



statec
LUXEMBOURG



La Ville de Luxembourg a commandité au CEPS/INSTEAD une étude de grande envergure à partir d'une enquête de terrain. Elle concerne les personnes âgées de 60 ans ou plus vivant à leur domicile au sein de la capitale. Les nombreux thèmes que l'enquête aborde doivent être restitués à une échelle infra-communale. Cette contrainte ne facilite ni la méthode d'échantillonnage, ni l'obtention de résultats et de données statistiques robustes à cette échelle. Cet article présente les moyens et méthodes dont nous disposons pour résoudre ces problèmes. Ainsi, la méthode utilisée fait appel à un échantillonnage spatial stratifié qui se fonde sur une agrégation raisonnée des quartiers administratifs de la Ville de Luxembourg, d'une part, et sur le contrôle a priori et a posteriori de variables auxiliaires (notamment démographiques), d'autre part. Sans s'attarder sur les résultats d'enquête eux-mêmes, faisant l'objet de prochaines éditions, cette étude expose la méthode utilisée et révèle les obstacles et les avantages de son application, tant à l'échelle de la ville que de ses quartiers résidentiels. Grâce à cette méthode, les élus locaux peuvent disposer désormais d'un outil d'aide à la décision en matière de planification intra-urbaine envers les personnes âgées.

Ville de Luxembourg et personnes âgées vivant à domicile : échantillonnage spatial et production de données d'enquête à l'échelle intra-urbaine

Philippe GERBER, avec la collaboration de Jean-Yves BIENVENUE* - CEPS/INSTEAD

Introduction

Quels que soient les pays de l'Union Européenne, le phénomène du vieillissement démographique de la population est patent. Ainsi, au Luxembourg, nous passons de 64 949 personnes âgées de 60 ans ou plus en 1981 à 81 511 personnes vingt ans plus tard¹. Ces changements démographiques s'accompagnent de mutations sociologiques et économiques majeures, comme la recherche d'une nouvelle image de l'âge, de meilleures conditions de ressources financières pour les plus de 60 ans, une meilleure santé, etc. Selon ces considérations, la Ville de Luxembourg a commandité au CEPS/INSTEAD une étude

de grande envergure, fondée en partie sur une enquête visant à prendre en compte l'avis des personnes âgées de 60 ans ou plus vivant à domicile. Les thèmes que l'enquête aborde sont nombreux : l'état du logement, le niveau de santé, l'appréciation du quartier ou encore les réseaux sociaux, pour ne citer que les plus importants. La capitale du Grand-Duché souhaite obtenir, à un niveau *infra-communal*, les informations permettant d'apprécier les besoins réels et les souhaits de ces personnes âgées, en vue de forger une nouvelle politique sociale, et en anticipant éventuellement la planification de nouvelles infrastructures.

* Nous souhaitons remercier la Ville de Luxembourg qui nous a confié l'étude portant sur les personnes âgées, notamment Claude Wiseler, Echevin de la Ville de Luxembourg, et Franca Riolino-Steichen, responsable du Service Seniors. Un rapport en a été tiré, effectué sous la coordination de Patrick Bousch et de Philippe Gerber, cf. [GER 2003]. Cette référence et les suivantes renvoient à la bibliographie située en fin de texte. Toute cette étude n'aurait pas pu s'effectuer sans la collaboration de plusieurs membres du CEPS/INSTEAD, notamment Stéphane Betti, Sylviane Breulheid, Jean-Marie Jungblut, Mathias Kuepie et Kristell Leduc. Qu'ils en soient ici remerciés.

¹ STATEC, recensements de 1981 et de 2001.

Dans ce cadre, les aspects méthodologiques de la mise en place de l'enquête doivent être rigoureusement contrôlés afin de pouvoir fournir aux décideurs institutionnels des données précises à l'échelle intra-urbaine, tout en voulant :

- restituer l'avis de la population âgée en interrogeant une partie de celle-ci (environ mille enquêtes possibles selon un type de sondage donné) ;
- assurer le maximum de représentativité au niveau infra-communal, donc avec une erreur d'échantillonnage² la plus faible possible, compte tenu des moyens budgétaires alloués ;
- assurer une totale comparabilité à ce niveau d'étude, en visant donc une erreur d'échantillonnage du même ordre d'une zone résidentielle à une autre.

Dans le but de respecter ces trois exigences, la méthode statistique adéquate est le sondage aléatoire spatial stratifié, plus efficace que la technique d'échantillonnage aléatoire simple. Toutefois, la technique de stratification proposée implique trois contraintes (liées entre autres aux impératifs financiers) :

- géographique (choix de la taille des unités spatiales) ;
- démographique (une population peu nombreuse des plus de 60 ans vivant à domicile et une

enquête reposant à la fois sur le ménage et l'individu) ;

- méthodologique (garantir la comparabilité spatiale et minimiser l'erreur d'échantillonnage en fonction des trois exigences précédentes).

Dans trois parties, nous explicitons plus précisément les exigences, contraintes et avantages de la méthode mise en œuvre : le contexte de l'étude, les modalités de tirage et, enfin, les résultats de la passation associés à leurs corrections.

1. Univers de référence et aires d'enquêtes

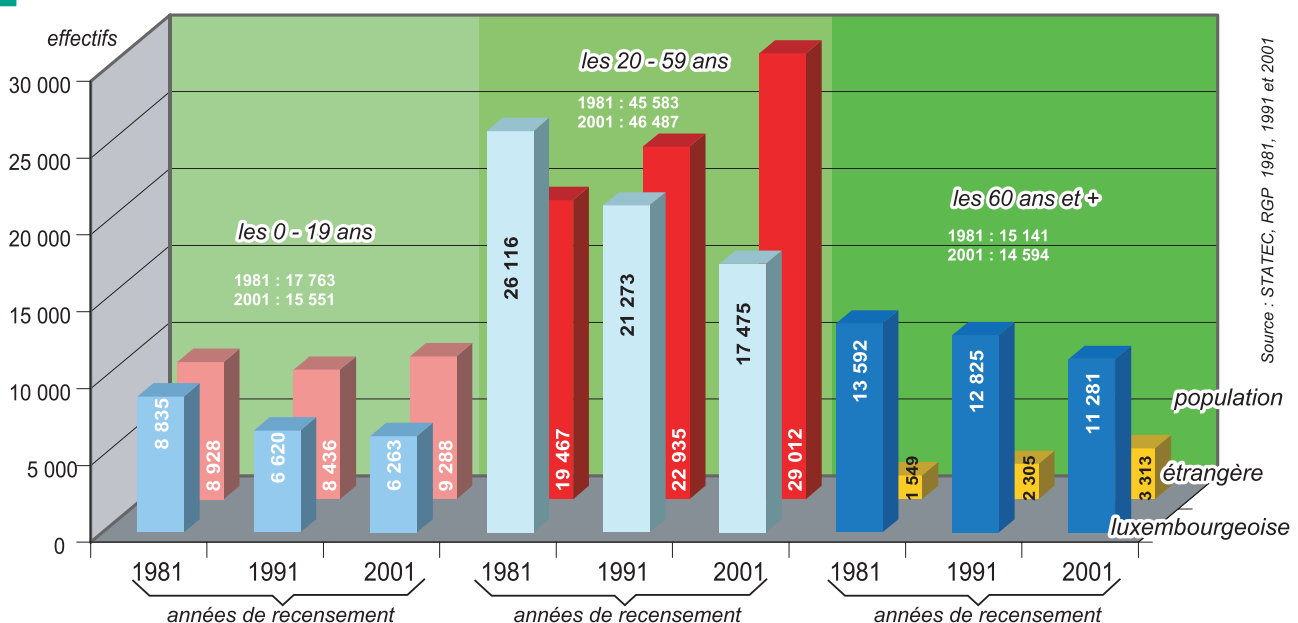
Le phénomène du vieillissement³ relaté auparavant ne doit pas masquer les disparités démographiques qui existent entre un pays et une ville, comme par exemple celle de Luxembourg. Les chiffres au niveau national mettent bien en évidence cette situation : il existe certes un vieillissement généralisé à l'échelle du pays, mais on note par ailleurs, une stagnation de la proportion de la population des plus de 60 ans au niveau de la capitale du Grand-Duché (cf. graphique 1). La part des personnes âgées de 60 ans ou plus passe de 19,2 % à 19 % dans la capitale et leur nombre absolu baisse entre 1981 et 2001.

