation sociale (accès au travail, à la formation, aux loisirs, éléments de niveaux de vie) et les facteurs environnementaux (entourage familial ou social, aides techniques aménagements du logement, accessibilité aux lieux de vie, discriminations ressenties). Ce volet complète un volet institution collecté en 2009.

Le document décrit la base de sondage, le tirage de l'échantillon, les redressements pour non-réponse et le calage, donc le calcul complet des poids. L'échantillon est tiré dans une base d'individus ayant répondu à l'enquête filtre « Vie Quotidienne et Santé, 2007 ». La stratification combine un critère géographique, tenant compte des six extensions départementales, et un critère de « sévérité présumée de la situation de handicap ». Ce critère comporte quatre modalités et a été défini grâce à l'enquête filtre. Il permet de surreprésenter les personnes en grande « sévérité présumée de la situation de handicap », peu nombreuses dans le champ. Il permet aussi d'inclure des personnes enquêtées en bonne santé, pour des éléments de comparaison, et en nombre suffisant pour limiter la dispersion des poids et l'impact des « faux négatifs ». Les redressements pour non-réponse consistent, classiquement, à diviser le poids de l'individu par sa probabilité de répondre. Cette probabilité est estimée par des modèles logits à partir de données démographiques, géographiques et sociales, et des données de l'enquête filtre. Cela permet d'inclure des critères de santé, qui s'avèrent très pertinents en relation avec les problématiques de l'enquête. Le calage est fait sur des marges démographiques, définies en cohérence avec le critère de stratification géographique. La dispersion finale des poids est élevée, le rapport inter-décile est de 40. Elle résulte d'abord du tirage de l'enquête filtre, à partir de l'enquête annuelle de recensement 2006 puis la présence des extensions y contribue. Le troisième facteur important résulte de la stratification suivant les groupes de « sévérité présumée de la situation de handicap ».

Summary

(Non-participation correction and adjustments of weights for the Household section of the Health and Disability survey)

The Household section of the Health and Disability survey was collected among people living in households in 2009. 29 931 persons living in mainland France and in overseas "départements", have responded to the survey, which correspond to a response rate of 76, 6 %. This major structural survey, includes many questions about respondents' functional state (diseases, deficiencies, limitations), social participation restrictions (access to the labour market, educational attainments or leisure, aspects of standards of living), environmental factors (like familial or social network, assistive technologies, home fittings, accessibility, discriminations (self-declared)). This section is the complement of an institution section, collected in 2009.

This document deals with the sampling framework, the sampling process, non-participation corrections and adjustments i.e. the full calculation of weights for the survey sample. The sample derived from those individuals having answered to the filter-survey "Daily Life and Health, 2007". Stratification aims to correct for geographical variation (and to take into accounts the over sample made in six "départements") and for presumed severity of disability situations. The latter criterion enabled an over-representation of people (presumably) suffering from severe handicaps. An index with four degrees has been built up with the filter-survey. This indicator also enabled to take into accounts individuals in good health, for comparisons, and a large enough sample population to limit the weight dispersion and the consequences of "false negatives". Conventionally, the non-participation corrections consist of dividing the individual weight by his response probability. This probability is computed with logistic models derived from demographic, geographical and social data, and data from the filter-survey. This allows health criteria, highly relevant for the subject matter of the survey to be included. A calibration was performed using demographic data chosen in accordance with the geographical criterion. Eventually, the weight dispersion was found to be high, the inter-decile ratio reaching 40. This happens to be the result first of the sampling process used in the filter survey, based on the annual census survey, 2006, secondly the presence of extensions, and thirdly the stratification, resulting from the variation in the disability situations.

Le volet Ménages de l'enquête Handicap-Santé :

Présentation, calcul des poids.

Partie I : Présentation du document et de l'enquête	5
I-A : Présentation des étapes de calcul des poids.....	5
I-B : Position du volet ménage au sein du dispositif d'ensemble.....	5
I-C : Quelques sigles et tableau récapitulatif	6
Partie II : L'échantillon, les poids de tirage.....	8
II-A : Synthèse	8
II-B : Répondants de l'enquête-filtre VQS : base de sondage et groupes.....	8
II-C : Taille de l'échantillon, les contraintes géographiques	10
II-D : L'échantillon, décomposition par groupe VQS.....	10
II-E : Grandes étapes de la construction de l'échantillon	11
II-F : Calcul du poids de tirage (enquête HSM)	14
Partie III : Les répondants	17
III-A : Résumé de la méthode suivie	17
III-B : Quelques éléments de cadrage pour les taux de réponses et la non-réponse	18
III-C : Intégration progressive des informations et effets.....	19
Partie IV : Calage sur marges : la pondération finale.....	23
IV-A : Synthèse	23
IV-B : Calage, décomposition géographique des sous-échantillons	23
IV-C : Pondération finale : les traitements des extensions en Guadeloupe et Martinique	29
Partie V : Autres essais de calage.	30
V-A : Calage sur CS.....	30
V-B : Autres calages	32
Partie VI : Synthèse sur la dispersion des poids.....	34
VI-A : Introduction	34
VI-B : Des poids de tirage de l'enquête filtre aux poids de tirage de l'enquête Handicap-Santé	34
Annexe 1 : détails des ajustements, calibrage de l'échantillon	37
Annexe 2 : Les grandes étapes du programme de tirage de l'échantillon de l'enquête Handicap-Santé, volet « ménages »	41
Annexe 3 : Statistiques descriptives pour la non-réponse	42
Annexe 4 : Effectifs des répondants par strate	47
Annexe 5 : Tableau des odds-ratios (modèle de non-réponse).....	49
Annexe 6 : Synthèse des travaux du CRIEM.....	50
Annexe 7 : Un élément d'appréciation de la pertinence des groupes VQS	56

Partie I : Présentation du document et de l'enquête

Ce document résume les travaux effectués pour établir les poids des individus interrogés au titre du volet ménage de l'enquête Handicap Santé¹. Ces poids permettent l'exploitation du volet de façon autonome. En effet, il a paru opportun de mettre à disposition des données aussi tôt que possible², sachant qu'elles sont représentatives de la population vivant en ménage ordinaire, soit près de 99 % de la population totale.

Il n'en demeure pas moins que l'objectif principal de l'enquête³ Handicap-Santé est de fournir des informations sur les situations de handicap ou de dépendance pour la population vivant en France, dans son ensemble. Toutefois, la chronologie des opérations statistiques et le volume des informations obtenues justifient des traitements séparés par volet.

I-A : Présentation des étapes de calcul des poids

L'organisation générale du document s'articule autour des grandes étapes du calcul des pondérations pour l'enquête Handicap-Santé, volet « ménages ». Après quelques jalons (suite de la partie I), le schéma général de ces calculs est posé.

La méthodologie générale suit un cheminement classique. Une attention particulière est portée à la description de l'échantillon ([partie II](#)) car il est issu d'une enquête filtre. Puis sont exposées les considérations relatives aux traitements pour non-réponse ([partie III](#)). Enfin, les calculs de calage font l'objet d'une [partie IV](#). Des indications sur des essais alternatifs de calage sont ajoutées dans une [partie V](#).

Les extensions de l'enquête dans plusieurs départements sont donc décrites en [partie II](#). Leurs existences, outre la structuration de l'échantillon, ont joué un grand rôle dans le choix final de la méthode de calage, les marges étant définies par une répartition de la population par sexe et âge, puis par zones géographiques. L'alternative principale consiste en un calage introduisant aussi une répartition par catégorie socioprofessionnelle.

Deux éléments caractéristiques des calculs doivent rester présents à l'esprit. D'abord, l'appui du volet ménage sur une enquête filtre induit des conséquences sur l'échantillon, mais autorise un affinement des traitements de non-réponse. Ensuite, la nécessité de surreprésenter des personnes en situation de handicap va se traduire par une dispersion notable des poids. Une synthèse sur cette dispersion complète le document ([partie VI](#)) avant quelques annexes techniques.

I-B : Position du volet ménage au sein du dispositif d'ensemble

En 1998-1999, l'enquête Handicap Incapacité Dépendance (HID) a permis, pour la première fois, d'évaluer les problèmes de handicap et de dépendance, ainsi que les difficultés rencontrées dans leur vie quotidienne par les personnes concernées. Cette enquête a été un grand succès (plus de 180 publications). L'enquête Handicap Santé a pour but de renouveler l'opération, en l'adaptant au nouveau contexte institutionnel et en tenant compte des enseignements de l'enquête HID.

¹ Une présentation générale de l'enquête Handicap-Santé est disponible aux e-adresses : <http://www.insee.fr/fr/methodes/default.asp?page=sources/ope-enq-handicap-sante-menages-hsm.htm> ;

http://www.cnis.fr/ind_enquetes.htm; <http://www.sante-sports.gouv.fr/handicap-sante.html>;

Ce dernier site présentant la documentation la plus complète sur l'ensemble du projet. L'enquête Handicap-Santé comporte un volet ménage et un volet institution. Un document analogue au présent document décrit les éléments relatifs au calcul des poids pour le volet institution. (http://www.insee.fr/fr/publications-et-services/docs_doc_travail/Docf1102.pdf).

Une note précisera les éléments relatifs aux poids permettant l'exploitation conjointe des deux volets.

² Le fichier de données a été mis à disposition en octobre 2009, quelques jours avant le début de la collecte du volet institution.

³ L'enquête (le volet ménage et le volet institution) a été réalisée en partenariat par l'INSEE et la DREES, avec l'appui de plusieurs partenaires : l'Association nationale pour la gestion du fonds pour l'insertion professionnelle des personnes handicapées (AGEFIPH), la Caisse nationale d'allocations familiales (CNAF), la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS), la Caisse d'assurance vieillesse (CNAV), la Caisse nationale de solidarité pour l'autonomie (CNSA), la Fédération nationale des mutuelles françaises (FNMF) et l'Institut de veille sanitaire (InVS).

Comme celle-ci, l'enquête Handicap Santé comprend un volet « Ménages » et un volet « Institutions ».

Le volet ménage est une enquête qui a permis de recueillir les informations pour un peu moins de 30 000 individus, l'enquête ayant eu lieu aux 2^{ème} et 3^{ème} trimestres 2008. Le volet institution est une enquête qui a permis de recueillir les informations pour un peu plus de 9 000 individus, l'enquête ayant eu lieu au 4^{ème} trimestre 2009.

Il y aura donc⁴ :

- 1) un traitement des données du volet ménage seul ;
- 2) un traitement des données du volet institution seul.

Le document présent est consacré aux travaux relatifs au volet ménage seul

I-C : Quelques sigles et tableau récapitulatif

Le tableau récapitulatif inclut les éléments relatifs à l'enquête filtre, Vie Quotidienne et Santé (VQS)⁵.

Les bases de sondage de l'enquête filtre (VQS) sont :

- L'enquête annuelle de recensement 2006 (EAR 2006) pour l'échantillon national et les échantillons des extensions en départements métropolitains. Ces départements sont le Nord (59), le Pas de Calais (62), le Rhône (69) et les Hauts de Seine (92) ;
- Identiques à celles des enquêtes emploi en départements d'outre-mer (EEDOM) pour les échantillons des extensions en départements d'outre mer (DOM, soit Guadeloupe (971) et Martinique (972)).

(Il y a eu des extensions VQS en Réunion et Guyane, mais pas d'extensions HSM).

Phase filtre (VQS) :

- A partir de ces bases : tirage des échantillons, enquête (VQS), redressements pour non réponse, partage des poids pour les départements avec extension (entre l'échantillon « national » et l'échantillon « en extension ») ;
- [L'enquête VQS étant exploitable par elle-même, le calage est également effectué, voir document référencé en note 5. Ce calage n'a pas d'incidences sur les calculs pour HSM.]

Cela permet d'obtenir les bases de sondages de l'enquête HSM (Handicap-Santé, volet Ménage).

Puis l'ensemble des étapes de l'enquête HSM :

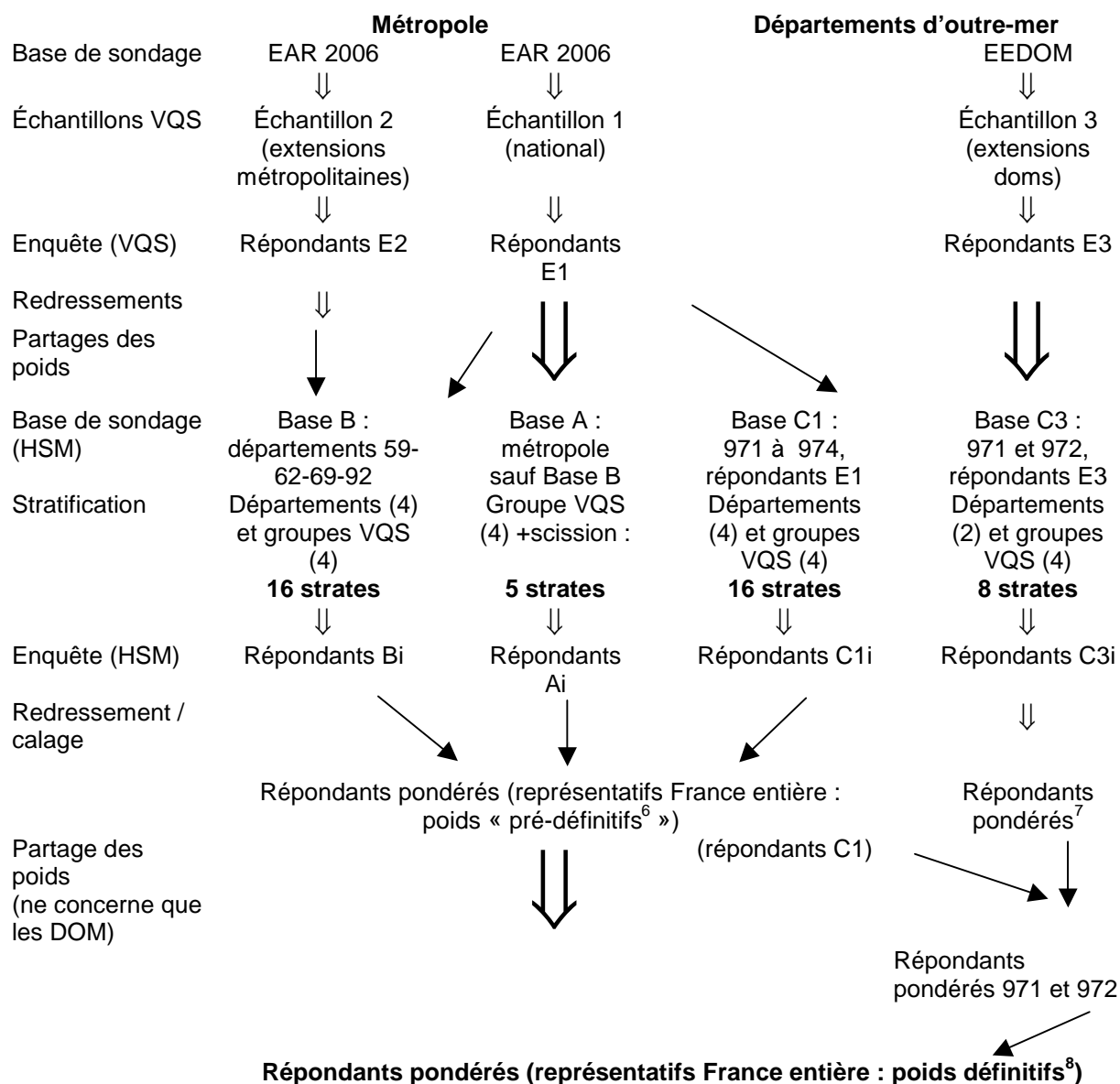
- Stratification et tirage de l'échantillon (HSM) : [partie II](#) de ce document ;
- Enquête (collecte) ;
- Redressements (traitements pour non réponse), [partie III](#) de ce document ;
- Calage des répondants hors extensions en DOM, [partie IV](#) de ce document ;
- Enfin, partage des poids compatible avec le calage précédent, voir aussi [annexe 6](#).

La stratification croise deux critères :

- Géographique, du fait des extensions ;
- Sévérité présumée de la situation de handicap, déterminée avec l'enquête filtre d'où 4 groupes de répondants à VQS (par sévérité croissante), dits « groupes VQS » ;
- Une particularité de l'échantillon national rajoute une strate (en scindant un groupe VQS en deux).

⁴ Apurement et pondérations du volet ménage menés d'octobre 2008 à septembre 2009, apurement et pondérations du volet institution menés de février 2010 à octobre 2010, travaux d'ajustements pour exploitation conjointe des deux volets effectués de septembre 2010 à mars 2011.

⁵ Voir présentation de l'enquête "Vie quotidienne et santé" de 2007, Loïc Midy : http://www.insee.fr/fr/publications-et-services/docs_doc_travail/Docf1001.pdf



⁶ Poids calculés au 1^{er} trimestre 2009, au sein de la division « Enquêtes et Études Démographiques ». Ces poids sont bien utilisables sur le champ France entière, 28 423 individus sont pondérés.

⁷ Poids calculés au 4^{ème} trimestre 2009 / début 2010, Centre de Ressources Interrégional des Enquêtes Ménages, voir aussi [annexe 6](#).

