

# Chapitre 1

## Description du plan de sondage de l'enquête HID

Le plan de sondage mis en œuvre par l'INSEE pour réaliser l'enquête HID ( Handicap- Incapacité-Dépendance) est un plan très complexe. En effet, le sondage a été effectué en deux temps :

- D'abord une enquête de filtrage VQS ( Vie Quotidienne et Santé ) selon un mode de tirage stratifié, à deux degrés et aréolaire.
- Puis l'enquête HID elle-même auprès d'un sous- échantillon des répondants à VQS, selon un mode de tirage stratifié et à allocation non proportionnelle.

Il s'agit donc d'un tirage en deux phases avec post - stratification.

### I.Le plan de sondage VQS.

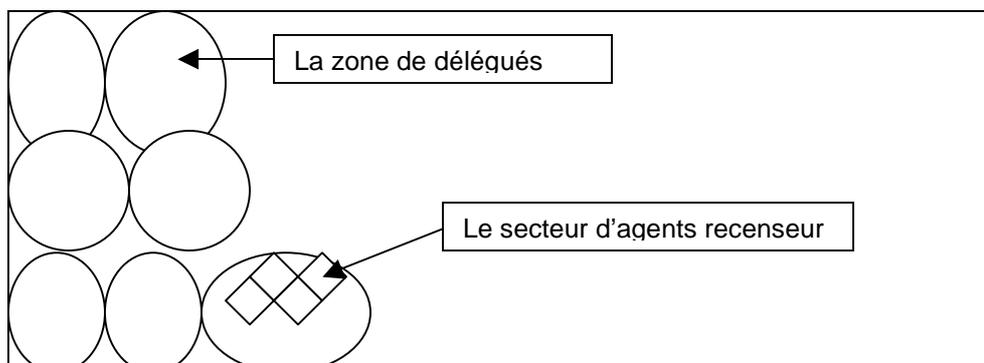
L'enquête VQS devait à l'origine être effectuée auprès d'un échantillon de 123 000 ménages ( tous les individus du ménage seraient interrogés). Certains départements ayant demandé des estimations locales, on a effectué des extensions d'échantillon sur ces départements, avec au total un échantillon de 180 000 ménages, ce qui représentait 443 000 individus.

La population française a été découpée en strates géographiques, de façon à permettre une bonne représentativité pour les grandes et les moyennes régions. Le tirage de l'échantillon des individus ( plus précisément des ménages ) pour le sondage VQS a été réalisé en deux étapes à l'intérieur de chaque strate :

- dans un premier temps, on a effectué le tirage de regroupements de districts appelés *zones de délégués* ;
- puis dans un deuxième temps, on a tiré *des secteurs d'agents recenseurs* à l'intérieur de chaque zone de délégués sélectionnés ;
- enfin, l'agent recenseur enquête tous les habitants de son secteur.

Le tirage pour une autre enquête a été en quelque sorte associé à celui du sondage VQS : l'Etude de l'Histoire Familiale ( EHF ). En fait, le tirage des zones de délégués est commun aux deux enquêtes, mais comme l'échantillon EHF était plus grand, on a tiré plus d'agents recenseurs pour EHF dans les zones de délégués sélectionnées.

La strate géographique



## **I.1.La stratification.**

Le découpage du territoire métropolitain en strates s'est effectué selon divers aspects.

Au début, une strate représentait une région administrative.

Mais pour tenir compte des extensions demandées par certains départements, le but était de tirer une plus grande proportion d'individus dans ces départements; il a donc fallu séparer le département en question du reste de sa région ; il constituait à lui seul une strate, les autres départements de la région constituant une deuxième strate.

Ensuite, certaines fractions de département ayant demandé une extension « langues » dans le cadre de l'enquête EHF, cette partie du département a constitué une strate, le reste du département et les autres départements de la région ont constitué une deuxième strate.

Prenons le cas de la Loire ( département 42 ) qui a décidé de financer une extension d'échantillon pour le sondage VQS. Une première strate est constituée par ce département et une deuxième strate est constituée par les départements 01, 07, 26, 38, 69, 73 et 74.

Ou encore le département des Pyrénées Atlantiques ( département 64 ) qui a décidé de financer une extension « langues ». La partie du département parlant le basque a constitué une strate et le reste du département 64 ainsi que les départements 47, 40, 33 et 24 ont constitué une deuxième strate.

