Rapport de méthodes

Pondérations du Panel Suisse de Ménages

PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la statistique OFS

Neuchâtel, 2009

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0	Statistische Grundlagen und Übersichten
1	Bevölkerung
2	Raum und Umwelt
3	Arbeit und Erwerb
4	Volkswirtschaft
5	Preise
6	Industrie und Dienstleistungen
7	Land- und Forstwirtschaft
8	Energie
9	Bau- und Wohnungswesen
10	Tourismus
11	Mobilität und Verkehr
12	Geld, Banken, Versicherungen
13	Soziale Sicherheit
14	Gesundheit
15	Bildung und Wissenschaft
16	Kultur, Medien, Informationsgesellschaft, Sport
17	Politik
18	Öffentliche Verwaltung und Finanzen
19	Kriminalität und Strafrecht
20	Wirtschaftliche und soziale Situation der Bevölkerung
21	Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0	Bases statistiques et produits généraux
1	Population
2	Espace et environnement
3	Vie active et rémunération du travail
4	Economie nationale
5	Prix
6	Industrie et services
7	Agriculture et sylviculture
8	Energie
9	Construction et logement
10	Tourisme
11	Mobilité et transports
12	Monnaie, banques, assurances
13	Protection sociale
14	Santé
15	Education et science
16	Culture, Médias, société de l'information, sport
17	Politique
18	Administration et finances publiques
19	Criminalité et droit pénal
20	Situation économique et sociale de la population
21	Développement durable et disparités régionales et internationales

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Pondérations du Panel Suisse de Ménages

PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés

Auteur **Eric Graf**
Office fédéral de la Statistique

Editeur **Office fédéral de la statistique**

Neuchâtel, 2009

Préambule

L'enquête PSM est une enquête réalisée auprès des ménages et des personnes résidant de manière permanente en Suisse. Son objectif principal est d'observer le changement social, notamment la dynamique de l'évolution des conditions de vie de la population en Suisse. Le Panel suisse de ménages est initialement (de 1998 à 2007) un projet commun du Fonds National Suisse de la Recherche Scientifique, de l'Office Fédéral de la Statistique (OFS) et de l'Université de Neuchâtel.

Depuis 2008, le projet est financé par le Fonds National Suisse de la Recherche Scientifique et est mené dans le cadre de FORS, le Centre de Compétences Suisse en Sciences Sociales, sis à l'Université de Lausanne. L'OFS a reçut le mandat de réaliser les pondérations des enquêtes PSM réalisées en 2007-2008 (vague 9 du PSM_I et vague 4 du PSM_II).

Ce rapport détaille et documente la manière dont ces données ont été pondérées. Les concepts relatifs à la pondération d'enquêtes panel et de la combinaison d'enquête panels sont assez ardus. Il se peut que le lecteur non averti rencontre quelques difficultés de compréhension dans certains passages. Si nécessaire, on trouvera dans la bibliographie une liste de documents traitant de ces sujets plus en détails.

Le travail présenté ici est essentiellement une mise à jour du même rapport paru l'année passée sur les pondérations des vagues précédentes du PSM.

Toute notre gratitude va à Monique Graf (METH) pour ses remarques avisées après une lecture approfondie de ce rapport ainsi qu'à Robin Tillmann (PSM) pour ses commentaires et son soutien.

Résumé

La structure du document est la suivante: la façon dont les données ont été collectées est brièvement décrite pour chacun des panels, puis, après avoir présenté les taux de réponse atteints, on définit les différents concepts liés à la pondération. Ensuite, les étapes de la pondération sont présentées, puis décrites une à une pour chacun des panels également. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite aussi séparément les pondérations au niveau ménage de celles au niveau individuel. On relève aussi les spécificités des pondérations transversales par rapport aux pondérations longitudinales. Les dernières sections font l'objet des pondérations combinées du PSM_I vague 9 et PSM_II vague 4. Finalement, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes mentionnées que brièvement dans le texte principal.

Mots-clé

Rapport de méthode; taux de réponse; incitatifs; méthode de segmentation; non-réponse; calage; plan d'échantillonnage; combinaison de panel; partage des poids; pondération longitudinale; pondération transversale.

Complément d'information: Eric Graf, tél. 032 713 64 01
Eric.Graf@bfs.admin.ch

Réalisation: Service de méthodes statistiques, OFS

Diffusion: Office fédéral de la statistique
CH-2010 Neuchâtel
Tél. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Numéro de commande: 338-0057-05

Prix: gratuit

Série: Statistique de la Suisse

Domaine: 0 Bases statistiques et produits généraux

Langue du texte original: Français

Graphisme/Layout: OFS

Copyright: OFS, Neuchâtel 2009
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

ISBN: 978-3-303-00424-1

Tables des matières

1 INTRODUCTION : LE PANEL SUISSE DE MÉNAGES (PSM) 5

1.1 INTRODUCTION 5
1.2 LE PROBLÈME DE L'ATTRITION 7
1.3 STRUCTURE DE L'ENQUÊTE PSM 8

2 PLAN D'ÉCHANTILLONNAGE ET DÉROULEMENT DE L'ENQUÊTE 10

2.1 PLAN D'ÉCHANTILLONNAGE ET TAILLES BRUTES DES PSM_I ET PSM_II 10
2.2 UNIVERS DE L'ENQUÊTE 11

3 TAUX DE RÉPONSES ATTEINTS EN 2007 13

3.1 RÈGLES DE SUIVI 14
3.1.1 Règles de contact pour les ménages *14*
3.1.2 Règles de suivi pour les individus *15*
3.1.3 Particularités dès 2007 (vagues 9 et 4) *15*

4 CONCEPTS RELATIFS À LA PONDÉRATION 18

4.1 CONCEPTS RELATIFS À LA PRODUCTION DE POIDS LONGITUDINAUX 18
4.1.1 Population longitudinale cible et observée *18*
4.1.2 Unité longitudinale *18*
4.1.3 Cohabitants *19*
4.1.4 Éligibilité longitudinale *19*
4.2 CONCEPTS RELATIFS À LA PRODUCTION DES POIDS TRANSVERSAUX 19
4.2.1 Population transversale cible et observée *19*
4.2.2 Unité transversale *20*
4.2.3 Unité dans le champ de l'enquête transversalement *20*
4.2.4 Éligibilité transversale *20*
4.2.5 Unités répondantes *20*
4.3 ILLUSTRATION DES CONCEPTS PAR UN EXEMPLE 20