⁸ Partage des poids, calculs faits par le Centre de Ressources Interrégional des Enquêtes Ménages. Intégration finale de tous les éléments par la division « Enquêtes et Études Démographiques ».

Partie II : L'échantillon, les poids de tirage

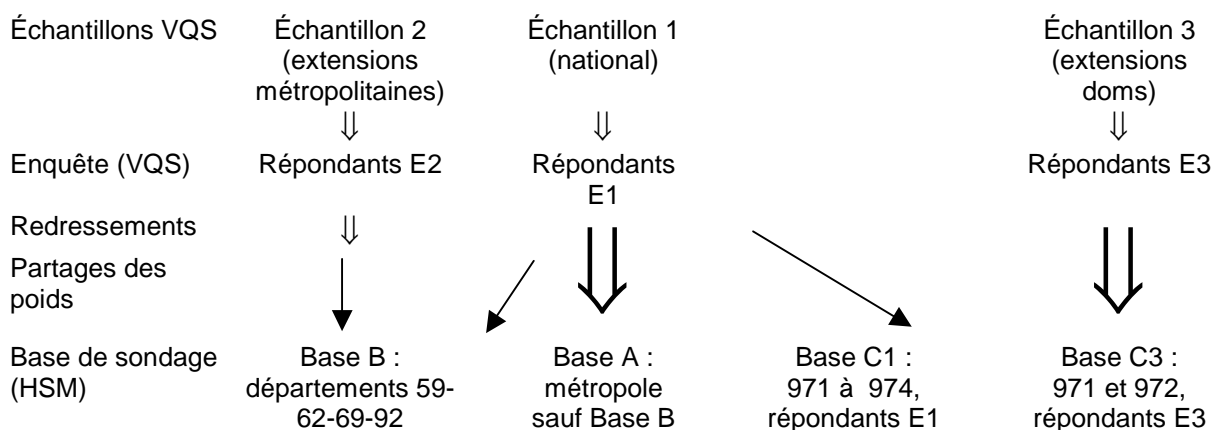
II-A : Synthèse

L'échantillon de l'enquête Handicap-Santé, volet « ménages » est constitué en deux temps. Un échantillon ménage a été construit suivant un mode complexe, permettant d'obtenir environ 260 000 répondants (individus) à une enquête filtre (Vie Quotidienne et Santé, VQS, menée en 2007). Cette complexité induit une dispersion assez élevée des poids. Ces répondants constituent la base de sondage de l'enquête Handicap-Santé, volet « ménages ». L'unité est donc bien, pour l'enquête Handicap-Santé, volet « ménages », l'individu.

Les répondants à l'enquête filtre ont été répartis en quatre groupes (les « groupes VQS » dits « de sévérité présumée de la situation de handicap »). La procédure de tirage stratifiée retenue vise à assurer des effectifs suffisants par groupe VQS que ce soit au niveau national ou dans chaque département ayant fait l'objet d'une extension. Il s'agit là de satisfaire les souhaits du groupe de conception et des co-maîtrises d'ouvrage. Cette procédure comprend la mise en œuvre de principes généraux, complétés par des ajustements.

La probabilité de tirage des individus échantillonnés pour le volet ménage est la probabilité (conditionnelle) d'être échantillonné sachant que l'on est répondant à l'enquête filtre. Cette dernière est elle-même la probabilité (conditionnelle) de répondre sachant que l'on est échantillonné pour l'enquête filtre. Les principes de calcul de ces deux dernières probabilités sont décrits dans le document référencé [note 5](#). On mettra donc l'accent sur le calcul de la première.

Tableau récapitulatif, extrait : des échantillons de l'enquête filtre (VQS) à la base de sondage de l'enquête HSM.



II-B : Répondants de l'enquête-filtre VQS⁹ : base de sondage et groupes

Pour obtenir des statistiques assez précises sur les personnes en situation de handicap ou de dépendance, il est donc important de disposer d'un nombre suffisant de personnes. A cet effet, comme pour l'enquête Handicap Incapacité Dépendance, une enquête filtre préliminaire est réalisée : l'enquête Vie quotidienne et Santé (VQS). Collectée d'avril 2007 à décembre 2007 auprès de 140 000 ménages (environ), par courrier, par téléphone ou en face-à-face, cette enquête demandait une description rapide, à partir de vingt-six questions, des personnes occupant le logement. Celles-ci ont ensuite été classées en fonction de leurs réponses en quatre strates, correspondant à des niveaux différents de sévérité présumée des situations de handicap. Cette enquête filtre est présentée de façon détaillée dans le document référencé en [note 5](#). On en rappelle seulement les éléments essentiels.

S'agissant de l'enquête filtre, Il y a 262 963 individus répondants. Ces répondants proviennent de ménages échantillonnés de trois façons :

⁹ Voir [note 5](#) pour une présentation détaillée.

Échantillon 1 : dit échantillon national, les ménages sont répartis dans tout le champ géographique, donc en France métropolitaine et dans les quatre départements d'outre mer. Ces ménages ont été tirés dans l'EAR 2006 ;

Échantillon 2 : extensions départementales en France métropolitaine. Quatre départements ont procédé à des extensions (Nord, Pas de Calais, Rhône et Hauts-de-Seine). Ces ménages ont également été tirés dans l'EAR 2006 ;

Échantillon 3 : extensions départementales en départements d'outre-mer. Ces extensions ont été adossées aux enquêtes emplois dans les départements d'outre-mer (EEDOM).

13 210 répondants sont dans l'échantillon 3 (extensions en DOM) et en Guyane ou à la Réunion. Ces deux départements pouvaient mener une extension (de l'enquête Handicap-Santé) mais ne l'ont finalement pas fait. Ces 13 210 répondants sont donc retirés de la base des répondants à l'enquête filtre qui sert de base de sondage pour l'enquête Handicap-Santé. Pour cette dernière, il reste 249 753 répondants (à l'enquête filtre). Il sera finalement tiré parmi ceux-ci 39 065 individus pour l'échantillon (et l'on obtiendra 29 931 répondants (à l'enquête Handicap Santé)).

Dans le détail, cela se décline :

Tableau des effectifs dans les bases de sondage, l'échantillon, données VQS (et l'ensemble des répondants, données HSM), suivant les champs.

	« national »	+ extensions « métro » (4 départements)	+ extensions Guadeloupe et Martinique	+ extensions « VQS », soit Guyane et Réunion
Base de sondage (répondants VQS)	186 863	238 813	249 753	262 963
Échantillon (pour HS-M)	28 127	37 074	39 065	-
Répondants (à HS-M)	21 696	28 423	29 931	-

Il convient d'insister sur le fait que l'enquête filtre ne devait servir que de filtre, et ne pas être exploitée en tant qu'enquête. Toutefois, le nombre de répondants est tel qu'il permet une exploitation selon des problématiques et indicateurs spécifiques¹⁰. Les données ont donc été traitées de façon complète : apurement, redressements et calage¹¹. On dispose ainsi des poids des répondants à cette enquête, que ce soient les poids de tirage, les poids redressés de la non-réponse et les poids calés.

A partir des réponses à l'enquête filtre VQS on calcule un « score » pour chaque répondant (de cette enquête). Ce score sera d'autant plus élevé que les réponses à l'enquête laissent présumer un mauvais état de santé ou une situation de handicap. Voir encore le document référencé [note 5](#) pour une présentation détaillée. Une normalisation permet de ramener ce score entre 0 (très bonne santé présumée) et 100. Il convient de noter que l'âge intervient dans la définition des groupes :

Score <	âge < = à 19	19 < âge <= 59	59 < âge <= 79	79 < âge	Groupe
	1	4	5	6	I

¹⁰ Par exemple, d'examiner d'éventuelles disparités régionales. Ce que la constitution et la taille de l'échantillon HS-M ne permettent pas. Une telle exploitation est présentée dans l'INSEE-Première 1254 de Loïc Midy : Limitations dans les activités et sentiment de handicap ne vont pas forcément de pair, <http://insee.fr/fr/ffc/ipweb/ip1254/ip1254.pdf>. Le sentiment de handicap est de plus un indicateur disponible dans l'enquête filtre, mais non repris dans l'enquête elle-même.

¹¹ Y compris les 13 210 observations qui ne seront pas utiles pour la constitution de l'échantillon de l'enquête Handicap-Santé. Toutefois, les pondérations de l'enquête filtre ont été calculées en deux temps (et deux lieux) : pondérations pour les échantillons 1 et 2 de l'enquête filtre en 2008, au sein de la division « Enquêtes et Études Démographiques », pondérations pour l'échantillon 3 fin 2009, au sein du Centre de Ressources Interrégionales des Enquêtes Ménages. L'étude mentionnée (note 10) ne prend donc pas en compte les répondants de ce dernier échantillon.

	10	12	25	40	II
	25	30	45	65	III
					IV

Note de lecture : si l'âge du répondant est compris entre 19 et 59 ans et que son score est compris entre 13 et 30, il sera classé dans le groupe III.

Champ : Les 262 963 répondants à l'enquête filtre VQS.

Quelques éléments pour apprécier la pertinence des groupes au regard de premiers résultats de l'enquête (Handicap-Santé) sont présentés en [annexe 7](#).

II-C : Taille de l'échantillon, les contraintes géographiques

L'objectif initial est d'avoir des données suffisamment précises pour des exploitations au niveau national. Cependant, comme pour l'enquête HID, des collectivités locales ont souhaité des extensions. Ces extensions sont faites pour les six départements : Nord, Pas-de-Calais, Rhône, Hauts-de-Seine, Guadeloupe et Martinique. Il a été décidé de faciliter d'éventuelles exploitations à un niveau infranational en effectuant des choix dans les méthodes de calages (qui seront présentés plus loin), dans la mesure où les tailles des sous-échantillons sont encore suffisantes. Ce niveau infranational est toutefois seulement celui des regroupements de régions¹².

La taille de l'échantillon est estimée lors de travaux préparatoires, reprenant largement ceux faits pour l'enquête HID et complétés par les résultats de celle-ci. Un effectif d'environ 20 000 répondants est visé, hors toutes extensions. Pour chaque extension métropolitaine, un effectif d'environ 2 400 répondants est visé, la moitié seulement pour les extensions ultra-marines. Le taux de réponse estimé est de 75 % (pour HID, il était légèrement inférieur à 78 %).

La décomposition par zones géographiques (avec division suivant l'échantillon en départements d'outre-mer) est alors la suivante :

Zones géographiques	Échantillon cible
Guadeloupe (national)	660
Guadeloupe (extension)	1 000
Martinique (national)	660
Martinique (extension)	1 000
Guyane	660
Réunion	660
Nord	3 200
Pas de Calais	3 200
Hauts de Seine	3 000
Rhône	3 000
Autres départements	22 000
Total	39 040

(Échantillon cible, soit nombre de personnes à enquêter y compris non-répondants).

II-D : L'échantillon, décomposition par groupe VQS

Le principe de surreprésenter les personnes en situation présumée sévère de handicap répond d'abord au besoin d'avoir un nombre suffisant de personnes en situation de handicap¹³, sachant que ces situations sont diverses et nombreuses. Mais l'un des constats d'HID avait été qu'il est nécessaire de garder suffisamment de personnes en bonne santé, pour d'une part avoir des éléments de comparaison, d'autre part, limiter la dispersion des poids et les conséquences sur la précision de « biens portant VQS - en mauvaise santé HSM » : ces personnes auraient un poids très fort¹⁴ parce

¹² C'est donc le niveau 2 de la nomenclature européenne (NUTS), ou les « ZEAT » françaises : soit Départements d'outre-mer, Île-de-France, Bassin Parisien sauf Île-de-France, Nord et Pas-de-Calais, Grand Est, Grand Ouest, Sud-ouest, Rhône-Alpes-Auvergne, Pourtour Méditerranéen.

¹³ L'expression est déjà lourde, mais imprécise : les thèmes de l'enquête incluent la dépendance et la santé, au-delà de la pluralité des situations de handicap.

¹⁴ Et ce d'autant plus qu'elles sont peu nombreuses dans l'échantillon de l'enquête HSM.

qu'échantillonnées comme « biens portantes VQS ». S'il s'avère qu'elles sont en mauvaise santé, par exemple affectées d'une déficience rare, la prévalence de cette déficience pourra être sensiblement surestimée.

Il avait donc été prévu de ne pas (trop) dépasser un ratio de poids de tirage de 10, entre les personnes « en bonne santé » et les personnes présumées comme les plus handicapées. Cela signifie aussi qu'il y aura donc plus de personnes « en bonne santé », relativement à ce qui s'était passé pour l'enquête HID.

II-E : Grandes étapes de la construction de l'échantillon

Étape 1 : On part donc d'une première estimation de la taille de l'échantillon, avec la décomposition par zone géographique, décrite dans II-B.

L'exploitation de l'enquête filtre (voir le document référencé en note 5) conduit à définir les « groupes VQS », au nombre de 4 :

Groupe I : pas de situation de handicap (présumée) ;

... ;

Groupe IV, situation de handicap présumée sévère.

Étape 2 : Au regard des effectifs et des objectifs de l'enquête, on définit alors des taux de sondages cibles par groupes, ce qui va conduire à une ré-estimation de la taille de l'échantillon, toutes zones.

Des 262 989 répondants à l'enquête filtre VQS, on retire les 13 210 répondants dits « E3 », soit les personnes échantillonnées dans les extensions de Réunion et Guyane (ces départements ne menant finalement qu'une extension VQS et non HSM).

Il y a donc 249 779¹⁵ répondants « utiles » pour l'enquête HSM.

On retient le principe de « tirer » toutes les personnes de groupe IV, une sur deux de groupe III, une sur cinq de groupe II, et de « solder » sur le groupe I.

Plus précisément : connaissant les effectifs de répondants à VQS et ayant fixé les taux cibles (suivant les contraintes d'effectif total et de répartition entre groupes), on peut calculer les effectifs « attendus » par groupe, soit le nombre de personnes que l'on souhaite échantillonner par groupe :

Nombre de personnes dans la base de sondage puis dans l'échantillon de l'enquête, suivant le groupe de sévérité présumée de handicap (donc fonction du taux cible) :

Groupe VQS	I	II	III	IV	total
Effectif dans la base de sondage	182 583	42956	15 627	8 613	249 779
Taux cible	0,074	0,2	0,49	1	
Effectif « attendu » de l'échantillon	13 511	8 591	7 657	8 613	38 372

La première cible était de 39 040. Elle est révisée (provisoirement) à 38 372.

Étape 3 : Ajustement, par la création d'un sous groupe VQS.

Le souci est de limiter la dispersion des poids. Or celle ci va provenir essentiellement de deux facteurs :

Une dispersion « initiale » due au mode de tirage dans l'EAR 2006¹⁶ ;

La dispersion du tirage à probabilités inégales suivant les groupes VQS.

Il est donc décidé de tirer avec probabilité 0,2 (au lieu de 0,074) les personnes dont le poids de tirage dans l'EAR 2006 est supérieur à 400. Cela augmente donc la taille de l'échantillon (d'environ 600 personnes).

On dispose alors des effectifs à échantillonner par zone et par groupe en appliquant les taux cibles aux effectifs de répondants VQS :

¹⁵ Après apurements finals, 262 963 et 249 753 exactement.

¹⁶ Ce problème n'existait pas pour HID, le filtre étant adossé au recensement.

Décomposition des répondants à l'enquête VQS :

Effectifs VQS						
<i>Par groupes...</i>	I-a	I-b	II	III	IV	Total
Guadeloupe (national)	3 275		868	285	199	4 627
Guadeloupe (extension)	4 855		620	210	144	5 829
Martinique (national)	3 366		929	303	205	4 803
Martinique (extension)	3 981		764	227	165	5 137
Guyane	3 788		877	291	160	5 116
Réunion	4 315		1 195	360	255	6 125
Nord	14 966		3 467	1 358	812	20 603
Pas-de-Calais	12 577		3 168	1 394	784	17 923
Hauts-de-Seine	14 155		2 654	862	434	18 105
Rhône	11 512		2 541	853	469	15 375
Autres départements	101 155	4 638	25 873	9 484	4 986	146 136
Total	177 945	4 638	42 956	15 627	8 613	249 779

Groupe I-a : groupe VQS I, poids de tirage dans l'EAR inférieur à 400.

Groupe I-b : groupe VQS I, poids de tirage dans l'EAR supérieur à 400.

On dispose maintenant des taux cibles :

Taux cibles	0,074	0,2	0,2	0,49	1
<i>Par groupes...</i>	I-a	I-b	II	III	IV

D'où :

Deuxième estimation des échantillonnés pour l'enquête HSM :

Cible¹⁷ révisée (1)	0,074	0,2	0,2	0,49	1		<i>Rappel :</i>
<i>Par groupes...</i>	I-a	I-b	II	III	IV	Total	Cible¹⁷ initiale
Guadeloupe (national)	242		174	140	199	755	660
Guadeloupe (extension)	359		124	103	144	730	1 000
Martinique (national)	249		186	148	205	788	660
Martinique (extension)	295		153	111	165	724	1 000
Guyane	280		175	143	160	758	660
Réunion	319		239	176	255	990	660
Nord	1 107		693	665	812	3 278	3 200
Pas-de-Calais	931		634	683	784	3 031	3 200
Hauts-de-Seine	1 047		531	422	434	2 435	3 000
Rhône	852		508	418	469	2 247	3 000
Autres départements	7 485	928	5 175	4 647	4 986	23 221	22 000
Total	13 168	928	8 591	7 657	8 613	38 957	39 040

Étape 4 : Ajustements des taux cibles par zone.

