

ENQUÊTES VQS ET CARE : OU COMMENT PRENDRE EN COMPTE LE TIRAGE D'ÉCHANTILLONS DÉPENDANTS DANS LE CADRE DU NCEE

Amélie Carrère¹ (**), Olivier Haag² (*), Noémie Soullier³ (**),

(*) Insee, Direction de la méthodologie et de la coordination statistique et internationale

(**) Drees, Sous-direction Observation de la solidarité

Résumé

L'enquête CARE (Capacités, Aides et REssources des seniors) a pour but d'éclairer le débat national sur la dépendance. La population âgée dépendante est surtout présente aux très grands âges, dans des proportions qui restent néanmoins faibles : environ 10 % des personnes âgées de 80 ans ou plus sont bénéficiaires de l'allocation personnelle d'autonomie (APA) (source : enquête annuelle sur les bénéficiaires de l'APA – DREES).

Pour interroger un nombre suffisant de personnes dépendantes vivant en logement ordinaire, il est donc nécessaire de les surreprésenter lors du tirage de l'échantillon. Or, aucune source administrative ne permet de repérer exactement cette population cible. Une première enquête permettant de disposer du niveau de dépendance des personnes de plus de 60 ans est donc nécessaire : c'est le rôle principal de l'enquête VQS (Vie Quotidienne et Santé). Les répondants à cette enquête serviront alors de base de sondage pour l'enquête CARE.

Ainsi l'échantillonnage de l'enquête VQS avait deux principaux objectifs :

- servir de base de sondage à l'enquête CARE qui sera une enquête réalisée en face à face par des enquêteurs de l'Insee. Cette base de sondage doit donc se limiter aux Zones d'Action Enquêteurs (ZAE) de l'échantillon maître (EM) de l'Insee ;
- assurer une représentativité départementale afin de pouvoir éclairer le débat local, ce qui suppose de couvrir également les communes situées en dehors des ZAE de l'EM.

La première partie de l'article présentera les travaux mis en œuvre pour définir le plan de sondage de VQS. Ces estimations se basent sur les résultats observés lors de la précédente édition des enquêtes VQS de 2007 et Handicap Santé Ménages 2008.

La suite de l'article concernera l'enquête CARE et présentera les hypothèses prises pour définir le plan de sondage CARE sans en connaître la base de sondage et ce afin de répondre aux besoins du Nouveau Cadre d'Emploi des Enquêteurs (NCEE). Enfin, la dernière partie présentera la méthodologie mise en œuvre pour le tirage de l'échantillon de CARE une fois la base de sondage réellement disponible, sous contrainte du respect des allocations par ZAE livrées en 2014.

¹ amelie.carrere@sante.gouv.fr

² olivier.haag@insee.fr

³ noemie.soullier@sante.gouv.fr



Abstract

The CARE Survey aims to collect information on capabilities, resources and care among the elderly. Since disabled old persons are relatively rare among the whole population, it is necessary to over-represent them in the sample through a “filter Survey”, called VQS: the every Life and Health Survey. VQS respondents’ constitute the population frame of the CARE Survey.

Besides, to organise data collection, the regional office has to know precisely, about 6 months before, the regional Survey allocation. Therefore, the main issue of the CARE survey was to define its allocation before the population frame is known.

After a brief presentation of VQS and CARE surveys, this paper will describe firstly the sampling design of VQS, secondly the way to calculate the anticipated CARE allocations and finally its sampling design.

Mots-clés N. B. 5 maximum

Échantillonnage, collecte, équilibrage



1.	Motivations d'une enquête filtre VQS.....	5
1.1.	Échantillonnage de l'enquête CARE	5
1.1.1.	Enquête CARE sans enquête filtre.....	6
1.1.2.	Echantillon complémentaire de bénéficiaires de l'APA dans l'enquête CARE.....	7
1.1.3.	Enquête Filtre VQS 2007.....	8
1.1.4.	Enquête Filtre VQS 2014.....	9
1.2.	Représentativité départementale.....	12
2.	Plan de sondage de VQS.....	14
2.1.	Constitution de la base de sondage	14
2.1.1.	- Restriction aux personnes de 60 ans et plus	14
2.1.2.	Conservation des logements foyers	14
2.1.3.	Retrait des personnes vivant dans des communautés hors champ de l'enquête.....	14
2.1.4.	Caractéristiques de la base finale de sondage	15
2.2.	Tirage de l'échantillon.....	15
2.2.1.	Calcul des allocations de l'échantillon principal de 140 000 logements.....	15
2.2.2.	Échantillon complémentaire de 60 000 logements.....	15
2.3.	Calcul des poids	17
2.3.1.	- Poids à utiliser pour CARE.....	17
2.3.2.	Poids à utiliser pour VQS	17
3.	- Principe du calcul anticipé des allocations pour chester	18
3.1.	Les allocations à livrer à MOAE	18
3.2.	Choix de la base de sondage pour les calculs	18
3.3.	Les hypothèses mises en œuvre.....	18
3.4.	Le calcul anticipé des quotités par ZAE pour la planification de l'activité 2015	19
4.	- Tirage de l'échantillon CARE	20
4.1.	Le principe général	20
4.2.	L'estimation du poids des logements répondants à l'enquête VQS.....	21
4.3.	Le poids des individus de la base de sondage CARE	22
4.4.	Les modifications à la marge des allocations théoriques	23
4.4.1.	- Les ZAE « cas simples »	23
4.4.2.	- Les ZAE « cas complexes »	24
4.5.	Le tirage proprement dit	25
4.6.	Les résultats en termes de dispersion des poids	26



Introduction

Le débat national de 2011 sur la dépendance des personnes âgées a fait émerger de nombreuses questions quant à l'évolution de la dépendance et aux moyens financiers, humains et techniques qu'elle requiert. Les systèmes d'information existants, notamment les enquêtes Handicap Santé⁴, ne permettaient pas d'apporter des réponses au niveau national sur ces thématiques. En particulier, il s'est révélé nécessaire de :

- lever une partie de l'incertitude qui entoure les hypothèses de développement de la dépendance. Ceci suppose de suivre l'évolution du nombre de personnes âgées en perte d'autonomie et, en particulier, de la comparer aux projections réalisées par la DREES en 2011.

Selon ces projections, 1,5 millions de personnes seraient dépendantes en 2025 (au sens où elles bénéficieraient de l'allocation personnalisée d'autonomie – APA), contre près de 1,2 millions en 2012, soit une hausse de 30 % (source : Groupe de travail présidé par J.M. Charpin « Perspectives démographiques et financières de la dépendance », 2011) ;

- mieux cerner les enjeux financiers auxquels font face les ménages confrontés à la dépendance, que la personne dépendante réside à domicile ou en établissement. Ces questions de financement de la perte d'autonomie et, plus spécifiquement, de reste à charge sont cruciales, d'autant plus que le coût de la prise en charge est appelé à s'accroître. Plusieurs facteurs peuvent provoquer cette hausse : la diversification des prestations, le renforcement des normes de sécurité et de qualité, les revalorisations des salaires et l'augmentation du niveau de qualification des personnels intervenants. À l'heure actuelle, les besoins liés à la perte d'autonomie sont assurés à la fois par l'entourage (proches aidants), par le biais de prestations (telles que l'APA ou l'action sociale des caisses de retraite) et par les ressources propres des personnes âgées.

Les enquêtes CARE (Capacités, Aides et REssources des seniors) ont pour but de répondre à ces besoins d'information (voir encadré). Or, la population âgée dépendante est surtout présente aux très grands âges, dans des proportions qui restent néanmoins faibles : en 2011, 8 % des personnes âgées de 60 ans ou plus sont bénéficiaires de l'allocation personnalisée d'autonomie (APA) (Bérardier, 2014) et environ 10 % des personnes âgées de 80 ans ou plus.

La question qui se pose est : comment en interroger suffisamment ?

La première partie mettra en évidence les motivations qui ont conduit à réaliser une enquête filtre (VQS) ainsi que les volumes nécessaires pour répondre aux objectifs de cette enquête. Pour interroger un nombre suffisant de personnes dépendantes vivant en logement ordinaire, il a été nécessaire de procéder à une surreprésentation des personnes estimées dépendantes via une enquête filtre : Vie Quotidienne et Santé (VQS). Les répondants à VQS constituent la base de sondage de l'enquête CARE.

La deuxième partie de l'article présentera les hypothèses mises en œuvre pour les prévisions d'allocations par ZAE, rendus nécessaires par le NCEE (Nouveau Cadre d'Emploi des Enquêteurs), et qui ont ici été réalisés alors même que la collecte de VQS n'avait pas encore commencé et que la base de sondage de l'enquête CARE n'est donc pas connue.

Une troisième partie s'intéressera à la méthodologie mise en œuvre pour le tirage de l'échantillon de CARE une fois la base de sondage réellement disponible, sous contrainte du respect des allocations par ZAE prévues.

⁴ Les enquêtes Handicap-Santé, qui ont un lien de parenté fort avec les enquêtes CARE, ont été conduites entre 2007 et 2009 et portaient sur l'ensemble de la population, vivant en métropole et dans les DOM, en logement ordinaire ou en institution.



Encadré : Dispositif d'enquêtes CARE

Le dispositif d'enquêtes « CARE » pour Capacités, Aides et REssources des seniors, en logements ordinaires (2014-2015) et en institutions (2016) est proche de celui des enquêtes Handicap-Santé de 2007-2009. Il comprend :

- L'enquête Vie Quotidienne et Santé (VQS) en 2014-2015, enquête filtre du volet « seniors » de l'enquête CARE en ménages,
- L'enquête CARE en ménages en 2015, volets « seniors » et « aidants »,
- L'enquête CARE en institutions en 2016, volets « seniors » et « aidants ».

Pour le volet « seniors » de l'enquête en logements ordinaires, les unités statistiques enquêtées sont les individus âgés de 60 ans ou plus à la date de début de collecte de l'enquête (2 mai 2015) vivant en logements ordinaires (y compris en logement-foyer) en France métropolitaine.

Le champ a été défini par une limite d'âge à 60 ans car cela correspond à l'âge minimum pour prétendre à des droits spécifiques à la dépendance (allocation personnalisée d'autonomie – APA, action sociale des caisses de retraite, etc.). Parmi la population interrogée, l'intérêt est principalement porté à la population âgée dépendante, c'est-à-dire ayant des difficultés à réaliser seule des activités de la vie quotidienne et pour qui l'aide d'un tiers est nécessaire. Mais les personnes âgées en bonne santé sont aussi interrogées afin de disposer de résultats sur l'ensemble des seniors, de calculer des prévalences et de permettre des comparaisons entre les situations des personnes dépendantes et celles des personnes non dépendantes.