5 ÉTAPES DE LA PONDÉRATION 23

5.1 INTRODUCTION 23
5.2 TYPES DE PONDÉRATIONS PRODUITES 24

6 ÉTAPES COMMUNES POUR TOUS LES TYPES DE POIDS 27

6.1 MODÉLISATION DE LA NON RÉPONSE PAR LA MÉTHODE DE SEGMENTATION 27
6.1.2 La méthode de segmentation *27*
6.1.2 Segmentation et régression logistique *28*
6.1.3 Application dans le cadre du PSM *29*
6.2 POIDS INITIAL DU PSM_I ET POIDS INITIAL DU PSM_II 30
6.3 CLASSEMENT DES INDIVIDUS: ÉTAPE 2 31
6.4 AJUSTEMENT POUR LA NON-RÉPONSE – GRILLE: ÉTAPE 3 34

7 PONDÉRATION TRANSVERSALE INDIVIDUELLE 36

7.1 PARTAGE DES POIDS: ÉTAPE 4 36
7.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE INDIVIDUEL: ÉTAPE 5 37
7.3 CALAGE: ÉTAPE 6 38

8 PONDÉRATION TRANSVERSALE MÉNAGE 40

8.1 PARTAGE DES POIDS: ÉTAPE 7 40
8.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE MÉNAGE: ÉTAPE 8 41

8.3 CALAGE: ÉTAPE 9 42
8.4 WINSORISATION DES POIDS APRÈS CALAGE: ÉTAPE 9BIS 43
9 PONDÉRATION LONGITUDINALE INDIVIDUELLE 45
9.1 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE INDIVIDUEL: ÉTAPE 10 45
9.2 CALAGE: ÉTAPE 11 46
10 PONDÉRATION TRANSVERSALE INDIVIDUELLE COMBINÉE PSM_I ET PSM_II 48
10.1 COMBINAISON TRANSVERSALE DE PANELS À PARTIR DES ÉCHANTILLONS TRANSVERSAUX DE CHACUN DES DEUX PANELS 48
10.2 COMBINAISON DES DEUX PANELS: ÉTAPE 13 50
10.3 CALAGE: ÉTAPE 14 51
10.4 WINSORISATION DES POIDS APRÈS CALAGE: ÉTAPE 14BIS 51
11 PONDÉRATION TRANSVERSALE COMBINÉE MÉNAGE PSM_I ET PSM_II 53
11.1 COMBINAISON DES DEUX PANELS: ÉTAPE 15 53
11.2 CALAGE: ÉTAPE 16 54
11.3 WINSORISATION DES POIDS APRÈS CALAGE: ÉTAPE 16BIS 54
12 PONDÉRATION LONGITUDINALE INDIVIDUELLE COMBINÉE DU PSM_I & PSM_II 56
12.1 AJUSTEMENT POUR LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL POUR LE PSM_I, ANNÉE DE RÉFÉRENCE 2004 56
12.2 COMBINAISON DES PANELS: ÉTAPE 17 58
12.3 CALAGE: ÉTAPE 18 59
13 CONCLUSIONS 61
13.1 LIGNES DIRECTRICES CONCERNANT L'APPLICATION DES PONDÉRATIONS PRODUITES 61
13.2 RECOMMANDATIONS CONCERNANT LES ESTIMATIONS DE LA PRÉCISION 62
14 BIBLIOGRAPHIE 63
ANNEXE A – FICHIERS DE DONNÉES POUR PSM_I ET PSM_II 65
ANNEXE B - DÉRIVATION DES VARIABLES POUR LE CLASSEMENT DES INDIVIDUS 71
ANNEXE C – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE GRILLE 74
ANNEXE D – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION LONGITUDINALE INDIVIDUELLE 79
ANNEXE E – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION TRANSVERSALE INDIVIDUELLE 84
ANNEXE F – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE MÉNAGE, PONDÉRATION TRANSVERSALE MÉNAGE 89
ANNEXE G – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION LONGITUDINALE INDIVIDUELLE DU PSM_I ET PSM_II COMBINÉS 94
ANNEXE H – CARACTÉRISTIQUES DES DISTRIBUTIONS DES PONDÉRATIONS PRODUITES 97
ANNEXE I – TAUX DE RÉPONSE 104
ANNEXE J – ALGORITHME D'APPARTENANCE À LA POPULATION CIBLE 106

1 Introduction : le Panel Suisse de Ménages (PSM)

1.1 Introduction

L'enquête PSM est une enquête réalisée auprès des ménages et des personnes résidant de manière permanente en Suisse. Son objectif principal est d'observer le changement social, notamment la dynamique de l'évolution des conditions de vie de la population en Suisse. Le Panel suisse de ménages est initialement (de 1998 à 2007) un projet commun du Fonds National Suisse de la Recherche Scientifique, de l'Office Fédéral de la Statistique (OFS) et de l'Université de Neuchâtel.
Depuis le début 2008, le projet PSM, toujours financé par le Fonds National Suisse de la Recherche Scientifique, est mené dans le cadre de FORS, le Centre de Compétences Suisse en Sciences Sociales, sis à l'Université de Lausanne.

Depuis sa création en 1999, l'enquête Vivre en Suisse porte sur un vaste ensemble de sujets et d'approches dans le domaine des sciences sociales. L'enquête est menée annuellement de septembre à février par l'institut M.I.S-Trend de Lausanne et Berne selon la technique de l'interview téléphonique assistée par ordinateur (CATI). Elle constitue une base de données longitudinale sans équivalent pour la Suisse.

En ce qui concerne les collaborations internationales (voir [1], [4], [5], [6]), le PSM a participé au Consortium of Household Panels for European Socio-Economic Research (CHER). Les enquêtes pilotes (2004 et 2005) du SILC (Statistics on Income and Living Conditions d'Eurostat) ont été menées par l'OFS en étroite collaboration avec le PSM. Actuellement, le PSM fait partie du Cross-National Equivalent File (CNEF).