La relative stabilité de la part de la population âgée se retrouve dans de nombreuses villes européennes, notamment en France⁴. Mais la figure 1 apporte également une autre précision : c'est le nombre de personnes âgées de nationalité luxembourgeoise qui diminue et celui des personnes non luxembourgeoises qui a augmenté le plus fortement (plus du double en vingt ans).

De plus, et dans l'optique d'une analyse infra-communale statistiquement comparable, nous remarquons que les 24 quartiers administratifs existants sont fortement hétérogènes, ne serait-ce qu'au niveau de la population (cf. figure 1). Les valeurs extrêmes sont à ce niveau éloquentes : le plus petit quartier (Pulvermühl) contient 340 habitants, alors que le plus grand (Bonnevoie Sud) dispose de près de 11 000 habitants.

Le deuxième problème tient au fait que ces quartiers ne présentent pas une répartition homogène des personnes âgées de 60 ans ou plus, comme le montre clairement la figure 2. La répartition relative varie entre moins de 13 % (quartier Pulvermühl) à plus de 37 % (quartier Ville Haute) de personnes âgées de 60 ans ou plus pour chaque quartier.

G₁ Evolution de la population de la Ville de Luxembourg par tranches d'âges et nationalité entre 1981 et 2001



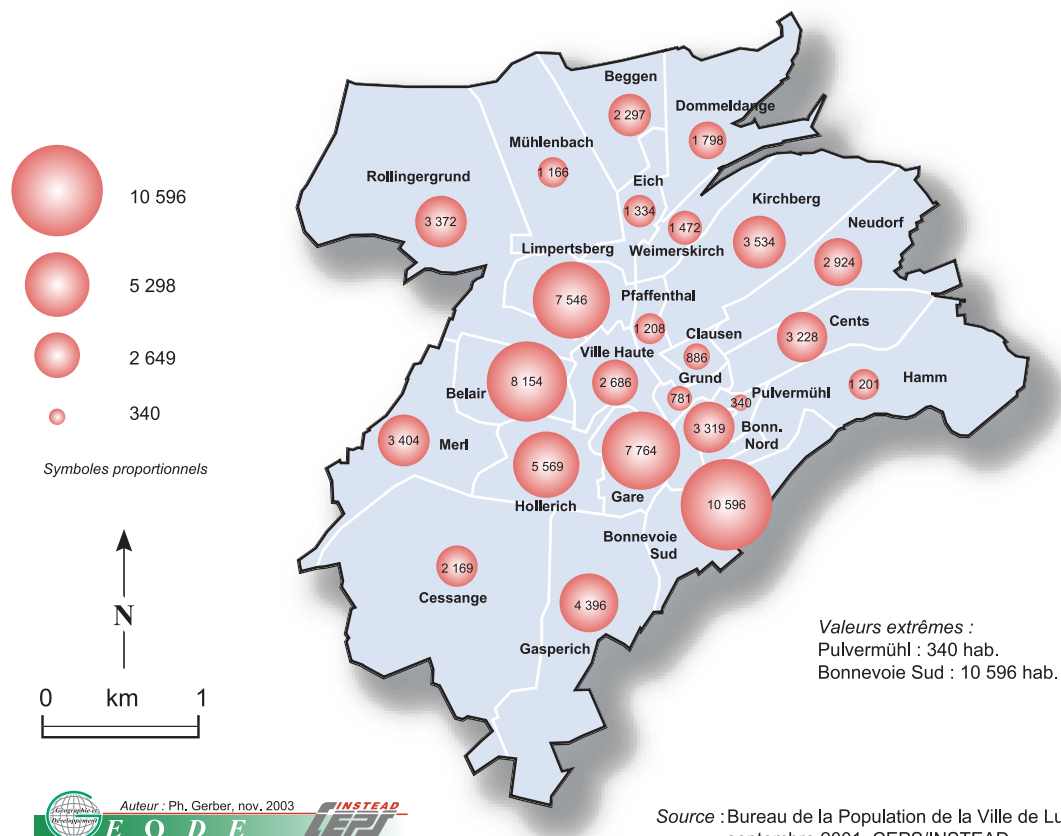
Source : STATEC, RGP 1981, 1991 et 2001

² Les termes scientifiques liés aux techniques d'enquête peuvent trouver leur définition sous la référence suivante : [ARD 1994].

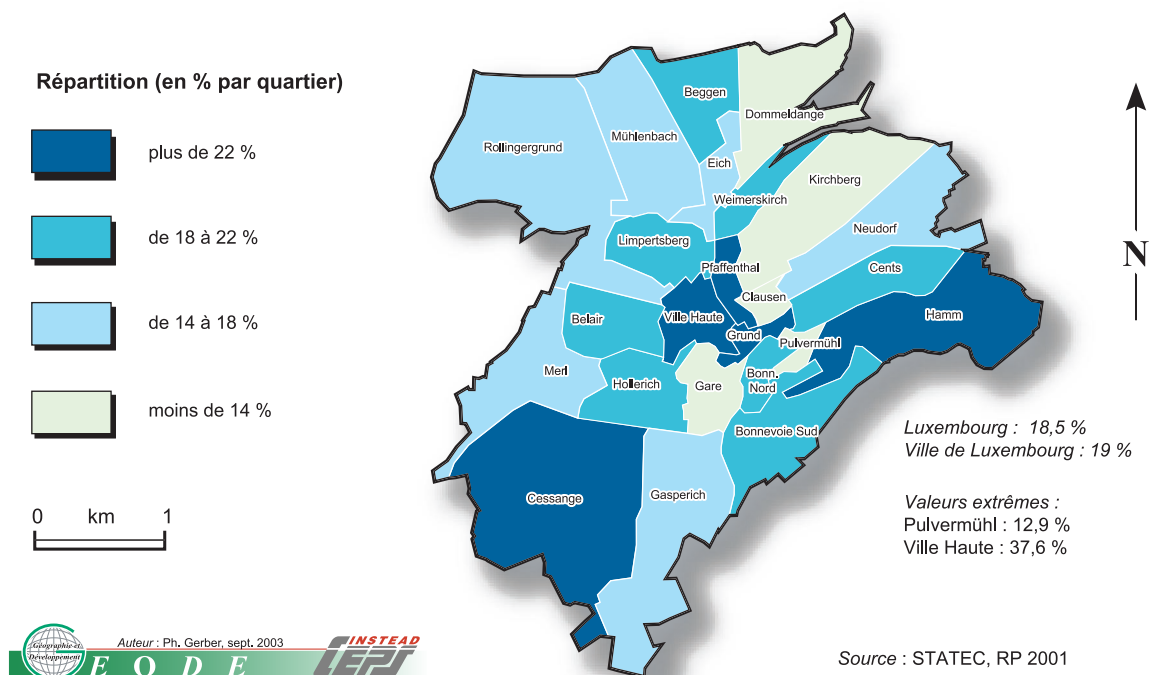
³ Nous pouvons citer, entre autres, [AGL 2002], [GUI 1999], [VIM 2000], [THE 1992], [ROC 2000].