1. Motivations d'une enquête filtre VQS

L'enquête CARE a pour objectif de mesurer l'évolution du nombre de personnes en perte d'autonomie. Il est donc nécessaire, pour mener des analyses robustes sur cette population, de disposer d'un nombre suffisant de personnes en perte d'autonomie dans l'échantillon de l'enquête CARE. C'est l'objectif principal de l'enquête VQS, qui sera présenté dans un premier temps.

En outre, les départements étant les chefs de file de l'action sociale et médico-sociale auprès des personnes âgées, le besoin d'information sur la dépendance au niveau départemental est important. Le mode de collecte de l'enquête VQS a permis d'envisager un élargissement de l'échantillon VQS afin de disposer de résultats par département. Le supplément de ménages nécessaire à réaliser ces analyses sera présenté dans un second temps.

1.1. Échantillonnage de l'enquête CARE

Bien que les prévisions estiment à la hausse la proportion de personnes dépendantes (Lecroart *et al.*, 2013), celles-ci constituent toujours une population rare. L'enquête Handicap-Santé (Insee-Drees, 2008) a estimé que 7 % de la population française des 60 ans ou plus vivant à domicile était dépendante au sens de la grille AGGIR⁵ (Eghbal-Téhérani et Makdessi, 2011). En supposant que cette proportion est identique en 2015, on peut alors estimer qu'environ 1,2 millions d'individus seraient dépendants en 2015 (voir Tableau 1). La même méthode a été appliquée pour les différentes sous-populations d'intérêt : personnes très dépendantes, personnes aidées, personnes atteintes de la maladie d'Alzheimer... Pour chaque sous-population, un seuil minimal de répondants qui a été défini (voir Tableau 1). Cet effectif doit permettre de réaliser les analyses prévues pour répondre aux

⁵ Les groupes iso-ressources (GIR), au nombre de 6, permettent de classer les personnes en fonction de leur degré de perte d'autonomie. Le classement dans un GIR s'effectue à l'aide de la grille Aggir (Autonomie gérontologie-groupe iso-ressources) qui prend en compte différentes dimensions de la perte d'autonomie (cohérence, orientation, toilette, communication...). Les personnes en GIR 1 sont les plus dépendantes. Les personnes en GIR 5-6 sont considérées comme autonomes, alors que le classement en GIR 1 à 4 ouvre droit à l'Allocation Personnalisée d'Autonomie (APA), allocation spécifique à la perte d'autonomie.



objectifs de l'enquête. L'objectif le plus contraignant est d'avoir au moins 600 répondants en GIR 1-2 (personnes les plus dépendantes).

Trois solutions étaient envisageables pour disposer de ce seuil minimal : soit tirer directement un large nombre de personnes pour constituer l'échantillon de l'enquête CARE, soit surreprésenter la population d'intérêt dans CARE en tirant un échantillon complémentaire de bénéficiaires de l'APA, soit surreprésenter la population d'intérêt dans CARE grâce à une enquête filtre.

Tableau 1. Estimation de la population en 2015 et objectifs fixés pour CARE par variable d'intérêt

	Estimation nationale à la date de CARE	Objectif pour l'enquête CARE
Personnes de 60 ans ou plus	17 013 000	10 000
Personnes estimées en GIR 1-4 ⁶	1 181 000 (7%)	2 000
Personnes estimées en GIR 1-2	314 000 (2%)	600
Personnes aidées par quelqu'un de l'entourage	3 920 000 (23%)	2 000
Personnes aidées par un aidant professionnel	2 616 000 (15%)	2 000
Personnes Alzheimer	461 000 (3%)	600

Lecture : en 2015, 1 181 000 personnes seront estimées en GIR 1-4, soit environ 7% de la population de 60 ans ou plus.

Sources : VQS 2007, HSM 2008, Omphale - projection 01/01/2013, Calculs Drees

L'INSEE réalise le tirage de l'échantillon dans les sources fiscales. Compte tenu de la taille importante de l'échantillon souhaité, l'échantillon ne pouvait pas être tiré dans la dernière EAR (enquête annuelle du recensement) sous Octopusse. Par ailleurs, le mode de collecte de l'enquête VQS (postal, internet et téléphonique) aurait nécessité une phase de saisie des adresses. Il n'était pas non plus envisageable de tirer dans les cinq derniers EAR, car cela aurait engendré trop de déchets vu la population enquêtée. La solution s'est donc arrêtée sur un tirage de l'échantillon dans les sources fiscales.

Le nouveau cadre d'emploi des enquêteurs de l'INSEE impose de connaître l'allocation par ZAE en septembre d'une année pour les enquêtes dont la collecte se déroule entre janvier et décembre de l'année suivante. Par ailleurs, les bases fiscales d'une année donnée ne sont disponibles qu'en mars de l'année suivante.

Les estimations de volumétrie réalisées par la suite utilisent ces informations, ainsi que les informations de suivi de collecte des enquêtes VQS 2007 et HSM 2008 au prorata de l'écart entre les bases de sondage et les collectes. En effet, la forte mobilité des personnes âgées (départ en institution, recohobitation avec les enfants, décès) conduit à être le plus précis possible concernant les hypothèses d'attrition de l'échantillon entre le tirage et la collecte.

1.1.1. Enquête CARE sans enquête filtre

Pour une collecte de l'enquête CARE en 2015 sans enquête filtre, il aurait été nécessaire de tirer l'échantillon en mars 2014 dans une base datant de 2013. En considérant une collecte tout début 2015, cela implique une ancienneté de la base de tirage d'environ 19 mois⁷. Le taux de réponse du Tableau 2 a été estimé en prenant en compte cet écart.

⁶ Personnes éligibles à l'APA

⁷ Les déclarations fiscales sont à renvoyer courant mai de chaque année. Il y a donc un écart de 19 mois entre juin 2013 et janvier 2015.



Tableau 2. Taux de réponse pour CARE 2015 estimés à partir d'HSM 2008, par tranche d'âge

	60-69 ans	70-79 ans	80 ans ou plus
Taux de réponse total	72%	66%	46%

Bien que le taux de réponse des personnes plus âgées soit plus faible, la proportion de personnes peu autonomes y est plus importante : dans Handicap-Santé, 1 % des personnes âgées de 60-79 ans sont classées en GIR 1-2, alors que c'est le cas de 5 % des 80 ans ou plus. Une surreprésentation des plus âgés pourrait donc permettre d'atteindre les objectifs fixés précédemment. Ainsi, en considérant un taux de sondage huit fois plus grand pour les 70 ans et plus⁸ que pour les 60 – 69 ans, l'échantillon serait constitué de 42 700 personnes tirées en tout, dont 3 900 âgées de 60 à 69 ans, 19 200 âgées de 70 à 79 ans et 19 600 âgées de 80 ans ou plus. Parmi les 25 000 répondants estimés, 2 300 seraient des personnes dépendantes (estimation GIR 1 à 4) et 600 des personnes avec une estimation GIR 1 et 2.

Outre, le coût d'une telle enquête en face-à-face, le tirage d'un échantillon de cette taille aurait épuisé l'échantillon-maître sur ces tranches de population et aurait pu difficilement être assuré uniquement par le réseau des enquêteurs de l'Insee.

Cette solution a donc été écartée.

1.1.2. Échantillon complémentaire de bénéficiaires de l'APA dans l'enquête CARE

Un échantillon complémentaire de bénéficiaires de l'APA aurait eu pour objectif d'augmenter le nombre de personnes dépendantes dans l'échantillon et plus spécifiquement le nombre de bénéficiaires de l'APA dans l'échantillon. En effet, l'un des objectifs de l'enquête CARE est de mesurer le reste à charge lié à la dépendance des personnes et donc notamment de recenser les aides publiques (principalement l'APA) dont elles disposent.

Ce tirage comporte le risque de tirer deux fois un même individu. Pour pouvoir repérer les individus qui sont dans les deux bases de sondage et ainsi effectuer le partage des poids, la date de la remontée des données APA doit être proche de celle de la base de sondage principale. Dans le cas contraire, l'échec d'appariement pour double compte est accru du fait des déménagements entre la date de collecte de la base principale et celle de la date de référence de l'extraction des données APA (6 % de déménagements en 9 mois environ). Ainsi, en complément des bases fiscales 2013 (validité au 1^{er} janvier), des remontées individuelles des données administratives des conseils généraux en date du 31 mars⁹ 2013 pourraient être utilisées. Une base consolidée en date de mars 2013 serait disponible environ 6 mois plus tard soit en octobre 2013.

Pour une collecte de l'enquête CARE en 2015, il aurait été nécessaire de tirer l'échantillon en mars 2014. Les deux bases de sondage (base fiscale et bases administratives des conseils généraux) dateraient du premier trimestre 2013. En considérant une collecte tout début 2015, cela implique une ancienneté de la base de tirage d'environ 19 mois, soit le même taux de réponse que précédemment pour l'échantillon tiré en population générale (voir Tableau 3).

Environ un tiers des bénéficiaires de l'APA en 2011 sont entrés dans le dispositif au cours de l'année écoulée (remontées individuelles APA 2011¹⁰). Or, le nombre de bénéficiaires de l'APA est à peu près stable, par conséquent environ un tiers sortent¹¹ également du dispositif chaque année. Ainsi, pour un échantillon constitué à partir de données 2013, on peut estimer qu'un tiers des bénéficiaires de l'APA seront encore dans le dispositif en 2015. En ajoutant un taux de mobilité hors

⁸ Étant donné la chute du taux de réponse sur les 80 ans et plus, une surreprésentation de cette population seule ne saurait être efficace.

⁹ Les extractions de données issues des conseils généraux se font par trimestre. Ainsi pour être au plus près des dates des déclarations fiscales, il est préférable de prendre le premier trimestre. Par ailleurs, un délai d'apurement des bases est aussi à prendre en considération.

¹⁰ Cette estimation est assez éloignée de celle de Debout 2010 (45%). Ceci s'explique principalement par le fait que les durées dans l'APA sont plus longues en 2011 qu'en 2007.

¹¹ Les sorties de l'APA correspondent principalement à des décès.



décès et un taux de non-réponse, environ un quart de l'échantillon de bénéficiaires de l'APA seraient répondants.

Tableau 3. Taux de réponse de l'échantillon complémentaire APA estimés à partir d'HSM 2008 et des remontées individuelles APA 2011, par tranche d'âge

	60-69 ans	70-79 ans	80 ans et plus
Taux de réponse total	53%	40%	18%

La taille de l'échantillon tiré en population générale serait d'environ 15 000 pour 9 400 répondants, ce qui permettrait d'obtenir environ 140 personnes estimées en GIR 1-2. En considérant un échantillon complémentaire d'environ 4 000 bénéficiaires de l'APA, environ 500 répondants seraient en GIR 1-2. Au total, on aurait donc environ 640 personnes estimées en GIR 1-2 pour un échantillon de 19 000 enquêtés et 12 000 répondants.