Globalement, les effectifs suivant les deux estimations sont donc très proches. Mais ce n'est pas le cas zone par zone, d'où un nouvel ajustement « ligne par ligne ».

Les principes sont de garder autant que possible l'effectif calculé (deuxième estimation) pour les groupes IV en priorité, groupes I ensuite, avec ajustements sur les groupes II et III. Ajoutons à cela des demandes de certains départements pour privilégier le groupe III, et le souhait du concepteur de garder les marges (effectifs par zone géographique et effectifs par groupe, totaux nationaux) du tableau précédent autant que possible.

On obtient la troisième estimation des effectifs (par zone et groupe) :

¹⁷ Il s'agit du nombre de personnes à échantillonner (avant non-réponse).

Effectifs HSM (3¹⁸)						
Par groupes...	I-a	I-b	II	III	IV	Total
Guadeloupe (national)	242		87	132	199	660
Guadeloupe (extension)	386		260	210	144	1 000
Martinique (national)	250		139	91	180	660
Martinique (extension)	318		290	227	165	1 000
Guyane	265		119	116	160	660
Réunion	270		100	100	190	660
Nord	1 110		678	600	812	3 200
Pas-de-Calais	930		683	803	784	3 200
Hauts-de-Seine	600		1 104	862	434	3 000
Rhône	852		826	853	469	3 000
Autres départements	7 400	900	4 650	4 064	4 986	22 000
Total	12 623	900	8 936	8 058	8 523	39 040

Étape 5 : Ajustement pour une contrainte de collecte

Une contrainte a été imposée : on n'interroge que deux personnes au plus par ménage. S'il y a plus de deux personnes échantillonnées dans un même ménage, il faut donc tirer au sort deux personnes parmi elles, les autres n'étant plus échantillonnées. Pour obtenir l'effectif par strate visé, il faut donc tirer à nouveau d'autres personnes (dans la même strate). Enfin, les strates dont le taux cible initial égal à 1 ne peuvent permettre un retraitage, donc des personnes de ces strates seront remplacées par des personnes d'une autre strate. Différentes précisions sont apportées en [annexe 1](#).

Les effectifs définitifs de l'échantillon, par groupes et par zones sont alors :

Effectifs HSM	(effectifs définitifs de l'échantillon)					
Par groupes...	I-a	I-b	II	III	IV	Total
Guadeloupe (national)	244		88	133	197	662
Guadeloupe (extension)	389		262	203	144	998
Martinique (national)	252		140	92	182	666
Martinique (extension)	321		292	222	158	993
Guyane	267		120	117	139	643
Réunion	272		101	101	192	666
Nord	1 119		684	605	791	3 199
Pas-de-Calais	938		689	810	766	3 203
Hauts-de-Seine	626		1 122	833	419	3 000
Rhône	859		833	818	455	2 965
Autres départements	7 463	908	4 690	4 099	4 910	22 070
Total	12 750	908	9 021	8 033	8 353	39 065

¹⁸ Il s'agit donc de la troisième estimation de la taille de l'échantillon.

Le poids réel par groupe et zone s'en déduit :

Poids de tirage : effectif répondant VQS / effectif échantillon HSM						
Par groupes...	I-a	I-b	II	III	IV	rappel effectif
poids cible (inverse des taux)	13,51	5,00	5,00	2,04	1,00	
poids réel						
Guadeloupe (N)	13,42		9,86	2,14	1,01	662
Guadeloupe (X)	12,48		2,37	1,03	1,00	998
Martinique (N)	13,36		6,64	3,29	1,13	666
Martinique (X)	12,40		2,62	1,02	1,04	993
Guyane	14,19		7,31	2,49	1,15	643
Réunion	15,86		11,83	3,56	1,33	666
Nord	13,37		5,07	2,24	1,03	3 199
Pas-de-Calais	13,41		4,60	1,72	1,02	3 203
Hauts-de-Seine	22,61		2,37	1,03	1,04	3 000
Rhône	13,40		3,05	1,04	1,03	2 965
Autres départements	13,55	5,11	5,52	2,31	1,02	22 070
ratio par groupe	13,96	5,11	4,76	1,95	1,03	39 065

Le ratio par groupe donne une valeur moyenne des poids par zone.

II-F : Calcul du poids de tirage (enquête HSM)

On note les poids des répondants à l'enquête filtre (VQS), respectivement :

Poids de tirage : WT_VQS ;
Poids de tirage, redressés de la non-réponse : WR_VQS ;
Poids calés : WC_VQS (ces poids n'interviennent pas dans le calcul).

Ces calculs de poids WR_VQS sont présentés dans le document référencé en [note 5](#). Ils ont été effectués au sein de la division « Enquêtes et Études Démographiques » (EED)¹⁹ pour les répondants dans les échantillons 1 et 2. Les calculs relatifs aux répondants de l'échantillon 3 ont été faits au sein du Centre de Ressources Interrégional des Enquêtes Ménages (CRIEM). Celui-ci a suivi d'aussi près que possible la méthode d'EED : même choix des variables explicatives de la non-réponse, mêmes codifications.

Les travaux et données présentés maintenant portent sur les 238 813 individus répondants (à VQS) au titre des échantillons 1 et 2 (donc hors toutes extensions en dom). L'échantillon pour l'enquête HSM correspondant à pour effectif 37 074 individus (et il y aura 28 423 répondants à l'enquête HSM).

Les 1 991 échantillonnés manquants (39 065 - 37 074) sont des individus dans l'échantillon C3, soit des personnes répondants au titre des extensions en Guadeloupe et Martinique. Il y a 10 940 répondants à l'enquête filtre, d'où les 1 991 échantillonnés. Voir [annexe 6](#). Pour ces échantillonnés (puis répondants), le CRIEM a déterminé les poids WT_VQS et WR_VQS après que l'ensemble des travaux présentés ici eussent été achevés²⁰. La dispersion des différents poids, qu'elle soit calculée pour les 39 065 ou 37 074 échantillonnés, ne change guère, à une petite déformation près due aux faibles poids de tirage dans les deux départements ultramarins.

La dispersion des poids WR_VQS est importante. Ainsi, à titre indicatif, la dispersion des poids, calculée sur les 238 813 répondants est donnée dans le tableau suivant. Cette dispersion résulte d'abord de la présence d'extensions : il y a environ un quart des répondants (VQS) dans les départements à extensions et métropolitains, de poids médian environ un cinquième du poids médian de tirage pour la zone « autres départements ». L'effet est analogue dans les départements d'outre-mer, et la prise en compte des 262 966 répondants aurait pour effet de déformer encore cette

¹⁹ Cette division fait partie de l'Unité « Études Démographiques et Sociales ». Les travaux ont été effectués courant 2008, alors que les calculs pour l'échantillon 3 ont été effectués fin 2009.

²⁰ Les poids WE_HS avaient cependant été calculés pour l'ensemble des échantillonnés pour l'enquête HS.

dispersion vers le bas, soit d'assez peu affecter les centiles au-delà de la médiane, mais de plus en plus les premiers centiles, puisque les poids de tirage des personnes en départements d'outre-mer sont, dès le tirage, dans le bas de la distribution, et de plus et en ordre de grandeur, divisés par 2,5 du fait que l'échantillon en extension est un peu plus important que la partie en dom de l'échantillon « national ». En second lieu, il y a aussi une dispersion attribuable aux procédés de tirage (dans l'EAR 2006). Dans chaque zone, le rapport interquartile de poids est de l'ordre de 1,3 et le rapport inter décile de l'ordre de 2. Enfin, ces procédés conduisent à avoir des individus de poids de tirage très élevés, habitant dans des départements peu urbanisés (Somme, Cher, les deux Savoie).

Ainsi, à titre indicatif, la dispersion des poids, calculée sur les 238 813 répondants au titre des échantillons 1 et 2 (donc hors toutes extensions en DOM) est donnée dans le tableau suivant.

Tableau des dispersions de poids **pour l'enquête-filtre**, sur 238 813 répondants.

Poids VQS	Tirage	Redressé non-réponse	<i>Calés</i>
(238 813 obs)	WT_VQS	WR_VQS	WC_VQS
Max	1 064	1409	1580
P99	712	852	879
P95	352	450	468
P90	320	415	431
P75	297	365	375
P50	227	267	276
P25	75	94	106
P10	53	66	73
P5	28,1	34,7	36,8
P1	20,9	26,1	27,6
Min	12,6	16,0	15,7
D9/D1	6	6	6
Q3/Q1	4	4	4

(Champ : répondants à l'enquête filtre VQS, hors extensions en DOM).

Il s'agit maintenant de définir un poids « initial » ou un poids de tirage des individus échantillonnés pour le volet ménage de l'enquête Handicap Santé que l'on note WT_HS (la probabilité correspondante étant notée PT_HS).

Le principe de tirage pour le volet ménage est le suivant :

- Un individu est échantillonné dans VQS ;
- Il répond à VQS (il est donc dans la base de sondage d'HS-M) ;
- Il est échantillonné dans HS-M.

On doit donc disposer de deux probabilités / poids :

La probabilité de répondre à VQS (poids redressé de la non-réponse à cette enquête) : PR_VQS
 La probabilité d'être échantillonné dans HS-M sachant que l'on est répondant à VQS : PE_HS

$$\text{Et l'on a la relation } \mathbf{PT_HS = PE_HS * PR_VQS.}$$

Cette relation se décline également en terme de poids (W....) :

$$\mathbf{WT_HS = WE_HS * WR_VQS}$$

Le calcul de WE_HS se comprend directement à partir des éléments présentés supra :

Un individu est échantillonné en deux temps :

- Tirage avec poids (ou probabilités) calculés zone par zone et groupe par groupe (soit 45 strates) ;
- Processus d'élimination-rattrapage.

Avec les approximations détaillées en [annexe 1](#), cela signifie que ce poids WE_HS est le produit de deux poids intermédiaires :

Le poids de tirage par groupe et zone donné par le tableau final de la partie II-E ;
 Un poids de retraitage valant 1 pour la très grande majorité des échantillonnés sauf si :

L'individu est dans un ménage ou N (avec $N > 2$) individus sont « tirés » lors de la phase précédente, avec retraitage de deux individus et élimination des autres, soit

Poids = $N / 2$ si l'individu est échantillonné ;

Poids = 0 s'il ne l'est plus.

[L'annexe 2](#) présente en outre les grandes étapes du programme de tirage de l'échantillon de l'enquête Handicap-Santé, volet « ménages ».

Synthèse : la dispersion des différents poids (sur 37 074 échantillonnés) :

37 074 obs	WT_VQS	WR_VQS	W_VQS->HS	WT_HS
	Poids de tirage pour enquête filtre	Poids redressé pour enquête filtre	Poids de tirage de l'échantillon HS parmi les répondants à VQS	Poids initial
Max	1 064	1 409	33,70	10 240
P99	797	936	22,47	6 073
P95	365	466	13,51	5 338
P90	321	422	13,51	4 765
P75	295	367	13,39	2 256
P50	214	264	5,05	846
P25	69,6	88,5	1,29	275
P10	53,3	67,4	1	102
P5	28,1	38,0	1	72
P1	20,9	28,1	1	36
Min	12,6	16,0	1	16
D9/D1	6	6	13,5	47
Q3/Q1	4	4	10	8

(Champ : échantillonnés HSM à partir des répondants à l'enquête filtre VQS, hors extensions en DOM).

Partie III : Les répondants

III-A : Résumé de la méthode suivie

La recherche d'un modèle de non-réponse a débuté par le constat que l'on dispose, pour les individus échantillonnés de deux types d'informations. D'une part, on connaît des informations sur le ménage (où vit l'individu) telles que l'implantation géographique, le type d'habitat, le statut d'occupation du logement, l'état matrimonial, la situation sur le marché du travail et le diplôme de la personne de référence²¹. D'autre part, s'ajoutent à ces informations, des données individuelles, soit les caractéristiques démographiques (sexe et âge), mais aussi des caractéristiques de santé : les réponses à VQS. Celles-ci peuvent s'utiliser directement ou par le truchement du « groupe VQS » (ou toute autre variable construite de façon similaire). Ces informations ont été testées et pour la plupart exploitées pour obtenir le modèle permettant de calculer les poids redressés de la non-réponse.

Une première étude de la non-réponse a été menée à partir des codes de résultats de l'enquête. Celle-ci conduit à garder les échantillonnés potentiellement hors-champ car faire un modèle de redressement de la non-réponse et un modèle de sortie du champ complique le traitement alors que :

- 1) L'utilisation des variables de VQS permet de redresser les non-réponses pour, par exemple, raison de décès ou de départ en institution.
- 2) Les codes de résultats de l'enquête ne garantissent pas l'exhaustivité du repérage de ces hors champ (par exemple des décédés)
- 3) Au final, le faible nombre de hors-champs repérés autorise un traitement unique.

Le principe est alors de faire un modèle de réponse à l'enquête, suivant une régression logistique sur des variables de différents types :

S : Santé (réponses à l'enquête filtre).

D : Démographie (âge et sexe)

H : Habitat (implantation géographique, type d'habitat, statut d'occupation du logement)

M : État matrimonial, situation sur le marché du travail et diplôme (information ménage)

Ces variables sont intégrées successivement, mais, évidemment, le meilleur modèle « S+D » n'est pas la simple juxtaposition du meilleur modèle S et du modèle D, etc.....

La dispersion induite par l'intégration de ces informations peut se résumer par :

	Poids initial	S	S+D	S+D+H	Final
	37 074 éch.	28 423 répondants			
Max	10 240	12 402	13 936	16 130	16 632
P99	6 073	7 873	8 041	9 229	9 220
P95	5 338	6 881	6 876	6 998	7 007
P90	4 765	6 146	6 110	5 885	5 882
P75	2 256	2 960	2 994	3 086	3 086
P50	846	1 131	1 126	1 103	1 098
P25	275	383	378	365	383
P10	102	145	143	142	142
P5	72	97	97	102	103
P1	36	55	54	53	55
Min	16	23	22	25	25
D9/D1	47	42	43	41	41
Q3/Q1	8	8	8	8	8

(Champ : échantillonnés pour HSM (37 074), répondants à HSM (28 423), hors extensions en DOM).

²¹ Données disponibles dans l'EAR pour l'échantillon national et les extensions en départements métropolitains.

III-B : Quelques éléments de cadrage pour les taux de réponses et la non-réponse

Le volet ménage a été collecté entre le 1^{er} avril et le 15 octobre 2008. L'échantillon initial comprend 39 065 personnes sur l'ensemble du territoire français, départements d'outre-mer compris. 76,6 % des enquêtés ont répondu. Au total, les fichiers portent donc sur 29 954 répondants.

Ces répondants sont jusqu'ici les individus pour lesquels on a les réponses au questionnaire, éventuellement de façon incomplète. Le travail d'apurement a conduit à distinguer des questionnaires complets de questionnaires incomplets, ces derniers au nombre de 73. Un questionnaire incomplet correspond essentiellement à un abandon en cours de questionnaire. Il est convenu que l'on réintègre les questionnaires parmi les répondants dès lors que l'on dispose des réponses aux modules de « cœur d'enquête » : au moins les 2/3 du questionnaire, incluant les questions sur santé, déficiences, ... restrictions d'activité). Cela permet de réintégrer 50 questionnaires sur l'ensemble des répondants (les 29 954), d'où les 29 931 répondants (= 29 954 - 73 + 50).

Hors extensions en DOM, les chiffres correspondants sont de 28 445 répondants avant apurement et 28 423 répondants après apurements.

Ces 29 931 répondants se répartissent géographiquement et par type d'échantillon :

Répartition des répondants (volet ménage)

	Départements	Tous sauf ...	59	62	69	92	DOM	
Échantillons								
National		17 101	829	571	838	371	1 986	21 696
Extensions en 59, 62, 69 et 92		0	1 749	2 085	1 372	1 521	0	6 727
Extensions en Guadeloupe et Martinique		0	0	0	0	0	1 508	1 508
Totaux		17 101	2 578	2 656	2 210	1 892	3 494	29 931

Dans un premier temps, le redressement a porté uniquement sur les répondants hors extensions en Guadeloupe et Martinique. Soit 37 074 échantillonnés et 28 423 répondants.

Ces redressements ont été repris à l'identique (autant que possible) pour les 1 991 autres échantillonnées et les 1 508 répondants correspondant : choix des mêmes variables et constructions des mêmes modalités pour « expliquer » la non-réponse.

Le redressement est donc décrit avec les données portant sur les 37 074 échantillonnés et 28 423 répondants. Les tableaux font donc référence à ces 37 074 ou 28 423 individus.

Une étude liminaire et exploratoire de la non-réponse a été menée à partir des codes de résultats de l'enquête. Elle figure en [annexe 3](#). Cette étude est menée à partir des codes de résultats d'enquête (d'où les 37 074 codes traités). Il en ressort que :

Le premier motif de non-réponse est constitué par les refus directs ou indirects ; ils sont plutôt le fait d'hommes de 20 à 30 ans ou de personnes vivant en Île-de-France, tandis que les personnes vivant en Nord ou en Pas de Calais ou déclarant un très mauvais état de santé à l'enquête filtre sont moins coutumières de ces pratiques.