En ce qui concerne l'enquête VQS, 7 départements et une région ont demandé une extension :

- région 23 ( Haute-Normandie ) avec les départements 27 ( Eure ) et 76 ( Seine-Maritime ),
- dép. 77 ( Seine-et-Marne ) rattaché à la région de collecte 21 ( Champagne-Ardenne ),
- dép. 95 ( Val-d'Oise ) rattaché à la région de collecte 23 ( Haute-Normandie ),
- dép. 62 ( Pas-de-Calais ) de la région 31 ( Nord-Pas-de-Calais ),
- dép. 35 ( Ille-et-Vilaine ) de la région 53 ( Bretagne ),
- dép. 42 ( Loire ) de la région 82 ( Rhône-Alpes ),
- dép. 34 ( Hérault ) de la région 91 ( Languedoc-Roussillon ),
- dép. 13 ( Bouches-du-Rhône ) de la région 93 ( Provence - Alpes - Côtes d'Azur ).

Les départements concernés par une extension « langue » sont :

- dép. 59 ( Nord ) de la région 31 ( Nord-Pas-de-Calais ),
- dép. 62 ( Pas-de-Calais ) de la région 31 ( Nord-Pas-de-Calais ),
- dép. 57 ( La Moselle ) de la région 41 ( Lorraine ),
- dép. 64 ( Les Pyrénées Atlantiques ) de la région 72 ( Aquitaine ),
- dép. 66 ( Les Pyrénées Orientales ) de la région 91 ( Languedoc-Roussillon ).

Le territoire métropolitain a donc été divisé en 36 strates, de taille régionale ou infra-régionale.

	Région de collecte	Départements	particularités
strate 01	11	75,78,92,93 et 94	
strate 02	21	8,10,51,52	
strate 03	21	77	extension VQS
strate 04	22	2,60,80	
strate 05	23	27	extension VQS
strate 06	23	76	extension VQS
strate 07	23	95	extension VQS
strate 08	24	18,28,36,37,41,45	
strate 09	24	91	rattaché à la région de collecte 24
strate 10	25	14,50,61	
strate 11	26	21,71,89	
strate 12	31	59	reste de la région 31
strate 13	31	59	extension langues
strate 14	31	62	extension VQS
strate 15	31	62	extension langues
strate 16	41	54,55,57	
strate 17	41	57	extension langues
strate 18	42	67,68	
strate 19	43	25,39,70	
strate 20	52	44,49,53,85	
strate 21	53	22,29,56	
strate 22	53	35	extension VQS
strate 23	54	16,17,79,86	
strate 24	72	24,33,40,47,64,	
strate 25	72	64	extension langues
strate 26	73	12,31,32,46,81,82	
strate 27	74	19,23,87	
strate 28	82	1,7,26,38	
strate 29	82	42	extension VQS
strate 30	83	3,15,63	
strate 31	91	30,48	
strate 32	91	34	extension VQS
strate 33	91	66	extension langues
strate 34	93	4,5,6,83,84	
strate 35	93	13	extension VQS
strate 36	94	20	

## I.2. Le tirage des zones de délégués ( ZD ).

Il y a au total 3553 zones de délégués, dont le nombre par région dépend bien-sûr de la taille de la région, avec un maximum de 464 pour l' Ile de France et un minimum de 31 pour la Corse.

Au cours de ce premier tirage, on a décidé de sélectionner environ 1 ZD sur 10 si la strate représentait une région ou le complément régional de départements en cas d'extension VQS, et un peu plus si la strate correspondait à un département (afin de traiter les extensions VQS ).

En fait pour pouvoir réaliser de bonnes estimations locales, le but était de tirer un échantillon d'au moins 20 000 personnes, en tous cas tirer au moins 8 ZD ; donc en moyenne, dans les départements demandant une extension VQS, on a tiré environ 1 ZD sur 4, avec un maximum pour les plus petits départements.

Pour la réalisation de ce tirage, on disposait du nombre de ZD dans chaque région et on a déterminé la quantité qu'il fallait sélectionner, c'est à dire environ un dixième. Pour chaque département, on a déterminé le nombre de ZD à tirer dans le cas où le département demanderait une estimation locale, c'est à dire un quart du nombre de ZD du département avec un minimum de 8 ZD. Il est à préciser qu'au moment où a été effectué le tirage des ZD, la liste des départements intéressés par l'extension n'était pas encore connue. Afin de pallier ce problème, on a décidé d'effectuer le tirage des zones de délégués d'abord au niveau départemental et de faire comme si tous les départements demandaient une extension. Ensuite on a regroupé les ZD ainsi sélectionnées par région et parmi elles, on a tiré les ZD qu'il fallait dans le cas où il n'y a pas d'extension. Ainsi, quand on connaîtrait la liste des départements intéressés par l'extension, il suffirait de conserver toutes les autres ZD sélectionnées au cours du premier tirage dans le département concerné par l'extension VQS.