⁴ Cf. notamment [CON 2001].

Répartition de la population résidente par quartier de la Ville de Luxembourg en 2001



Répartition des personnes de 60 ans ou plus par quartier en 2001



En tenant compte de ces deux facteurs, à savoir l'hétérogénéité du poids démographique et la répartition différenciée des personnes âgées au sein des 24 quartiers administratifs, la technique d'échantillonnage aléatoire simple ne donnerait pas de résultats suffisamment précis. En reprenant l'exemple du quartier de Pulvermühl, il n'y a « que » 12,9 % de personnes âgées en 2001, soit 44 personnes ! A moins de faire un recensement de toutes les personnes âgées de ce quartier, il n'y a pas de solution efficace pour produire une évaluation chiffrée précise à cette échelle par la procédure d'échantillonnage aléatoire simple. Ces répartitions posent alors un problème pour le plan de sondage : si le découpage administratif de 24 quartiers est conservé pour construire des strates d'un échantillon aléatoire stratifié, la qualité des résultats ne pourrait être assurée pour chaque strate en raison de l'inférence statistique. Ces quartiers doivent donc être adaptés et redéfinis afin de faciliter l'échantillonnage et de contrôler l'analyse future des données.

En partant de tous les ménages de la Ville de Luxembourg (univers de référence), soit au total 47 068 ménages fiscaux⁵, représentant 81 144 individus, la subdivision en strates géographiques de la capitale s'effectue alors en respectant une répartition équilibrée de la population, et se fonde sur plusieurs critères :

- un minimum de 2 000 personnes âgées de plus de 60 ans par strate, seuil établi à partir de critères statistiques de contrôle. Ces critères sont notamment liés aux découpages démographiques envisagés (trois catégories d'âge par exemple) dans l'exploitation future des résultats ;
- une contiguïté géographique des quartiers ;
- et une certaine homogénéité sociale de ces strates.

A la suite d'une série de tests, nous arrivons à un découpage en 6 zones ou strates, dont la répartition de la population varie peu, de 11 053 à 15 585 habitants (*cf. tableau 1*). Ces zones doivent maintenant faire l'objet d'un tri entre univers de référence (ménages et habitants de la Ville de Luxembourg) et population mère, à savoir les ménages dont une personne a au moins 60 ans ou plus. Nous pouvons résumer cette étape par la figure 3.

La population mère comptabilise en tout 14 984 individus répartis dans 11 769 ménages. Mais elle ne correspond pas encore à la population ciblée de l'enquête : doivent en être éliminés les ménages collectifs. Nous arrivons ainsi à une population ciblée de 10 567 ménages privés, soit 13 736 individus de 60 ans ou plus en septembre 2001, date de référence du fichier de population.

Nous finalisons alors le plan de sondage en répartissant ces ménages, avec leurs individus respectifs, en fonction des 6 strates spatiales pré-définies dans le tableau 1.

Ainsi, le lien de référence entre les quartiers, les adresses, les zones d'échantillonnage, les ménages et les individus étant établi, l'étape suivante consiste à tirer un échantillon aléatoire d'adresses au sein de chaque sous-ensemble territorial (plan de sondage en 6 strates) qui sera détaillé dans la prochaine partie.

2. Tirage et caractéristiques de l'échantillon

L'échantillonnage s'effectue en plusieurs étapes qu'il convient de bien préciser afin de comprendre le cheminement méthodologique nécessaire pour assurer des résultats de qualité, tant au niveau des zones résidentielles que des ménages et des personnes âgées.

2.1. La méthode de tirage

La méthode de tirage de l'échantillon utilisée, dans le cadre de cette enquête, repose sur une technique d'échantillonnage à deux phases⁶ (ou étapes) :

- le tirage d'un échantillon aléatoire stratifié (en 6 zones résidentielles) de **ménages privés** possédant au moins une personne âgée de 60 ans ou plus ;
- la **sélection aléatoire** d'une personne âgée de 60 ans ou plus au sein de chaque ménage de l'échantillon.

Cette procédure de sélection permet de respecter les conditions décrites ci-après :

- le caractère représentatif des 10 567 ménages ciblés répartis dans les 6 zones résidentielles retenues dans l'étude ;
- l'estimation sans biais de tous les membres des ménages privés concernés (16 772 individus tous âges confondus) ;
- l'estimation sans biais des 13 736 personnes âgées de 60 ans ou plus résidant dans un des 10 567 ménages privés.

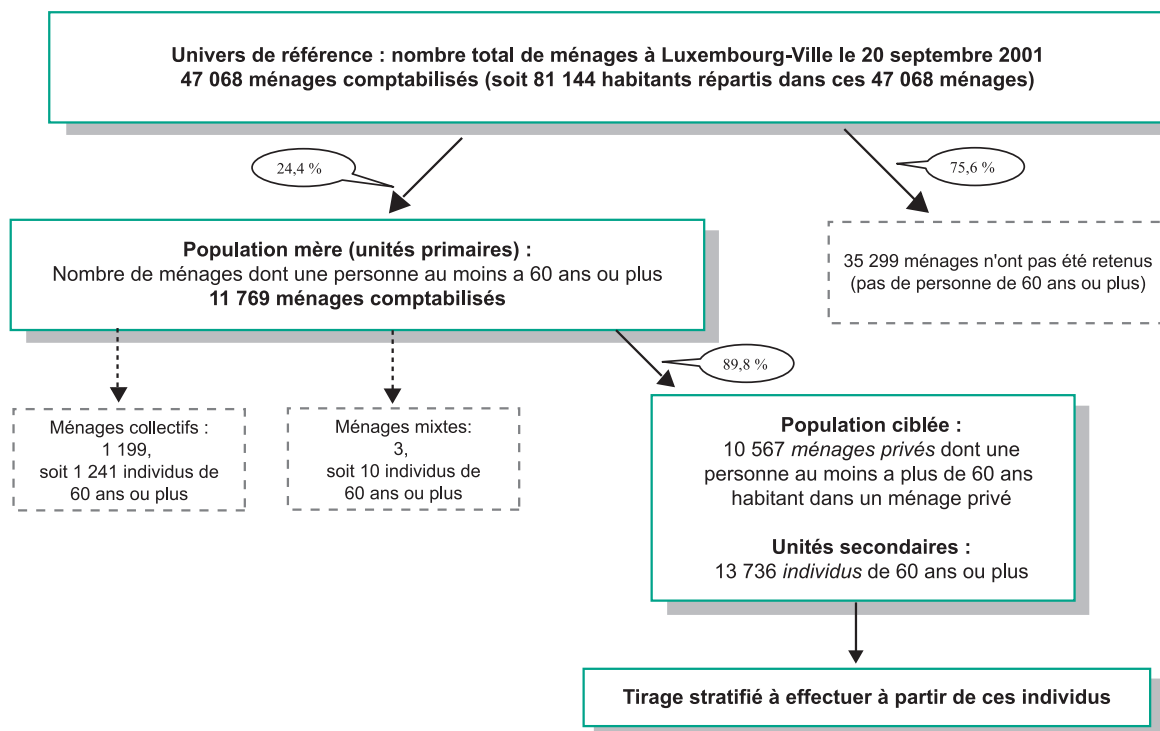
Ces conditions sont complétées en croisant ces variables par des variables auxiliaires telles que l'âge, le sexe, la nationalité et l'état civil. Nous allons détailler ces étapes de méthode de tirage en trois phases successives.