Par ailleurs, les remontées individuelles APA faites par la Drees ne sont pas exhaustives sur la France entière. Elles ne sont pas non plus un échantillon aléatoire puisque la remontée des données est basée sur le volontariat des départements. La remontée de ces données nécessite un travail conséquent pour les conseils généraux, c'est pourquoi certains refusent de mettre à disposition leurs données. Mais, outre la complexité de la collecte de ces bases administratives, l'enquête CARE vise une acception plus large de la perte d'autonomie, que le bénéfice d'une allocation comme l'APA ne permet pas de recouvrir, ne serait-ce que parce que le bénéfice de cette prestation suppose une démarche de la part des personnes. De plus, avec un échantillon complémentaire de bénéficiaires de l'APA l'estimation d'un taux de pseudo non-recours à l'APA via la mise en perspective du degré de dépendance (mesuré à partir du questionnaire CARE) et du bénéfice de l'APA (mesuré via un enrichissement avec les données administratives) serait caduque.

Aucune autre base administrative ne permet de repérer les personnes dépendantes. Ainsi pour la surreprésentation de ces personnes, nous avons envisagé la reconduite d'une enquête filtre, comme cela avait été le cas en 2007.

1.1.3. Enquête Filtre VQS 2007

L'objectif de l'enquête Vie Quotidienne et Santé (VQS) de 2007 était d'assurer un échantillon de personnes handicapées de taille suffisante dans le volet ménages de l'enquête Handicap-Santé (2008). Les strates d'intérêt de VQS 2007 étaient déterminées selon l'âge et le groupe d'autonomie. Le degré d'autonomie est estimé sous la forme d'un score, qui utilise les réponses à l'enquête VQS. La construction du score s'appuyait sur les relations observées entre VQS 1999 et HID 2000. Le groupe d'autonomie est ensuite alloué à chaque individu selon son âge et son score : le groupe I correspond aux personnes les plus autonomes (pour lesquelles le taux de sondage est faible) et le groupe IV aux personnes les plus handicapées (strate exhaustive).

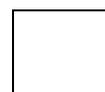
Les taux de réponses sont estimés en considérant une base de sondage de 2013 et une collecte de l'enquête filtre en septembre 2014 (soit un délai de 15 mois environ), puis une enquête CARE en mai 2015 (soit 8 mois après VQS), voir Tableau 4.

Tableau 4. Taux de réponse estimés à partir de VQS 2007 et HSM 2008, par tranche d'âge

	60-69 ans	70-79 ans	80 ans et plus
Taux de réponse VQS	64%	62%	47%
Taux de réponse CARE seniors	69%	66%	55%

Un tirage de 120 000 ménages pour l'enquête filtre permettrait d'obtenir près de 70 000 ménages répondants grâce à une collecte postale et téléphonique. Parmi ces répondants, 15 000 seraient sélectionnés pour l'enquête CARE en face-à-face (1 personne par ménage), en surreprésentant les groupes les plus dépendants. On peut alors espérer 11 000 répondants dont 1 800 personnes estimées en GIR 1 à 4 et 515 personnes estimées en GIR 1-2, ce qui est en deçà de l'objectif visé.

Cependant, l'efficacité de l'enquête filtre peut être améliorée. En effet, le groupe d'autonomie de VQS 2007, bien que déterminé différemment selon l'âge, n'était pas spécifique à la population âgée et



encore moins au repérage de la perte d'autonomie due au vieillissement. Or, contrairement à HSM, l'enquête CARE a un champ restreint aux personnes âgées de 60 ans ou plus, et il est donc nécessaire que la classification en groupe VQS capte plus spécifiquement les problèmes d'autonomie liés au vieillissement.

Il a donc été décidé d'amplifier le rôle de l'enquête filtre en adaptant le questionnaire et la stratification à la perte d'autonomie des seniors.

1.1.4. Enquête Filtre VQS 2014

Pour le plan de sondage du volet ménages de Handicap-Santé (HSM), un score décrivant le degré de handicap a été construit à partir des réponses à VQS 2007 (Midy 2010, Bouvier 2011). Ce score a été défini en analysant les liens existant entre les réponses à VQS 1999 et le degré de handicap repéré dans HID et défini en trois catégories, en prenant en compte 6 aspects : limitations et déficiences sensorielles, limitations motrices, limitations cognitives, restrictions dans les activités essentielles de la vie quotidienne, restrictions dans les activités instrumentales de la vie quotidienne, restrictions de participation sociale (formation et emploi).

La construction du score se fait par incrémentation : chaque réponse ajoute 0, 1, 3 ou 6 points au score selon que la modalité représente l'absence de limitation, une faible limitation, une limitation moyenne ou une forte limitation. Ce score est ensuite divisé par le score maximal possible pour chaque individu, afin de tenir compte de la non-réponse partielle : le score maximal n'est pas incrémenté si la personne n'a pas répondu à la question. Le score est enfin normalisé : il est compris entre 0 et 100.

La classification dans un groupe de handicap était définie par des seuils de score variant selon l'âge de la personne (voir Tableau 5).

Tableau 5. Détermination du groupe d'autonomie selon le score et l'âge dans VQS 2007

	0 – 19 ans	20 – 59 ans	60 – 79 ans	80 ans ou plus	Groupe VQS
Score	0	[0 – 3]	[0 – 4]	[0 – 5]	I
	[1 – 9]	[4 – 11]	[5 – 24]	[6 – 39]	II
	[10 – 24]	[12 – 29]	[25 – 44]	[40 – 64]	III
	[25 – 100]	[30 – 100]	[45 – 100]	[65 – 100]	IV

Afin d'améliorer le rôle filtre de l'enquête VQS, une étude des liens entre VQS 2007 et HSM est effectuée. Deux objectifs sous-tendent cette analyse :

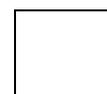
- Recentrer la construction du score sur la perte d'autonomie liée au vieillissement et ainsi améliorer le pouvoir discriminatoire des groupes VQS par rapport aux objectifs de l'enquête CARE ;
- Prendre en compte la précision accrue du questionnaire VQS, qui a en 2007 et en 2014, une modalité supplémentaire pour qualifier les difficultés.

La population d'intérêt est définie sous trois angles : l'âge, le niveau de dépendance et le bénéfice de l'APA. Ces trois variables définissent des populations qui se recoupent, mais qui ne se superposent pas :

- Les plus âgés sont plus dépendants mais ils ne le sont pas tous ;
- Les plus dépendants bénéficient plus souvent de l'APA, mais il existe du non-recours.

Par conséquent, la stratification doit prendre en compte ces trois éléments, qui seront soit collectés dans l'enquête (degré de dépendance et bénéfice de l'APA) soit présents dans la base de sondage (âge). Les strates tiendront compte du degré d'autonomie résumé par le groupe VQS et estimé sous la forme d'un score (du même type que pour VQS 2007) à partir des réponses à l'enquête filtre.

Les variables d'intérêt choisies pour représenter la perte d'autonomie liée au vieillissement sont le GIR estimé à partir des réponses à l'enquête HSM (regroupé en trois catégories : 1-2, 3-4, 5-6) ainsi qu'un indicateur de la maladie d'Alzheimer (également en 3 modalités). Les variables explicatives sont les questions de VQS 2007. Les analyses sont effectuées via une modélisation logit et séparément sur les moins de 80 ans et les 80 ans ou plus, afin de prendre en compte les spécificités de ces sous-populations quant à la perte d'autonomie.



L'analyse montre que les variables relatives aux déficiences sensorielles et au handicap ne sont pas discriminantes pour les variables d'intérêt de CARE et sont donc exclues du score. Par ailleurs, les restrictions sur l'activité essentielle de la vie quotidienne « se laver » sont importantes dans la définition de la perte d'autonomie due au vieillissement : la variable correspondante de HSM a été ajoutée pour simuler l'effet de l'ajout de cette question dans VQS 2014. De la même manière, une question sur le bénéfice de l'APA est ajoutée dans VQS 2014 afin de mieux cibler cette population d'intérêt.

Comme pour VQS 2007, le score est incrémenté d'un certain nombre de points selon la modalité répondue (ici de 1 à 12 points, voir Annexe) ; il est ensuite ramené à un score sur 100 en prenant en compte la non-réponse partielle.

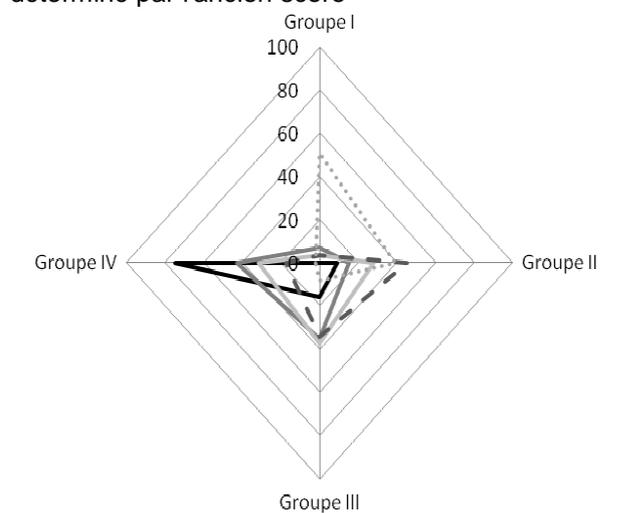
La distribution du nouveau score différant de celle du score défini pour VQS 2007, les seuils définissant les groupes ont été modifiés (voir Tableau 6).

Tableau 6. Détermination du groupe d'autonomie selon le score et l'âge dans VQS 2007 et 2014

	60 – 79 ans		80 ans ou plus		Groupe
	VQS 2007	VQS 2014	VQS 2007	VQS 2014	
Score	[0 – 4]	[0 – 7]	[0 – 5]	[0 – 9]	I
	[5 – 24]	[8 – 23]	[6 – 39]	[10 – 34]	II
	[25 – 44]	[24 – 39]	[40 – 64]	[35 – 59]	III
	[45 – 100]	[40 – 100]	[65 – 100]	[60 – 100]	IV

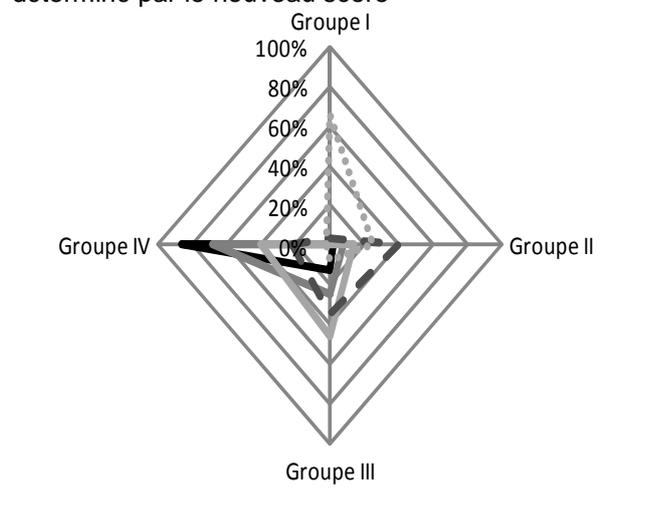
Ces nouveaux seuils ont été choisis de manière à avoir une bonne discrimination sur le critère « GIR estimé ». Le nouveau score permet ainsi d'avoir une meilleure discrimination dans les groupes : le groupe IV capte plus majoritairement des individus dépendants (GIR 1-4) et le groupe I plus exclusivement des individus autonomes (GIR 5-6) (voir Graphiques 1 et 2). Cela permet d'avoir une meilleure optimisation du plan de sondage sur ce critère.