Pour réaliser l'extension langue sur une fraction de département, deux mesures ont été prises. D'abord, étant donné que la sélection des ZD a été effectuée par région (ou par département en cas d'extension VQS), rien ne garantissait que des ZD soient tirées dans la fraction de département concernée. C'est pourquoi a-t-on rajouté des régions « fictives » constituées par les fractions de département intéressées par l'extension langue d'une même région. Par exemple, dans la région 31, la partie du département 59 et la partie du département 62 qui ont bénéficié d'une extension langue ont formé la région 30. Ces fractions de département sont devenues respectivement les départements 3A et 3B. Donc les strates 13 et 15 sont constituées respectivement des départements 3A et 3B et forment la région 30.

Au total, quatre régions fictives ont été rajoutées aux 22 régions de collecte :

- région 30 constituée des départements 3A et 3B ;
- région 40 constituée de la fraction du département 57 (région 41), elle est donc constituée de la strate 17 ; cette partie du département était appelée département 4A ;
- région 70 constituée de la fraction du département 64 (région 72), elle est donc constituée de la strate 25 ; cette partie du département était appelée département 7A ;
- région 90 constituée du département 66 (région 91), elle est donc constituée de la strate 33 .

La deuxième mesure prise pour réaliser l'extension langue était de sélectionner plus de ZD dans la fraction de département concerné. Donc dans les 4 régions fictives, on a tiré une proportion plus grande en ZD que celle de 1 ZD sur 10.

En pratique, on a utilisé le fichier des districts établi après le recensement de la population en 1990. Un district représente environ 180 habitants. Pour chaque district, on a recherché la structure socio-économique ainsi que la structure familiale. A l'intérieur de chaque strate ( donc ici le département ), on a classé les districts selon une nomenclature socio-économique.

Soient NDVQS le nombre de ZD à tirer dans la strate et Xsp90 la population de la strate au recensement de la population de 1990. On choisit de tirer deux ZD supplémentaires pour les réserves, on constitue ainsi un matelas ( de deux zones ) par strate.

L'algorithme de tirage employé est celui du tirage systématique.

On calcule le pas en nombre d'individus qui vaut :

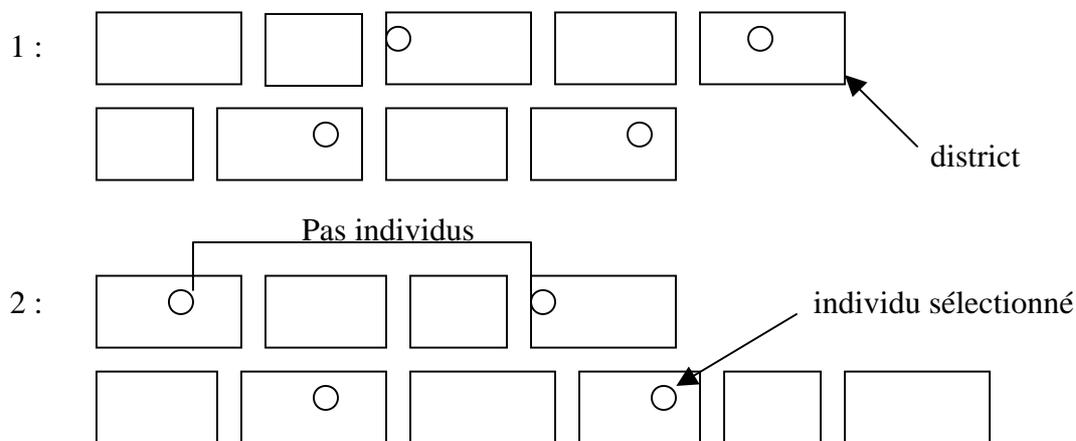
$$\text{Pas} = \text{int} (X_{sp90} / NDVQS + 2 ) \text{ où } \text{int} ( a ) \text{ désigne la partie entière du réel positif } a.$$

Le premier individu est tiré au hasard en début de fichier de la manière suivante :

on détermine un nombre au hasard entre 0 et 1 appelé aléa et son rang dans le fichier vaut : aléa \* Pas.

Partant de celui-ci, le principe consiste à descendre ensuite le long du fichier en retenant un individu tous les Pas individus.