2.2. Sélection d'un premier échantillon aléatoire stratifié de ménages privés

La première étape consiste donc en la sélection d'un premier échantillon d'unités primaires de 1 500 ménages, puis de 3 000 ménages en prévision de l'élaboration d'une liste d'attente. Le tirage s'effectue selon un plan de sondage stratifié en fonction des six zones géographiques définies plus haut. Cela signifie que chaque strate possède son propre lot d'adresses tirées aléatoirement

⁵ Source : fichier de population de la Ville de Luxembourg, septembre 2001. Au moment de la mise en place de la méthode d'échantillonnage, nous ne disposions pas encore des chiffres du dernier recensement.

⁶ Source générale : [SAR 1992].



Source : Bureau de Population de la Ville de Luxembourg (BPVdL), septembre 2001, CEPS/INSTEAD, 2003

au sein de chaque strate. En fonction de deux facteurs, à savoir l'objectif d'atteindre 1 000 répondants et la passation de l'enquête, nous avons retenu au total 1 803 adresses effectivement utilisées sur le terrain.

Dès lors, afin de tester le degré de représentativité du premier échantillon, deux listes de variables de contrôle ont été retenues selon que l'on se situe au niveau du ménage privé ou au niveau des individus.

Par rapport aux ménages privés, nous avons privilégié les variables liées à la zone résidentielle et à la taille du ménage. Prenons l'exemple de la zone résidentielle (cf. tableau 2).

Dans cette table, la variance d'échantillonnage est nulle (donc non affichée), du fait que les strates assurent la base de contrôle (et de sondage) de l'enquête : l'échantillon tiré aléatoirement est donc parfaitement représentatif des zones résidentielles.

T₁

Plan de sondage : répartition de la population mère et de la population ciblée en 6 strates selon les 24 quartiers administratifs

Les 6 strates d'échantillonnage (zones résidentielles)	Univers de référence	Population mère		Population ciblée ou plan de sondage	
		ménages	individus	ménages	individus
1 Cessange, Hollerich, Merl	11 053	1 654	2 596	1 629	2 114
2 Belair, Rollingergrund	11 489	1 722	2 629	1 632	2 135
3 Bonnevoie N, Gare, Ville-Haute	13 731	2 198	2 926	1 575	1 945
4 Bonnevoie S, Gasperich	15 020	2 013	3 173	2 006	2 637
5 Beggen, Dommeldange, Eich, Limpertsberg, Mühlenbach, Weimerskirch	15 665	2 104	3 337	2 040	2 656
6 Cents, Clausen, Grund, Hamm, Kirchberg, Neudorf, Pfaffenthal, Pülvemuhl	14 186	2 078	3 364	1 685	2 249
Total	81 144	11 769	18 025	10 567	13 736

Source : BPVdL, sept. 2001, CEPS/INSTEAD, 2003

Au niveau des unités secondaires (tous les membres du ménage), l'âge, le sexe, la nationalité et l'état civil ont servi de variables de contrôle. Sans les détailler toutes, nous pouvons nous focaliser sur les deux premières d'entre elles (cf. tableau 3).

Malgré les variations dues aux erreurs d'échantillonnage⁷, le tirage respecte les valeurs observées au niveau du fichier du Bureau de la Population de la Ville de Luxembourg car les bornes inférieure et supérieure intègrent les valeurs observées. Ce constat s'observe également pour les autres variables de contrôle utilisées (la nationalité et l'état civil).

Nous pouvons ainsi considérer que le tirage respecte les proportions et répartitions démographiques des individus (unités secondaires) et des ménages âgés (unités primaires). Il n'existe pas de biais dans l'échantillon car les bornes inférieure et supérieure des intervalles de confiance figurant dans les tableaux incluent les chiffres réels du fichier de population de la Ville de Luxembourg.

2.3. Sélection d'un second échantillon d'individus âgés à partir du premier échantillon

Pour des raisons de rapidité et de coût de la passation de l'enquête, nous ne voulons interroger qu'une seule personne de 60 ans ou plus dans le ménage sélectionné lors de la première étape de l'échantillonnage. Or, comment sélectionner, de manière aléatoire, une seule personne de 60 ans ou plus, lorsqu'il existe par exemple deux personnes de 60 ans ou plus dans le même ménage ? La deuxième étape consiste alors en un échantillon aléatoire d'unités secondaires (ou individus âgés de 60 ans ou plus).

Ainsi, à partir de l'échantillon d'unités primaires (les ménages sélectionnés dans la première étape), nous constituons une population d'unités secondaires composée des

T2 Degré de représentativité de l'échantillon au niveau du ménage privé selon la zone résidentielle

Zones résidentielles ou strates	Strates des ménages privés (BPVdL)	Echantillon (nombre de ménages)	Poids de l'échantillon	Taux de sondage	Estimation pondérée (contrôlée)
1	1 628	277	5,88	0,170	1 628
2	1 631	279	5,85	0,171	1 631
3	1 569	269	5,83	0,171	1 569
4	2 008	342	5,87	0,170	2 008
5	2 044	348	5,87	0,170	2 044
6	1 687	288	5,86	0,171	1 687
Total	10 567	1 803		0,171	10 567

Source : BPVdL, sept. 2001, enquête Personnes Agées Ville de Luxembourg (PAVdL), 2002, CEPS/INSTEAD, 2003

T3 Degré de représentativité de l'échantillon au niveau de tous les individus selon l'âge et le sexe

Sexe	Age	Nb de personnes des ménages (BPVdL)	Echantillon (nombre d'individus)	Estimation pondérée	Erreur d'échantillonnage		
					Borne inférieure	Borne supérieure	Coefficient de variation
Masculin	- de 60 ans	1 670	283	1 659	1 488	1 829	5,14
	60-64 ans	1 792	309	1 811	1 634	1 988	4,89
	65-74 ans	3 255	561	3 289	3 062	3 515	3,44
	75 ans ou +	2 955	508	2 977	2 759	3 195	3,66
Féminin	- de 60 ans	1 366	241	1 413	1 255	1 571	5,59
	60-64 ans	1 648	277	1 623	1 454	1 792	5,20
	65-74 ans	2 579	438	2 567	2 361	2 772	4,01
	75 ans ou +	1 507	253	1 484	1 322	1 646	5,46
	Total	16 772	2 870	16 821			

Source : BPVdL, 2001, enquête PAVdL, 2002, CEPS/INSTEAD, 2003

2 346 personnes âgées présentes dans les 1 803 ménages du premier échantillon. Afin de garder le caractère aléatoire de la base de sondage, une personne âgée de 60 ans ou plus a été sélectionnée sur la base d'une procédure ou clé appelée Kish⁸ au sein de chaque ménage du premier échantillon. La règle principale en est la suivante : est interrogée la personne d'un ménage du premier échantillon, sachant que cette personne doit avoir au moins 60 ans passés et que sa date d'anni-

versaire doit correspondre à celle qui est la plus proche de la date d'exécution du tirage.

Grâce au fichier de population de la Ville de Luxembourg, contenant la date de naissance de la personne, il a été possible de trier et de déterminer les personnes concernées⁹. Au total, 1 803 personnes âgées résidant dans l'un des 1 803 ménages privés ont été listées (sur les 2 346 personnes âgées de 60 ans ou plus) et contactées sur le terrain.