Comme son nom l'indique, le PSM est un panel, i.e. que les mêmes personnes et ménages sont réinterrogés année après année et répondent, à quelques variations près, aux mêmes questions. C'est un panel simple, par opposition à un panel rotatif, il n'y a donc pas d'activations régulières de nouvelles unités à enquêter, ni d'abandons volontaires de certaines autres unités. L'enquête PSM voit pour l'instant évoluer deux échantillons en parallèle au fil du temps, nous les noterons dans la suite par :

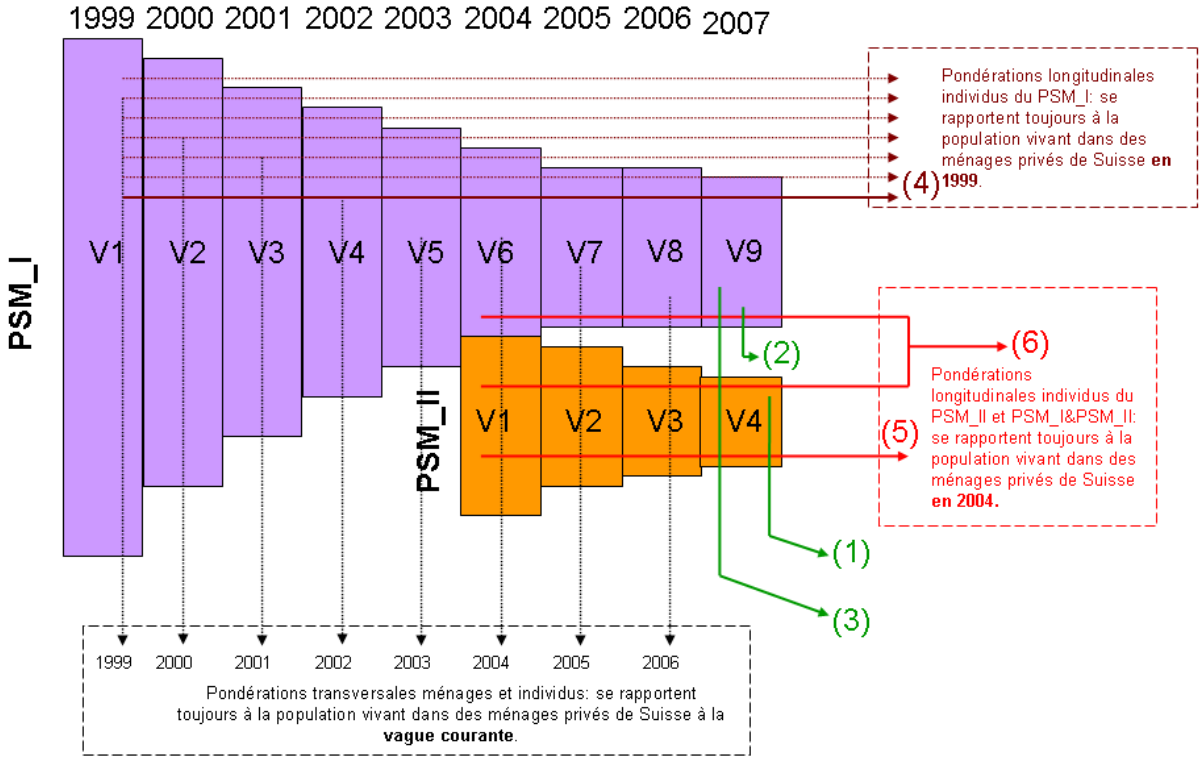
- ***PSM_I vague 9*** : il s'agit de l'échantillon des ménages et personnes sélectionnés en 1999 et interrogés pour la première fois cette année-là. Ce document traite de la pondération de la 9e vague d'interviews de ces unités,
- ***PSM_II vague 4*** : il s'agit de l'échantillon des ménages et personnes sélectionnés et interrogés pour la première fois en 2004. Ce document traite de la pondération de la 4e vague d'interviews de ces unités.

En 2000, la vague 1 du PSM_I (données récoltées en 1999) fut pondérée par des méthodologues de l'OFS (cf. [2]). Après quoi, jusqu'en 2004 (vagues 6 du PSM_I et 1 du PSM_II), le calcul des pondérations a été réalisé entièrement par Statistique Canada (cf. [4]). En 2005 (vagues 7 et 2) il fut effectué en collaboration entre l'auteur de ce document et Statistique Canada. (cf. [3]). Pour 2006 (cf. [24]) et 2007 c'est l'auteur qui a mené à bien cette tâche. L'objet de ce rapport est une version mise à jour[1] du même rapport paru l'année passée concernant 2006.

[1] Pour faciliter la comparaison entre les pondérations effectuées pour 2006 et pour 2007, on a gardé dans ce rapport la même numérotation des tableaux que pour l'année passée. C'est pourquoi la numérotation des tableaux de ce document passe de 8 à 16...

Le document dont il est question ici décrit de façon détaillée le processus de production des poids pour l'enquête PSM de 2007[2], à savoir, celles du PSM_I vague 9, du PSM_II vague 4 ainsi que la pondération combinée des échantillons PSM_I et PSM_II pour 2007. La figure 1 donne un aperçu des neuf différents jeux de poids produits.

Figure 1 Schéma des neuf différents jeux de poids produits en 2007 : trois pondérations transversales individus, trois pondérations transversales ménages et trois pondérations longitudinales individus.



Légende

(1) Pondérations transversale individuelle et transversale ménage du PSM_II
(2) Pondérations transversale individuelle et transversale ménage du PSM_I
(3) Pondérations transversale individuelle et transversale ménage combinées des PSM_I et PSM_II réunis
(4) Pondération longitudinale individuelle du PSM_I
(5) Pondération longitudinale individuelle du PSM_II
(6) Pondération longitudinale individuelle combinée des PSM_I et PSM_II réunis

Les pondérations décrites dans ce document sont, à quelques détails près, très semblables à celles concernant l'année 2006 et à celles qui furent effectuées par Statistique Canada pour les années précédentes et parfaitement compatibles avec ces dernières.

[2] Pour alléger le texte, nous dirons « réalisé en 2007 » pour l'enquête qui s'est déroulée de septembre 2007 à février 2008.

1.2 Le problème de l'attrition

La figure 2 et le tableau 3 montrent les pertes d'individus et de ménages que le PSM a subi depuis son commencement.
C'est pour compenser cette attrition de l'échantillon initial qu'en 2004 on a débuté un deuxième panel (le PSM_II), parallèlement au premier. On a aussi, dès 2004, pris toutes sortes de mesures pour lutter contre l'attrition (incitatifs, formations supplémentaires des interviewers, réactivation de ménages n'ayant plus participé à l'enquête depuis plus de deux années, etc.). Pour plus de détails concernant l'évolution de l'enquête sur le terrain, nous renvoyons le lecteur aux références [5] et [6].
Pour les règles de suivi des personnes, voir le paragraphe 3.1, et pour les taux de réponses atteints en fin de période de terrain, voir le chapitre 3 et l'annexe I.

Figure 2 Attrition des enquêtes PSM_I et PSM_II. Nombre de questionnaires individuels et ménages complétés chaque année.