Graphique 1. Répartition des répondants à HSM selon le GIR estimé et le groupe d'autonomie VQS déterminé par l'ancien score



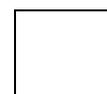
— GIR 1-2 — GIR 3 — GIR 4 - - GIR 5 GIR 6
 Note de lecture : 75 % des répondants à HSM estimés en GIR 1-2 sont classés dans le groupe IV avec l'ancien score.

Graphique 2. Répartition des répondants à HSM selon le GIR estimé et le groupe d'autonomie VQS déterminé par le nouveau score

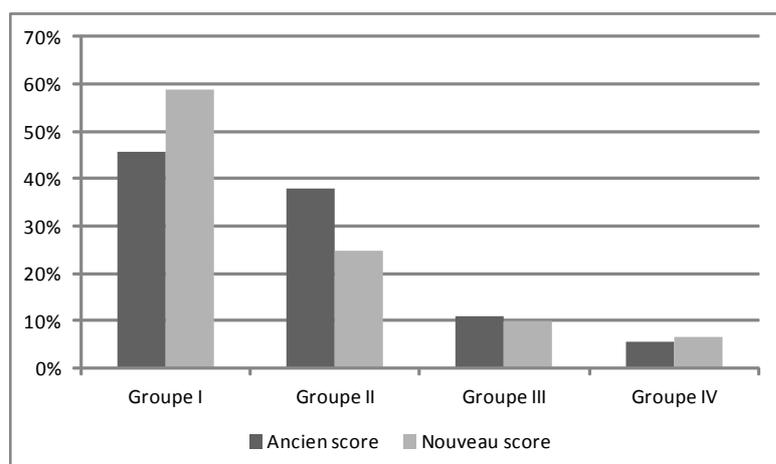


— GIR 1-2 — GIR 3 — GIR 4 - - GIR 5 GIR 6
 Note de lecture : 86 % des répondants à HSM estimés en GIR 1-2 sont classés dans le groupe IV avec le nouveau score.

Cette optimisation modifie la répartition par groupe d'autonomie des répondants à VQS : par exemple, la meilleure correspondance entre « GIR 6 » et « Groupe I » augmente l'effectif attendu du groupe I (voir Graphique 3). Mais les effectifs des populations d'intérêt n'en sont pas altérés, car la meilleure discrimination assurée par le nouveau score confère une meilleure efficacité au plan de sondage.



Graphique 3. Répartition des répondants selon le groupe, avec l'ancien et le nouveau score



Les taux de réponses sont ceux mentionnés au tableau 4.

À taille d'échantillon égale (15 000 seniors), les effectifs attendus par sous-population croissent sensiblement avec le nouveau score : 35 % de personnes estimées en GIR 1-2 supplémentaires et 28 % de personnes estimées en GIR 1-4 supplémentaires. Le nouveau score optimisé par rapport au GIR permet également d'avoir des effectifs de répondants suffisants pour les autres populations d'intérêt, car celles-ci sont bien corrélées à l'indicateur GIR (voir tableau 7).

La solution de l'enquête filtre comporte le risque d'apparition de faux négatifs, en cas d'écart trop important entre les enquêtes VQS et CARE. Un faux négatif correspond à un individu classé en bonne santé à l'issue de VQS mais qui s'avèrerait en mauvaise santé dans CARE. Lors de l'enquête HID 1999, l'échantillon des personnes en bonne santé était trop petit. Le poids des personnes en bonne santé était alors élevé et un « faux négatif » a fortement perturbé les analyses (Bouvier, 2012). Pour limiter ce risque, un plus grand nombre de personnes en bonne santé a été tiré dans Handicap-Santé volet ménages (2008) et le rapport des poids entre les strates a été limité. Ces précautions seront également mises en œuvre pour le tirage de CARE seniors 2015, afin de se prémunir d'une influence excessive des « faux négatifs ».

L'estimation du nombre de répondants par population d'intérêt pour chacune des solutions est présentée dans le tableau 7. La conduite d'une enquête VQS avec un filtre amélioré par rapport à 2007 a été retenue.

Afin de se prémunir d'un taux de réponse en dessous des estimations, une réserve de 20 000 ménages supplémentaires a été ajoutée à cet échantillon principal. Ceci permettra de tirer l'échantillon CARE dans de bonnes conditions.

Tableau 7. Synthèse des résultats

	Sans enquête filtre	Avec échantillon APA	Avec score VQS 2007	Avec score VQS 2014
Échantillon VQS	-	-	120 000	120 000
Répondants VQS	-	-	69 700	69 700
Échantillon CARE	42 700	19 000	15 000	15 000
Répondants CARE	24 500	10 300	10 700	10 700
GIR 1-4	2 270	3 000	1 900	2 400
GIR 1-2	620	640	515	700
Personne aidée par entourage	7 200	3 000	4 200	5 600
Personne aidée par un professionnel	5 200	2 300	3 000	4 000
Personne Alzheimer	930	710	720	910



1.2. Représentativité départementale

La gestion de l'APA est décentralisée au niveau départemental et les bases de bénéficiaires de l'APA ne sont disponibles qu'au sein des Conseils Généraux. La DREES envisage des extractions de ces bases départementales pour les apparier avec l'enquête CARE, et ainsi bénéficier d'une information plus précise sur la population des bénéficiaires de cette allocation. Ce type de travaux (réalisé dans les conseils généraux) mobilise des moyens importants. En contrepartie de cet investissement, les conseils généraux participants peuvent souhaiter un retour d'informations permettant d'estimer des résultats sur leur territoire.

De façon générale, le besoin d'informations sur la dépendance est important au niveau départemental, les départements étant chefs de file sur l'action sociale et médico-sociale auprès des personnes âgées. Les seuls indicateurs concernant les personnes âgées, dont dispose le groupe Indicateurs socio-départementaux, sont des indicateurs de bénéfice d'une allocation ou prestation (APA, ACTP, PCH, ASH). Une mise en regard de ces indicateurs et de l'estimation de l'état de santé, et en particulier de la perte d'autonomie, de la population au niveau local permettra d'éclairer les acteurs locaux.

Une première solution serait de réaliser des extensions départementales dans l'enquête CARE. Cette solution ne pourrait pas être mise en place pour tous les départements car cela serait trop coûteux. Pour une estimation au niveau local, la conduite d'extensions sur CARE demanderait la mise en place d'une approche « petits domaines ». Pour que celle-ci soit possible, il faut un échantillon suffisamment important dans chaque région, ce qui implique une stratification géographique (Le Guenec 2012). L'échantillon VQS devrait aussi être plus conséquent. Pour réaliser une extension départementale avec un objectif d'environ 1 500 individus répondants pour un département, il faudrait disposer dans l'échantillon CARE de 2 200 individus, soit 9 500 logements à tirer sur le département dans VQS. En considérant que l'échantillon CARE est réparti de façon homogène sur le territoire métropolitain, on peut considérer qu'environ 1 300 logements sont tirés par département pour VQS. De ce fait, une extension sur un seul département conduirait à tirer environ 8 700 logements supplémentaires sur VQS et 2 000 individus supplémentaires sur CARE (voir Tableau 8).

Tableau 8. Échantillon par département dans le cas d'extensions départementales dans l'enquête CARE

	Sans extensions sur CARE	Avec extensions sur CARE	Différence
Échantillon VQS	1 300	9 500	8 200
Répondants VQS	740	5 500	4 800
Échantillon CARE	160	2 200	2 040
Répondants CARE	120	1 500	1 420

Par ailleurs, des extensions sur CARE sur quelques départements s'avèreraient peu « rentables » compte tenu de l'écart important entre la complexité du questionnaire et les souhaits des conseils généraux. En effet, lors du bilan des extensions départementales menées dans l'enquête Handicap-Santé en ménages, les conseils généraux ont relaté que le coût engendré par la collecte de cette enquête longue en face-à-face était plus important que le bénéfice qu'ils en ont tiré. Notamment, la complexité du questionnaire a rendu son exploitation difficile et, conjuguée à un échantillon qui ne permettait pas d'isoler chaque sous-population d'intérêt, n'a que partiellement répondu à leurs préoccupations locales.

La deuxième solution serait donc de réaliser ces extensions sur l'enquête VQS. Le format court de l'enquête VQS a l'avantage de pouvoir être réalisé sans face-à-face¹², ce qui en diminue le coût et laisse la possibilité d'un large échantillon. Cela permet de répondre facilement à un objectif de représentativité départementale pour tous les départements, permettant ainsi des comparaisons entre

¹² Une analyse des réponses de VQS 2007 sans face-à-face et avec face-à-face a permis de mettre en évidence qu'il ne risquait pas d'y avoir de biais de sélection sans ce mode de collecte.



territoires. Le questionnaire simple garantit, de plus, une utilisation optimale par les acteurs locaux. Il explore les divers aspects de la perte d'autonomie et permet d'estimer des prévalences sur chaque territoire.

L'introduction d'extensions au niveau de l'enquête filtre semble être un bon compromis et la solution la plus rentable pour tous les acteurs. En effet, l'enquête filtre peut aisément être stratifiée géographiquement, et le coût d'un enquête supplémentaire est moins élevé pour cette enquête courte. De plus, le nombre d'interrogés pour l'enquête CARE, en face à face, ne sera pas impacté, puisque les extensions locales ne seront réalisées qu'au niveau de l'enquête filtre et ne seront pas prises en compte dans la base de tirage de CARE.

L'effectif enquêté dans VQS a alors été élargi et stratifié de sorte à assurer la représentativité de chaque département.

L'objectif pour ces extensions départementales était de disposer d'une moyenne de 2 000 individus de 60 ans ou plus répondants par département sur le territoire métropolitain¹³.

En utilisant les taux de réponses du Tableau 4 et le nombre de personnes de 60 ans ou plus par logement, le nombre de logements supplémentaires devant être tirés est de 60 000 (voir Tableau 9).

Par ailleurs, 10 000 logements ont été tirés dans les DOM (hors Mayotte), afin de produire des résultats également sur ces territoires.