Les deuxième et troisième motifs sont les déménagements (hors institutions) et échecs à différents moments du contact. Cela concerne plutôt les personnes de 20 à 30 ans ou en bonne santé. Inversement les personnes âgées sont moins concernées.

Ces trois motifs concernent 73 % des non-répondants. Il reste alors 2 362 échantillonnés n'ayant pas répondu pour d'autres motifs dont :

Les décès : surtout des personnes de plus de 80 ans (et plutôt des hommes) ou des personnes en mauvaise santé ;

Les départs en institutions : même chose, mais plutôt des femmes ;

Inaptitudes diverses à l'entretien.

Par strate, on a les taux de réponses suivants :

Taux de réponses (%)	I	II	III	IV	Tous groupes
Guadeloupe (N)	73,4	77,3	83,5	70,6	75,1
Guadeloupe (X)	72,2	80,5	81,8	77,1	77,1
Martinique (N)	78,2	79,3	83,7	76,9	78,8
Martinique (X)	70,7	76,0	77,9	74,1	74,4
Guyane	66,3	69,2	65,8	60,4	65,5
Réunion	79,4	79,2	88,1	82,8	81,7
Nord	80,3	84,2	83,0	76,1	80,6
Pas-de-Calais	83,2	83,5	84,7	80,5	83,0
Hauts-de-Seine	63,3	67,8	59,1	59,2	63,2
Rhône	73,9	77,3	74,0	71,9	74,6
Autres	77,4	80,0	78,4	74,3	77,5
Toutes zones	76,5	78,7	77,1	74,1	76,6
Hors extensions DOM	76,8	78,7	76,9	74,1	76,7

Champ : échantillonnés pour l'enquête HSM, France métropolitaine et DOM (toutes zones et extensions).

Voir en [annexe 4](#) les effectifs de répondants. Dans cette annexe figurent également les effectifs avec la distinction échantillon national / échantillons pour extensions, pour les quatre départements métropolitains.

III-C : Intégration progressive des informations et effets

Le redressement pour non-réponse consiste à modéliser une probabilité de réponse (d'une personne dans l'échantillon). Il s'agit de modèles économétriques (« logit »), où la variable dichotomique (répond / ne répond pas) est « expliquée » par des variables disponibles, qui vont être détaillées. On cherche un modèle aussi prédictif que possible puis on garde l'inverse de la probabilité de réponse comme poids de redressement (pour les répondants).

On présente maintenant les familles de modèles testés. Pour tous les tableaux de répartition des poids et de dispersion des ratio poids redressé sur poids initial, le champ est constitué par les 37 074 échantillonnés, puis les 28 423 répondants, soit hors extensions en DOM.

III-C-1 : Modèles utilisant les informations de type « santé »

La première étape a été la recherche d'un modèle utilisant les variables individuelles VQS, soit de santé. Toutes les variables VQS ont été testées (il y en a 26), le modèle S n'en retient que quelques-unes. Par ailleurs, il s'est avéré pertinent de regrouper les modalités, soit essentiellement de créer une variable dichotomique distinguant un mauvais état de santé ou une forte limitation, par opposition aux autres modalités.

Répartition des poids (le poids total des répondants est de 59,2 millions)

Poids	initial	redressé	
	37 074 échantillonnés / 28 423 répondants		
	(a)	(b)	(b)/(a)
Max	10 240	12 402	1,21
P99	6 073	7 873	1,30
P95	5 338	6 881	1,29
P90	4 765	6 146	1,29
P75	2 256	2 960	1,31
P50	846	1 131	1,34
P25	275	383	1,39
P10	102	145	1,42
P5	72	97	1,35
P1	36	55	1,53
Min	16	23	1,44
D9/D1	47	42	
Q3/Q1	8	8	

Et la dispersion du ratio poids redressé sur poids initial (de l'enquête HSM, résultat de la partie II) est la suivante :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	1,198	1,218	1,218	1,254	1,262	1,305	1,305	1,389	1,502	1,768	2,073	1,107	1,034

III-C-2 : Intégration des informations de type démographique

Il s'agit de rajouter les variables individuelles démographiques : sexe et âge. Après quelques essais, on a retenu un découpage par sexe et tranche d'âge de 10 ans (jusqu'à 80 ans).

Le modèle S+D combine alors les variables de santé et démographiques.

Répartition des poids (le poids total des répondants est de 59,4 millions²²)

Poids	initial	redressé	
	37 074 échantillonnés / 28 423 répondants		
	(a)	(b)	(b)/(a)
Max	10 240	13 936	1,36
P99	6 073	8 041	1,32
P95	5 338	6 876	1,29
P90	4 765	6 110	1,28
P75	2 256	2 994	1,33
P50	846	1 126	1,33
P25	275	378	1,37
P10	102	143	1,40
P5	72	97	1,35
P1	36	54	1,50
Min	16	22	1,38
D9/D1	47	43	
Q3/Q1	8	8	

Et la dispersion du ratio poids redressé sur poids initial de l'enquête HSM est la suivante :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	1,164	1,190	1,211	1,223	1,247	1,269	1,336	1,467	1,517	1,812	2,293	1,199	1,071

III-C-3 : Intégration des informations sur le logement et sa localisation

Ces modèles intègrent les variables ménages concernant l'habitat : département, région voire régions regroupées, avec notamment l'essai dans les modèles de la variable de stratification (géographique), et des variables sur le type de logement, sa localisation dans le tissu urbain et le statut d'occupation. Concernant le découpage géographique, le dilemme suivant est apparu : créer des variables sur la localisation avec une pertinence géographique (soit départements ou régions connexes) ne donne pas ou peu d'amélioration, donc il est tentant d'utiliser un regroupement basé sur les taux de non-réponse en dehors de tout autre critère pertinent. Cette seconde possibilité a paru meilleure (meilleur accord entre valeur réelle et valeur prédite). De même que pour les variables de VQS, des regroupements de modalités ont apporté de substantielles améliorations.

²² Le total est donc légèrement différent du total obtenu par le modèle « S », ou les suivants. Ces légers écarts proviennent de la méthode suivie. Le modèle « prédit » qu'un individu est répondant ou non suivant ses caractéristiques. La « prédiction » du modèle diffère du comportement observé dans quelque cas (ordre de grandeur : 1 %). Il y a donc quelques substitutions entre vrais répondants et répondants d'après le modèle. Les poids des échantillonnés étant assez dispersés, une substitution ne s'opère pas nécessairement à poids constant. La sommation des poids sur les vrais répondants peut donc s'écarter de la sommation des poids sur les répondants « prédits », qui est elle toujours égale à la somme des poids initiaux des échantillonnés.

Modèle SDH : regroupant les modèles précédents.

Répartition des poids (le poids total des répondants est de 59,4 millions)

Poids	initial	redressé	
	37 074 échantillonnés / 28 423 répondants		
	(a)	(b)	(b)/(a)
Max	10 240	16 130	1,58
P99	6 073	9 229	1,52
P95	5 338	6 998	1,31
P90	4 765	5 885	1,24
P75	2 256	3 086	1,37
P50	846	1 103	1,30
P25	275	365	1,33
P10	102	142	1,39
P5	72	102	1,42
P1	36	53	1,47
Min	16	25	1,56
D9/D1	47	41	
Q3/Q1	8	8	

Et la dispersion du ratio poids redressé sur poids initial de l'enquête HSM est la suivante :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	1,072	1,097	1,120	1,141	1,178	1,255	1,388	1,568	1,698	2,027	3,621	1,375	1,178

III-C-4 : Intégration des informations la personne de référence du logement

Des modèles utilisant les variables ménages concernant l'information disponible sur la personne de référence du logement : état matrimonial, situation sur le marché du travail et diplôme sont enfin testés. Après essai, le modèle SDH n'est amélioré que par l'ajout d'une variable dichotomique (marié vs autres état matrimoniaux).

Le modèle F (final) est donc obtenu en combinant les modèles précédents.

Répartition des poids (le poids total des répondants est de 59,3 millions)

Poids	initial	redressé	
	37 074 échantillonnés / 28 423 répondants		
	(a)	(b)	(b)/(a)
Max	10 240	16 632	1,62
P99	6 073	9 220	1,52
P95	5 338	7 007	1,31
P90	4 765	5 882	1,23
P75	2 256	3 086	1,37
P50	846	1 098	1,30
P25	275	383	1,39
P10	102	142	1,39
P5	72	103	1,43
P1	36	55	1,53
Min	16	25	1,56
D9/D1	47	41	
Q3/Q1	8	8	

Avec la dispersion du ratio poids redressé sur poids initial de l'enquête HSM :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	1,067	1,094	1,118	1,136	1,176	1,251	1,388	1,569	1,701	2,055	3,944	1,381	1,181

Quelques remarques sur les poids extrêmes :

Une lecture des différents poids (du tirage VQS au poids redressé retenu) sur les (environs) 400 observations extrêmes (les 200 poids les plus élevés et les 200 poids les moins élevés) montre que ces résultats extrêmes ne peuvent être évités, mais ne remettent pas en cause la possibilité d'exploiter l'enquête.

Poids les plus élevés : essentiellement des répondants des départements d'Île-de France connus pour leur taux de réponse inférieurs à la moyenne, en bonne santé, d'âge 20-40 ans. Les deux cas les plus extrêmes correspondent de plus à des personnes appartenant à un ménage dont plus de deux personnes ont été échantillonnées. Il s'agit donc de cas où se cumulent tous les facteurs défavorables (pour être répondants), mais comme le fait d'être « en bonne santé » fait partie de ceux-ci, ils n'interviennent pas dans les modalités de réponses d'intérêt. On constate enfin de façon plus générale la grande robustesse des résultats donnés par les différents modèles sur les questions test (choisies dans les différents modules de l'enquête).

Poids les moins élevés : essentiellement répondants des départements d'outre-mer et en mauvaise santé.

Voir en [annexe 5](#) le tableau des odds-ratios du modèle retenu

Partie IV : Calage sur marges : la pondération finale

IV-A : Synthèse

L'étape finale, pour parvenir à la pondération, consiste en un calage sur marges. On a cherché à définir des marges de type démographique et de type social. Cependant l'essentiel du travail est guidé par la recherche d'une décomposition pertinente de l'échantillon en sous-échantillons. La taille de l'échantillon n'est pas prévue pour permettre des exploitations infranationales. Cependant, les départements métropolitains avec extensions sont tels qu'il était tentant de tester des calages avec un critère géographique. Par ailleurs, l'organisation du travail pour les extensions en département d'outre-mer a été facilitée du fait d'avoir séparé ces derniers départements dans la méthode retenue²³. Il s'avère que des calages sur marges de population par région et par sexe-âge sont déjà trop contraignants. Mais une décomposition géographique intégrant les regroupements de régions (zeat, zone économique d'aménagement du territoire) n'induit que peu de dispersion supplémentaire (au regard de celle constatée après calage sur les marges calculées à partir du découpage à partir des départements à extension). On retient donc un découpage de l'échantillon suivant les ZEAT, et si nécessaire, distinguant dans celles-ci les départements à extension. Dans un premier temps, on a effectué les calages par zones géographiques sur les variables démographiques seules. Puis on a testé la possibilité d'imposer des marges par catégories socioprofessionnelles : la dispersion augmente beaucoup sauf à regrouper ces catégories. On retient finalement la méthode de calage sur marges démographiques seulement. On présente donc le calage sur variables démographiques puis en partie suivante, les autres essais.

IV-B : Calage, décomposition géographique des sous-échantillons

IV-B-1 : Introduction

Pour déterminer la méthode de construction des marges de type démographiques, on a testé 4 jeux de pondérations :

- 1) Calage sur France entière (une zone)
- 2) Calages par zeat (groupement de régions), soit 9 zones
- 3) Calages par strate géographique (l'échantillon a été tiré suivant une stratification géographique combinée à la sévérité présumée du handicap, certaines strates correspondent aux extensions), soit également 9 zones.
- 4) Calages par zone géographique combinant les deux critères précédents : il y aura 14 zones.

Les répondants sont des individus vivant en ménages ordinaires, en France, et qui ont répondu à l'enquête filtre (VQS) en 2007, donc qui sont âgés d'au moins un an au moment de l'enquête HSM.

L'enquête HSM sera donc représentative de :

la population vivante en 2008 et déjà vivante en 2007 ;
habitant en logement ordinaire aussi bien en métropole que dans les départements d'outre mer.

Les marges démographiques sont alors construites de la façon suivante :

On dispose²⁴ des données de population en ménage au 1^{er} janvier 2008, par sexe et âge (annuel). On dispose des données de population par sexe-âge (tranche d'âge de 5 ans) par régions, département au 1^{er} janvier 2007. On a alors défini pour les calages les effectifs de population en ménage au 1^{er} janvier 2008, par département, sexe et tranche d'âge quinquennal. Ceci s'est calculé

²³ Le calcul des pondérations de l'échantillon national avec calage incluant une contrainte géographique (calages sur Guadeloupe-Martinique-Ensemble des quatre DOM) assure d'une part la possibilité d'exploiter l'enquête au niveau national avec les poids provisoires, donc dès octobre 2009. D'autre part, il a suffi de recalculer (voir [annexe 6](#)) les poids pour les départements de Guadeloupe et Martinique, en prenant les individus des échantillons en extensions en compte, sans avoir besoin de modifier les poids des autres individus pour finaliser les pondérations.

²⁴ Données fournies par la section « Comptabilité démographique », au sein d'EED.

en considérant que les évolutions de population entre 2007 et 2008 ne dépendent pas des départements. De plus la part des ménages dans le total est considérée comme constante.

L'ensemble des variantes testées balaye les possibilités en termes de découpage géographique, en mélangeant ou adaptant la taille des tranches d'âge (5, 10 ou 20 ans). Les calages peuvent simplement échouer (pas de répondants dans certains croisements trop fins, par exemple région et tranche de 10 ans). Ces calages peuvent aussi converger mais au prix de dispersions des poids considérables. On a testé les résultats en comparant avec quelques variables VQS, la robustesse interne des 4 méthodes et analysé les informations usuelles : répartition des poids et dispersion du ratio poids calé / poids redressé. Un problème apparaît pour les départements d'outre-mer, où il y a environ 500 répondants seulement par département, soit certains effectifs nuls dès que l'on descend à un découpage par tranche d'âge décennale. Il n'est donc pas possible d'effectuer un calage sur marge avec ces croisements. De même, un calage par région apparaît comme trop contraignant au vu de certains effectifs, et il ne semble pas utile de faire des regroupements ad hoc.

IV-B-2 : Données pour marges

La répartition de population entre les répondants à l'enquête et la population en ménage ordinaire est différente ; cela tient d'abord à la méthode de tirage qui, privilégiant les personnes dont on suppose qu'elles seront dans une situation lourde de handicap, surreprésente d'abord les personnes âgées. Il faut souligner que le champ étant les personnes ayant répondu à l'enquête filtre puis à l'enquête elle-même, il y a un déficit de très jeunes personnes (moins d'un an) parmi les répondants à l'enquête.

Tableau : les effectifs et pourcentage par sexe et âge

Tranche	Répondants	Enquête	Population	Tranche	Répondants	Enquête	Population
Femmes, 0-4	418	1,47	3,09	Hommes, 0-4	456	1,60	3,23
Femmes, 5-9	631	2,22	3,09	Hommes, 5-9	807	2,84	3,24
Femmes, 10-14	672	2,36	2,99	Hommes, 10-14	748	2,63	3,13
Femmes, 15-19	735	2,59	3,14	Hommes, 15-19	823	2,90	3,27
Femmes, 20-24	484	1,70	3,13	Hommes, 20-24	495	1,74	3,10
Femmes, 25-29	510	1,79	3,21	Hommes, 25-29	436	1,53	3,12
Femmes, 30-34	629	2,21	3,25	Hommes, 30-34	532	1,87	3,17
Femmes, 35-39	812	2,86	3,61	Hommes, 35-39	754	2,65	3,51
Femmes, 40-44	970	3,41	3,68	Hommes, 40-44	858	3,02	3,53
Femmes, 45-49	1 130	3,98	3,58	Hommes, 45-49	981	3,45	3,40
Femmes, 50-54	1 264	4,45	3,47	Hommes, 50-54	1 080	3,80	3,25
Femmes, 55-59	1 478	5,20	3,45	Hommes, 55-59	1 319	4,64	3,25
Femmes, 60-64	969	3,41	2,77	Hommes, 60-64	886	3,12	2,59
Femmes, 65-69	820	2,88	2,13	Hommes, 65-69	623	2,19	1,89
Femmes, 70-74	998	3,51	2,17	Hommes, 70-74	726	2,55	1,74
Femmes, 75-79	1 133	3,99	2,06	Hommes, 75-79	727	2,56	1,46
Femmes, 80-84	877	3,09	1,59	Hommes, 80-84	490	1,72	0,96
Femmes, 85-89	552	1,94	0,87	Hommes, 85-89	239	0,84	0,45
Femmes, 90-94	198	0,70	0,24	Hommes, 90-94	66	0,23	0,10
Femmes, 95-...	77	0,27	0,09	Hommes, 95-...	20	0,07	0,03
	15 357	54,03	51,59		13 066	45,97	48,41

Note de lecture : il y a 970 répondants dans l'enquête qui sont des femmes de 40 à 44 ans. Soit 3,41 % des répondants ; dans la population en ménages ordinaires utilisée pour le calage, cette part est de 3,68%.