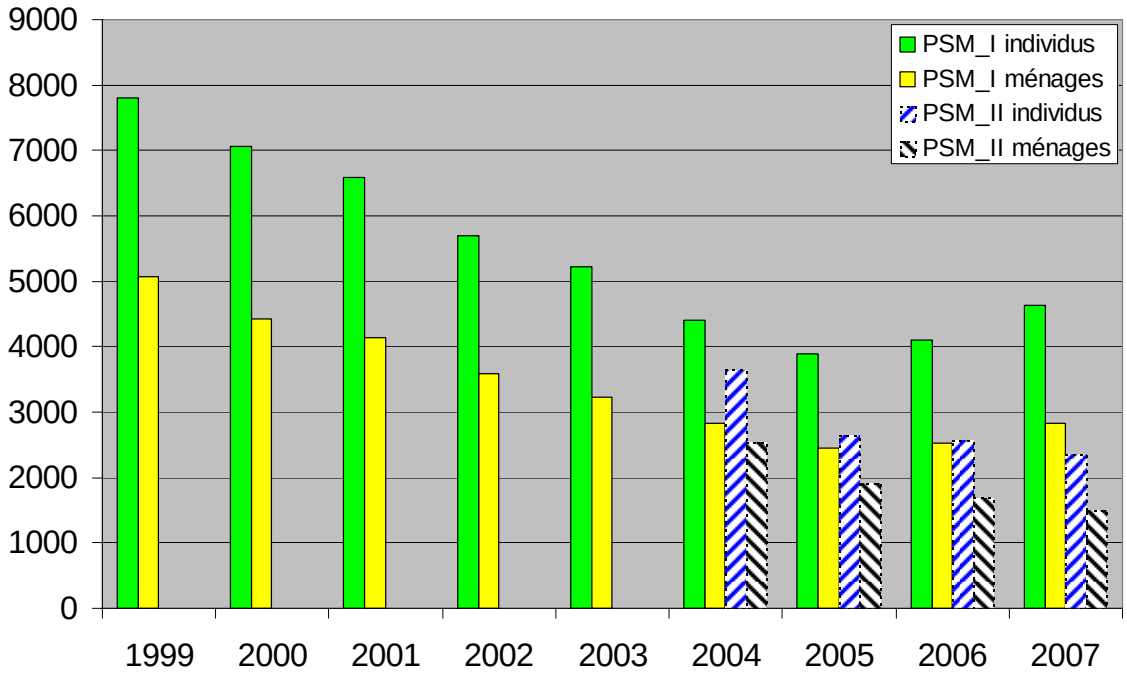


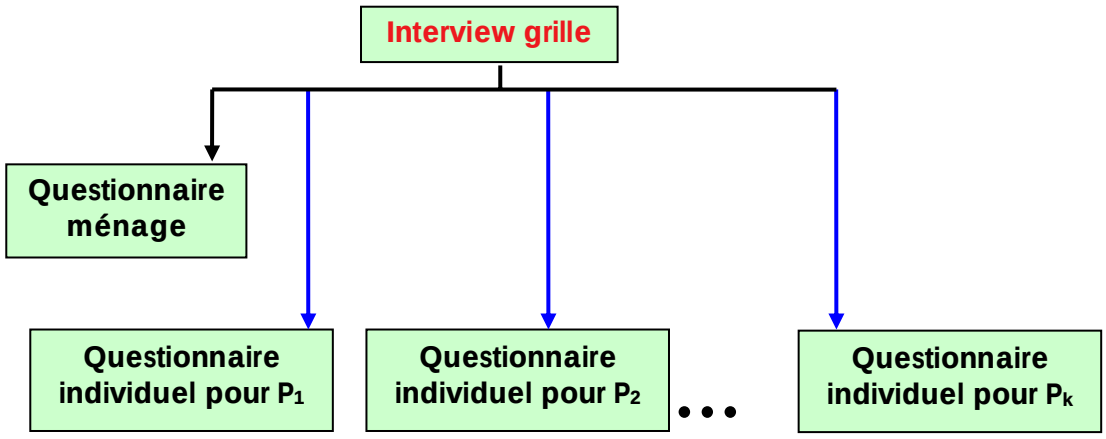
Tableau 3 Nombre de questionnaires individuels/ménages complétés dans l'enquête PSM. L'attrition relative est la proportion du nombre de cas interrogés l'année n sur le nombre de cas l'année (n-1).
Notons qu'il s'agit ici des nombres utilisés pour le calcul des pondérations. Ces totaux peuvent quelque peu différer de ceux publiés par le PSM. En effet, les pondérations se basent d'une part sur les statistiques et des règles de suivi et, d'autre part, leur calcul débute souvent avant que les données de l'enquête soient considérée comme définitives pour toutes les variables,

	Niveau individus			*Niveau ménage*		
Année	PSMI	Attrition relative	Attrition absolue	PSMI	Attrition relative	Attrition absolue
1999	7799	1.00	1.00	5074	1.00	1.00
2000	7071	0.91	0.91	4425	0.87	0.87
2001	6593	0.93	0.85	4139	0.94	0.82
2002	5691	0.86	0.73	3581	0.87	0.71
2003	5220	0.92	0.67	3227	0.90	0.64
2004	4406	0.84	0.56	2831	0.88	0.56
2005	3885	0.88	0.50	2457	0.87	0.48
2006	4093	1.05	0.52	2534	1.03	0.50
2007	4634	1.13	0.59	2822	1.11	0.56
	PSMII			PSMII		
2004	3648	1.00	1.00	2538	1.00	1.00
2005	2648	0.73	0.73	1909	0.75	0.75
2006	2570	0.97	0.70	1682	0.88	0.66
2007	2351	0.91	0.64	1495	0.89	0.59

1.3 Structure de l'enquête PSM

Le PSM est une enquête qui comporte 3 questionnaires : la grille, le questionnaire ménage et le questionnaire individuel.

Figure 4 Structure de l'enquête PSM



Le questionnaire grille consiste en une sorte de screening (ou inventaire) du ménage, il dure moins de 10 minutes et permet, entre autre, de récolter les informations sociodémographiques de base sur tous les membres du ménage ainsi que de

déterminer leur statut d'éligibilité aux interviews. Ni le questionnaire ménage, ni aucun des questionnaires individuels ne peuvent être complétés avant que le questionnaire grille ne soit réalisé. En principe c'est la *personne de référence*[3] qui répond au questionnaire grille, mais si cette dernière n'est pas disponible au premier contact, un autre adulte (de 14+ ans) peut le faire.
Une fois que la grille est complétée l'interviewer peut passer aux questionnaires ménage et individuel(s). Le questionnaire ménage est de préférence rempli par la personne de référence ou par un adulte qui « connaît bien son ménage ». Toute personne de 14 ans ou plus vivant dans le ménage est éligible pour le questionnaire individuel.

Une description des fichiers de données utilisés pour la pondération de la vague 9 du PSM_I et de la vague 4 du PSM_II ainsi que leur combinaison se trouve dans l'annexe A.

[3] La personne de référence doit être éligible et « bien » connaître son ménage. C'est en principe elle qui répond au questionnaire ménage année après année.