Ensuite on a recherché dans quels districts étaient les individus tirés et ces districts ont constitués des « points d'entrée », les ZD retenues étant celles incluant ces districts.



.....

1 : signifie que les districts ont une structure socio-économique de classe 1

Enfin, on a déterminé au hasard les deux ZD qui seraient matelas parmi celles ainsi tirées.

Le fait d'avoir trié le fichier des districts a assuré une bonne représentativité de l'échantillon selon les caractéristiques sociales.

On montrera dans le chapitre suivant que les zones de délégués n'étant pas toutes de même taille, les probabilités de chacune d'être sélectionnée à l'intérieur d'une même strate diffèrent.

Pour se ramener à la strate régionale, dans laquelle il fallait tirer une ZD sur 10, on regroupe par région les « points d'entrée » tirés précédemment, en excluant les zones matelas, puis on les a trié selon une nomenclature de structure familiale.

Soient NDEHF le nombre de ZD à tirer dans la strate et Xsp90 sa population au recensement de la population de 1990. On rajoute un matelas de deux ZD par strate.

On a tiré les NDEHF zones de délégués parmi les NDVQS zones de délégués ( moins les zones matelas ) que l'on a sélectionné par le premier tirage.

La procédure de tirage est la même que précédemment, le pas de tirage en nombre d'individus étant égal à

$$\text{Pas2} = \text{int} ( X_{sp90} / NDEHF + 2 ).$$

Dans cette étape du tirage, on affecte aux districts un poids différent de celui qu'ils avaient au premier tirage. En effet, au cours du tirage départemental, on a tiré des individus dans le département, chaque district étant constitué de sa population de 1990 ; ainsi, le total des populations des districts était égal à la population du département en 1990.

Au cours du deuxième tirage, on affecte à chacun des NDVQS districts tirés dans le

département une population qui vaut :  $\frac{\text{population du département}}{\text{NDVQS}}$ , ainsi le total des populations donne la population de la région en 1990.

Et comme pour le premier tirage, le premier individu est sélectionné au hasard en début de fichier, puis on descend le long du fichier en sélectionnant un individu tous les Pas2 individus.

Les matelas sont également sélectionnés au hasard.

Une fois que les départements voulant une extension pour l'enquête VQS ont été connus, on a conservé toutes leurs ZD sélectionnées au premier tirage ; ainsi on avait la configuration d'une ZD sur 10 par strate et au moins 8 ZD en cas d'extension.

On a ainsi tiré un échantillon de 425 zones de délégués.

Le fait d'avoir trié le fichier des districts a assuré à l'échantillon une bonne représentativité selon les caractéristiques sociales et démographiques de la population française.

### **I.3.Le tirage des secteurs d'agents recenseurs ( AR ).**

Un agent recenseur est responsable en moyenne d'une zone de 530 habitants. Il est chargé de distribuer et de relever les questionnaires VQS à l'ensemble de la population vivant en domicile ordinaire dans sa zone de recensement.

Une ZD comporte en moyenne 30.8 AR .

Comme on l'a expliqué au paragraphe précédent, le tirage des zones de délégués pour l'enquête EHF est commun à celui de l'enquête VQS.

Dans une ZD sélectionnée, on a décidé de tirer 7.7 AR sur les 30.8 pour les deux enquêtes, mais chaque AR tiré ne s'occupera que d'une seule enquête.

Le nombre d'AR à sélectionner pour l'enquête VQS dépendait de la nature de la zone de délégués, à savoir s'il s'agissait :

1. d'une zone sans aucune extension ;
2. d'une zone avec extension VQS ;
3. d'une zone avec extension « langue » ;
4. d'une zone avec extension VQS et extension « langue » ( cas du département 62 ).

Dans le premier cas, sur les 7.7 AR à sélectionner, 1.6 s'occupaient de l'enquête VQS.

Dans le deuxième cas, la part VQS était plus grande afin d'avoir un échantillon plus important et surtout de taille suffisante. Et dans le troisième cas, moins de secteurs d' AR pour VQS ont été sélectionnés afin d'interroger plus de personnes dans l'enquête EHF.

On a calculé les différentes probabilités  $txv$  de tirage des secteurs d'AR ; ce calcul sera présenté au chapitre suivant.  $txv$  prend des valeurs qui vont de 0.02 à 0.17 avec bien-sûr un maximum pour les zones de délégués tirées dans un département demandant une extension VQS.

Pour chaque ZD, on connaissait le nombre exact  $N$  de secteurs d'AR au moment du tirage ; on a donc calculé le nombre  $n$  de secteurs d'AR à tirer pour l'enquête VQS en appliquant le taux de sondage  $txv$  correspondant à la nature de la zone.