28 423 répondants (hors extensions DOM)

Tableau : les effectifs de répondants par critères géographiques (zeat, « zones », régions).

zeat	Code		Régions	zeat	
Départements d'outre-mer	0	1 986	Alsace	4	554
Île-de-France	1	4 187	Aquitaine	7	879
Bassin parisien	2	3 845	Auvergne	8	498
Nord et Pas-de-Calais	3	5 234	Basse-Normandie	2	557
Est	4	2 008	Bourgogne	2	481
Ouest	5	2 706	Bretagne	5	830
Sud-ouest	7	2 161	Centre	2	791
Rhône-Alpes-Auvergne	8	3 771	Champagne	2	705
Bassin méditerranéen	9	2 525	Corse	9	80
			Départements d'outre-mer	0	1 986
			Franche-Comté	4	516
			Haute-Normandie	2	425
zone			Île-de-France sauf 92	1	2 295
Guadeloupe	GUA	497	Languedoc	9	837
Guyane	GUY	421	Limousin	7	282
Hauts de Seine	HDS	1 892	Lorraine	4	938
Martinique	MAR	525	Midi-Pyrénées	7	1 000
Nord	NRD	2 578	Provence-Alpes-Côte d'Azur	9	1 608
Pas-de-Calais	PDC	2 656	Pays de Loire	5	1 253
Reste de la France	RDF	17 101	Picardie	2	886
Réunion	REU	543	Poitou-Charentes	5	623
Rhône	RHO	2 210	Rhone-Alpes	8	3 273

28 423 répondants (hors extensions DOM)

Zone : les départements intervenant dans l'échantillonnage

Zeat : zone économique d'aménagement du territoire, regroupement de régions

IV-B-3 : Calage 1 : pas de décomposition (France entière, tranches d'âges quinquennales)

Il n'y pas lieu de considérer que certaines tranches d'âges sont d'effectifs insuffisants dans l'enquête²⁵. Le calage s'effectue par méthode de raking ratio. Voici la dispersion des poids (28 423 répondants (hors extensions DOM)).

Poids	Redressé	Calé	
	(a)	(b)	(b)/(a)
Max	16 632	20 210	1,22
P99	9 220	10 295	1,12
P95	7 007	7 515	1,07
P90	5 882	6 133	1,04
P75	3 086	3 172	1,03
P50	1 098	1 104	1,01
P25	383	359	0,94
P10	142	142	1,00
P5	103	103	1,00
P1	55	56	1,02
Min	25	25	1,00
D9/D1	41	43	
Q3/Q1	8	9	

28 423 répondants (hors extensions DOM)

²⁵ En fait, une variante avec regroupement des 90-94 et plus de 95 ans a été testée.

Pour ce calage, la dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,744	0,828	0,829	0,863	0,966	1,027	1,070	1,155	1,287	1,386	1,614	1,338	1,107

28 423 répondants (hors extensions DOM)

IV-B-4 : Calage 2 : décompositions suivant zeat

L'échantillon est décomposé en zeat, les marges sont en tranches d'âge décennales, pour chaque zeat. Le calage s'effectue par méthode de raking ratio.

Ratio median par zeat (ratio poids_final / poids redressé)

ZEAT	Ratio	Rappel intitulé zeat
0	1,39	Dom
1	0,96	Île-de-France
2	0,89	Bassin parisien
3	1,02	Nord et Pas-de-Calais
4	0,94	Est
5	1,15	Ouest
7	1,06	Sud-ouest
8	1,17	Rhône-Alpe-Auvergne
9	1,02	Bassin méditerranéen

Dispersion des poids pour ce calage (28 423 répondants (hors extensions DOM))

Poids	Redressé	Calé	
	(a)	(b)	(b)/(a)
Max	16 632	18 802	1,13
P99	9 220	10 387	1,13
P95	7 007	7 437	1,06
P90	5 882	6 045	1,03
P75	3 086	3 220	1,04
P50	1 098	1 127	1,03
P25	383	373	0,97
P10	142	148	1,04
P5	103	109	1,06
P1	55	69	1,25
Min	25	29	1,16
D9/D1	41	41	
Q3/Q1	8	9	

Et dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,643	0,696	0,767	0,829	0,926	1,043	1,167	1,344	1,464	1,690	2,682	1,622	1,261

IV-B-5 : Calage 3 : décompositions suivant échantillons (= zone)

L'échantillon est décomposé en DOM (4), départements à extensions en métropole (4) puis reste de la France. Les marges sont alors adaptées aux tailles : tranches duodécennales en DOM, décennales en départements à extensions en métropole et quinquennales dans le reste de la France.

Ratio median par zeat (ratio poids_final / poids redressé)

zone	Ratio	Rappel intitulé zone
GUA	2,44	Guadeloupe
MAR	2,48	Martinique
GUY	1,59	Guyane
REU	0,99	Réunion
NRD	1,02	Nord
PDC	1,01	Pas-de-Calais
HDS	0,98	Hauts-de-Seine
RHO	1,25	Rhône
RDF	1,00	Reste de la France

Dispersion des poids pour ce calage (28 423 répondants (hors extensions DOM))

Poids	Redressé (a)	Calé (b)	(b)/(a)
Max	16 632	19 877	1,20
P99	9 220	10 152	1,10
P95	7 007	7 397	1,06
P90	5 882	6 046	1,03
P75	3 086	3 141	1,02
P50	1 098	1169	1,06
P25	383	380	0,99
P10	142	155	1,09
P5	103	114	1,11
P1	55	79	1,44
Min	25	25	1,00
D9/D1	41	39	
Q3/Q1	8	8	

Et dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,545	0,679	0,800	0,863	0,934	1,020	1,128	1,310	1,546	2,537	3,731	1,518	1,207

IV-B-6 : Calage 4 : décompositions combinant les deux précédentes.

L'échantillon est décomposé en 14 zones, soit le découpage le plus petit englobant les zeat (calage 2) et les zones (calage 3). Comme il n'y a pas eu d'extensions en Guyane et Réunion, ces deux départements sont regroupés en une seule zone de calage. On a donc environ 2 000 répondants par zone, mais seulement 500 en Guadeloupe et Martinique, et il ya bien (un peu moins de) 2000 dans les départements des Hauts-de-Seine et Rhône-Alpes-Auvergne sans Rhône. Les tranches sont décennales, dans toutes les zones.

Ratio median par zeat (ratio poids_final / poids redressé)

zone	Ratio	Rappel intitulé zone	Effectif
GUA	2,44	Guadeloupe	497
MAR	2,48	Martinique	525
REGY	1,05	Réunion et Guyane	964
NRD	1,02	Nord	2 578
PDC	1,01	Pas-de-Calais	2 656
HDS	0,98	Hauts-de-Seine (« 92 »)	1 892
RHO	1,25	Rhône (« 69 »)	2 210
2	0,89	Bassin parisien	3 845
4	0,94	Est	2 008
5	1,15	Ouest	2 706
7	1,06	Sud-ouest	2 161
9	1,02	Bassin méditerranéen	2 525
IDFR	0,98	Île-de-France sans « 92 »	2 295
RASR	1,17	Rhône-Alpes Auvergne sans « 69 »	1 561

Dispersion des poids pour ce calage (28 423 répondants (hors extensions DOM))

Poids	Redressé	Calé	
	(a)	(b)	(b)/(a)
Max	16 632	19 877	1,20
P99	9 220	10 152	1,10
P95	7 007	7 397	1,06
P90	5 882	6 046	1,03
P75	3 086	3 141	1,02
P50	1 098	1169	1,06
P25	383	380	0,99
P10	142	155	1,09
P5	103	114	1,11
P1	55	79	1,44
Min	25	25	1,00
D9/D1	41	39	
Q3/Q1	8	8	

Et dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,545	0,666	0,782	0,834	0,945	1,063	1,249	1,772	2,418	3,197	5,812	2,124	1,322

IV-B-7 : Quelques résultats :

Une question sur l'état de santé en général est posée quasiment à l'identique dans l'enquête filtre et dans l'enquête elle-même (etat_sant et bsante).

Le premier tableau indique les résultats d'abord sur tous les répondants à l'enquête filtre (VQS-final) puis les résultats sur les répondants à l'enquête (HSM) (on rappelle que l'on a conservé leurs réponses à l'enquête filtre). Soit, en terme de comparaison avec les résultats de VQS :

	VQS	HS, volet ménage		HS, volet ménage, calage no			
poids	final	initial	redressé	1	2	3	4
Question VQS : etat_sant. Comment est votre état de santé en général ?							
1. Très bon	40,90	40,24	39,83	41,40	41,39	41,24	41,39
2. Bon	39,84	40,14	40,37	40,09	40,15	40,13	40,01
3. Assez bon	15,32	15,54	15,63	14,63	14,64	14,72	14,68
4. Mauvais	3,21	3,31	3,41	3,17	3,19	3,19	3,19
5. Très mauvais	0,73	0,77	0,74	0,70	0,71	0,71	0,71

Le second tableau montre seulement la robustesse (répondants à la question posée dans l'enquête). Toutefois on constate également une bonne cohérence avec les résultats donnés par l'enquête filtre. Soit, en terme de robustesse :

	HS, volet ménage		HS, volet ménage, calage no			
poids	initial	redressé	1	2	3	4
BSANTE. Comment est votre état de santé en général ?						
1. Très bon	37,67	37,76	39,32	39,18	39,16	39,15
2. Bon	36,26	36,07	35,92	35,95	35,95	35,93
3. Assez bon	18,11	17,89	17,07	17,12	17,15	17,15
4. Mauvais	6,75	6,98	6,50	6,56	6,54	6,57
5. Très mauvais	1,13	1,21	1,12	1,11	1,12	1,12

De façon plus générale, on a testé les questions suffisamment proches dans l'enquête filtre et dans l'enquête elle-même de la même manière : comparaison et robustesse, pour des résultats plutôt satisfaisants.

On a également vérifié sur certaines variables que les résultats obtenus sont conformes à ce qu'on obtient par d'autres sources lorsque c'était possible. Ces vérifications sont rassurantes en général (voir partie suivante pour un cas particulier).

En conclusion, on retient le principe du traitement 4 : découpage des répondants suivant un zonage géographique comprenant les départements à extension et grands regroupements de régions. Un découpage par région aurait été trop fin au regard du nombre de répondants.

IV-C : Pondération finale : les traitements des extensions en Guadeloupe et Martinique

Les marges sont identiques pour les deux échantillons. Le CRIEM a alors effectué le partage des poids entre les deux échantillons pour produire les poids des répondants dans ces départements. Ceci n'affecte pas les poids des autres répondants.

Partie V : Autres essais de calage.

V-A : Calage sur CS

A partir de la méthode retenue (calage 4), on a intégré les contraintes socioprofessionnelles²⁶.

Données d'effectifs parmi les répondants suivant les CS

CS en 8		CS en 24		CS en 42		Suite...	
1	244	10	244	11	85		
2	528	21	251	12	32	62	553
3	1 195	22	243	13	127	63	628
4	2 018	23	34	21	251	64	239
5	3165	31	118	22	243	65	180
6	2 597	32	459	23	34	67	578
7	9 661	36	618	31	118	68	305
8	9 015	41	844	33	140	69	114
		46	588	34	240	71	757
		47	392	35	79	72	800
		48	194	37	326	74	896
		51	1 164	38	292	75	1448
		54	606	42	260	77	2 852
		55	468	43	419	78	2 908
		56	927	44	5	81	204
		61	1 600	45	160	84	4 747
		66	883	46	588	85	3 090
		69	114	47	392	86	974
		71	757	48	194	CS en 42	
		72	800	52	996		
		73	2 344	53	168		
		76	5 760	54	606		
		81	204	55	468		
		82	8 811	56	927		

Au vu des effectifs, différents essais sont menés. Les principes généraux constatés sont :

- 1) Une complication des méthodes qui deviennent très «cas par cas», car les calages ne réussissent pas ou seulement au prix d'un ajout de contraintes (méthode de calage par logit qui permet de borner a priori la dispersion du ratio poids calés sur poids non calés). De même, le regroupement de CS est ad hoc, CS en 8 modalités puis regroupement des «agriculteurs» et «chefs d'entreprise» pour permettre le calage ;
- 2) Malgré les précautions prises, la dispersion supplémentaire induite est très forte ;
- 3) Les tests ne montrent cependant pas de substantielles modifications des résultats (cohérence des différents calages, contrôle avec d'autres données).

Rien ne permet donc d'affirmer que de tels calages apportent une amélioration. Par contre, la complexité des calculs et la dispersion sont nettement augmentées.

Synthèse du traitement, résultats

L'essai le plus abouti est le suivant : pour chaque zone, on maintient la distinction hommes-femmes et on introduit les découpages par tranche d'âge, soit de 10 soit de 20 ans. On effectue également un calage sur marges de CS seulement dans les zones métropolitaines. Enfin, comme les dispersions de poids sont très grandes, la méthode de contrôle de dispersion évoquée (logit, l'étendue de dispersion a priori est donnée dans le tableau) est employée dans certaines zones géographiques. De nombreux essais ont été faits en faisant varier les bornes (arbitrages entre déformation des poids et

²⁶ Avec les données de recensement pour définir les marges.

dispersion, essentiellement). On présente seulement l'exemple finalement retenu. Cependant un calage alternatif avec pour la zone d'Île-de-France, hors Hauts-de-Seine, des bornes un peu plus rapprochées, ne peut se faire.

Les résultats sont alors (calculs effectués pour les 28 423 répondants) :

Tableau des méthodes employées par zone :

	Tranche d'âge	CS	
Guadeloupe	20 ans H/F	non	Raking ratio
Martinique	20 ans H/F	non	Raking ratio
Autres dom	20 ans H/F	non	Raking ratio
Nord	10 ans H/F	oui	Logit (0,35 - 6)
Pas de Calais	10 ans H/F	oui	Logit (0,35 - 6)
Hauts de Seine	10 ans H/F	oui	Logit (0,35 - 6)
Rhône	10 ans H/F	oui	Logit (0,35 - 6)
Bassin parisien	10 ans H/F	oui	Raking ratio
Est	10 ans H/F	oui	Raking ratio
Ouest	10 ans H/F	oui	Raking ratio
Sud-ouest	10 ans H/F	oui	Raking ratio
Bassin méditerranéen	10 ans H/F	oui	Raking ratio
Île-de-France sans 92	10 ans H/F	oui	Logit (0,25 - 7)
Rhône Alpe Auvergne sans 69	10 ans H/F	oui	Logit (0,35 - 6)

Répartition des poids

Poids	Redressé	Calé	
	(a)	(b)	(b)/(a)
Max	16 632	65 082	3,91
P99	9 220	10 805	1,17
P95	7 007	7 467	1,07
P90	5 882	6 035	1,03
P75	3 086	3 087	1,00
P50	1 098	1 121	1,02
P25	383	370	0,97
P10	142	146	1,03
P5	103	97	0,94
P1	55	49	0,89
Min	25	12	0,48
D9/D1	41	41	
Q3/Q1	8	8	

Et dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,250	0,348	0,497	0,662	0,834	1,027	1,207	1,515	2,110	3,771	7,000	2,290	1,447

Poids très faibles : (inférieurs à 30) : individus en mauvaise santé, poids de tirage faible (département à extension), plutôt autour de 50-60 ans : 14 personnes.

Poids très forts : (supérieurs à 20 000) : individus en bonne santé, poids de tirage assez fort, tous en Île-de-France sauf 92, plutôt autour de 50-60 ans : 17 personnes. Les personnes dont le ratio de calage / redressement est très élevé (=7) ont entre 50 et 60 ans. Il n'y a que 4 poids au-dessus de 30 000.

Quelques vérifications

En terme de comparaison avec les résultats de VQS :

	VQS	HS, volet ménage			HS, volet ménage, calage no			
poids	final	initial	redressé	2	3	4	5	
Question VQS : etat_sant. Comment est votre état de santé en général ?								
1. Très bon	40,90	40,24	39,83	41,39	41,24	41,39	41,30	
2. Bon	39,84	40,14	40,37	40,15	40,13	40,01	39,60	
3. Assez bon	15,32	15,54	15,63	14,64	14,72	14,68	14,90	
4. Mauvais	3,21	3,31	3,41	3,19	3,19	3,19	3,36	
5. Très mauvais	0,73	0,77	0,74	0,71	0,71	0,71	0,76	

En terme de robustesse :

	HS, volet ménage			HS, volet ménage, calage no			
poids	initial	redressé	2	3	4	5	
BSANTE. Comment est votre état de santé en général ?							
1. Très bon	37,67	37,76	39,18	39,16	39,15	39,17	
2. Bon	36,26	36,07	35,95	35,95	35,93	35,60	
3. Assez bon	18,11	17,89	17,12	17,15	17,15	17,19	
4. Mauvais	6,75	6,98	6,56	6,54	6,57	6,76	
5. Très mauvais	1,13	1,21	1,11	1,12	1,12	1,20	

V-B : Autres calages

Calage alternatif avec CS : en regroupant les cs 7 et 8, on réduit les poids très importants : le maximum revient à 25 800.