2 Plan d'échantillonnage et déroulement de l'enquête

2.1 Plan d'échantillonnage et tailles brutes des PSM_I et PSM_II

Les échantillons de PSM_I et PSM_II sont stratifiés par grande région géographique (NUTS II), proportionnellement au nombre de ménages par strate. Les adresses de l'échantillon brut furent réparties dans les proportions suivantes (recensements 1990 et 2000):

Tableau 5 Stratification de l'échantillon brut. Les proportions sont données par $P_h^{99} = \frac{N_h^{99}}{N^{99}}$ et $P_h^{04} = \frac{N_h^{04}}{N^{04}}$ (voir aussi tableau 6).

Strates *h*	Cantons	Proportion d'adresses P_h^{99} (en %) PSM_II	Proportion d'adresses P_h^{04} (en %) PSM_II
1: Région lémanique	VD, VS, GE	18.45	18.22
2: Espace Mitteland	BE, FR, SO, NE, JU	23.25	22.92
3: Suisse du nord-ouest	BS, BL, AG	13.44	13.86
4: Zurich	ZH	17.51	18.22
5: Suisse orientale	GL, SH, AR, AI, SG, GR, TG	15.68	13.70
6: Suisse centrale	LU, UR, SZ, OW, NW, ZG	7.20	8.75
7: Tessin	TI	4.47	4.33
Total		100%	100%

Le tirage est donc **uniquement** proportionnel au nombre de ménages par grande région et **ne tient pas compte** du nombre moyen de personnes dans les ménages par région. On n'augmente pas la taille du Tessin.
Le cadre de sondage était le SRH[4] actualisé trimestriellement en mars, juin, septembre et décembre. Le PSM_I a été tiré dans le SRH de 02/99, les PSM_II dans le SRH 02/04. Les tailles des strates au moment du tirage pour PSM_I et PSM_II étaient les suivantes :

Tableau 6 Tailles (SRH) N_h des strates h (h = 1,..,7) au moment des deux tirages

Strates *h*	PSM_I, 1999 N_h^{99}	PSM_II, 2004 N_h^{04}
1: Région lémanique	714'725	648'590
2: Espace Mitteland	837'452	784'266
3: Suisse du nord-ouest	484'667	455'833
4: Zurich	646'469	587'850
5: Suisse orientale	531'731	493'606
6: Suisse centrale	313'548	306'605
7: Tessin	180'623	160'123
TOTAL (N=)	**3'709'215**	**3'436'873**

[4] Registre pour les enquêtes auprès des ménages de l'OFS.

Le tableau 7 indique la répartition de l'échantillon brut, i.e. le nombre d'adresses activées pour l'enquête par strate, noté S_h, en vague 1.

Tableau 7 Tailles S_h de l'échantillon brut par strates. Les S_h représentent le nombre d'adresses activées pour l'enquête. Ces adresses ne correspondent pas forcément à des ménages privés, mais pour la plupart c'est le cas. Nous renvoyons le lecteur à l'annexe H pour une description détaillée du statut final attribué à chaque adresse.

Strates *h*	PSM_I, v1 1999 S_h^I	PSM_II, v1 2004 S_h^{II}
1: Région lémanique	2615	912
2: Espace Mitteland	3296	1146
3: Suisse du nord-ouest	1904	692
4: Zurich	2482	912
5: Suisse orientale	2223	686
6: Suisse centrale	1021	436
7: Tessin	633	216
TOTAL	**14174**	**5000**

2.2 Univers de l'enquête

Le PSM couvre toutes les personnes résidant dans des ménages privés de Suisse et possédant un raccordement téléphonique (fixe ou mobile) inscrit dans l'annuaire téléphonique. Les personnes vivant dans des homes, dans des institutions, dans des ménages collectifs, en prison ne sont pas couvertes par l'enquête. Les personnes non joignables par téléphone ou possédant un raccordement téléphonique non inscrit ne sont pas couvertes par l'enquête (sous-couverture).
On estime qu'en 2004, lors de la sélection de l'échantillon du PSM_II, 98.5% des ménages privés possédaient un raccordement téléphonique. Par ailleurs le SRH (base de sondage pour les enquêtes auprès des ménages) couvrait quant à lui environ 93% de ces ménages. En 1999, lors de la sélection de l'échantillon du PSM_I, la couverture du SRH était d'environ 95%.

La base de sondage SRH est sujette aux erreurs suivantes :

- *sous-couverture* : certains ménages ne figurent pas dans l'annuaire au temps de la sélection. Par exemple les ménages dont aucun de leurs numéros sont inscrits, ou les ménages non joignables par téléphone. Ce problème peut produire des biais, c'est-à-dire des différences entre les estimations produites à partir de la population observée (l'enquête PSM) et celles qu'on aurait produites à partir de la population cible (l'ensemble des ménages privés résidant en Suisse à un certain moment fixé dans le temps).
- *doublons :* malgré les procédures rigoureuses de nettoyage du SRH visant à ne garder qu'un numéro de téléphone par ménage, certains ménages figurent plus d'une fois dans la base de sondage. Le problème conduit à une détermination erronée de certaines des probabilités initiales de sélection. Malgré cela, on ne calcule pas de facteur de correction pour les ménages

avec plusieurs lignes téléphoniques, l'information serait disponible, mais l'effet est négligeable.

- *sur-couverture* : sélection d'unités hors de la population cible (entreprises, homes, prisons, ménages collectifs, résidences secondaires, ...). Notons que pour un panel, ce problème n'est rencontré qu'à la vague 1 et qu'en principe ces adresses-là sont considérées comme hors échantillons (*non sample cases*).

3 Taux de réponses atteints en 2007

Le tableau 8 résume les taux de réponse obtenus par rapport à l'échantillon brut, i.e. par rapport au nombre d'adresses activées pour les enquêtes indépendamment de leur statut final (validité). Relevons qu'une vague 1 se comporte de manière très différente qu'une vague subséquente. Pour plus de détails quant aux taux de réponse, se référer à l'annexe I.

Tableau 8 Taux de réponse aux différents niveaux de collecte des données pour le PSM_I en 1999 (vague 1) et 2007 (vague 9), ainsi que pour le PSM_II en 2004 (vague 1) et 2007 (vague 4) (source: statistiques finales M.I.S-Trend SA)[5].