Le tirage des  $n$  secteurs d'AR parmi les  $N$  de la ZD a été effectué par sondage aléatoire simple.

Toutefois, il a été décidé qu'en raison de la taille de la strate 13 ( fraction du département 59 concernée par l'extension langue ), aucun secteur d'AR ne serait tiré pour l'enquête VQS. Donc on ne prendra pas en compte les ZD de cette strate, le taux  $txv$  vaut 0 dans toutes ses ZD. Par ailleurs, à cause des effets d'arrondi lors du calcul du nombre de secteurs d'AR à tirer dans une ZD, il est arrivé qu'aucun secteur n'aient été tiré dans certaines ZD bien que le taux  $txv$  ne soit pas nul.

Au final, le nombre de ZD concernées par l'enquête VQS est de 391.

Sur le territoire métropolitain, 763 secteurs d'agents recenseurs ont ainsi été sélectionnés, 2 275 districts ont été interrogés appartenant aux 391 zones de délégués.

Les départements avec extension VQS ont enquêté beaucoup plus de secteurs que la moyenne car non seulement plus de zones de délégués faisaient partie de l'échantillon, mais plus de secteurs ont été tirés dans chacune de leurs zones. Par exemple, le département 77 avait 40 secteurs de sélectionnés alors que le département 75 en avait 18.

Tous les ménages habitant les secteurs tirés à titre de résidence principale ont donc reçu un questionnaire. L'échantillon de l'enquête VQS a ainsi été déterminé.

Dans les faits, 416 000 individus ont été concernés par le sondage VQS, parmi lesquels 359 000 ont fournis une réponse suffisamment complète pour pouvoir être exploitée pour l'enquête HID , soit un taux d'échec de 14% .

#### **I.4. Résumé.**

Le tirage systématique a assuré à chaque individu d'une strate la même probabilité de tirage. On peut donc dire que le tirage des districts s'est effectué proportionnellement à la taille de leur population en 1990. On montrera au chapitre suivant que le premier degré de tirage peut-être assimilé à un tirage de zones de délégués proportionnellement à la taille de leur population en 1990.

En résumé, la première phase du plan de sondage est un sondage stratifié à deux degrés et en grappes, le premier degré de tirage étant le tirage de zones de délégués proportionnellement à la taille de leur population en 1990, le second étant un tirage d'agents recenseurs par sondage aléatoire simple.

Les probabilités de tirage pour les deux degrés diffèrent entre les deux strates.

## **II. Le plan de sondage HID.**

A l'origine, l'enquête HID devait être réalisée auprès d'un échantillon de 20 000 personnes parmi les répondants à l'enquête VQS. L'Hérault qui avait déjà demandé une extension VQS pour son département a en plus demandé une extension d'échantillon pour l'enquête HID. C'est la raison pour laquelle, on a cherché à ce que l'échantillon compte 20 000 personnes pour l'enquête nationale plus 1 800 dans le département de l'Hérault.

Pour cette enquête, on a décidé de déterminer un échantillon qui surreprésenterait fortement les personnes les plus certainement et sévèrement atteintes par un handicap, permettant ainsi d'en décrire les situations avec suffisamment de précision. L'analyse des résultats de VQS a permis de construire un indicateur synthétique des réponses en 6 modalités de handicap croissant :

Groupe 1 ( 78.6 % ) : personnes déclarant ne souffrir d'aucune difficulté ;

Groupe 2 ( 6.6 % ) : personnes déclarant une seule difficulté ;

Groupe 3 ( 4.4 % ) : personnes déclarant « avoir un handicap » ou « avoir demandé une reconnaissance » ou souffrir d'une « limitation d'activité » ou dépendre d'une aide humaine ou souffrir de plusieurs autres difficultés.

Groupe 4 ( 2.6 % ) : personnes déclarant « avoir un handicap » ou « avoir demandé une reconnaissance », et personnes déclarant souffrir d'une « limitation d'activité », déclaration appuyée par des items d'aide humaine ou technique ou plusieurs autres.

Groupe 5 ( 3.8 % ) : personnes déclarant « avoir un handicap » ou « avoir demandé une reconnaissance », déclaration fortement appuyée par d'autres items ;

Groupe 6 ( 4 % ) : personnes déclarant avoir obtenu une reconnaissance de leur handicap ( plus, pour les moins de 16 ans : enfants et adolescents inscrits dans une classe ou un établissement spécialisé ).