Tableau 9. Estimation de l'échantillon VQS dans le cas d'extensions départementales dans l'enquête VQS

	Nombre total France métropolitaine	Nombre moyen par département
Échantillon total VQS	200 000	2 080
<i>dont principal hors réserve</i>	120 000	1 250
<i>dont réserve</i>	20 000	210
<i>dont extensions</i>	60 000	620
Nombre de ménages répondants à VQS	116 200	1 210
<i>dont principal hors réserve</i>	69 700	730
<i>dont réserve</i>	11 600	120
<i>dont extensions</i>	34 900	360
Nombre d'individus répondants à VQS	186 600	1 940
<i>dont principal hors réserve</i>	111 900	1 170
<i>dont réserve</i>	18 700	190
<i>dont extensions</i>	56 000	580

Ainsi l'échantillonnage de l'enquête VQS avait deux objectifs :

- servir de base de sondage à l'enquête CARE qui sera une enquête réalisée en face à face par des enquêteurs de l'Insee. Cette base de sondage doit donc se limiter aux Zones d'Action Enquêteurs (ZAE) de l'échantillon maître (EM) de l'Insee ;
- assurer une représentativité départementale afin de pouvoir éclairer le débat local, ce qui suppose de couvrir également les communes situées en dehors des ZAE de l'EM.

¹³ Le tirage a ensuite pris en compte les disparités géographiques, en tirant un échantillon plus important pour les départements accueillant plus de personnes âgées de 60 ans ou plus.



2. Plan de sondage de VQS

2.1. Constitution de la base de sondage

Le champ des enquêtes correspond aux logements où sont présents au moins une personne âgée de 60 ans et plus à la date de l'enquête CARE¹⁴. Les communautés dont les maisons de retraite et les EHPAD¹⁵ sont exclues du champ. En revanche, les logements foyers non EHPAD, assimilés à des communautés au sens du RSL (Répertoire Statistique des Logements), sont inclus car ces logements sont équipés d'une cuisine qui en fait des logements autonomes. En effet, une deuxième enquête, dite « CARE institutions » sera menée par la DREES auprès des personnes âgées de 60 ans ou plus séjournant de manière permanente dans des établissements d'hébergement pour personnes âgées. Il était donc nécessaire de disjoindre au mieux les bases de sondage de ces deux enquêtes différentes dont les résultats pourront être agrégés.

La base de sondage a été constituée à partir des sources fiscales (taxe d'habitation, impôt sur le revenu et fichier d'imposition des personnes) et du RSL (Haag O, Loonis V., Mamari C., Vinet L., 2013). Les différentes étapes de constitution ont été les suivantes :

2.1.1. - Restriction aux personnes de 60 ans et plus

En recherchant les personnes nées avant le 2 mai 1955, le nombre d'individus présents dans le champ s'élève à 17 156 678 personnes.

2.1.2. Conservation des logements foyers

La DREES a fourni deux bases d'adresses : la première contient les adresses de logements foyers, qui ne sont pas des communautés ou des EHPAD, et qu'elle souhaite donc conserver ; la deuxième contient une liste d'adresses d'EHPAD qui doivent être au contraire retirés.

Un appariement à l'adresse postale (n° de voie, indice de répétition, libellé de voie, commune) a été fait pour conserver les individus présents dans la base d'adresses de logements foyers. 29 276 individus ont ainsi été repérés à 807 adresses. La base de logements foyers contenant 2 132 adresses, le taux de recouvrement est de 37,9%.

2.1.3. Retrait des personnes vivant dans des communautés hors champ de l'enquête

- Le premier retrait concerne les maisons de retraite et EHPAD identifiables directement dans les sources fiscales. Il s'agit des individus pour lesquelles le complément de l'adresse contient les mots clés « MDR », « RETRAITE », ou « EHPA », les individus sont retirés de la base, soit ici 100 746 individus ;
- Le deuxième retrait concerne les maisons de retraite et EHPAD identifiés par la DREES. Comme pour les logements foyers, un appariement à l'adresse postale est fait, pour retirer les adresses communes. 2 801 adresses ont aussi été repérées (le fichier en contient 8 053, soit un taux d'appariement de 34,8%). A ces adresses sont rattachés 48 319 individus, qui sont donc retirés.
- Le troisième retrait concerne les foyers hors champ. Cette fois-ci on recherche les mots « ADOMA » et « SONACOTRA ». Cela fait enlever 3 738 unités supplémentaires ;
- Le quatrième retrait concerne des individus repérés comme appartenant à une communauté selon le RSL. Ils étaient au nombre de 222 005 ;

¹⁴ Soit l'ensemble des individus ayant au moins 60 ans le 2 mai 2015.

¹⁵ Établissement d'Hébergement pour Personnes Âgées Dépendantes.



- Enfin, malgré toutes les précautions prises dans les filtres précédents, il peut rester des communautés non identifiées. Ainsi si un logement¹⁶ hébergeait au moins 5 personnes âgées d'au moins 75 ans au 01/05/2015, ces individus ont été supprimés. 239 individus ont été retirés de la sorte.

2.1.4. Caractéristiques de la base finale de sondage

En incluant les logements foyers repérés aux unités restantes suite à cet écrémage, la base de sondage contient 11 690 259 unités de tirage, réparties comme suit :

- 11 424 001 logements ordinaires
- 21 079 autres logements
- 245 179 foyers fiscaux non rattachés

2.2. Tirage de l'échantillon

2.2.1. Calcul des allocations de l'échantillon principal de 140 000 logements

Comme l'échantillon principal doit servir de base de sondage pour CARE, il a été décidé de le tirer comme tout échantillon d'enquête dans l'échantillon maître (EM), à savoir en cherchant à obtenir une équipondération des logements interrogés (Faivre, 2009). Aucun minimum par ZAE n'a été défini.

L'allocation par ZAE est donc la suivante :
$$n_{ZAE} = \frac{n_{princ} * N_{ZAE}}{\pi_{ZAE} * \sum_{ZAE \in EM} \frac{N_{ZAE}}{\pi_{ZAE}}}$$

où

- n_{princ} est la taille de l'échantillon principal ;
- N_{ZAE} est le nombre de logements dans le champ de l'enquête, pour chaque ZAE ;
- π_{ZAE} est la probabilité d'inclusion de la ZAE dans l'EM ;
- $\sum_{ZAE \in EM} \frac{N_{ZAE}}{\pi_{ZAE}}$ est l'estimateur du nombre total de logements du champ de l'enquête¹⁷.

L'allocation par département s'en déduit, par sommation des allocations des ZAE du département¹⁸.

Il est à noter que 3 départements ne sont pas présents dans l'EM¹⁹ :

- Alpes-de-Haute-Provence (04)
- Lozère (48)
- Haute-Marne (52)

2.2.2. Échantillon complémentaire de 60 000 logements

Cet échantillon permet de compléter l'échantillon principal pour assurer une diffusion départementale de l'ensemble de l'échantillon VQS.

¹⁶ Dans les faits, il s'agissait soit d'un logement au sens de la TH soit d'un foyer fiscal non rattaché à un logement au sens de la TH.

¹⁷ La base de sondage compte a priori 12 162 622 logements dans le champ et cet estimateur vaut 11 958 087, soit un écart de 1,7%.

¹⁸ Il est à noter qu'une ZAE a été affectée au département de sa commune pivot. Or il existe des ZAE multi départementales. Les comptages obtenus ici diffèrent donc à la marge de ceux réellement observés sur l'échantillon final.

¹⁹ La représentativité départementale ne faisait pas partie des contraintes lors de la définition de l'EM. Le tirage des ZAE de l'EM n'a été équilibré qu'au niveau régional.



Il doit permettre d'avoir au moins 1 000 logements de l'échantillon VQS par département. Cette contrainte se traduira bien entendu in fine par une dispersion des poids plus importante que pour celle observée dans les enquêtes ménages « traditionnelles ».

Comme il ne donnera pas lieu à une interrogation en face à face, il a été décidé de le disjoindre le plus possible de l'EM.

Il n'était toutefois pas possible de le tirer totalement en dehors de l'EM car le complémentaire de l'EM ne contenant aucune très grande commune, ne peut être considéré comme représentatif du territoire français.

Ainsi, la base de sondage contient :

- l'ensemble des ZAE ayant une probabilité d'inclusion à l'EM supérieure à 0,5²⁰ ;
- les ZAE hors EM ayant une probabilité d'inclusion à l'EM strictement inférieure à 0,5.

A partir du minimum départemental (fixé à 1 000 logements), le reliquat R qu'il restait à répartir par département en fonction du minimum à tirer par département et de l'allocation départementale de l'échantillon principal a été calculé de la façon suivante :

$$R = 200\,000 - 96 * n_{dep}^{\min} - \sum_{dep} \max(0, n_{dep}^{princ} - n_{dep}^{\min})$$

où

- n_{dep}^{\min} est le minimum de logements à tirer par département (1 000 dans notre cas) ;
- n_{dep}^{princ} est le nombre de logements tirés dans l'échantillon principal pour le département.

Ce reliquat est ensuite réparti au prorata du nombre de logements dans le champ par département.

L'allocation départementale à tirer dans l'échantillon complémentaire devient alors

$$n_{dep}^{compl} = \text{arrondi} \left(\max(n_{dep}^{princ}, n_{dep}^{\min}) + R * \frac{N_{dep}}{\sum_{dep} N_{dep}} - n_{dep}^{princ} \right)$$

où N_{dep} est le nombre de logements dans le champ, par département.

On déduit ensuite l'allocation par ZAE de cette allocation départementale de la façon suivante :

$$n_{ZAE}^{compl} = \frac{n_{dep}^{compl} * N_{ZAE}}{(1 - \pi_{ZAE}) * k_{\pi_{ZAE} < 0,5} * \sum_{ZAE \in bds_{compl}} \left(\frac{N_{ZAE}}{(1 - \pi_{ZAE}) * \mathbb{I}_{\pi_{ZAE} < 0,5}} * \mathbb{I}_{ZAE \in dep} \right)}$$

où

- $\mathbb{I}_{ZAE \in dep} = \begin{cases} 1 \text{ si la ZAE est dans le département "dep"} \\ 0 \text{ sinon} \end{cases}$
- $k_{\pi_{ZAE} < 0,5} = \begin{cases} 1 \text{ si } \pi_{ZAE} < 0,5 \\ \frac{1}{1 - \pi_{ZAE}} \text{ sinon} \end{cases}$

²⁰ Il s'agit des grandes communes. Ce seuil de 0,5 a été retenu afin d'avoir une bonne représentativité de la base de sondage tout en évitant de tirer trop de FA dans les grandes communes. En effet, ce nombre est proportionnel

à $\frac{1}{1 - \pi_{ZAE}}$.



2.3. Calcul des poids

Deux jeux de poids sont calculés :

- Un pour l'échantillon principal seul. Ce poids sera utile pour le calcul du poids de sondage des individus tirés dans CARE ;
- Un sur l'ensemble de l'échantillon VQS (principal + complémentaire) qui sera utile pour l'exploitation de l'enquête VQS.