		PSM_I vague 1, 1999				PSM_I vague 9, 2007			
Niveau ménages	Echantillon brut (nombre d'adresses activées)	14174	100%			3976	100%		
	Nombre de grilles réalisées	7063	50%	100%		2898	73%	100%	
	Grilles + quest. Ménage réalisés	6001	42%	85%	100%	2822	71%	97%	100%
	Nombre de ménages complets[6]	5074	36%	72%	85%	2770	70%	96%	98%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	10293	100%			5974	100%		
	Nombre de quest. Individuel réalisés	7799	76%			4634	78%		

		PSM_II vague 1, 2004				PSM_II vague 4, 2007			
Niveau ménages	Echantillon brut (nombre d'adresses activées)	5000	100%			2037	100%		
	Nombre de grilles réalisées	2704	54%	100%		1549	76%	100%	
	Grilles + quest. Ménage réalisés	2535	51%	94%	100%	1495	73%	97%	100%
	Nombre de ménages complets	2335	47%	86%	92%	1457	71%	94%	97%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	5384	100%			3108	100%		
	Nombre de quest. Individuel réalisés	3652	68%			2351	76%		

Comme en 2006, on a utilisé des incitatifs et les mesures suivantes pour gagner des répondants:

- un **remerciement** pour les répondants: tous les répondants au questionnaire individuel ont pu choisir entre un bon CFF de 15.- francs (18%), un bon pour

[5] Comme mentionné au chapitre précédant, les totaux présentés ici peuvent légèrement différer des nombres publiés par le PSM.

[6] Un ménage est dit « complet » si les questionnaires grille, ménage et au moins un questionnaire individuel ont été réalisés.

une entrée de cinéma (33%) ou un don de 15.- francs à une association caritative (49%),
- préalablement aux appels de refus, envoi d'un courrier adressé soit au ménage dans le cas d'un refus ménage, soit à une personne en particulier si le refus a eu lieu au niveau individuel,
- possibilité pour les enquêteurs d'interroger toutes les personnes du foyer au niveau individuel, indépendamment de leur décision des années précédentes de participer ou non à l'enquête,
- deux primes pour les enquêteurs en fonction d'objectifs atteints ou non, l'une basée sur le taux de participation général et l'autre sur le taux de conversion des refus,
- informations transmises régulièrement aux enquêteurs sur l'avancée des travaux.

De plus, on a réactivé 1'354 anciennes adresses de répondants pour la dernière fois aux vagues 2, 3 et 4, tout comme on l'a fait en 2006 en réactivant d'anciennes adresses de répondants pour la dernière fois en vague 5, ce procédé s'étant révélé fructueux tant du point de vue du nombre d'interrogés final que pour mettre un frein à l'attrition.

D'autre part, tous les ménages dont l'adresse n'a pas été activée au démarrage de la prise d'information, mais plus tard et directement en appel de refus, ont reçu un bon CFF de 50.- francs avec le courrier préliminaire. Alors que l'an dernier le taux de réussite pour les ménages W5 réactivés étaient de 35 % de ménages complets, il est passé à 52 % pour les ménages perdus en W2, W3 ou W4, la différence étant probablement due non seulement au fait que nous avons laissé plus longtemps tranquille ces foyers, mais également justement au bon CFF de 50.- francs qui accompagnait le courrier préliminaire.

3.1 Règles de suivi

Bien que le panel suive des personnes et non des ménages, le premier contact se fait toujours avec (une personne dans) un ménage! Dès la vague 2, lorsqu'on retente un premier contact sur les base des informations récoltées l'année précédente, on ne sait pas encore si le ménage qu'on contacte contient encore des longitudinaux ou pas. C'est pourquoi, on a des règles de suivi pour les ménages déterminant quels ménages recontacter par rapport à l'année précédente, et des règles de suivi pour les personnes pour savoir qui interroger réellement.

3.1.1 Règles de contact pour les ménages

La règle générale est d'interroger tous les ménages qui ont rempli au moins une grille à la vague précédente. On procède aux interviews pour autant qu'un individu longitudinal soit d'accord de remplir le questionnaire ménage et son questionnaire individuel (il est toujours possible de rattraper les autres individus lors des vagues futures). Mais:

1. On abandonne définitivement pour les vagues suivantes les ménages non-contactés du tout lors de la 1ère vague, ou ceux n'ayant fourni aucune information à la 1ère vague (même pas une grille), ou ceux n'ayant rempli qu'un questionnaire non-réponse à la vague 1.

2. ***Pour le PSM_I*** on a abandonné définitivement tous les ménages n'ayant répondu qu'à la grille de la vague 1. ***Pour le PSM_II***, on a changé cette règle et abandonné définitivement un ménage qui, à la vague 1, n'a fait qu'une grille, que si on ne collecte pas plus que sa grille à la vague 2.
3. On abandonne définitivement les ménages qui ne répondent pas pour deux vagues successives, ceux qui donnent un « refus final » (ménages où personne n'est prêt à répondre à l'interview de ménage après toutes les tentatives de reconversion), ceux qui déménagent hors de Suisse et ceux qui sont entièrement et définitivement institutionnalisés.

3.1.2 Règles de suivi pour les individus

4. Les répondants longitudinaux (OSM[7]) et les enfants issus de ceux-ci sont suivis en tous temps, tandis que les cohabitants[8] ne sont (ré-)interrogés que tant qu'ils vivent avec un OSM.
5. La limite d'éligibilité est fixée à 14 ans.
6. En règle générale, les répondants longitudinaux (OSM) sont toujours suivis jusqu'à ce qu'ils meurent ou soient définitivement institutionnalisés (ou sortent de la population cible pour une autre raison).
7. On abandonne définitivement les individus donnant un « refus final » à deux vagues consécutives ainsi que ceux communiquant un refus écrit.

3.1.3 Particularités dès 2007 (vagues 9 et 4)

Il a été décidé de libérer, pour la vague 9, les adresses de tous les ménages complets interrogés en vague 2 (2001), 3 (2002) et 4 (2003), y compris ceux qui avaient été abandonnés depuis pour cause de double refus (i.e. refus à deux vagues successives). On n'a cependant pas essayé de recontacter les ménages qui avaient été abandonné suite à un refus écrit. Plus de 1500 ménages ont ainsi été «réintégrés» à l'échantillon brut, 40% d'entre eux ont accepté de participer à nouveau à l'enquête. Cette opération a donc été un succès.

De plus, tout comme en 2006 la règle numéro 7 ci-dessus a été abolie. Ainsi on a donné la possibilité aux enquêteurs d'interroger tous les membres d'un ménage, sans tenir compte de leurs refus antérieurs.

En outre, en ce qui concerne le suivi des adresses et des numéros invalides, M.I.S-Trend SA a eu recours aux services de la société *Dcl.data care* spécialisée dans le traçage des déménagements et autres changements d'adresses.

[7] OSM = Original Sample Member.
[8] Pour plus de précisions sur les concepts de longitudinaux et de cohabitants, voir au chapitre 4.