Le tirage des individus HID dans l'échantillon des répondants VQS est stratifié par VQS et par l'âge ( en deux modalités ) avec des probabilités fortement inégales . En effet, les personnes appartenant au groupe 6 ont été tirées selon un taux de sondage élevé, alors que celles du groupe 1, à plus grand effectif dans l'ensemble de la population, ont eu au contraire un taux de sondage minimal.

On a tenu à garder un groupe témoin ( groupe 1 ) et d'y mener l'enquête détaillée même si les répondants ont déclaré n'avoir aucune difficulté à signaler.

L'éventail des probabilités de tirage de cette étape varie presque de 1 à 100.

## II.1.La stratification « HID ».

Il fut intéressant de prendre en compte l'âge des personnes dans la stratification car les réponses à VQS ont permis d'estimer que dans la population totale, la répartition des individus dans les différents groupes VQS était très différente selon qu'il s'agissait d'une personne de plus de 70 ans ou de moins de 70 ans.

Voici la répartition par groupe de l'échantillon des répondants VQS et celle de la population totale obtenue par pondération.

Chez les moins de 70 ans			Chez les 70 ans et plus		
	Echant. VQS	Popul. totale		Echant. VQS	Popul. totale
Groupe 1	83.8 %	83,7	Groupe 1	33.6 %	34,2
Groupe 2	6.0 %	6,1	Groupe 2	12.1 %	12,1
Groupe 3	3.3 %	3,4	Groupe 3	13.4 %	13,2
Groupe 4	1.3 %	1,4	Groupe 4	13.6 %	13,2
Groupe 5	2.1 %	2,1	Groupe 5	18.4 %	18,4
Groupe 6	3.4 %	3,4	Groupe 6	9.1 %	8,8

C'est pourquoi dans l'échantillon HID, on a cherché à avoir des proportions par groupe différentes selon la classe d'âge. L'échantillon HID présentait la répartition suivante :

Chez les moins de 70 ans :

13% issus du groupe 1 ;  
 7% issus du groupe 2 ;  
 13% issus du groupe 3 ;  
 7% issus du groupe 4 ;  
 20% issus du groupe 5 ;  
 40% issus du groupe 6 .

Chez les 70 ans et plus :

15% issus du groupe 1 ;  
 24% issus du groupe 2 ;  
 16% issus du groupe 3 ;  
 11% issus du groupe 4 ;  
 10% issus du groupe 5 ;  
 25% issus du groupe 6 .

Le tirage a donc été effectué selon 10 strates croisant les six groupes VQS et un ou deux groupes d'âge selon les groupes VQS.

Strate 01 : groupe VQS 1 et âge < 70 ans ;  
 Strate 02 : groupe VQS 1 et âge >= 70 ans ;  
 Strate 03 : groupe VQS 2 et âge < 70 ans ;  
 Strate 04 : groupe VQS 2 et âge >= 70 ans ;  
 Strate 05 : groupe VQS 3 de tout âge ;  
 Strate 06 : groupe VQS 4 de tout âge ;  
 Strate 07 : groupe VQS 5 et âge < 70 ans ;  
 Strate 08 : groupe VQS 5 et âge >= 70 ans ;  
 Strate 09 : groupe VQS 6 et âge < 70 ans ;  
 Strate 10 : groupe VQS 6 et âge >= 70 ans .

## **II.2.Le tirage de l'échantillon HID.**

### **II.2.1.Les probabilités de tirage.**

Cette stratification a été effectuée au sein de chaque strate géographique de VQS ; on rappelle que ces strates sont des regroupements de départements de taille régionale ou infra - régionale. Dans cette deuxième phase du plan de sondage, on n'a pas tenu compte des spécificités propres à l'enquête EHF puisque les deux enquêtes n'avaient plus rien en commun ; le nombre de strates géographiques a donc été ramené à 31 sur l'ensemble de la métropole.

La première étape a été de déterminer les probabilités de tirage au sein de chaque strate HID ( au nombre de 10 ), ce pour chacune des strates VQS ( au nombre de 31 ). En plus du fait que les individus soient surreprésentés par handicap croissant, deux autres raisons ont fait que les probabilités de tirage étaient très inégales.

Tout d'abord, selon la strate VQS.