2.3.1. - Poids à utiliser pour CARE

Il s'agit du poids traditionnel d'un échantillon de logements en 2 phases tiré dans l'EM.

On note $\omega_{\log ech p}^{ZAE}$ le poids d'un logement échantillonné dans l'échantillon principal dans une ZAE. Ce poids est égal à

$$\omega_{\log ech p}^{ZAE} = \frac{N_{ZAE}}{\pi_{ZAE} * n_{ZAE}^{princ}}$$

où

- N_{ZAE} est le nombre de logements du fichier TH dans le champ, pour la ZAE ;
- π_{ZAE} est la probabilité d'inclusion de la ZAE dans l'EM ;
- n_{ZAE}^{princ} est le nombre de logements échantillonnés dans l'échantillon principal pour la ZAE.

2.3.2. Poids à utiliser pour VQS

L'échantillon complémentaire n'est pas tiré dans une base totalement disjointe de celle de l'échantillon principal. Il est donc nécessaire de procéder à un partage des poids pour le calcul du poids VQS définitif.

Le poids d'un logement de l'échantillon complémentaire, noté, $\omega_{\log ech c}^{ZAE}$, vaut :

- $\omega_{\log ech c}^{ZAE} = \frac{N_{ZAE}}{n_{ZAE}^{compl}}$ pour la strate des ZAE telles que $\pi_{ZAE} > 0,5$;
- $\omega_{\log ech c}^{ZAE} = \frac{N_{ZAE}}{(1 - \pi_{ZAE}) * n_{ZAE}^{compl}}$ pour les autres.

La technique du partage des poids permet ensuite de réunir ces deux échantillons (principal et complémentaire) et pondérer l'ensemble obtenu pour pouvoir inférer sans biais à la population France métropolitaine (ou sur toutes sous-populations, comme par exemple pour des études départementales).

Le principe est le suivant : $\omega_{\log ech tot}^{ZAE} = \alpha * \omega_{\log ech p}^{ZAE} + (1 - \alpha) * \omega_{\log ech c}^{ZAE}$

Comme la variance des estimations est inversement proportionnelle à la taille de l'échantillon, le coefficient α optimum introduit dans le partage des poids peut être approximé par le ratio suivant : nombre de logements via l'EM (soit 140 000) divisé par le nombre de logements total de l'échantillon (soit 200 000).



3. - Principe du calcul anticipé des allocations pour chester21

Dans le cadre du NCEE, il a été nécessaire de fournir pour le mois de septembre 2014 les allocations par ZAE des enquêtes qui auront lieu en 2015, afin de planifier l'activité des enquêteurs pour l'année à venir.

L'échantillon d'individus CARE sera collecté à partir de mai 2015. Comme présenté ci-dessus, cet échantillon de 15 000 individus est particulier car il a été tiré parmi les ménages répondant à l'enquête VQS. La « base de tirage » de l'enquête n'a donc été connue qu'en début d'année 2015. Et, contrairement aux enquêtes plus traditionnelles, on ne dispose pas au moment du calcul des allocations d'une base de sondage proche de celle attendue.

Cette partie a pour objet de présenter la méthodologie et les hypothèses faites pour la mise en œuvre de ces calculs d'allocations.

3.1. Les allocations à livrer à Chester

La base de sondage de CARE constituée des personnes répondantes à VQS âgées de 60 ans ou plus au 1^{er} mai 2015 était disponible à la fin de l'enquête VQS en février 2015. Or, pour planifier le travail des enquêteurs, deux allocations différentes au niveau ZAE ont dû être calculées en septembre 2014. En effet, 2 temps de passation du questionnaire en face-à-face différents selon le degré de dépendance des individus ont été estimés à partir des durées d'interrogation du test CAPI pour le calcul de la charge de travail des enquêteurs. Ainsi, les allocations suivantes ont été fournies :

- une allocation des individus des groupes VQS I et II (les personnes les plus autonomes et pour lesquelles le temps d'interrogation provisionné dans chester est moins important) ;
- une allocation des individus des groupes VQS III et IV (les personnes les plus dépendantes).

3.2. Choix de la base de sondage pour les calculs

Deux solutions étaient envisageables :

- partir de l'échantillon VQS, simuler la non-réponse pour obtenir une base de ménages répondants, sélectionner un individu par ménage répondant et calculer les allocations sur cette base en cherchant à minimiser la dispersion des poids des individus par strate ;
- partir du champ complet de l'enquête CARE issu des fichiers fiscaux²², sélectionner un individu par ménage comportant plusieurs individus dans le champ et calculer les allocations sur cette base en cherchant à minimiser la dispersion des poids des individus par strate.

C'est finalement cette deuxième solution qui a été mise en œuvre. Elle permet de limiter le nombre de paramètres à estimer dans nos modèles de calcul des allocations car elle ne nécessite pas de connaître les taux de réponse à VQS. En effet, on connaît mal les déterminants de la non-réponse à VQS et dans ces conditions il devient difficile d'estimer une probabilité de réponse fiable par ménage.

3.3. Les hypothèses mises en œuvre

On ne dispose pas du groupe d'autonomie des individus dans la base de sondage complète de CARE issue des sources fiscales (cette information n'est en effet disponible qu'à l'issue de la collecte de VQS). Les seules données dont on dispose sont les résultats des enquêtes VQS/HSM de 2007-2008, qui fournit au niveau national une évaluation de la répartition de la population par groupe VQS et par tranches d'âge (60-69 ans, 70-79 ans et 80 ans ou plus). On suppose ici que ces proportions sont toujours valables en 2015. Ces proportions sont les suivantes :

²¹ Outil de planification pour la répartition des charges entre les enquêteurs

²² Il s'agit des individus de 60 ans ou plus au 1er mai 2015.



Tableau 10. Répartition des individus selon leur âge et leur groupe VQS

Groupes VQS \ âge	60-69	70-79	80+	Total
groupe I (autonomes)	76%	58%	30%	59%
groupe II	18%	25%	38%	25%
groupe III	4%	10%	21%	10%
groupe IV (dépendants)	2%	8%	11%	6%

En outre, pour la sélection d'un individu par ménage, comme on ne connaît pas a priori le résultat du processus de sélection de l'individu CARE au sein des ménages répondants à VQS, on prend comme référence l'âge de la personne la plus âgée dans le ménage, de façon à avoir la prévision la plus prudente possible en termes de temps d'enquête²³. En effet, comme le montre le tableau ci-dessus, la proportion de personnes très dépendantes augmente avec l'âge.

Ainsi, le nombre de personnes dans le champ pour chaque ZAE*strate est obtenu en appliquant la répartition des groupes VQS par tranche d'âge (du tableau 10 ci-dessus) au nombre d'individus²⁴ par tranche d'âge de la ZAE.

3.4. Le calcul anticipé des quotités par ZAE pour la planification de l'activité 2015

Pour un groupe VQS et une tranche d'âge donnés, l'allocation optimale par ZAE est calculée à partir de l'allocation totale à tirer pour cette strate et du nombre de personnes « potentiellement interrogeables » dans ce groupe VQS estimé dans chaque ZAE. L'objectif de l'échantillonnage est de minimiser la dispersion des poids par ZAE pour une strate donnée.

Ainsi, la formule de calcul de l'allocation par ZAE est la suivante :

$$n_h^{ZAE} = \frac{N_h^{ZAE} * n_h}{N_h * \pi_{ZAE}}$$

avec - h le groupe VQS ;

- π_{ZAE} la probabilité d'inclusion de la ZAE dans l'échantillon maître ;

- n_h le nombre d'individus à tirer dans la strate « h ». Ce nombre a été fourni par la DREES ;

- N_h^{ZAE} le nombre d'individus (i.e. uniquement le plus âgé du logement) de la base de sondage de la ZAE appartenant à la strate « h » ;

$$- N_h = \sum_{ZAE} N_h^{ZAE} .$$

Afin d'éviter tout biais, une borne minimum d'un individu par strate est nécessaire à partir du moment où la strate contient au moins un individu au départ. Ceci permet en effet d'éviter que des individus aient une probabilité nulle d'être dans l'échantillon.

Tableau 11 : Répartition de l'échantillon alloué à partir des spécifications de la DREES.

	60-69	70-79	80+	Total
groupe I	1 639	809	353	2 802
groupe II	1 120	982	1 275	3 378
groupe III	836	1 387	2 533	4 755
groupe IV	728	1 498	1 839	4 065
Total	4 323	4 677	6 000	15 000

²³ Cette décision a été prise afin de se placer dans la situation la plus prudente vis à vis du calcul des quotités dans chester vu que le temps d'entretien des individus les plus vieux sera plus long. L'objectif était d'éviter autant que possible de se trouver en sous-quotité dans une ZAE où le hasard du tirage Kish ferait qu'on sélectionne souvent la personne la plus dépendante.

²⁴ Seul l'individu de 60 ans ou plus le plus âgé d'un ménage est retenu pour ces calculs.



Enfin, pour le calcul de la charge de travail par enquêteur, deux temps de remplissage du questionnaire sont prévus selon le niveau de dépendance des personnes (personnes des groupes VQS I et II versus III et IV). Ainsi, les allocations fournies par ZAE seront les suivantes

$$n_1^{ZAE} = \sum_{gvqs=1}^2 n_{gvqs}^{ZAE} \quad \text{et} \quad n_2^{ZAE} = \sum_{gvqs=3}^4 n_{gvqs}^{ZAE}$$

En fournissant ces allocations, on fixe donc le temps de traitement de CARE à allouer par ZAE. Ce temps vaudra :

$$T_{ZAE}^{CARE} = T^{rep} * (n_1^{ZAE} + n_2^{ZAE}) + n_1^{ZAE} * T_1 + n_2^{ZAE} * T_2$$

où T^{rep} est le temps de repérage et de prise de contact.
 T_1 est le temps de passation des questionnaires des personnes fortement dépendantes (y compris le temps de passation des aidants en face-à-face).
 T_2 est le temps de passation des questionnaires des personnes moins dépendantes (y compris le temps de passation des aidants en face-à-face)

Pour l'allocation finale, cette contrainte de temps total par ZAE sera respectée. Sous cette contrainte, et au vu des répondants à VQS, les allocations n_1^{ZAE} et n_2^{ZAE} pourront être modifiées par rapport au calcul prévisionnel actuel pour limiter au maximum la dispersion des poids, tout en respectant bien entendu en parallèle les allocations par strate fournies par la DREES.