Tests et essais de calages avec données administratives de bénéficiaires d'allocations.

Les données concernent les différentes allocations et ont été fournies par la DREES, par exploitation d'enquêtes (enquête Aide Sociale) ou fichiers administratifs. Les comparaisons sont faites avec des données de 2007 et 2008.

Les traitements paraissent satisfaisants si on se réfère aux bénéficiaires de l'AAH (Allocation d'adultes Handicapés) : 825 000 selon l'enquête contre 830 000 en données DREES. Il en va de même pour l'AEH (Allocation d'éducation de l'enfant handicapé) : 190 000 contre 160 000, et pour l'Allocation compensatrice pour tierce personne : 95 000 contre 93 000. Par contre, cette cohérence manque pour l'Allocation personnalisée d'autonomie (APA) et la prestation de compensation du Handicap (PCH). Pour la première l'enquête donne une estimation à 350 000, soit la moitié seulement des données exogènes. Pour la PCH, les effectifs sont faibles et très variables une année sur l'autre (de l'ordre de 20 000 à 40 000).

Il a donc été effectué un calage sur une estimation du nombre de bénéficiaires de l'APA. Le taux a été fixé à 1% . En terme de dispersion des poids et du ratio des poids avant et après calage, cela paraît satisfaisant, mais malheureusement, les résultats sur les questions tests sont radicalement changés. Par exemple on a 49 % de personnes actives suivant les pondérations présentées plus haut et 33 % de retraitées. Les pondérations avec APA donnent respectivement 33 % et 35 %.

Répartition des poids

Poids	Redressé	Calé	
	(a)	(b)	(b)/(a)
Max	16 632	17 816	1,07
P99	9 220	10 462	1,13
P95	7 007	7 537	1,08
P90	5 882	6 121	1,04
P75	3 086	3 096	1,00
P50	1 098	1 161	1,06
P25	383	371	0,97
P10	142	153	1,08
P5	103	112	1,09
P1	55	75	1,36
Min	25	20	0,80
D9/D1	41	40	
Q3/Q1	8	8	

Et dispersion du ratio poids calé sur poids redressé :

Quantile	0	1	5	10	25	50	75	90	95	99	100	D9/D1	Q3/Q1
poids	0,218	0,408	0,684	0,811	0,920	1,052	1,171	1,336	1,500	2,715	3,720	1,647	1,273

(Champ : 28 423 répondants (hors extensions DOM)).

Partie VI : Synthèse sur la dispersion des poids

VI-A : Introduction

La phase de filtre a déjà produit une dispersion significative, soit, concernant les poids « VQS » un rapport inter décile de 6 et un rapport interquartile de 4. Rappelons (voir [partie II-D](#)) que cette dispersion doit être attribuée pour l'essentiel au mode de tirage²⁷. La constitution des groupes VQS et le procédé de tirage de l'échantillon pour le volet ménage induisent une dispersion supplémentaire très importante, soit un rapport inter décile de 13,5 et un rapport interquartile de 10. Les deux dispersions sont heureusement assez indépendantes, la combinaison des deux phases produit un rapport inter décile de 47 et un rapport interquartile de 8.

Les traitements de la non-réponse diminuent un peu la dispersion. Plus précisément, les queues de distributions s'étirent : le ratio poids maximal sur poids minimal augmente, avec quelques individus cumulant les facteurs de non-réponse. Cependant ces queues de distributions s'affinent²⁸ : les comportements exceptionnels sont de plus en plus rares. On distingue ici les traitements utilisant l'enquête filtre des traitements utilisant les informations plus usuelles (sexe, âge, lieu d'habitation, ...) : ce sont davantage les effets de santé (identifiables grâce à l'enquête filtre) qui influent. Cette observation est confirmée par les relevés lors des bilans²⁹ : les enquêtés ont d'autant plus tendance à répondre qu'ils sont concernés par les thèmes de l'enquête, mais leur état de santé est au contraire une limitation. Au total, les limitations l'emportent³⁰, donc les personnes avec faible poids initial du fait du tirage sont davantage repondérées du fait de la non-réponse.

Au final, après traitements pour non-réponse, on arrive à un rapport inter décile de 42 et un rapport interquartile de 8. Le calage change assez peu la dispersion, soit au final un rapport inter décile de 40 et un rapport interquartile de 9. Toutefois, les queues de distributions s'étirent encore. Il a été envisagé de contrebalancer cet effet de façon classique : redressements par classe plutôt que par individu. Toutefois, tant au niveau des redressements pour non-réponse que pour les calages, il a été préféré de suivre des méthodes assez uniformes pour ne pas rajouter à la complexité du traitement.

VI-B : Des poids de tirage de l'enquête filtre aux poids de tirage de l'enquête Handicap-Santé

La dispersion des poids « de tirage » de l'enquête Handicap-Santé est donc assez grande, le ratio extrême pour l'enquête HSM est de 640, à comparer à 84, la valeur correspondante pour l'enquête filtre VQS. C'est bien évidemment le tirage avec probabilités inégales suivant la sévérité présumée de la situation de handicap qui explique cette augmentation. Comme on calcule cette dispersion sur les échantillonnés, soit 37 074 individus, la dispersion des poids de l'enquête filtre est légèrement différente de celle de ces mêmes poids sur la base de sondage. Par nature du processus de tirage, une dispersion importante va résulter du mode de tirage de l'échantillon de l'enquête parmi les répondants à l'enquête filtre ([partie II](#)).

²⁷ Ce qui renvoie, une fois de plus au document référencé [note 5](#).

²⁸ Ou, pour voir les choses autrement, les comportements exceptionnels de non-réponse sont plutôt le fait de personnes dont les poids de tirage ne sont pas exceptionnels.

²⁹ Il s'agit des rencontres physiques entre concepteurs et enquêteurs, qui apportent une information précieuse car qualitative, donc différente des informations tirées de l'examen des tables de données.

³⁰ Ces limitations vont de la simple incapacité à répondre à un questionnaire long à l'absence de l'enquêté, qu'il soit parti en institution ou même décédé.

La dispersion des différents poids :

Premier tableau (pour résumer la partie II, l'échantillon) : les pondérations initiales, du tirage pour l'enquête filtre au tirage de l'enquête proprement dite :

37 074 obs	Poids tirage VQS	Poids redressé VQS	Poids tirage HSM	Poids initial
Max	1 064	1 409	33,70	10 240
P99	797	936	22,47	6 073
P95	365	466	13,51	5 338
P90	321	422	13,51	4 765
P75	295	367	13,39	2 256
P50	214	264	5,05	846
P25	69,6	88,5	1,29	275
P10	53,3	67,4	1	102
P5	28,1	38,0	1	72
P1	20,9	28,1	1	36
Min	12,6	16,0	1	16
D9/D1	6	6	13,5	47
Q3/Q1	4	4	10	8

Second tableau : la dispersion induite par les différents modèles de **non-réponse** (résumé de la partie III) :

Modèle S : régression sur variables de type « santé » ;

Modèle S+D : régression sur variables de type « santé » et de type « démographiques » ;

Modèle S+D+H : régression sur variables de type « santé », de type « démographiques » et de type « emplacement / statut d'occupation du logement ».

Sur 37 074 échantillonnés pour les poids initiaux, 28 423 répondants pour les poids redressés.

	Poids initial	S	S+D	S+D+H	Final
Max	10 240	12 402	13 936	16 130	16 632
P99	6 073	7 873	8 041	9 229	9 220
P95	5 338	6 881	6 876	6 998	7 007
P90	4 765	6 146	6 110	5 885	5 882
P75	2 256	2 960	2 994	3 086	3 086
P50	846	1 131	1 126	1 103	1 098
P25	275	383	378	365	383
P10	102	145	143	142	142
P5	72	97	97	102	103
P1	36	55	54	53	55
Min	16	23	22	25	25
D9/D1	47	42	43	41	41
Q3/Q1	8	8	8	8	8

En italique, les dispersions pour modèles partiels de redressement.

Troisième tableau : la dispersion induite par le **calage** (sur les 28 423 répondants).

	Poids initial	Poids redressés de la non-réponse	<i>Calage France entière</i>	<i>Calage sur zones définies par extensions</i>	<i>Calage sur « zeat »</i>	Calage final (calage mixte)
Max	10 240	16 632	20 210	19 877	18 802	17 388
P99	6 073	9 220	10 295	10 152	10 387	10 290
P95	5 338	7 007	7 515	7 397	7 437	7 450
P90	4 765	5 882	6 133	6 046	6 045	6 091
P75	2 256	3 086	3 172	3 141	3 220	3 156
P50	846	1 098	1 104	1 169	1 127	1 169
P25	275	383	359	380	373	377
P10	102	142	142	155	148	154
P5	72	103	103	114	109	113
P1	36	55	56	79	69	77
Min	16	25	25	25	29	24
D9/D1	47	41	43	39	41	40
Q3/Q1	8	8	9	8	9	8

En italique, les alternatives de calage (partie IV, hors calages non retenus).

Annexe 1 : détails des ajustements, calibrage de l'échantillon

Cette annexe présente des éléments détaillant la partie II-E (ajustements de l'échantillon).

Rappel : Deuxième estimation des échantillonnés pour l'enquête HSM :

Effectifs HSM (2)	0,074	0,2	0,2	0,49	1		<i>Rappel :</i>
Par groupes...	I-a	I-b	II	III	IV	Total	Cible initiale
Guadeloupe (national)	242		174	140	199	755	660
Guadeloupe (extension)	359		124	103	144	730	1 000
Martinique (national)	249		186	148	205	788	660
Martinique (extension)	295		153	111	165	724	1 000
Guyane	280		175	143	160	758	660
Réunion	319		239	176	255	990	660
Nord	1 107		693	665	812	3 278	3 200
Pas-de-Calais	931		634	683	784	3 031	3 200
Hauts-de-Seine	1 047		531	422	434	2 435	3 000
Rhône	852		508	418	469	2 247	3 000
Autres départements	7 485	928	5 175	4 647	4 986	23 221	22 000
Total	13 168	928	8 591	7 657	8 613	38 957	39 040

Il s'agit de comparer la cible initiale, soit la décomposition de l'échantillon ciblé par zone géographique, avec le nombre de répondants par groupe et par zones (les taux de sondages cibles par groupes étant fixés).

La troisième estimation résulte des ajustements nécessaires pour obtenir des échantillons par zone, les effectifs des groupes I et IV sont aussi peu modifiés que possible, l'arbitrage étant fait sur les groupes II et III :

Troisième estimation des échantillonnés pour l'enquête HSM, après corrections :

Effectifs HSM (3)							
Par groupes...	I-a	I-b	II	III	IV	Total	
Guadeloupe (national)	242		87	132	199	660	
Guadeloupe (extension)	386		260	210	144	1 000	
Martinique (national)	250		139	91	180	660	
Martinique (extension)	318		290	227	165	1 000	
Guyane	265		119	116	160	660	
Réunion	270		100	100	190	660	
Nord	1 110		678	600	812	3 200	
Pas-de-Calais	930		683	803	784	3 200	
Hauts-de-Seine	600		1 104	862	434	3 000	
Rhône	852		826	853	469	3 000	
Autres départements	7 400	900	4 650	4 064	4 986	22 000	
Total	12 623	900	8 936	8 058	8 523	39 040	

Différences (corrections des effectifs entre les estimations 2 et 3).

(3)-(2)	Effectifs à échantillonner pour HSM					
	I-a	I-b	II	III	IV	Total
Par groupes...						
Guadeloupe (national)	0	0	-87	-8	0	-95
Guadeloupe (extension)	27	0	136	107	0	270
Martinique (national)	1	0	-47	-57	-25	-128
Martinique (extension)	23	0	137	116	0	276
Guyane	-15	0	-56	-27	0	-98
Réunion	-49	0	-139	-76	-65	-330
Nord	3	0	-15	-65	0	-78
Pas-de-Calais	-1	0	49	120	0	169
Hauts-de-Seine	-447	0	573	440	0	565
Rhône	0	0	318	435	0	753
Autres départements	-85	-28	-525	-583	0	-1 221
Total	-545	-28	345	401	-90	83

On précise ces corrections pour deux exemples « extrêmes ». Dans le département du Nord, l'écart entre effectif attendu (3 278) et effectif à échantillonner (3 200) est relativement faible. Le principe décrit peut s'appliquer, soit :

Nord						
Par groupes...	I-a	I-b	II	III	IV	Total
Effectifs VQS	14 966		3 467	1 358	812	20 603
Effectifs HSM (2)	1 107		693	665	812	3 278
Effectifs HSM (3)	1 110		678	600	812	3 200

Réunion						
Par groupes...	I-a	I-b	II	III	IV	Total
Effectifs VQS	4 315		1 195	360	255	6 125
Effectifs HSM (2)	319		239	176	255	990
Effectifs HSM (3)	270		100	100	190	660
Ajustement	-49		-139	-76	-65	-330

La somme des effectifs de groupes³¹ I et IV est de 574 (soit 319+255), pour 660 personnes à échantillonner, ce qui signifie que les 86 personnes à échantillonner parmi les 1 555 répondants de groupes II et III ne peuvent l'être qu'avec un taux de sondage plus petit³² que celui du groupe I. On diminue donc les effectifs des groupes I et IV, ce qui permet de fixer les effectifs des autres groupes. On obtient donc peut obtenir des taux réels assez différents des taux cibles, notamment si l'effectif de la strate géographique est relativement faible.

³¹ « de strates constituées de personnes de groupe X », pour être précis.

³² Ce qui est doublement gênant : cela signifie une augmentation de la dispersion et conduit à tirer avec probabilité moindre des personnes dont on présume pourtant qu'elles ont une plus mauvaise situation de handicap (que les personnes de groupe I).

D'où les taux après le premier ajustement par zone et groupes :

ratio ajusté : effectif répondant VQS / effectif échantillon HSM après 1er ajustement						
Par groupes...	I-a	I-b	II	III	IV	rappel effectif
poids cible	13,51	5,00	5,00	2,04	1,00	
ajusté						
Guadeloupe (national)	13,53		9,98	2,16	1,00	662
Guadeloupe (extension)	12,58		2,38	1,00	1,00	998
Martinique (national)	13,46		6,68	3,33	1,14	666
Martinique (extension)	12,52		2,63	1,00	1,00	993
Guyane	14,29		7,37	2,51	1,00	643
Réunion	15,98		11,95	3,60	1,34	666
Nord	13,48		5,11	2,26	1,00	3 199
Pas-de-Calais	13,52		4,64	1,74	1,00	3 203
Hauts-de-Seine	23,59		2,40	1,00	1,00	3 000
Rhône	13,51		3,08	1,00	1,00	2 965
Autres départements	13,67	5,15	5,56	2,33	1,00	22 070
ratio par groupe	14,10	5,15	4,81	1,94	1,01	39 065

Le ratio par groupe donne une valeur moyenne des poids par zone.

Deuxième ajustement (pour qu'il n'y ait pas plus de deux personnes interrogées par ménage) :

Il s'agit d'un processus de retraitage-élimination. Le détail du tirage est en annexe 2. Bien qu'il s'agisse d'une approximation, on considérera que ce processus se résume ainsi :

Tirage de davantage de personne par strates (soit au total un supplément de 180 personnes de groupe I-a, 9 de groupe I-b, 135 de groupe II, 105 de groupe III et 16 de groupe IV, ces dernières forcément dans les départements où la première correction avait conduit à ne pas tirer exhaustivement dans ce groupe).

Poids supplémentaire :

Il faut multiplier par un poids « de second tirage » valant :

1 si la personne est « tirée » initialement et n'appartient pas à un ménage ou plus de 2 personnes ont été tirées initialement (cas très majoritaire) ;

0 si la personne est tirée au cours du processus de tirage initial ou au cours du processus de retraitage-élimination, mais « éliminée » ;

« N / 2 » si la personne est tirée initialement ou au cours du processus de retraitage-élimination, dans un ménage où N personnes ont été tirées³³ et N-2 éliminées.

Cela modifie encore les taux (en fait les poids, qui sont légèrement diminués). Par exemple, l'effectif de la strate « Nord-Groupe II » est attendu à 678. L'effectif VQS correspondant est de 3467, d'où un poids « ante » de 3467 / 678, soit 5,11. Il faut retirer 9 personnes. Ainsi un total de 678+9=687 personnes ont été tirées, d'où le poids « post » de 3467 / 687 soit 5,05.

On peut remarquer ici que ce processus conduit à interroger dans des ménages des personnes « en bonne santé » en remplacement de personnes de groupes IV ou III. En collecte, cette situation a été pointée comme difficile : les enquêtés adhèrent d'autant mieux à l'enquête qu'ils sont en mauvaise santé, au vu des thèmes abordés. Il leur paraît particulièrement malvenu d'interroger une personne en bonne santé alors qu'il y a dans ce même ménage une personne lourdement handicapée.

³³ Le flou pudique sur le tirage résulte du fait que l'on a déterminé N au moment du tirage initial et non comme résultant de tous les tirages de rattrapage.

On obtient alors les effectifs définitifs par zone et par groupe de l'échantillon HSM.