⁷ Par rapport aux calculs de variance, cf. [DEV 1999].

⁸ Du nom d'un statisticien. Cf. [KIS 1965].

⁹ Cela facilitera d'ailleurs le travail de terrain de l'enquêteur, qui saura exactement quelle personne du ménage interroger.

Au niveau de l'échantillon Kish des personnes âgées, des défauts de représentativité n'ont pas été détectés, tant au niveau du sexe, de l'âge, que de la nationalité et de l'état civil. Nous reprenons l'exemple de l'âge et du sexe en annexe 1. Certes, il existe des différences mais elles sont sans gravité car les bornes inférieure et supérieure incluent les valeurs observées du fichier de population.

Les deux étapes méthodologiques exposées restent encore relativement théoriques : il convient maintenant de les confronter au terrain, en y ajoutant quelques règles de passation, notamment en fonction de cas particuliers que les enquêteurs ou enquêtrices peuvent rencontrer auprès des ménages concernés.

2.4. Application de la procédure Kish sur le terrain

Il est possible de répartir la liste de personnes âgées à interroger dans chaque ménage privé en fonction des différents enquêteurs et enquêtrices. Cette liste spécifie les nom, prénom, adresse et date de naissance de la personne à interroger dans le ménage privé. Cependant, il a été prévu d'adapter l'utilisation de la méthode Kish en présence de 2 cas de figure de manière à ne pas créer de biais de représentativité et de tenir compte des handicaps éventuels que connaissent certaines personnes âgées.

1. Cas où le ménage à enquêter est composé de 2 ou 3 personnes :

- a) si la personne de 60 ans ou plus, choisie selon la méthode Kish, n'est pas **apte à répondre** (absence, maladie, handicap), les enquêteurs ont pour consigne d'interroger une autre personne **vivant dans** le ménage : cette autre personne constitue alors une personne « d'approximation » (ou proxy), c'est-à-dire qu'elle répond **à la place** de la personne choisie au départ au hasard ;

- b) par contre, si la personne âgée, choisie selon la méthode Kish, **ne veut pas répondre** (refus), dans ce cas, **le ménage ne sera pas enquêté** (aucune personne), même si une autre personne (conjoint par exemple) est prête à y participer (résultats dans la partie suivante) ;

- c) si aucun membre du ménage n'est apte à répondre, alors qu'une autre personne le veut bien (un enfant qui se trouve en visite par exemple, ou une assistante sociale ou une infirmière... qui connaissent la personne tirée selon la procédure Kish), cette dernière personne, consentante, sera à son tour considérée comme personne « d'approximation » et répondra **à la place** de la personne tirée au hasard.

2. Cas où le ménage est composé d'une seule personne :

- a) si la personne de 60 ans ou plus **ne veut pas répondre** (refus), **le ménage ne sera pas enquêté**, même si une autre personne ne faisant pas partie du ménage (frère ou infirmière par exemple) est prête à participer ;
- b) si la personne de plus de 60 ans n'est pas **apte à répondre** (absence, maladie, décès), prendre si possible une personne « d'approximation ».

Ces contraintes déterminées, toutes les conditions sont maintenant réunies pour que les enquêteurs et enquêtrices se rendent sur le terrain. Nous pouvons ainsi développer les résultats de la passation de l'enquête par interview en vis-à-vis et les redressements afférents.

3. Passation de l'enquête et corrections statistiques

Une enquête de terrain suppose l'existence de certains aléas ou biais que l'organisateur et les enquêteurs et enquêtrices ne peuvent pas forcé-

ment maîtriser : il s'agit notamment des mauvaises adresses, des personnes qui refusent de répondre, etc. Ces biais doivent être répertoriés, contrôlés et apurés afin de garantir la fiabilité des résultats statistiques de l'enquête. Ainsi, après avoir développé les résultats de la passation de l'enquête, le deuxième paragraphe s'attardera sur les causes des principaux biais et le dernier abordera quelques exemples estimés de résultats d'enquête.

3.1. Résultats de la passation de l'enquête de terrain

En résumé, sur les 1 803 envois de courriers, trois grandes familles de réponses ont été répertoriées :

- *les hors champ*, c'est-à-dire les personnes décédées, en maison de retraite ou qui ont déménagé hors de la ville. Ils représentent 94 ménages¹⁰, soit 5,2 % des 1 803 envois ;
- *les répondants*. 903 ménages/individus kish constituent les répondants de l'enquête, soit 50,1 % des courriers. Ils correspondent aux répondants à partir desquels les résultats de l'enquête pourront être extrapolés ;
- *les non répondants*. 806 enquêtes non réalisées se sont partagées entre : 537 refus volontaires d'enquêtes (29,7 %) ; 67 mauvaises adresses (3,7 %) ; 31 refus pour incapacité de répondre (1,7 %) ; et 172 ménages toujours absents (9,6 %).

Il est nécessaire d'observer plus en détail les caractéristiques des ménages formant les non répondants. Car, s'il existe des défauts de représentativité entraînant des biais dans les futurs résultats de l'enquête, il est important de les corriger. La démarche adoptée est de détecter en premier lieu les variables qui interviennent directement ou indirectement dans le processus de non-réponse d'abord, et de dresser le profil des répondants ensuite.

¹⁰ Ces hors champ (hors du champ de l'enquête) seront assimilés dans le processus de repondération à des répondants. En effet, ils ne sont plus concernés par le questionnaire, donc ils peuvent être considérés comme ne faisant plus partie de notre échantillon. De ce fait, ces hors champ nous donnent une information supplémentaire par rapport au fichier de population initial et à la population de référence, bien qu'ils n'interviennent pas dans l'analyse des résultats d'enquête.

3.2. Détection des principaux biais

En fonction de la problématique posée, nous nous intéressons d'abord à la répartition des non répondants selon les zones résidentielles. Ces strates ou zones, ayant servi de base de contrôle, à la fois de la procédure d'échantillonnage et de la passation du questionnaire, ne montrent guère de différences entre elles (cf. annexe 2). Cette remarque justifie à elle seule la méthode d'échantillonnage spatial utilisée car nous avons pu contrôler et assurer une comparabilité au niveau infra-communal. Malgré un taux de non-réponse observé variant entre 38,5 et 48,7 %, l'un des buts de l'enquête est d'ores et déjà atteint : disposer de résultats d'enquête pondérés comparables à l'échelle infra-communale.

Nous avons associé cette variable à d'autres variables auxiliaires, à savoir l'âge, le sexe, la nationalité, l'état civil, la taille du ménage, ainsi que les groupes d'enquêteurs classés par taux de réponse, au sein d'une régression logistique afin d'analyser méthodiquement le comportement de (non) réponse¹¹ de l'enquête.

Cette méthode statistique permet d'identifier et de comparer entre elles les différentes variables qui influencent la propension à appartenir au groupe des répondants ou à celui des non répondants. Les résultats (cf. annexe 3) montrent que la variable « groupes d'enquêteurs par classe de taux de réponse » est celle qui présente le plus haut degré de signification. Ainsi, toutes choses égales par ailleurs, les modalités liées aux enquêteurs expliquent la propension la plus forte des enquêtés à ne pas répondre au questionnaire. Les autres variables ne sont pas significatives par rapport au modèle de régression.