- Les 23 strates où l'échantillon VQS a été tiré sans surreprésentation locale représentaient environ 79.5 % de la population métropolitaine habitant en domiciles ordinaires ( soit 45 600 000 personnes au 1<sup>er</sup> décembre 1999, date centrale de la collecte ). Elles ont regroupé 216 000 réponses exploitables à VQS, soit un taux de sondage de 1 / 211. On y a recueilli 12 436 réponses à HID, soit un taux HID / VQS de 1 / 17 et un taux global de sondage pour HID d'environ 1 / 3670.
- Les 7 strates où VQS seule a été surreprésentée ( mais pas HID ) comptaient environ 10 960 000 habitants en domiciles ordinaires. Elles ont regroupé environ 128 000 réponses exploitables à VQS, soit un taux de sondage de 1 / 86. On y a recueilli 3 034 réponses à HID, soit un taux HID / VQS de 1 / 42 et un taux global de sondage pour HID de 1 / 3612, voisin donc des 23 premières strates.
- L' Hérault ( 875 000 habitants en domiciles ordinaires ), qui a bénéficié d'une extension d'échantillon pour VQS et pour HID, a compté pour sa part 16 150 réponses exploitables à VQS ( soit un taux de sondage de 1 / 54 ) et 1 480 réponses à HID, soit un taux HID / VQS de 1 / 10.9 et un taux global de sondage pour HID de 1 / 590, plus de six fois supérieur aux précédents.

Enfin, le nombre de personnes à enquêter par région dépendait en partie de la capacité locale du réseau d'enquête, car une des différences avec l'enquête précédente est que les questionnaires VQS étaient remis aux personnes au moment du recensement de la population de mars 1999 puis récupérés ultérieurement, alors que pour l'enquête HID, des enquêteurs allaient sur place interroger les personnes.

Ce point sera expliqué plus en détail et sera accompagné d'un exemple dans le paragraphe suivant.

Toutes ces raisons ont donc fait que les probabilités de tirage à l'intérieur de chaque strate HID étaient très inégales.

On présentera les moyennes, extrêmes et dispersion des taux de sondage par strate HID dans le chapitre suivant.

Toutefois, dans chaque zone d'enquête ( et donc dans l'ensemble du tirage HID dans l'échantillon VQS ), les probabilités de tirage dans les 10 strates étaient proportionnelles à une échelle fixe ( les coefficients de tirage ). C'est-à-dire que d'une zone à l'autre on n'avait pas la même probabilité de tirage pour la même strate HID, mais le vecteur des dix probabilités pour une zone était colinéaire au vecteur ci-dessous :

	Coefficients de tirage
strate 01	0,68
strate 02	18,75
strate 03	5,62
strate 04	19,44
strate 05	20
strate 06	30
strate 07	52
strate 08	32
strate 09	65
strate 10	40

### II.2.2.La réalisation concrète du tirage.

On a réalisé ce tirage en partant du découpage des régions en zones de délégués que l'on avait utilisé pour le sondage VQS. Les zones qui ont été enquêtées restent celles qui ont été sélectionnées au cours du tirage VQS ; par contre, comme seul un département a demandé une extension d'échantillon pour l'enquête HID, on n'a pas tenu compte de toutes les zones de délégués qui avait été tirées en surplus dans certains départements dans le but de réaliser l'extension VQS. Par exemple, dans la région PACA, 27 ZD avaient été tirées à cause de l'extension d'échantillon demandée par les Bouches du Rhône ; mais on a proposé de tirer des individus pour l'échantillon HID que dans 22 ZD.

Les répondants VQS ont été regroupés par *zones d'enquête*. Le plus souvent, une zone d'enquête correspondait aux répondants VQS d'une ZD, mais elle pouvait aussi regrouper deux ZD. Dans la région 54, il n'y avait qu'une seule zone d'enquête.

Dans chacune de ces zones d'enquête, le but était de tirer un nombre déterminé d'individus et, selon l'effectif, un ou plusieurs enquêteurs serait chargé d'aller à leur domicile les interroger.

Le responsable de l'enquête à l'INSEE de Paris a déterminé le nombre d'individus qu'il voudrait dans chaque zone d'enquête en fonction du nombre de répondants VQS obtenus dans la zone, puis a demandé l'avis des directions régionales ( DR ).

En effet il était nécessaire de négocier avec les DR car le nombre de personnes à enquêter dans une zone dépendait de la charge de travail, de la compétence de l'enquêteur, de sa rapidité ; s'il n'avait pas suffisamment d'expérience, on réduisait son effectif et quand cela était possible, d'autres enquêteurs étaient chargés d'interroger davantage de personnes. De plus, il fallait être suffisamment formé pour pouvoir aborder un sujet aussi délicat que les incapacités physiques.