Du point de vue de la gestion des adresses, en 2007, les règles de suivi étaient les suivantes :

Enquête	Description	Décision
PSM_I	1. Ménages complets en vague 8 ou seulement le questionnaire ménage ou seulement un questionnaire individuel	Activer
	2. Grilles seulement en vague 8 ne provenant pas des *ex- répondants réactivés de vague 5* et avec au moins le questionnaire ménage ou un questionnaire individuel en vague 8	Activer
	3. Grilles seulement en vague 8 et provenant des *ex-répondants réactivés de vague 5*	Abandonner définitivement
	4. Non-répondants totaux en vague 8 et provenant des *ex- répondants réactivés de vague 5*	Abandonner définitivement
	5. Non-répondants totaux en vague 8 mais au moins le questionnaire ménage ou un questionnaire individuel en vague 7	Activer en appel de refus avec interviewers spécialisés
	6. Non-répondants totaux en vague 8 et seulement questionnaire grille réalisée en vague 7	Activer en appel de refus avec interviewers spécialisés
	7. Non-répondants totaux en vague 8 et non-répondants totaux en vague 7	Abandonner définitivement
	8. Ménages complets pour la dernière fois en vague 2, 3 ou 4 ET : pas déjà pris dans les catégories 3 ou 4 ET : pas abandonnés à cause d'un refus écrit	Activer en appel de refus avec interviewers spécialisés
PSM_II	9. Ménages complets en vague 3 ou seulement le questionnaire ménage ou seulement un questionnaire individuel	Activer
	10. Questionnaire grille seulement en vague 3 et avec au moins le questionnaire ménage ou un questionnaire individuel en vague 2	Activer
	11. Non-répondants totaux en vague 3 mais au moins le questionnaire ménage ou un questionnaire individuel en vague 2	Activer en appel de refus avec interviewers spécialisés
	12. Non-répondants totaux en vague 3 et seulement questionnaire grille en vague 2	Activer en appel de refus avec interviewers spécialisés
	13. Non-répondants totaux en vague 3 et non-répondants totaux en vague 2	Abandonner définitivement

Tous les refus écrits reçus de chaque année, les décès ainsi que les adresses non-retrouvées à la suite des retours du courrier préliminaire n'ont par contre jamais été réactivées.

La règle de suivi n°8 a, tout comme l'année précédente, eu un effet favorable sur le nombre absolu d'interviews réalisées aux niveaux ménages et individus (voir figure 2 et tableau 3). C'est ce qui explique les nombre de personnes et de ménages interrogés en 2007 supérieurs à ceux de 2006 (tout comme ceux de 2006 étaient supérieurs à ceux de 2005 pour les mêmes raisons)!

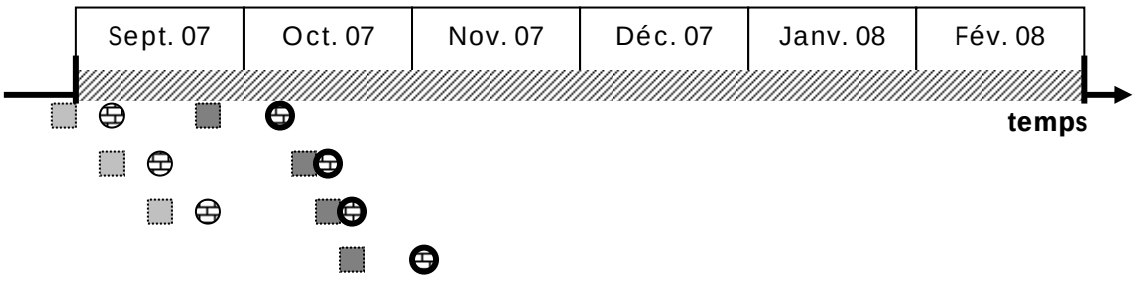
3.2 Déroulement des enquêtes PSM_I vague 9 et PSM_II vague 4

La période de terrain a duré du 3 septembre 2007 au 1er mars 2008, soit 6 mois. Comme lors de la vague précédente, le fichier d'adresses à activer dans l'enquête a été divisé en trois fichiers comprenant grosso modo le même nombre d'adresses afin de contacter un maximum de foyers dans la semaine qui suit l'envoi du courrier préliminaire. Tous les ménages ont reçu préalablement un courrier émanant du PSM leur annonçant l'appel imminent d'un enquêteur de M.IS-Trend (voir tableau 16).
L'an dernier, une brochure expliquant aux ménages l'importance de leur participation répétée chaque année avait été jointe au courrier préliminaire. Ce ne fut pas le cas cette année, sauf pour les courriers destinés aux refusants des années précédentes ainsi qu'aux ménages perdus il y a quelques années et réactivés pour cette vague.
D'autre part, tous les ménages dont l'adresse n'a pas été activée au démarrage de la prise d'information, mais plus tard et directement en appel de refus (voir paragraphe 3.1.3), ont reçu un bon CFF de 50.- francs avec le courrier préliminaire.

Pour une description plus précise et complète du déroulement des enquêtes et du travail de pilotage avec l'institut de sondage, nous renvoyons le lecteur aux références [5], [6] et [25].

Tableau 16 Dates d'envois de courrier et d'activation d'adresses. Les envois du courrier préliminaire ont eu lieu à chaque fois environ une semaine avant la libération du nouveau bloc d'adresses. Ces courriers ont été envoyés avant chaque activation d'adresses (aussi pour les appels de refus).

Envois du courrier préliminaire	Activation des échantillons d'adresses	Envois des courriers appels de refus	Activation des échantillons d'adresses appels de refus
31 août 2007	5 sept. 2007	26 sept. 2007	4 oct. 2007
7 sept. 2007	13 sept. 2007	10 oct. 2007	16 oct. 2007
14 sept. 2007	19 sept. 2007	16 oct. 2007	23 oct. 2007
		23 oct. 2007	1er nov. 2007



4 Concepts relatifs à la pondération

On a ici brièvement résumé la terminologie et la philosophie relatives à la manière de construire les pondérations actuelles du PSM. Les lecteurs non-familiers avec ces concepts peuvent se référer, par exemple, à [9], [15], [16], [17] et [18] pour étudier le sujet de manière plus approfondie.

4.1 Concepts relatifs à la production de poids longitudinaux

Les pondérations transversales et longitudinales ont été effectuées essentiellement de la même manière que ce qu'a fait Statistique Canada les années précédentes. Les concepts relatifs à la pondération sont présentés ici dans le contexte du PSM. A la fin de ce chapitre on trouvera un exemple illustrant les aspects théoriques abordés ci-dessous et difficiles à assimiler pour le lecteur non-familier.