La correction de tous ces biais va s'effectuer à partir d'une procédure de re-pondération. Les facteurs intervenants d'extrapolation vont

tenir compte des erreurs observées sur le terrain après l'enquête, d'une part, et de la représentativité des estimations selon certaines caractéristiques des ménages, des individus membres de ces ménages et des individus Kish, d'autre part. Ces biais vont être corrigés en fonction de différentes variables auxiliaires grâce à la méthode de calage sur marges, procédure de pondération utilisée à l'origine pour pondérer l'enquête du recensement français.

3.3. Détermination des résultats d'enquête par un estimateur de calage

Le redressement de l'échantillon vise non seulement à réduire le biais inhérent à la présence d'un comportement de réponse non homogène au sein de la population ciblée, mais également à améliorer la représentativité de l'échantillon des 903 répondants vis-à-vis d'un certain nombre de critères non pris en compte *a priori* lors du tirage de l'échantillon (sexe, âge, nationalité, état civil).

En tenant compte des biais détectés et en voulant nous rapprocher des informations liées au fichier de population, nous souhaitons caler nos estimations en fonction de ces deux types d'information.

De manière générale, le plan d'échantillonnage et cette méthode de re-pondération définissent les formules qui devront être utilisées pour les estimations des résultats de l'enquête¹² (cf. annexe 4). A titre indicatif, nous donnons quelques résultats estimés d'enquête illustrant l'aboutissement de ce cheminement méthodologique.

3.4. Des résultats estimés, extrapolés et situés dans des fourchettes d'évaluation

Alors que, selon le fichier de population, il existe 13 736 personnes âgées vivant à domicile au niveau de la Ville de Luxembourg en septembre 2001, le total défini après calage est extrapolé et estimé à 12 615 individus, selon un coefficient de variation général de 0,9 % (cf. tableau 4).

La différence entre les deux nombres représente l'estimation (environ 0,9 % d'écart) du nombre d'observations erronées du fichier de population (les hors champ estimés à l'échelle de la population âgée). L'estimation avancée peut donc varier de 2*0,9 % de la valeur des 12 615 individus, que ce soit de manière inférieure ou supérieure, mais l'avantage (et la rigueur statistique sous-jacente) est de pou-

T4 Estimation du total des individus de 60 ans ou plus et selon les strates

	Effectifs			Pourcentages			Coefficient de variation (%)
	Borne inférieure	Estimation	Borne supérieure	Borne inférieure	Estimation	Borne supérieure	
Zone 1	1 898	1 967	2 037	96,5	100	103,6	1,8
Zone 2	1 754	1 854	1 954	94,6	100	105,4	2,7
Zone 3	1 642	1 739	1 835	94,4	100	105,5	2,8
Zone 4	2 321	2 413	2 506	96,2	100	103,9	1,9
Zone 5	2 520	2 606	2 693	96,7	100	103,3	1,7
Zone 6	1 959	2 036	2 113	96,2	100	103,8	1,9
Total	12 401	12 615	12 830	98,3	100	101,7	0,9

Source : BPvDL, 2001, enquête PAVdL, 2002, CEPS/INSTEAD, 2003

¹¹ Source : [DEV 1993a].

¹² Pour les explications de cette formule et de celles qui en découlent (notamment pour les calculs de variance), se référer au site du CEPS/INSTEAD, www.ceps.lu.

voir estimer cette marge d'erreur. Dans ce tableau, sont également fournis les chiffres de la population âgée par strate. Les variations y sont très faibles et équivalentes ; c'était d'ailleurs l'un des objectifs de l'échantillonnage spatial réalisé. Ainsi, malgré des effectifs relativement réduits (en moyenne 150 interrogés par strate), nous arrivons à un coefficient de variation qui oscille entre 1,7 et 2,8 %.

En guise de comparaison, lorsque nous calculons les intervalles de confiance au niveau de trois classes d'âge (cf. *tableau 5*), ils sont équivalents à ceux des six strates précédentes (environ 2,3 % en moyenne), alors que les effectifs interrogés par classe d'âge sont naturellement plus nombreux (en moyenne 300 interrogés par classe).

Les deux exemples précédents proviennent de variables auxiliaires sous contrôle, c'est-à-dire que les chiffres sont connus *a priori*, grâce au fichier de population. Qu'en est-il pour les variables d'intérêt de l'enquête elle-même, d'une part, et pour des cas où les effectifs interrogés sont peu nombreux, d'autre part ? En fait, plus les effectifs se réduisent, plus les variations sont fortes. Une bonne illustration en est l'indicateur de dépendance, construit à partir de plusieurs questions de l'enquête¹³. L'indicateur de groupes iso-besoins (GIB, cf. *tableau 6*) est constitué de 4 niveaux : 1) GIB 0 : groupe de personnes entièrement valides, 2) GIB 1 : groupe de personnes présentant une faible dépendance ; GIB 2 : groupe de personnes de moyenne dépendance ; 4) GIB 3 : groupe de personnes de forte à très forte dépendance.

Les personnes du groupe iso-besoins 3 sont les moins nombreuses. Le coefficient de variation qui lui est associé est bien plus conséquent : il est de 14,2 % pour ce dernier groupe, alors qu'il n'est que de 3,9 % pour le GIB 0. En précisant maintenant ces résultats par tranche d'âge, l'une des variables explicatives du phénomène de dépendance,

T5 Estimation du total des individus de 60 ans ou plus selon les classes d'âge

	Effectifs			Pourcentages			Coefficient de variation (%)
	Borne inférieure	Estimation	Borne supérieure	Borne inférieure	Estimation	Borne supérieure	
60 à 64 ans	3 121	3 272	3 423	24,3	25,9	27,6	2,3
65 à 74 ans	5 125	5 351	5 576	39,9	42,4	44,9	2,1
75 ans et +	3 805	3 993	4 181	29,7	31,7	33,7	2,4
Total	12 401	12 615	12 830	98,3	100	101,7	0,9

Source : BPVdL, 2001, enquête PAVdl, 2002, CEPS/INSTEAD, 2003

T6 Estimation du total des individus de 60 ans ou plus selon les groupes iso-besoins (GIB)

Groupes iso-besoins	Effectifs			Pourcentages			Coefficient de variation (%)
	Borne inférieure	Estimation	Borne supérieure	Borne inférieure	Estimation	Borne supérieure	
1) GIB 0	4 820	5 226	5 631	37,6	41,4	45,4	3,9
2) GIB 1	4 702	5 116	5 529	36,7	40,6	44,6	4,0
3) GIB 2	1 319	1 606	1 892	10,3	12,7	15,3	8,9
4) GIB 3	478	669	859	3,7	5,3	6,9	14,2
Total	12 401	12 615	12 830	98,3	100	101,7	0,9

Source : BPVdL, 2001, enquête PAVdl, 2002, CEPS/INSTEAD, 2003

T7 Estimation du total des individus de 75 ans ou plus selon les groupes iso-besoins (GIB)

Population des 75 ans ou +	Effectifs			Pourcentages			Coefficient de variation (%)
	Borne inférieure	Estimation	Borne supérieure	Borne inférieure	Estimation	Borne supérieure	
GIB 0	1 065	1 295	1 525	25,5	32,4	40,1	8,9
GIB 1	1 360	1 612	1 865	32,5	40,4	49,0	7,8
GIB 2	256	711	896	6,1	17,8	23,6	13,0
GIB 3	239	375	510	5,7	9,4	13,4	18,0
Total	3 805	3 993	4 181	29,8	31,6	33,7	2,4

Source : BPVdL, 2001, enquête PAVdl, 2002, CEPS/INSTEAD, 2003

le coefficient de variation sera de 18 % pour le GIB 3 et les 75 ans ou plus (cf. *tableau 7*). Or, ce dernier coefficient se situe à la limite de l'intervalle acceptable des données, selon les critères standards de qualité statistique¹⁴.