Effectifs HSM	(effectifs définitifs de l'échantillon)					
	I-a	I-b	II	III	IV	Total
Par groupes...						
Guadeloupe (national)	244		88	133	197	662
Guadeloupe (extension)	389		262	203	144	998
Martinique (national)	252		140	92	182	666
Martinique (extension)	321		292	222	158	993
Guyane	267		120	117	139	643
Réunion	272		101	101	192	666
Nord	1 119		684	605	791	3 199
Pas-de-Calais	938		689	810	766	3 203
Hauts-de-Seine	626		1 122	833	419	3 000
Rhône	859		833	818	455	2 965
Autres départements	7 463	908	4 690	4 099	4 910	22 070
Total	12 750	908	9 021	8 033	8 353	39 065

Le poids réel par groupe et zone s'en déduit :

Poids de tirage : effectif répondant VQS / effectif échantillon HSM						
Par groupes...	I-a	I-b	II	III	IV	rappel effectif
poids cible (inverse des taux)	13,51	5,00	5,00	2,04	1,00	
poids réel						
Guadeloupe (national)	13,42		9,86	2,14	1,01	662
Guadeloupe (extension)	12,48		2,37	1,03	1,00	998
Martinique (national)	13,36		6,64	3,29	1,13	666
Martinique (extension)	12,40		2,62	1,02	1,04	993
Guyane	14,19		7,31	2,49	1,15	643
Réunion	15,86		11,83	3,56	1,33	666
Nord	13,37		5,07	2,24	1,03	3 199
Pas-de-Calais	13,41		4,60	1,72	1,02	3 203
Hauts-de-Seine	22,61		2,37	1,03	1,04	3 000
Rhône	13,40		3,05	1,04	1,03	2 965
Autres départements	13,55	5,11	5,52	2,31	1,02	22 070
ratio par groupe	13,96	5,11	4,76	1,95	1,03	39 065

Le ratio par groupe donne une valeur moyenne des poids par zone.

Annexe 2 : Les grandes étapes du programme de tirage de l'échantillon de l'enquête Handicap-Santé, volet « ménages »

Notation : on résume par **contrainte** la contrainte de ne pas interroger plus de deux personnes dans un même ménage.

Première étape (tirage³⁴ sans **contrainte**) :

On tire dans chaque strate un effectif égal à la cible.
On a donc un poids de tirage (égal au poids cible). D'où une table (echhs1).

Deuxième étape : correction suivant la **contrainte**

On va « éliminer » des individus tirés en première étape pour respecter la **contrainte**.
On constitue donc une table d'individus à partir de echhs1, ces individus étant au moins 3 par ménages, on tire 2 individus par ménages, d'où une nouvelle table (ssechhs1_p2) et un poids qui ne peut prendre que des valeurs demi-entières.

Il se peut alors que l'on ait éliminé des individus de groupe 4, alors que ces individus sont dans des strates à taux cible égal à 1.

Troisième étape : rattrapage des individus perdus

On va retirer des individus pour remplacer ceux qui sont éliminés. On ne peut le faire dans les strates où les taux cibles sont de 1 (tous les individus ont été pris dans echhs1). On répartit le manque dans les autres strates, pour définir les effectifs (à rattraper) pour chaque strate (somme des individus à remplacer et de la partie idoine du manque). On reprend les répondants de VQS, sauf les individus tirés la première fois et dans un ménage où plus d'un individu a été tiré. On effectue le tirage, d'où une table (echhs1_ratrap).

On constitue alors la table (echhs2) des individus échantillonnés lors des trois étapes. Celle-ci peut encore comporter des ménages ayant plus de deux individus (si par exemple des individus rattrapés en troisième étape sont dans un ménage où il y a un échantillonné en première étape).

Quatrième étape : second tour d'élimination / rattrapage

On procède comme à l'étape 3 : après avoir défini les individus à remplacer (ils sont dans la table echhs1_ratrap et dans un ménage à plus de trois individus échantillonnés). Ils seront remplacés par un individu dans la même strate.

Cinquième étape : sorties et vérifications

³⁴ Usage de la proc surveyselect de sas, tirages à probabilités égales.

Annexe 3 : Statistiques descriptives pour la non-réponse

Données (statistiques descriptives sur la non-réponse au volet ménage de l'enquête Handicap-Santé. Échantillon : national + extensions en métropole (37074 échantillonnés, 28446 répondants).

Données descriptives suivant les codes de résultats d'enquête.

Les modalités sont regroupées, soit 8 modes, dont voici la répartition globale.

Ensemble	Effectifs	Répartition (%)	Code
Répondants	28 225	76,4	ERV
Échecs à différents moments du contact	1 863	5,0	Cont
Déménagements, ..., (mais hors institutions)	2 021	5,5	Demg
Entretiens partiels	473	1,3	Part
Refus directs ou indirects	2 583	7,0	Refu
Inaptes (entretien impossible)	786	2,1	IMP
Décédés	687	1,9	DCD
Partis en institutions	327	0,9	INS
Total	36 965	100	
(valeurs manquantes)	109		

Tableau 1 : la non-réponse suivant le genre.

Effectifs		H	F	ratios	F/H
Répondants	28 225	13 018	15 207		1,003
Echecs à différents moment contact	1 863	893	970		0,93
Déménagements, ..., hors institutions	2 021	974	1 047		0,92
Entretiens partiels	473	217	256		1,01
Refus directs ou indirects	2 583	1 185	1 398		1,01
Inaptes (entretien impossible)	786	321	465		1,24
Décédés	687	354	333		0,81
Partis en institutions	327	111	216		1,67
	36 965	17 073	19 892		

Lecture des ratios : la situation de référence est l'ensemble de la catégorie : les femmes ont 19 % de chances de moins que les hommes d'être décédées, mais 67 % de chances de plus d'être parties en institutions.

Tableau 2 : la non-réponse suivant le groupe VQS.

Effectifs		I	II	III	IV		III/I	III/II	IV/I
Répondants	28 225	9 938	6 607	5 790	5 890		1,02	0,99	0,95
Echecs à différents moment contact	1 863	749	416	358	340		0,85	0,81	0,73
Déménagements, ..., hors institutions	2 021	911	394	355	361		0,66	0,66	0,64
Entretiens partiels	473	78	98	120	177		1,92	2,62	3,66
Refus directs ou indirects	2 583	1 002	638	510	433		0,97	0,87	0,70
Inaptes (entretien impossible)	786	208	170	197	211		1,25	1,61	1,63
Décédés	687	20	88	177	402		6,73	15,07	32,39
Partis en institutions	327	10	36	79	202		5,50	13,45	32,55
	36 965	12 916	8 447	7 586	8 016				

Dans la suite, on calcule les ratios taux par mode / taux pour l'ensemble (il ne s'agit pas d'odds-ratios : il n'y a pas de situation de référence et les ratios ne sont pas ceteribus partibus).

Annexe 3-A : Les décès.

Décédés	Effectifs	%tage	ratio
Total	687	1,86	1
Par groupes			
I	20	0,15	0,1
II	88	1,04	0,6
III	177	2,33	1,3
IV	402	5,01	2,7
Par genre			
H	354	2,07	1,1
F	333	1,67	0,9
Par genre et âge			
Fem 50-60	22	0,64	0,3
Fem 60-70	23	1,08	0,6
Fem 70- 80	72	2,64	1,4
Fem 80 et +	197	8,24	4,4
Hom 50-60	38	1,23	0,7
Hom 60-70	45	2,48	1,3
Hom70- 80	88	4,62	2,5
Hom 80 et +	158	13,63	7,3
Par régions			
Île-de France	107	1,70	0,9
Nord et Pas-de-Calais	130	2,03	1,1
Rhône-Alpes	66	1,50	0,8
PACA	54	2,44	1,3
Suivant santé (VQS)			
TB	9	0,11	0,1
B	36	0,33	0,2
AB	207	1,87	1,0
M	268	5,40	2,9
TM	162	10,79	5,8

Note de lecture : 402 des 687 non-répondants pour cause de décès étaient classés en groupe IV. Ce nombre est 2,7 fois plus élevé que si la part des décédés classés en groupe IV avait été la même que la part des décédés dans l'ensemble des échantillonnés

Pour se donner un ordre de grandeur, on a calculé les décès attendus par rapport aux effectifs, en utilisant les taux de décès (par tranche de 5 ans, données démographiques), à comparer aux décédés répertoriés au moment de l'enquête. Il ne s'agit évidemment que d'ordre de grandeurs (l'écart entre VQS et HSM n'est pas d'un an et la variable DCD n'est pas un état civil).

Ainsi, le tableau suivant se lit : il y a 9 140 hommes âgés de moins de 50 ans et échantillonnés. Leur répartition par âge permettait de prévoir que 12 d'entre eux seraient décédés au moment de la collecte. Or on a constaté que 25 d'entre eux étaient décédés. Si l'on précise par groupe VQS, pour le groupe I, la prédiction est de 6, pour 3 décès constatés, tandis que les chiffres correspondant pour le groupe IV sont de 2 décès prédits pour 17 constatés. Le constat est donc rassurant, la classification faite à partir de l'enquête filtre relie de la façon attendue la sévérité de la situation (présumée) de handicap et la fréquence des décès. On remarque enfin que ce lien semble moins fort pour les femmes que pour les hommes (de même que l'âge). Finalement, la surreprésentation de personnes en « mauvaise santé » conduit bien à constater davantage de décès que la seule structure par sexe et âge permettait de prévoir.

Tableau comparatif des décès attendus / constatés :

Hommes		Ensemble			Groupe I		Groupe IV	
		Effectif	Attendu	HSM	Attendu	HSM	Attendu	HSM
moins	de 50	9 140	12	25	6	3	2	17
50	à 70	4 906	49	83	13	4	12	43
+ de	70	3 069	185	246	15	3	55	141
Femmes		Effectif	Attendu	HSM	Attendu	HSM	Attendu	HSM
		moins	de 50	9 269	6	19	3	2
50	à 70	5 557	24	45	6	5	5	23
+ de	70	5 133	237	269	11	3	84	166

Annexe 3-B : Les départs en institutions.

Institutions	Effectifs	%tage	ratio
Total	327	0,88	1
Par groupes			
I	10	0,08	0,1
II	36	0,43	0,5
III	79	1,04	1,2
IV	202	2,52	2,8
Par genre			
H	111	0,65	0,7
F	216	1,09	1,2
Par genre et âge			
Fem 50-60			
Fem 60-70			
Fem 70- 80	43	1,58	1,8
Fem 80 et +	146	6,10	6,9
Hom 50-60			
Hom 60-70			
Hom 70- 80	29	1,52	1,7
Hom 80 et +	47	4,06	4,6
Par régions			
Île-de France	51	0,81	0,9
Nord et Pas-de-Calais	36	0,56	0,6
Rhône-Alpes	26	0,59	0,7
PACA	23	1,04	1,2
Suivant santé (VQS)			
TB			
B			
AB	111	1,00	1,1
M	125	2,52	2,8
TM	61	4,06	4,6

A noter les relations inverses (comparativement aux décès) entre sexe, âge et départ en institutions : au-delà de 80 ans, sur 4 personnes parties en institutions, 3 sont des femmes, et cela semble résulter à la fois de leur plus grande espérance de vie (elles sont plus nombreuses à avoir dépassé 80 ans) que de leur propension à aller en institution.

Annexe 3-C : Les refus.

refus			
	Effectifs	%tage	ratio
Total	2583	6,99	1
Par groupes			
I	1002	7,76	1,1
II	638	7,55	1,1
III	510	6,72	1,0
IV	433	5,40	0,8
Par genre			
H	1185	6,94	1,0
F	1398	7,03	1,0
Par genre et âge			
Fem 50-60	79	5,58	0,8
Fem 60-70	97	5,58	0,8
Fem 70- 80	98	6,82	1,0
Fem 80 et +	123	6,41	0,9
Hom 50-60	227	8,36	1,2
Hom 60-70	240	7,02	1,0
Hom70- 80	155	7,28	1,0
Hom 80 et +	211	7,74	1,1
Par régions	168	7,02	1,0
Île-de France	91	5,56	0,8
Nord et Pas-de-Calais	115	5,81	0,8
Rhône-Alpes	121	9,02	1,3
PACA	123	6,97	1,0
Suivant santé (VQS)	184	7,70	1,1
TB	238	7,71	1,1
B	108	5,96	0,9
AB	133	6,98	1,0
M	72	6,21	0,9
TM			
Total	665	10,55	1,5
Par groupes	311	4,86	0,7
I	366	8,32	1,2
II	190	8,58	1,2
III			
IV	603	7,30	1,0
Par genre	837	7,60	1,1
H	745	6,74	1,0
F	307	6,18	0,9
Par genre et âge	77	5,13	0,7

Annexe 3-D : Les déménagements (de logements ordinaires à logements ordinaires)³⁵.

déménagements			
	Effectifs	%tage	ratio
Total	2021	5,47	1
Par groupes			
I	911	7,05	1,3
II	394	4,66	0,9
III	355	4,68	0,9
IV	361	4,50	0,8
Par genre			
H	974	5,70	1,0
F	1047	5,26	1,0
Par genre et âge			
Fem 50-60	98	6,92	1,3
Fem 60-70	92	5,29	1,0
Fem 70- 80	194	13,51	2,5
Fem 80 et +	150	7,82	1,4
Hom 50-60	131	4,82	0,9
Hom 60-70	157	4,59	0,8
Hom70- 80	82	3,85	0,7
Hom 80 et +	78	2,86	0,5
Par régions	65	2,72	0,5
Île-de France	126	7,70	1,5
Nord et Pas-de-Calais	117	5,91	1,2
Rhône-Alpes	159	11,86	2,4
PACA	156	8,84	1,8
Suivant santé (VQS)	120	5,02	1,0
TB	141	4,57	0,9
B	63	3,48	0,7
AB	63	3,31	0,7
M	29	2,50	0,5
TM			
Total	488	7,74	1,4
Par groupes	272	4,25	0,8
I	282	6,41	1,2
II	139	6,28	1,1
III			
IV	617	7,47	1,4
Par genre	662	6,01	1,1
H	459	4,15	0,8
F	210	4,23	0,8
Par genre et âge	63	4,19	0,8

³⁵ Il n'était pas prévu de suivre les personnes ayant déménagé.

Annexe 4 : Effectifs des répondants par strate

Annexe 4-A : Les effectifs de répondants par strate.

Par strate (zone par groupes VQS), on a les effectifs de répondants suivants :

Effectifs de répondants	I	II	III	IV	Tous groupes
Guadeloupe (N)	179	68	111	139	497
Guadeloupe (X)	281	211	166	111	769
Martinique (N)	197	111	77	140	525
Martinique (X)	227	222	173	117	739
Guyane	177	83	77	84	421
Réunion	216	80	89	159	544
Nord	899	576	502	602	2 579
Pas-de-Calais	780	575	686	617	2 658
Hauts-de-Seine	396	761	492	248	1 897
Rhône	635	644	605	327	2 211
Autres	5 773	4 480	3 214	3 646	17 113
Toutes zones	9 760	7 811	6 192	6 190	29 953
Hors extensions dom	9 252	7 378	5 853	5 962	28 445

Annexe 4-B : Éléments sur l'échantillon : détails dans les départements métropolitains avec extensions.