4. - Tirage de l'échantillon CARE

4.1. Le principe général

L'objectif recherché est toujours d'essayer autant que faire se peut d'aboutir in fine à une équipondération des individus. La difficulté ici est en plus de respecter les deux contraintes suivantes :

- le nombre d'individus par groupe VQS fourni par la DREES
- le temps total alloué dans chester par ZAE. En revanche, le respect stricto sensu des allocations n_1^{ZAE} et n_2^{ZAE} calculées de façon anticipée pour la prévision des quotités (cf. partie III) n'était pas nécessaire. Dans les faits, ces allocations plus fines ont été respectées quand ceci était possible.

Dans le cas de CARE, le poids d'un individu dépend :

- du poids de sondage de son logement dans l'échantillon VQS ;
- du poids de sondage issu de la sélection de l'individu parmi les répondants de VQS.

Ainsi, afin de se rapprocher de l'équipondération des individus tirés, l'idée est de sélectionner l'individu dans CARE par le biais d'un tirage à probabilité inégale.

L'objectif d'équipondération s'écrit directement au niveau individu, en introduisant les notations suivantes : π_{BS}^i qui est la probabilité que l'individu i soit présent dans la base de sondage CARE, et qui est le produit de la probabilité que le logement soit tiré et réponde à VQS par la probabilité que l'individu i soit sélectionné dans son logement ; et $\pi_{CARE/BS}^i$ qui est la probabilité que l'individu soit tiré pour CARE sachant qu'il est dans la base de sondage. Dans ce cas



$$\omega^i = \frac{1}{\pi_{BS}^i} \cdot \frac{1}{\pi_{CARE/BS}^i} = \frac{N_{groupe_j}}{n_{groupe_j}}$$

D'où on en déduit que :

$$\pi_{CARE/BS}^i = \frac{\hat{\omega}_i * n_{groupe_j * ZAE}}{\sum_{i \in ZAE} \hat{\omega}_i * \Pi_{i \in groupe_j}}$$

Où finalement

- $\hat{\omega}_i$ est le poids de l'individu i répondant à VQS et qui a été sélectionné dans CARE. C'est-à-dire le poids du logement de VQS après traitement de la non-réponse que multiplie le nombre d'individus du champ de CARE de ce logement ;

- $n_{groupe_j * ZAE}$ est l'allocation par strate calculée dans la partie précédente.

On a donc bien in fine un tirage proportionnel au poids des individus répondants à VQS.

Les différentes étapes de ce tirage et notamment l'estimation de $\hat{\omega}_i$ sont décrites en détail dans les points suivants.

4.2. L'estimation du poids des logements répondants à VQS

Afin de disposer d'un poids des logements répondants à l'enquête VQS, poids qui sera utilisé pour le tirage à probabilité inégale décrit ci-dessous, la non-réponse dans VQS a été traitée par repondération selon la méthode des Groupes de Réponse Homogène (GRH) (Caron N., 2005).

Une première régression logistique a permis de déterminer les variables les plus explicatives de la non-réponse parmi les variables auxiliaires disponibles²⁵. Six variables ont finalement été retenues :

- la nature du logement ;
- le revenu total du foyer (en tranche) ;
- l'âge de la personne la plus âgée ;
- la tranche d'unité urbaine ;
- la région ;
- le type d'occupation.

Ce modèle a ensuite permis d'estimer une probabilité de réponse de chaque logement de la base de sondage. Cette dernière a ensuite été triée selon cette probabilité de réponse. Les 10 groupes de réponse homogènes (GRH) sont ensuite constitués en regroupant les 1 400 premiers logements de la base triée dans le 1^{er} GRH, les 1 400 suivants dans le 2^e etc.

Tableau 12. Médiane, moyenne et écart-type de la probabilité de réponse estimée pour chacun des 10 groupes homogènes de réponse.

GRH	moyenne	médiane	écart-type
1	0,3759	0,3913	0,0548
2	0,4681	0,4689	0,0165
3	0,5151	0,5150	0,0122
4	0,5520	0,5517	0,0090
5	0,5850	0,5855	0,0089
6	0,6129	0,6116	0,0083
7	0,6407	0,6413	0,0073
8	0,6665	0,6657	0,0086
9	0,6946	0,6945	0,0080
10	0,7357	0,7313	0,0186

²⁵ Il s'agit de diverses variables issues de la base de sondage.



Une fois ces groupes homogènes de réponse constitués, les probabilités de réponse sont supposées indépendantes d'un groupe à l'autre et **constantes** au sein de chaque groupe²⁶. La probabilité de réponse $\hat{\pi}_i^{rep}$ d'un logement i appartenant à un groupe h donné est alors estimée à partir des poids de sondage initiaux $\omega_{log\ ech\ p}^{ZAE}$ (cf. Point 2.3 de cette note), par :

$$\hat{\pi}_i^{rep} = \hat{\pi}_h^{rep} = \frac{\sum_{i=1}^{140000} (\omega_{log\ ech\ p}^{ZAE} * \mathbb{I}_{i \in ZAE} * \mathbb{I}_{i \in h} * \mathbb{I}_{i \in \text{répondant}})}{\sum_{i=1}^{140000} (\omega_{log\ ech\ p}^{ZAE} * \mathbb{I}_{i \in ZAE} * \mathbb{I}_{i \in h})}$$

avec $\mathbb{I}_{i \in ZAE} = \begin{cases} 1 & \text{si } i \in ZAE \\ 0 & \text{sin on} \end{cases}$

$$\mathbb{I}_{i \in h} = \begin{cases} 1 & \text{si } i \in \text{strate "h"} \\ 0 & \text{sin on} \end{cases}$$

$$\mathbb{I}_{i \in \text{répondant}} = \begin{cases} 1 & \text{si } i \text{ a répondu à VQS} \\ 0 & \text{sin on} \end{cases}$$

À l'issue de cette étape, les poids des logements répondants ont été ajustés de façon à prendre en compte le mécanisme de non-réponse. La pondération d'une unité i donnée suite à cette correction de la non-réponse totale (CNRT) vaut donc :

$$\hat{\omega}_i^{VQS} = \frac{\omega_{log\ ech\ p}^{ZAE} * \mathbb{I}_{i \in ZAE}}{\hat{\pi}_i^{rep}} * \mathbb{I}_{i \in \text{répondant}}$$

4.3. Le poids des individus de la base de sondage CARE

Une des contraintes était de ne pas interroger pour CARE plus d'une personne dans un même logement. En effet, le protocole CARE prévoit d'interroger en plus d'une personne de 60 ans ou plus du logement tous ses aidants de l'entourage, dont certains peuvent résider dans le même logement. Interroger plusieurs personnes d'un même logement aurait pu conduire à des doubles interrogations de personnes d'une part et à un temps de collecte extrêmement long d'autre part.

Pour respecter cette contrainte, la première étape du tirage de l'échantillon a été de sélectionner, à probabilité égale, dans chaque logement un individu parmi les individus répondants à VQS et faisant partie du champ de l'enquête CARE. On obtient alors une base d'individu et non plus de logements.

- À ce niveau on peut considérer que la probabilité de tirage d'un individu à l'issue de cette première phase du tirage est : $\pi_i^{log} = \frac{1}{\sum_{j \in \text{log}} (\mathbb{I}_{j \in \text{rep_vqs}} * \mathbb{I}_{j \in \text{champ_care}})}$

Où $\mathbb{I}_{j \in \text{rep_vqs}} = \begin{cases} 1 & \text{si l'individu } j \text{ du ménage a répondu à VQS} \\ 0 & \text{sin on} \end{cases}$

²⁶ De façon à limiter la sensibilité des estimations au modèle de non-réponse.



$$\mathbb{1}_{j \in \text{champ_care}} = \begin{cases} 1 & \text{si l'individu } j \text{ du ménage appartient au champ de CARE} \\ 0 & \text{sinon} \end{cases}$$

On en déduit donc le poids des individus de la base de sondage de CARE qui sont obtenus en faisant le produit du poids estimé du logement répondant à VQS par le nombre d'individus du champ de CARE dans ce logement. En reprenant les notations précédentes, on obtient :

$$\hat{\omega}_i = \frac{\hat{\omega}_{\log}^{\text{VQS}}}{\pi_i^{\log}}$$

Afin de se rapprocher de l'équipondération des logements, le tirage de CARE se fera donc proportionnellement à ce poids dans chaque strate de tirage.

4.4. Les modifications à la marge des allocations théoriques

Comme, on l'a vu au chapitre 3, l'allocation de l'échantillon a été calculée à partir d'une base de sondage qui n'est pas celle dans laquelle est réalisé le tirage et ce afin de répondre aux contraintes du NCEE.

Cette contrainte a deux conséquences importantes et plutôt négatives quant à l'objectif d'équipondération :

- l'allocation à respecter n'est plus optimale par rapport à celle qui aurait été calculée avec la base de sondage définitive ;
- il existe des strates pour lesquelles la taille de l'échantillon à tirer est supérieure au nombre d'individus dans la strate.

Si le premier problème n'a pas été corrigé²⁷, le second, lui, l'a été de la façon suivante.

Sur les 2 268 strates de tirage (567 ZAE * 4 groupes d'autonomie), 226 étaient sujettes à ce problème. Ce manque d'individus dans la base de sondage ne concernait que des strates d'individus dépendants (groupes VQS III et IV).

Ces strates peuvent être classées en deux catégories :

- les ZAE pour lesquelles un seul groupe VQS était déficitaire et pour lesquelles le groupe équivalent au sens chester²⁸ permettait de combler ce déficit (136 cas). Ces ZAE seront appelées « cas simple » par la suite ;
- les autres ZAE (68 ZAE dont 22 qui avaient un déficit pour les 2 groupes VQS). Ces ZAE seront appelées « cas complexes ».

4.4.1. - Les ZAE « cas simples »

Pour résoudre le problème de ces ZAE, l'idée était :

- de tirer dans l'échantillon l'ensemble des individus du groupe VQS à problème
- d'augmenter l'allocation du groupe VQS équivalent au sens chester de la ZAE afin de respecter la contrainte chester
- de sélectionner une ZAE dite partenaire, « sans problème » a priori, dans le même département si possible sinon dans la même région et de faire le traitement inverse (i.e. augmenter l'allocation du groupe VQS à problème de la ZAE à problème du déficit observé dans la première ZAE et de réduire d'autant l'allocation du groupe chester équivalent). Ce traitement permet de conserver l'allocation par groupe d'autonomie « France entière » imposé par la DREES. Était choisie comme ZAE partenaire celle ayant

²⁷ Il était en effet très difficile de converger vers une allocation optimale qui permette à la fois de respecter les contraintes chester par ZAE et la taille de l'échantillon par groupe d'autonomie voulue par la DREES.

²⁸ Il s'agissait du groupe III (respectivement IV) si le groupe IV (respectivement III) était déficitaire



l'allocation la plus forte pour le groupe VQS à problème. Cette ZAE est choisie de façon à réduire les poids extrêmes. Cependant dans certains cas où le déficit était trop grand (au-delà de 3 pour une strate donnée), plusieurs ZAE étaient sélectionnées afin de ne pas déséquilibrer trop fortement l'une d'elles et entraîner une forte augmentation des poids. La répartition des allocations se faisait alors équitablement entre les ZAE de soutien.