Ces résultats illustratifs incitent à la prudence quant à la manipulation des petites populations où les coefficients de variation afférents peuvent connaître des écarts relativement

importants. Toutefois, la pondération pratiquée au sein de cette enquête, accompagnée de son processus de redressement, engendre le calcul d'une variance qui permet de contrôler ces écarts d'estimation et ainsi d'en évaluer son degré de précision. Les résultats d'enquête y gagnent en fiabilité et en efficacité car la variance permettrait, par exemple, de fournir une limite d'exploitation géographique des découpages infra-communaux proposés.

¹³ Pour la construction de cet indicateur, cf. annexe 4, [GER 2003].

¹⁴ Au sujet de la qualité statistique, cf. [STA 2002]

Conclusion

A travers l'exemple d'une enquête sur les personnes âgées à domicile au niveau de la Ville de Luxembourg, la technique de l'échantillonnage spatial stratifié révèle de nombreux atouts. Cependant, elle suppose au préalable plusieurs exigences : des sources d'information de cadrage, une connaissance du terrain (et du sujet d'étude), certaines conditions préalables, et un suivi au niveau des strates.

Ainsi, la Ville de Luxembourg a fourni le fichier administratif de la population. Malgré la présence de certaines erreurs dans ce fichier, cet outil a permis de créer, de contrôler et de gérer l'échantillon, en partant d'informations très récentes : ces données de cadrage ont permis notamment le découpage infra-communal.

Au niveau de la connaissance du terrain, notre commanditaire a fait appel au CEPS/INSTEAD qui dispose de l'expérience et du savoir-faire nécessaires pour mener à bien ce type d'enquête. Par rapport au sujet d'étude, les strates spatiales trouvent d'abord leur justification dans le but de l'enquête : dans notre cas, il était bien souligné la volonté d'obtenir des données comparables au niveau infra-communal ; cela suppose un rôle explicite des strates spatiales dans l'analyse postérieure des résultats dont nous avons fourni quelques illustrations. Néanmoins, il est évident que la thématique des personnes âgées devait être maîtrisée dès le départ, quel que soit le mode de tirage.

Une condition préalable est que ces strates, et leur population ciblée associée, doivent contenir un nombre de personnes à interroger suffisamment important pour l'inférence statistique (approximation normale...). En outre, les observations constituant une strate doivent tendre vers une certaine homogé-

néité (sociale par exemple), alors que les observations entre chaque strate doivent être hétérogènes. Ces deux conditions nous ont obligés à agréger les 24 quartiers administratifs de la Ville de Luxembourg en 6 zones résidentielles ou strates spatiales.

Ensuite, la stratification obtenue suppose un suivi au moment de la passation de l'enquête. Le but est de contrôler au fur et à mesure les réponses ou non-réponses par strate. Si une strate avait été, par exemple, sous-représentée au niveau de ses effectifs de répondants, il aurait été possible de gérer ce biais en augmentant la taille de l'échantillon (adresses) grâce à la liste d'attente établie au moment du tirage de l'échantillon. Cette option n'a cependant pas été nécessaire.

Malgré ces exigences, la méthode de stratification spatiale apporte indéniablement son lot d'avantages. Ainsi, grâce à la création des six zones résidentielles, la technique de l'échantillonnage aléatoire simple est dépassée et inefficace. Ceci tant au niveau de la variance¹⁵ calculée *a priori*, que du biais (une erreur d'observation) que l'on essaie de maîtriser *a posteriori*. Nous améliorons sensiblement la précision des résultats grâce à la procédure retenue.

De plus, sous la contrainte d'un nombre limité d'observations (pas plus de mille enquêtes), cette méthode offre un degré de précision - de même ordre au sein de chaque zone - qui a permis de faciliter la comparabilité des résultats au niveau infra-communal. Il n'existe donc pas de déséquilibre spatial, et les résultats sont géographiquement valides. Par exemple, nous souhaitons connaître quels sont les commerces et services dont la personne âgée dispose à proximité de son domicile. Sans passer en

revue tous les commerces concernés, nous notons un déficit évident pour la pharmacie, notamment dans les zones résidentielles est (n°6) et sud-ouest (n°1). Selon nos résultats statistiques, découlant de l'appréciation des personnes interrogées, moins de 20 % des personnes de ces zones résidentielles disposent en 2002 d'une pharmacie à proximité de leur domicile, contre plus de 70 % pour les quatre autres secteurs. Après vérification sur le terrain, nous remarquons effectivement que les zones est et sud-ouest sont les moins bien pourvues en pharmacies. Ainsi, l'analyse des résultats d'enquête permet de détecter non seulement les commerces et services qui peuvent manquer au niveau de la Ville de Luxembourg, mais également leur répartition spatiale inégale à l'échelle des quartiers résidentiels.

Cet exemple illustre bien le genre de résultats susceptibles d'intéresser la Ville de Luxembourg, ou toute autre municipalité urbaine. Les données fiables transmises à l'échelle infra-communale sont essentielles pour les élus, à l'heure où les politiques cherchent de plus en plus à trouver des réponses locales à des phénomènes globaux, comme le vieillissement démographique de la population. Des solutions par rapport à la planification urbaine peuvent être étudiées et développées grâce à cette méthode d'échantillonnage.

La fiabilité et le contrôle des chiffres exposés tendent à démontrer que les résultats de l'enquête, lorsqu'ils s'appuient sur une population interrogée relativement nombreuse, sont tout à fait comparables entre les strates, donc entre les zones résidentielles. En ce sens, la stratification spatiale n'est pas une fin en soi : elle permet de mieux contrôler les résultats *a priori* et *a posteriori* par rapport à un échantillon aléatoire simple à l'échelle intra-urbaine.

¹⁵ Par la méthode de stratification spatiale, nous obtenons une erreur d'échantillonnage inférieure à la méthode d'échantillonnage simple entre les zones résidentielles.