Il pouvait également arriver que sur une zone d'enquête il n'y ait pas d'enquêteur parce que le précédent était parti à la retraite et que la DR n'a pas pu le remplacer. Dans ce cas la DR a proposé à l'INSEE de Paris une ou deux autres zones d'enquête parmi celles qui avaient été tirées pour l'enquête VQS et que l'on avait mis de côté.

Prenons le cas de la région PACA. L'INSEE a proposé à la DR de tirer un échantillon de 26 personnes dans la zone de PERI et 26 personnes dans la zone de TALLONE. Mais la DR ne disposant pas d'enquêteur dans ces zones pour la collecte HID, elle a alors proposé de tirer les 52 personnes HID attendues plutôt dans l'une des zones d'Aix en Provence.

L'effectif HID par zone convenu après discussion oscillait autour de 50 individus, avec un minimum de 26 et pouvait aller jusqu'à 90 dans des régions comme l'île de France.

Dans des zones rurales très dispersées, il n'était pas possible d'interroger beaucoup de personnes car cela aurait été beaucoup trop coûteux à cause des frais de déplacement notamment. Prenons le cas du Limousin. L'ensemble de ses zones d'enquête n'a couvert qu'un tiers de sa superficie et a représenté 90% de sa population.

En ce qui concerne l'Hérault, 8 zones de délégués avaient été tirées pour le sondage VQS, 7 ont été conservées pour le sondage HID, et chacune présentait un très gros effectif HID ( 260 individus en moyenne ) afin d'obtenir un échantillon de 1800 personnes et réaliser les estimations locales. Dans ce cas, la zone de délégués comprenait plusieurs zones d'enquête et un enquêteur était chargé d'interroger 20 à 70 personnes.

Pour se fixer les idées : il y avait un peu plus de 400 zones d'enquête avec un effectif moyen de 50 individus présents dans l'échantillon HID ; globalement, une zone d'enquête représentait donc 130 000 personnes, avec évidemment nettement moins dans les zones rurales et plus dans les régions à forte densité de population, sauf dans le département de l'Hérault où une zone d'enquête représentait plutôt 15 000 personnes.

Pour la réalisation concrète du tirage, dans une zone d'enquête, on a classé les répondants à VQS selon les 10 strates HID. On a calculé le nombre d'individus à tirer dans chaque strate en appliquant les différents taux de sondage et on a tiré ces individus suivant une méthode de sondage aléatoire simple dans chacune des 10 strates.

Diverses particularités ont provoqué quelques rectifications des taux de sondage dans de rares zones d'enquête :

- d'abord, il pouvait arriver que dans une zone le nombre d'individus du groupe 6 ( les plus sévèrement handicapés ) soit insuffisant pour fournir le nombre d'individus HID voulu ; on a dû compenser en augmentant arbitrairement le taux de sondage dans les groupes 1 et 2 de la même zone d'enquête ;

- ensuite, après contrôle, lorsque plus de deux individus tirés faisaient partie du même ménage ( en fait le nombre maximal s'est établi à trois ), le moins handicapé des trois était retiré de l'échantillon HID ;
- enfin, les effectifs tirés par strate HID étant généralement faibles dans chaque zone d'enquête, les effets d'arrondi ont assez sensiblement modifié les taux réels de tirage.

Par exemple dans la Champagne - Ardenne, la zone de Montry ne contenait pas suffisamment de répondants VQS du groupe 6, donc on a été obligé de tirer 15 personnes supplémentaires dans le groupe 1. Après contrôle, on s'est rendu compte qu'il y avait des cas où il y avait trois personnes tirées dans le même ménage et ce cas on a gardé celles de groupe le plus haut et d'âge le plus faible.

L'échantillon de l'enquête était finalement constitué de 21 760 personnes vivant dans 20 116 ménages. On a recueilli 16 945 réponses à HID, soit un taux de déchet de 22.2 % en raison des refus, déménagements à une adresse inconnue, absence pendant toute la période de collecte, décès et entrée en institution sanitaire et sociale depuis la date du recensement (donc hors du champ de l'enquête ).

### **II.3. Résumé.**

La deuxième phase du plan de sondage est un sondage stratifié par VQS dans chaque zone d'enquête, à probabilités inégales, avec un tirage d'individus par sondage aléatoire simple dans chaque croisement zone d'enquête \* strate HID.

Le schéma de la page suivante résume les différentes étapes du plan de sondage de l'enquête HID.

PPT = tirage avec une probabilité proportionnelle à la taille ;  
 SAS = tirage par sondage aléatoire simple.

### III. Synthèse.