Exemple :

ZAE à problème dans un cas simple

	Groupe III	Groupe IV	Total
Nombre d'individus dans la base de sondage	5	10	15
Allocation chester initiale	7	6	13
Allocation tirée	5 (-2)	8 (+2)	13

ZAE « partenaire » sans problème a priori

	Groupe III	Groupe IV	Total
Nombre d'individus dans la base de sondage	15	17	32
Allocation chester initiale	9	8	17
Allocation tirée	11 (+2)	6 (-2)	17

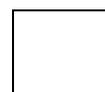
Pour les deux ZAE

	Groupe III	Groupe IV	Total
Allocation chester initiale	16	14	30
Allocation tirée	16	14	30

4.4.2. - Les ZAE « cas complexes »

Pour résoudre le problème de ces ZAE, l'idée était :

- de tirer dans l'échantillon l'ensemble des individus des deux groupes d'autonomie III et IV ;
- d'augmenter l'allocation des groupes I et II dans la ZAE afin de respecter la charge de travail totale pour la ZAE. Sachant que pour Chester, en termes de temps d'enquêtes, 2 questionnaires des groupes III et IV sont équivalents à 3 questionnaires des groupes I et II, c'est cette clé de répartition qui a été mise en œuvre pour définir le nombre de questionnaires à rajouter dans les groupes I et II. Entre les groupes I et II, le choix du nombre d'unités à rajouter s'est fait au prorata du poids de ces groupes VQS dans la base de sondage.
- de sélectionner une ZAE dite partenaire, « sans problème » a priori, dans le même département si possible, sinon dans la même région, et de faire le traitement inverse (i.e. augmenter l'allocation des groupes d'autonomie à problème du déficit observé dans la première ZAE et réduire l'allocation des groupes I et II afin de conserver la charge de travail prédéfinie). Ce traitement permet de conserver l'allocation par groupe VQS « France entière » imposé par la DREES. Était choisi comme ZAE partenaire celle ayant la somme des marges la plus grande entre les allocations pour les groupes à problème et le nombre de répondants dans les strates correspondantes. Cependant dans certains cas où le déficit était trop grand (au-delà de 3 pour une strate donnée), plusieurs ZAE étaient sélectionnées afin de ne pas déséquilibrer trop fortement l'une d'elles et entraîner une explosion au niveau des poids. La répartition des allocations se faisait alors équitablement entre les ZAE de soutien. Enfin, la réduction du nombre d'individu dans les groupes 1 et 2 pour les ZAE partenaires a également été contrôlée afin de ne pas trop augmenter leur poids.



Exemple :

ZAE à problème dans un cas complexe

	Groupe I	Groupe II	Groupe III	Groupe IV	Total
Nombre d'individus dans la base de sondage	53	74	3	3	133
Allocation chester initiale	3	3	5	4	15
Allocation tirée	5 (+2)	6 (+3)	3 (-2)	3 (-1)	17

ZAE « partenaire » sans problème a priori

	Groupe I	Groupe II	Groupe III	Groupe IV	Total
Nombre d'individus dans la base de sondage	41	35	11	8	95
Allocation chester initiale	4	5	6	5	20
Allocation tirée	2 (-2)	2 (-3)	8 (+2)	6 (+1)	18

Pour les deux ZAE

	Groupe I	Groupe II	Groupe III	Groupe IV	Total
Allocation chester initiale	7	8	11	9	35
Allocation tirée	7	8	11	9	35

4.5. Le tirage proprement dit

Le tirage s'est ensuite fait à probabilité inégale au sein de chaque strate.

La probabilité théorique de sélection d'un individu dans la strate ZAE*groupe VQS j était la suivante :

$$\pi_{CARE/BS}^i = \frac{\hat{\omega}_i * n_{\text{groupe}_j * ZAE}}{\sum_{i \in ZAE} \hat{\omega}_i * \mathbb{I}_{i \in \text{groupe}_j}} \quad (\text{cf. paragraphe 4.1 et 4.3 pour les notations}).$$

Toutefois, dans certains cas, cette probabilité dépassait 1. Dans ces cas, l'individu était d'office dans l'échantillon avec une probabilité de tirage égale à 1. Il y avait 1 096 individus dans ce cas.

De même, comme on l'a vu au paragraphe précédent, certaines strates étaient sélectionnées de façon exhaustive. Dans ces cas aussi, l'individu était d'office dans l'échantillon avec une probabilité de tirage égale à 1. Il y avait 3 328 individus dans ce cas.

Au final, la probabilité de tirage était donc de

$$\begin{cases} 1 \text{ pour les individus de l'exhaustif} \\ \frac{\hat{\omega}_i * n_{\text{groupe}_j * ZAE \setminus exh}}{\sum_{i \in ZAE \setminus exh} \hat{\omega}_i * \mathbb{I}_{i \in \text{groupe}_j}} \end{cases}$$



4.6. Les résultats en termes de dispersion des poids

Tableau 13. Distribution des poids de sondage CARE par groupe VQS

Groupe VQS	Taille de l'échantillon	Fractiles de la distribution des poids								
		minimum	5 ^e percentile	1 ^{er} décile	1 ^{er} quartile	médiane	3 ^e quartile	9 ^e décile	95 ^e percentile	Maximum
I	2804	1063	2350	2646	3060	3550	4062	4653	5202	9907
II	3377	295	557	647	839	1005	1193	1397	1639	7079
III	4323	99	148	168	219	297	381	476	548	2135
IV	4496	99	137	151	180	234	305	376	413	2583

Comme on peut le constater l'équipondération des individus CARE sélectionnés par groupe VQS n'est pas totalement atteinte. Toutefois la distribution des poids obtenue est proche de celle d'une enquête ménage traditionnelle tirée dans l'échantillon maître par le biais de l'outil Octopusse. Tous les traitements mis en œuvre pour se rapprocher de l'équipondération ont donc été importants car sans eux la dispersion finale aurait été bien plus importante.

La principale raison qui explique le non-respect l'équipondération in fine vient bien entendu de la différence entre la base de sondage théorique utilisée pour le calcul des allocations et la base de sondage définitive issue de la collecte de VQS. Ainsi, le calcul des allocations par strate de tirage (ZAE*groupe VQS) sur la base de sondage définitive, est souvent différente de celles estimées pour Chester. Dans 63% des cas l'écart par strate de tirage entre les 2 allocations est certes inférieur à 1 individu mais pour 5% des strates cet écart dépasse les 5 individus. Ceci a bien entendu un impact direct sur les poids finaux. En outre, les modifications d'allocations nécessaires pour respecter les contraintes Chester et décrites au point 4.4 ci-dessus ne font qu'amplifier ces écarts de poids entre ZAE.



Bibliographie

- [1] Bérardier M., « Les bénéficiaires de l'allocation personnalisée d'autonomie à domicile et leurs ressources en 2011 », *Études et Résultats*, n°876, Drees, février 2014.
- [2] Bouvier G., « L'enquête Handicap-Santé : présentation générale », *Document de travail*, N° F1109, INSEE, 61 p. (RAP), 2011.
- [3] Bouvier G., « Les faux négatifs du volet ménage de l'enquête Handicap-Santé 2008 », *Actes des Journées de Méthodologies Statistiques*, INSEE, 2012.
- [4] Groupe de travail présidé par Charpin J.M. « Perspectives démographiques et financières de la dépendance », 2011.
- [5] Debout C. « Durée de perception de l'Allocation personnalisée d'autonomie (APA) », *Dossier de travail, Série Sources et Méthodes*, n°15, Drees, juillet 2010.
- [6] Debout C. « La durée de perception de l'APA : 4 ans en moyenne », *Études et résultats*, n° 724, Drees, avril 2010.
- [7] Eghbal-Téhéran S. et Makdessi Y., « Les estimations GIR dans les enquêtes Handicap-Santé 2008-2009 », *Dossier de travail, Série Sources et Méthodes*, n°26, Drees, septembre 2011.
- [8] Lecroart A., Froment O., Marbot C. et Roy D., « Projection des populations âgées dépendantes : deux méthodes d'estimation », *Dossier Solidarité Santé*, n°43, Drees, septembre 2013.
- [9] Le Guennec J. « Estimations locales du handicap dans l'enquête Handicap-Santé 2008 », *Document de travail*, INSEE, H2012/02, 200p. (RAP), 2012.
- [10] Midy L., « Présentation de l'enquête « Vie quotidienne et Santé » de 2007 », *Document de travail*, INSEE, F1001, 55 p. (RAP), 2010.
- [11] Haag O., Loonis V., Mamari C., Vinet L., « La source fiscale : une base de sondage qui s'impose pour les enquêtes auprès des logements et des individus », Recueil du Symposium de Statistique Canada « Produire des estimations fiables à partir de bases imparfaites », octobre 2013.
- [12] Faivre S., « Le projet OCTOPUSSE de nouvel Échantillon-Maître de l'Insee », actes des journées de méthodologie statistique 23, 24 et 25 mars 2009.
- [13] Caron N., « La correction de la non-réponse par repondération et par imputation », chapitre IV du document de travail M0502, 2005



Annexe

Variables VQS ²⁹	Ancien score	Nouveau score	
		< 80 ans	≥ 80 ans
Aide de l'entourage	6/3	6	6
Aide d'un professionnel		6	6
Aide technique	1	1	1
Aménagement du logement	2	0	1
Difficultés à attraper un objet	3/1	3/1	1
Difficultés pour entendre	3/1	0	0
Difficultés pour comprendre	3/1	3/1	3/1
Difficultés pour se concentrer	3/1	3/1	6/3
État général de santé	6/3/1	3/1	3/1
Reconnaissance officielle d'un handicap	6	0	0
Impossibilité à faire au moins une activité	6	3	1
Difficultés à prendre des initiatives	3/1	0	0
Limitation depuis plus de 6 mois	6/3	6/3	6/3
Difficultés pour se servir de ses mains	3/1	0	3/1
Maladie chronique	1	0	0
Difficultés pour marcher	3/1	1	3/1
Difficultés pour se souvenir de certaines choses	3/1	0	6/3
Difficultés pour parler	3/1	0	0
Difficultés à résoudre les problèmes quotidiens	3/1	12/6	9/4
Difficultés pour se pencher	3/1	3/1	1
Reconnaissance officielle d'un handicap	6	0	0
Difficultés pour sortir du logement	3/1	6/3	6/3
Difficultés pour voir de loins	3/1	0	0
Difficultés pour voir de près	3/1	0	0
Difficultés pour se laver	0	6/3/1	6/3/1
Bénéfice de l'APA	0	6	6

²⁹ Les correspondances entre les variables et les questions VQS sont données dans l'annexe 4.