Bibliographie

- [AGL 2002] - AGLIETTA M., BLANCHET D., HERAN F. (2002, coord.) : *Démographie et économie*. Rapport établi dans le cadre de la création du Conseil d'Analyse Economique, Ed. La Doc. Française, Paris, 343 p.
- [ARD 1994] - ARDILLY P. (1994) : *Les techniques de sondage*. Ed. Technip, Paris, 393 p.
- [CON 2001] - CONSEIL ECONOMIQUE ET SOCIAL (2001) : *Les personnes âgées dans la société*. Avis adopté par le C.E.S. au cours de la séance du 23 mai 2001, rapporteur Maurice BONNET, téléchargeable (novembre 2003) à : http://www.conseil-economique-et-social.fr/ces_dat2/24kiosq/rapan2001/Som1.htm. 60 p.
- [DEV 1992] - DEVILLE J.-C., SARNDAL C.-E. (1993) : Calibration estimators in survey sampling. *Journal of the American statistical Association*, n° 87, pp. 376-382.
- [DEV 1993a] - DEVILLE J.-C., DUPONT F. (1993) : Non réponse : principes et méthodes. In Actes des journées de méthodologie statistique des 15 et 16 déc. 1993. *INSEE Méthodes*, Ed. INSEE, Paris, pp. 53-70.
- [DEV 1993b] - DEVILLE J.-C., SARNDAL C.-E., SAUTORY O. (1993) : Generalized raking procedure in survey sampling. *Journal of the American statistical Association*, n° 88, pp. 1 013-1 020.
- [DEV 1999] - DEVILLE J.-C. (1999) : Estimation de la variance pour des statistiques et des estimateurs complexes : techniques de résidus et de linéarisation. *Techniques d'enquêtes*, n° 25, pp. 219-230.
- [GER 2003] - GERBER P., BOUSCH P. (2003, coord.) : *Personnes Agées Ville de Luxembourg. Rapport établi à la demande de la Ville de Luxembourg. Volume 1 : analyses. Volume 2 : annexes*. Décembre 2003, Differdange.
- [GUI 1999] - GUINCHARD-KUNSTLER P. (1999) : Vieillir en France. Enjeux et besoins d'une nouvelle orientation de la politique en direction des personnes âgées en perte d'autonomie. Rapport à Monsieur le Premier Ministre. Document téléchargeable (novembre 2003) à l'adresse : [http://www.inserm.fr/serveur/vieil.nsf/\(Web+Ressources+Attachments\)/vieillir%20en%20France/\\$File/vieillir%20en%20France.pdf](http://www.inserm.fr/serveur/vieil.nsf/(Web+Ressources+Attachments)/vieillir%20en%20France/$File/vieillir%20en%20France.pdf). 102 p.
- [KIS 1965] - KISH L. (1965) : *Survey Sampling*, Ed. J. Wiley and Sons, New York.
- [ROC 2000] - ROCHEFORT R. (2000) : *Vive le papy-boom*. Ed. O. Jacob, Paris, 291 p.
- [SAR 1992] - SARNDAL C.-E., SWENSSON B., WRETMAN J.H. (1992) : *Model assisted survey sampling*. Ed. Springer Verlag, New York, 695 p.
- [STA 2002] - STATISTIQUE CANADA (2002) : *Le cadre d'assurance de la qualité de Statistique Canada*. Ed. Statistique Canada, 37 p. Téléchargeable (novembre 2003) : <http://www.statcan.ca/francais/freepub/12-586-XIF/12-586-XIF02001.pdf>.
- [THE 1992] - THEVENET A. (1992) : *Le quatrième âge*. Ed. PUF, Coll. Que Sais-Je ?, Paris, 128 p.
- [VIM 2000] - VIMONT C. (2001) : *Le nouveau Troisième Age. Une société active en devenir*. Ed. Economica, Paris, 117 p.

Annexes

A₁ Degré de représentativité de l'échantillon au niveau des individus Kish selon l'âge et le sexe

Sexe	Age	Nb de personnes des ménages (BPVdL)	Echantillon (nombre de ménages)	Estimation pondérée (individus)	Erreur d'échantillonnage		
					Borne inférieure	Borne supérieure	Coefficient de variation
Masculin	- de 60 ans						
	60-64 ans	1 792	232	1 834	1 608	2 060	6,17
	65-74 ans	3 255	431	3 336	3 051	3 620	4,26
	75 ans ou +	2 955	454	3 036	2 793	3 279	4,00
Féminin	- de 60 ans						
	60-64 ans	1 648	237	1 641	1 445	1 837	5,98
	65-74 ans	2 579	292	2 508	2 238	2 779	5,40
	75 ans ou +	1 507	157	1 396	1 182	1 609	7,64
Total		13 736	1 803	13 750			

Source : BPVdL, 2001, enquête PAVdL, 2002

A₂ Répondants et non répondants au niveau des zones résidentielles

Classement par zones résidentielles	Population censée être interrogée pondérée (%)	Population effectivement interrogée pondérée (%)	Non répondants (effectifs)	Répondants (effectifs)	Taux de non-réponse observé (%)
Zone 1	15,41	14,28	135	142	48,7
Zone 2	15,43	15,31	126	153	45,2
Zone 3	14,85	14,07	128	141	47,6
Zone 4	19	20,70	136	206	39,8
Zone 5	19,34	17,89	170	178	48,9
Zone 6	15,96	17,74	111	177	38,5
Total	100	100	806	997	44,7

Source : BPVdL, 2001, enquête PAVdL, 2002

Variable ¹	Modalité	Modalité estimée	Erreur standard	Wald	Degré de signification
Groupes d'enquêteurs par classe de taux de réponse (r)	r < 0,5	-1,27	0,20	39,64	***
	0,5 < r < 0,53	-0,85	0,21	16,78	***
	0,53 < r < 0,573	-0,69	0,22	10,14	**
	0,573 < r < 0,61	-0,60	0,22	7,68	**
	0,61 < r < 0,66	-0,39	0,21	3,44	*
	r > 0,66	0	/	/	/
Zones résidentielles	zone 1	0,00	0,19	0,00	n.s.
	zone 2	-0,17	0,18	0,88	n.s.
	zone 3	-0,06	0,19	0,11	n.s.
	zone 4	0,08	0,18	0,19	n.s.
	zone 5	-0,11	0,18	0,40	n.s.
	zone 6	0	/	/	/
Taille des ménages	1 personne	-0,16	0,18	0,83	n.s.
	2 personnes	-0,05	0,17	0,09	n.s.
	3 personnes et +	0,00	/	/	/
Sexe par tranches d'âge	F de 60 à 64 ans	-0,09	0,22	0,17	n.s.
	F de 65 à 74 ans	0,00	0,20	0,00	n.s.
	F 75 ans et +	-0,19	0,20	0,93	n.s.
	M de 60 à 64 ans	-1,33	0,23	1,95	n.s.
	M de 65 à 74 ans	-0,15	0,21	0,55	n.s.
	M 75 et +	0,00	/	/	/
Nationalité	luxembourgeoise	0,09	0,12	0,58	n.s.
	autres nationalités	0	/	/	/
Etat civil	célibataires, divorcés	0,07	0,15	0,21	n.s.
	mariés, remariés	0,03	0,20	0,02	n.s.
	veufs	-1,27	0,20	39,64	/
Constante		1,12	0,32	12,19	0,0005

Source : BPVdL, 2001, enquête PAVdL, 2002

Notes de lecture :

¹ : variable dépendante : répondants/non répondants au questionnaire

*** : modalité significative au seuil de 1 %

** : modalité significative au seuil de 5 %

* : modalité significative au seuil de 10 %

n.s. : modalité non significative

Estimateur de calage

L'estimateur de calage [DEV 1992] [DEV 1993b] est asymptotiquement équivalent à un estimateur de régression. Ainsi, l'estimation d'un effectif Y est donné par :

$$\hat{Y}_{cal} = \sum_{k=1}^{903} w_k y_k = \hat{Y}_{HT} + \hat{\beta}'(X - \hat{X}_{HT})$$

où :

 $\hat{Y}$ = Y estimé et calé w_k = poids d'estimation y_k = variable mesurée au niveau de l'enquête

$$\hat{\beta} = (X'DX)^{-1}(X'DY) \quad \text{et} \quad D = \text{diag}\left(d_k \frac{1803}{903}\right) \quad \text{calcul matriciel où X est la}$$

matrice regroupant les variables qui ont servi à caler les poids.

CEPS/INSTEAD

B.P. 48

L-4501 Differdange

Tél. : 58 58 55-513

e-mail : isabelle.bouvy@ceps.lu

[http:// www.ceps.lu](http://www.ceps.lu)**statec**

B.P. 304

L-2013 Luxembourg

Tél. : 478-4276/4250

[http:// www.statec.lu](http://www.statec.lu)**Ville de Luxembourg**

Service Seniors

28, place Guillaume II

L-1648 Luxembourg

Tél. : 4796-2757

[http:// www.luxembourg-city.lu](http://www.luxembourg-city.lu)