Répondants à l'enquête filtre (base de sondage)

	Groupe	« national »	« extension »	Total
Nord	I	5009	9957	14966
	II	1113	2354	3467
	III	410	948	1358
	IV	258	554	812
Pas-de-Calais	I	2798	9779	12577
	II	653	2515	3168
	III	280	1114	1394
	IV	170	614	784
Hauts-de-Seine	I	2896	11259	14155
	II	512	2142	2654
	III	189	673	862
	IV	83	351	434
Rhône	I	4232	7280	11512
	II	950	1591	2541
	III	329	524	853
	IV	174	295	469
4 départements		20 056	51 950	72006

Échantillonnés :

	Groupe	« national »	« extension »	Total
Nord	I	373	746	1119
	II	219	465	684
	III	184	421	605
	IV	253	538	791
Pas-de-Calais	I	213	725	938
	II	137	552	689
	III	164	646	810
	IV	167	599	766
Hauts-de-Seine	I	126	500	626
	II	220	902	1122
	III	183	650	833
	IV	82	337	419
Rhône	I	309	551	860
	II	307	526	833
	III	316	501	817
	IV	167	288	455
4 départements		3420	8947	12367

Répondants :

	Groupe	« national »	« extension »	Total
Nord	I	298	601	899
	II	187	389	576
	III	151	351	502
	IV	194	408	602
Pas-de-Calais	I	179	601	780
	II	116	459	575
	III	138	548	686
	IV	138	479	617
Hauts-de-Seine	I	81	315	396
	II	142	619	761
	III	100	392	492
	IV	49	199	248
Rhône	I	228	407	635
	II	244	400	644
	III	241	364	605
	IV	126	201	327
4 départements		2 612	6 733	9 345

Annexe 5 : Tableau des odds-ratios (modèle de non-réponse)

Tableau des odds-ratios, modèle de non-réponse (à l'enquête HSM), version finale :

Etat matrimonial de la personne de référence du logement	Odds ratio	
Autres états matrimoniaux	0,8	
Mariés	ref	
Zone géographique (regroupements de départements)		
zone 1	0,3	
zone 2	0,5	
zone 3	0,6	
zone 4	0,7	
zone 5	ref	
Statut d'occupation		
Ménage propriétaire du logement occupé	1,2	
Autres statuts	ref	
Type de logement		
Maison	1,4	
Autres logements	ref	
Tranche d'âge, par sexe		alternative
femmes < 9 ans	2	1,3
femmes de 10 à 19 ans	2,1	1,3
femmes de 20 à 29 ans	1,2	0,7
femmes de 30 à 39 ans	1,7	1,1
femmes de 40 à 49 ans	1,7	1,1
femmes de 50 à 59 ans	1,8	1,2
femmes de 60 à 69 ans	2	1,2
femmes de 70 à 79 ans	1,8	1,1
femmes de plus de 80 ans	1,2	0,8
hommes < 9 ans	2,1	1,3
hommes de 10 à 19 ans	1,9	1,2
hommes de 20 à 29 ans	1,1	0,7
hommes de 30 à 39 ans	1,5	1
hommes de 40 à 49 ans	1,6	ref
hommes de 50 à 59 ans	1,7	1
hommes de 60 à 69 ans	1,7	1,1
hommes de 70 à 79 ans	1,5	1
hommes de plus de 80 ans	ref	0,6
Variables VQS		
Etat de santé		
Très bon, bon, moyen	1,3	
Mauvais ou très mauvais	ref	
Maladie chronique		
Non	0,8	
Oui	ref	
Concentration (avez vous des difficultés à ...)		
Autres modalités	1,2	
Oui, beaucoup	ref	
Initiative (avez vous des difficultés à ...)		
Autres modalités	1,2	
Oui, beaucoup	ref	
problèmes quotidiens (avez vous des difficultés à ...)		
Autres modalités	1,1	
Oui, beaucoup	ref	
Sorties (avez vous des difficultés à ...)		
Autres modalités	1,2	
Oui, beaucoup	ref	
Comprendre (avez vous des difficultés à ...)		
Autres modalités	1,1	
Oui, beaucoup	ref	
Recevoir l'aide d'une autre personne		
Autres modalités	1,1	
Oui, beaucoup	ref	
Reconnaissance officielle d'un handicap		
Non	0,8	
Oui	ref	

Annexe 6 : Synthèse des travaux du CRIEM

Calcul de la pondération de l'enquête HSM pour la Guadeloupe et la Martinique³⁶

Introduction

Les modalités de tirage, donc les travaux de redressements et de pondérations sont quelque peu différents pour les sous-échantillon en extension dans les départements de Guadeloupe et Martinique. Pour ces personnes, la base de sondage de l'enquête filtre n'est pas l'enquête annuelle de recensement 2006 mais les répondants à l'enquête emploi spécifique en département d'outre-mer (EEDOM). Cette enquête se déroule dans les quatre départements d'outre-mer (Martinique, Guadeloupe, Guyane et Réunion), au même moment et sur une période de 13 semaines, de mars à juin. Dans chaque ménage tiré au sort, toute personne âgée de 15 ans et plus est interrogée trois années de suite. La première interrogation est réalisée par visite, les deux suivantes par téléphone. Les personnes sont tirées dans le tiers « médian » et le tiers « sortant » pour le filtre VQS. Ce filtre ayant lieu en 2007, le tiers médian est donc constitué de personnes interrogées une première fois au titre de l'EEDOM en 2006 tandis que le tiers sortant concerne des personnes entrées dans le dispositif en 2005.

Étape 1 : Calcul du poids de tirage des réinterrogés EEDOM 2007 (enquête filtre des échantillons extension DOM = VQS adossé à l'EEDOM 2007 pour les réinterrogés)

- Chacun des 3 sous échantillons de l'EEDOM étant tiré de manière à être représentatif de la population, pour obtenir le poids de tirage des ménages réinterrogés de l'EEDOM 2007, on utilise le poids de tirage de l'échantillon EEDOM³⁷ divisé par deux pour calculer le poids de tirage des ménages réinterrogés WT_EEDOMmen.
- Étape de calage en utilisant les marges utilisées pour le calage de 2007 sur les variables logements et sexe-âge pour transformer le poids ménage WT_EEDOMmen en poids individu WT_EEDOM (le calage n'était pas utile à ce stade là mais les difficultés³⁸ pour passer du poids ménage au poids individus nous ont fait opter pour cette solution).

Étape 2 : Prise en compte de la non réponse VQS : Calcul du poids de tirage REDRESSE de l'enquête filtre des échantillons extension DOM

- Tous les répondants EEDOM2007 ne sont pas répondants VQS : 5829 individus répondants VQS (pour 6257 répondants EEDOM, soit environ 7% de non répondants) pour la Guadeloupe et 5140 (pour 5556, soit environ 7% de NR) pour la Martinique.
- Pour prendre en compte cette non-réponse de façon grossière, on fait une "règle de trois" (rapport entre la somme des poids des individus répondants EEDOM et celle des individus répondants VQS).

Étape 3 : Constitution de la table des échantillons extensions HSM et intégration des poids de tirage de l'échantillon HSM parmi les répondants à VQS-EEDOM 2007

- Sélection des individus de l'échantillon HSM faisant partie des extensions DOM. Intégration du poids de tirage de l'échantillon HSM parmi les répondants à VQS-EEDOM 2007.

³⁶ Seuls ces deux départements ont réalisé des extensions en outre-mer.

³⁷ Les poids EXTRL ont été recalculés pour intégrer des ménages qui avaient été supprimés des tables EEDOM 2007 par erreur (mauvaise gestion des éclatements de logement).

³⁸ La solution de multiplier le poids ménage par le ratio redressé (nombre d'individus du ménage / nombre d'individus du ménage interrogés pour HSM) ne donnait pas des résultats satisfaisants (somme des poids sur l'échantillon HSM bien supérieure à la population totale de Guadeloupe ou de Martinique) :

$WT_EEDOM = WT_EEDOMmen * (nbind / nbindHSM) * moy1 / moy2$, moy1 étant le nb d'ind moyen dans les ménages répondants réinterros EEDOM et moy 2 le nb d'ind moyen dans les ménages échantillonnés HSM.

Étape 4 : Intégration des poids de tirage redressé VQS-EEDOM dans les échantillons HSM et calcul des poids de tirage initiaux HSM

Calcul du poids de tirage initial extension HSM : $WT_{HS} = WR_{EEDOM} * W_{EEDOM_{HS}}$.

La dispersion des différents poids sur les échantillons extension HSM :

Guadeloupe :

982 obs	WT_EEDOM	WR_EEDOM	W_EEDOM_HS	WT_HS
	poids de tirage pour enquête filtre	poids redressé pour enquête filtre	poids de tirage de l'échantillon HSM parmi les répondants à VQS	poids initial HSM
Max	173	190	18,63	2286
P99	142	157	12,42	1858
P95	119	131	12,42	1433
P90	107	118	12,42	1234
P75	89	98	12,42	672
P50	62	68	2,36	158
P25	33	37	1,00	73
P10	13	14	1,00	31
P5	11	12	1,00	21
P1	11	12	1,00	12
Min	11	12	1,00	12
D9/D1	8,5	8,5	12,4	39,9
Q3/Q1	2,7	2,7	12,4	9,2

Martinique :

988 obs	WT_EEDOM	WR_EEDOM	W_EEDOM_HS	WT_HS
	poids de tirage pour enquête filtre	poids redressé pour enquête filtre	poids de tirage de l'échantillon HSM parmi les répondants à VQS	poids initial HSM
Max	179	199	18,43	3314
P99	162	180	12,29	2194
P95	160	178	12,29	1722
P90	141	157	12,29	1277
P75	110	122	12,29	392
P50	56	63	2,61	167
P25	27	30	1,00	77
P10	27	30	1,00	30
P5	27	30	1,00	30
P1	27	29	1,00	30
Min	27	29	1,00	29
D9/D1	5,3	5,3	12,3	42,6
Q3/Q1	4,1	4,1	12,3	5,1

Étape 5 : Redressement de la Non-réponse HSM des extensions Guadeloupe et Martinique

- Construction des variables nécessaires au traitement de la NR : variables VQS, variables démographiques, variables géographiques et autres ;
- Calcul du poids après redressement de la non réponse :
$$\text{poidsHS_red} = \text{WT_HS} / \text{proba de répondre à HSM.}$$

Étape 6 : Calage HSM des échantillons extensions Guadeloupe et Martinique

- Le calage est fait sur le même modèle que le calage national, à savoir :
 - o uniquement sur la variable démographique sexe * âge, avec des tranches décennales pour l'âge,
 - o en utilisant les marges fournies par la division études et enquêtes démographiques ;
- En sortie de l'étape 6, on a les tables calHSM.poidsHSM97Xx avec en observation les répondants extension HSM (768 individus dans poidsHSM971x, 736 dans poidsHSM972x) et en variable les mêmes que celles présentes précédemment + poidsHS_fin.

La dispersion des différents poids sur les répondants extension HSM :

Guadeloupe :

768 obs	WT_HS	poidsHS_red	poidsHS_fin
	poids initial HSM	poids redressé de la non réponse HSM	poids redressé de la non réponse et calé HSM
Max	2041	3682	4210
P99	1826	2750	2998
P95	1386	1919	1975
P90	1184	1509	1532
P75	619	766	757
P50	150	197	199
P25	72	89	89
P10	31	40	42
P5	21	27	23
P1	12	14	13
Min	12	13	9
D9/D1	38,0	37,4	36,5
Q3/Q1	8,7	8,6	8,5

Martinique³⁹ :

736 obs	WT_HS	poidsHS_red	poidsHS_fin
	poids initial HSM	poids redressé de la non réponse HSM	poids redressé de la non réponse et calé HSM
Max	2450	3688	4336
P99	2194	3294	3271
P95	1722	2387	2324
P90	1174	1583	1703
P75	369	530	622
P50	156	211	232
P25	77	96	91
P10	30	39	47
P5	30	35	32
P1	30	33	21
Min	29	32	19
D9/D1	39,2	40,6	36,2
Q3/Q1	4,8	5,5	6,8

Comparaison de la variable « État de santé » entre les répondants VQS et les répondants HSM en utilisant les jeux de pondérations finaux :

<u>Guadeloupe :</u>	<u>Martinique :</u>
1-très bon : 34% pour HSM vs 40% pour VQS	1-très bon : 26% pour HSM vs 32% pour VQS
2-bon : 44 vs 43	2-bon : 45 vs 44
3-assez bon : 18 vs 14	3-assez bon : 24 vs 20
4-mauvais : 3.5 vs 2.8	4-mauvais : 4.0 vs 3.3
5-très mauvais : 0.5 vs 0.5	5-très mauvais : 0.7 vs 0.7

Différences notables sur les modalités "très bon" et "assez bon".

Étape 7 : Partage des poids des échantillons extension et national Guadeloupe et Martinique

Récupération des échantillons national dans la table envoiGB.aurore_02fev pour récupérer la variable de poids après traitement de la non réponse (poids_hs_red) : 498 répondants 971, 525 répondants 972.

³⁹ Sur les répondants extension HSM Martinique, un individu a posé problème avec un poids final très important (poidsHS_fin = 6900). C'est un homme de 25 ans, en bonne santé, dont la probabilité de tirage dans l'EEDOM est importante (logement dans une petite adresse de petite commune, dans la zone « sud » de la Martinique). Il a été décidé de lui attribuer le poidsHS_red de son « plus proche voisin » (le voisin étant entendu comme un homme entre 20 et 30 ans), à savoir 914 à la place de 3130.

La dispersion du poids final sur les répondants national HSM :

Guadeloupe :

497 obs	poidsHS_ini	poidsHS_red	poidsHS_fin
	poids initial HSM	poids redressé de la non réponse HSM	poids redressé de la non réponse et calé HSM
Max	499	807	3474
P99	499	744	3153
P95	499	648	2115
P90	463	589	1790
P75	417	504	1343
P50	79	141	464
P25	38	56	150
P10	31	42	100
P5	31	39	77
P1	31	36	54
Min	31	35	44
D9/D1	14,8	14,2	17,9
Q3/Q1	11,1	8,9	9,0

Martinique :

525 obs	poidsHS_ini	poidsHS_red	poidsHS_fin
	poids initial HSM	poids redressé de la non réponse HSM	poids redressé de la non réponse et calé HSM
Max	758	963	2611
P99	505	760	2237
P95	505	703	1939
P90	505	658	1683
P75	447	544	1253
P50	221	271	593
P25	41	77	184
P10	36	47	108
P5	36	44	82
P1	34	41	63
Min	34	39	50
D9/D1	13,9	14,0	15,6
Q3/Q1	10,8	7,1	6,8

- Concaténation des échantillons extension et national ;
- Partage des poids selon le modèle suivant :

$$poidsHS_final = (nx * poidsHS_finx) / (nx + nn) + (nn * poidsHS_finn) / (nx + nn)$$
avec nx et nn le nombre d'individus répondants des échantillons extension et national respectivement, poidsHS_finx et poidsHS_finn le poids après calage des échantillons extension et national respectivement.

La dispersion du poids final sur les répondants HSM (extension + national) :

1265 obs	poidsHS_final	1261 obs	poidsHS_final
<u>Guadeloupe</u>	poids redressé de la non réponse et calé HSM après partage des poids	<u>Martinique</u>	poids redressé de la non réponse et calé HSM après partage des poids
Max	2556	Max	2531
P99	1749	P99	1598
P95	1084	P95	1095
P90	813	P90	774
P75	520	P75	434
P50	127	P50	181
P25	56	P25	62
P10	31	P10	34
P5	21	P5	25
P1	9	P1	15
Min	5	Min	11
D9/D1	26,1	D9/D1	23,1
Q3/Q1	9,3	Q3/Q1	7,0

Annexe 7 : Un élément d'appréciation de la pertinence des groupes VQS

La méthode de construction (score et classification) des « groupes VQS » revient à privilégier une approche statistique et monodimensionnelle du handicap, donc excessivement simplifiée par rapport aux objectifs de l'enquête⁴⁰. Le questionnaire de l'enquête filtre n'est pas repris dans l'enquête et il n'est donc pas possible de comparer le groupe VQS selon l'enquête filtre et ce même groupe estimé à partir des données de l'enquête elle-même. Toutefois, des comparaisons sur des indicateurs que l'on peut construire de façon identique ou très proche suivant les deux enquêtes confirment que globalement la correspondance est bonne, mais qu'il y a des divergences dès que l'on affine les analyses.

De façon assez générale, comme risque de la méthode suivie, se pose le problème des « faux négatifs », soit pour résumer, une personne « en bonne santé » selon le filtre et en mauvaise santé selon l'enquête. Un faux négatif est donc quelqu'un qui a été affecté en groupe I selon l'enquête filtre, car celle-ci ne détecte pas (pour lui) des problèmes de santé ou de handicap. Son poids est donc élevé car il a été multiplié (au moins) par le facteur 13,5 en cours de calcul. Il s'avère avoir des problèmes de santé ou de handicap selon l'enquête. Donc il contribue à un indicateur de ce type, beaucoup plus fortement que s'il avait été correctement identifié dans l'enquête filtre. D'où un risque de surestimation de l'indicateur.

On a procédé à quelques analyses plus fines que le simple examen des personnes de poids très élevés. Pour certains indicateurs, portant sur une altération possible de l'état fonctionnel de la personne, l'examen a porté sur les personnes (environ 300 pour chaque indicateur) qui sont « positive⁴¹ » et de poids élevés (les 5 % des positifs de plus grand poids).

L'examen au cas par cas montre qu'il y a très peu de situations où l'explication serait un changement de situation de la personne entre l'enquête filtre et l'enquête elle-même. La situation la plus fréquente (soit environ les trois-quarts des cas) est celle de personnes qui cumulent des déficiences et limitations que l'on soupçonne de « faible⁴² » intensité, par exemple en raison de leur âge. Se pose alors bien la question de savoir si leurs poids traduisent correctement l'existence d'une part assez importante de personnes dans cette situation.

On conclut⁴³ donc : on ne peut exclure l'existence de faux négatifs, soit comme « erreur » soit comme changement d'état, mais ils sont sans doute assez peu nombreux. La présence de personnes « négatives » selon le filtre et « positives » selon l'enquête résulte plutôt du mode de questionnement (détaillé et individuel) qui permet de faire émerger finement déficiences et limitations, d'où des situations « intermédiaires ».

⁴⁰ Ce qui acte une évidence : l'enquête filtre n'est pas suffisante pour atteindre les objectifs du dispositif « Handicap-Santé ».

⁴¹ D'état fonctionnel altéré, donc, il peut s'agir d'un mauvais état de santé général ou de « handicap », moteurs, cognitifs, tels qu'ils ont été définis et étudiés dans l'IP 1308, http://www.insee.fr/fr/themes/document.asp?reg_id=0&ref_id=ip1308 Les discriminations liées au handicap et à la santé. L'état fonctionnel altéré résulte de la conjonction d'une déficience déclarée dans l'enquête et d'une limitation fonctionnelle en relation avec la déficience.

⁴² Le statisticien espère que son indicateur « de handicap » se calcule avec une variable latente et un seuil. Une idée qui se révèle très rapidement totalement inappropriée au regard des sujets de l'enquête. L'analyse repose bien sur un examen exhaustif des personnes, de leurs déficiences et limitations.

⁴³ Provisoirement, en attendant les exploitations détaillées de l'enquête.