4.1.1 Population longitudinale cible et observée

La *population longitudinale cible* est composée de tous les individus vivant dans des ménages privés de Suisse en septembre 1999 pour le PSM_I, en septembre 2004 pour le PSM_II, date à laquelle la collecte des données a commencé. La base de sondage utilisée est le SRH (voir chapitre 2).
La *population longitudinale observée* est composée de tous les individus vivant dans des ménages privés de Suisse qui avaient un numéro de téléphone valide, inscrit et répertorié dans le SRH 02/99 pour le PSM_I, respectivement dans le SRH 02/04 pour le PSM_II, date à laquelle l'échantillon a été sélectionné.

4.1.2 Unité longitudinale

À la sélection de l'échantillon, toutes les personnes faisant partie des ménages choisis et étant âgées de 14 ans et plus lors de la sélection, deviennent des membres de l'échantillon longitudinal, indépendamment du fait qu'ils soient répondants ou non répondants au questionnaire individuel, et sont appelés « longitudinaux ». Ces personnes sont considérées comme des membres de l'échantillon longitudinal pendant la durée de vie du panel.
En pratique, la règle de suivi des longitudinaux était que leur ménage ait au moins répondu au questionnaire grille en vague 1. Par souci de simplification on a, dans les procédures de pondérations, appelé longitudinaux, les individus de l'échantillon longitudinal qui vivaient dans des ménages ayant au moins répondu à la grille en vague 1.

Si une personne longitudinale déménage, décède ou quitte la population cible, elle fera toujours partie de la population longitudinale cible et aura un poids longitudinal. En effet, les totaux de contrôle utilisés pour le calage considèrent toutes les unités de la population de la vague 1 indépendamment de ce qui peut leur arriver dans le futur. On se doit donc d'être conséquent et d'assigner un poids à toutes les unités même si elles n'ont aucune utilité analytique. De plus, il est avantageux que toutes les estimations produites soient cohérentes avec celles provenant d'autres sources. De cette manière, il est possible d'évaluer la représentativité de l'échantillon au fil du temps.

En principe, aucune autre personne ne devient membre de l'échantillon longitudinal du panel[9]. Ainsi, l'échantillon longitudinal est constitué au moment de l'introduction du panel, à la vague 1, et reste tel quel pendant toute la durée du panel. La population ciblée par l'échantillon longitudinal demeurera la même pour toute la durée de vie du panel. Du point de vue longitudinal, l'unité d'intérêt est l'individu; le ménage pouvant difficilement servir d'outil d'analyse longitudinale à cause de sa nature dynamique dans le temps. De plus, le calcul d'un poids longitudinal au niveau du ménage demanderait un développement théorique supplémentaire. En effet, un grand nombre de définitions de ménage longitudinal sont à considérer avant de faire un choix.

4.1.3 Cohabitants

On appelle *cohabitants* les individus qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante. Les cohabitants sont classés en deux catégories:

- les cohabitants *initialement présents*, qui faisaient partie de la population transversale cible au moment de la sélection du panel,
- les cohabitants *initialement absents*, qui se sont joints à la population transversale cible entre la première vague et la vague courante. Ces derniers sont des immigrants.

4.1.4 Éligibilité longitudinale

Une personne de l'échantillon original est éligible pour la *pondération longitudinale* si elle faisait partie de la population cible lors de la sélection de l'échantillon original (en 1999 pour le PSM_I, en 2004 pour le PSM_II) et a rempli le questionnaire individuel pour l'année courante.
Notons que cette règle présente l'inconvénient pour les utilisateurs de données qu'un individu longitudinal non-répondant en vague 1 reçoit un poids de zéro en vague 1, mais s'il répond à la vague 2, il aura un poids longitudinal en vague 2.

4.2 Concepts relatifs à la production des poids transversaux

4.2.1 Population transversale cible et observée

La *population transversale cible* pour la vague 8 du PSM_I et la vague 3 du PSM_II est composée des ménages privés de Suisse existant en septembre 2006.
La *population transversale observée* pour la vague 8 du PSM_I est composée des ménages privés vivant en Suisse en septembre 2006 et dont au moins un membre faisait partie de la population longitudinale observée. En particulier, cette population exclut donc les ménages composés exclusivement d'immigrants qui sont arrivés en Suisse après septembre 1999 pour le PSM_I, après septembre 2004 pour le

[9] Cependant, comme cela a été fait l'année passée, on a enfreint à la règle et considéré les enfants de longitudinaux devenus adultes comme des longitudinaux et non comme des cohabitants initialement absent. Ceci pour ne pas abandonner et perdre les ménages où des enfants de moins de 14 ans à la vague 1 qui ont maintenant plus de 14 ans sont les seuls répondants ou représentants longitudinaux d'un ménage.

PSM_II, ainsi que certaines personnes qui étaient institutionnalisées ou non sélectionnables pour une autre raison en septembre 2006.

4.2.2 Unité transversale

Du point de vue transversal, on s'intéresse aussi bien aux individus âgés de 14 ans et plus qu'aux ménages. Le PSM interviewe toutes les personnes habitant avec au moins un individu longitudinal. L'échantillon transversal pour une année donnée est alors constitué de toutes les personnes longitudinales dans le champ de l'enquête à l'année de référence correspondante, et de toutes les personnes qui habitent avec elles au moment de l'entrevue et qui sont dans le champ de l'enquête.

4.2.3 Unité dans le champ de l'enquête transversalement

Un ménage de l'échantillon est considéré comme étant dans le champ de l'enquête pour la vague en cours s'il est situé dans une des sept grandes régions échantillonnées et fait partie de la population cible.
Toute personne longitudinale qui est dans le champ de l'enquête transversalement fera aussi partie de l'échantillon transversal.

4.2.4 Éligibilité transversale

Une personne dans l'échantillon est éligible à la *pondération transversale* si elle fait partie de la population cible pour l'année de référence traitée et a répondu au questionnaire individuel pour l'année courante. De même, un ménage dans l'échantillon est éligible à la pondération transversale s'il fait partie de la population cible pour l'année de référence traitée et a répondu au questionnaire ménage ainsi qu'(obligatoirement) à la grille pour l'année courante.

4.2.5 Unités répondantes

Une personne est considérée comme *répondante* si elle a complété le questionnaire individuel.
Un ménage répondant est considéré comme tel s'il a complété la grille et le questionnaire ménage.

Il arrive qu'un ménage n'ait complété que la grille et son questionnaire ménage, et aussi, pour un nombre restreint de cas, qu'un ménage ait complété la grille ainsi qu'un ou plusieurs questionnaires individuels sans avoir complété le questionnaire ménage.

4.3 Illustration des concepts par un exemple

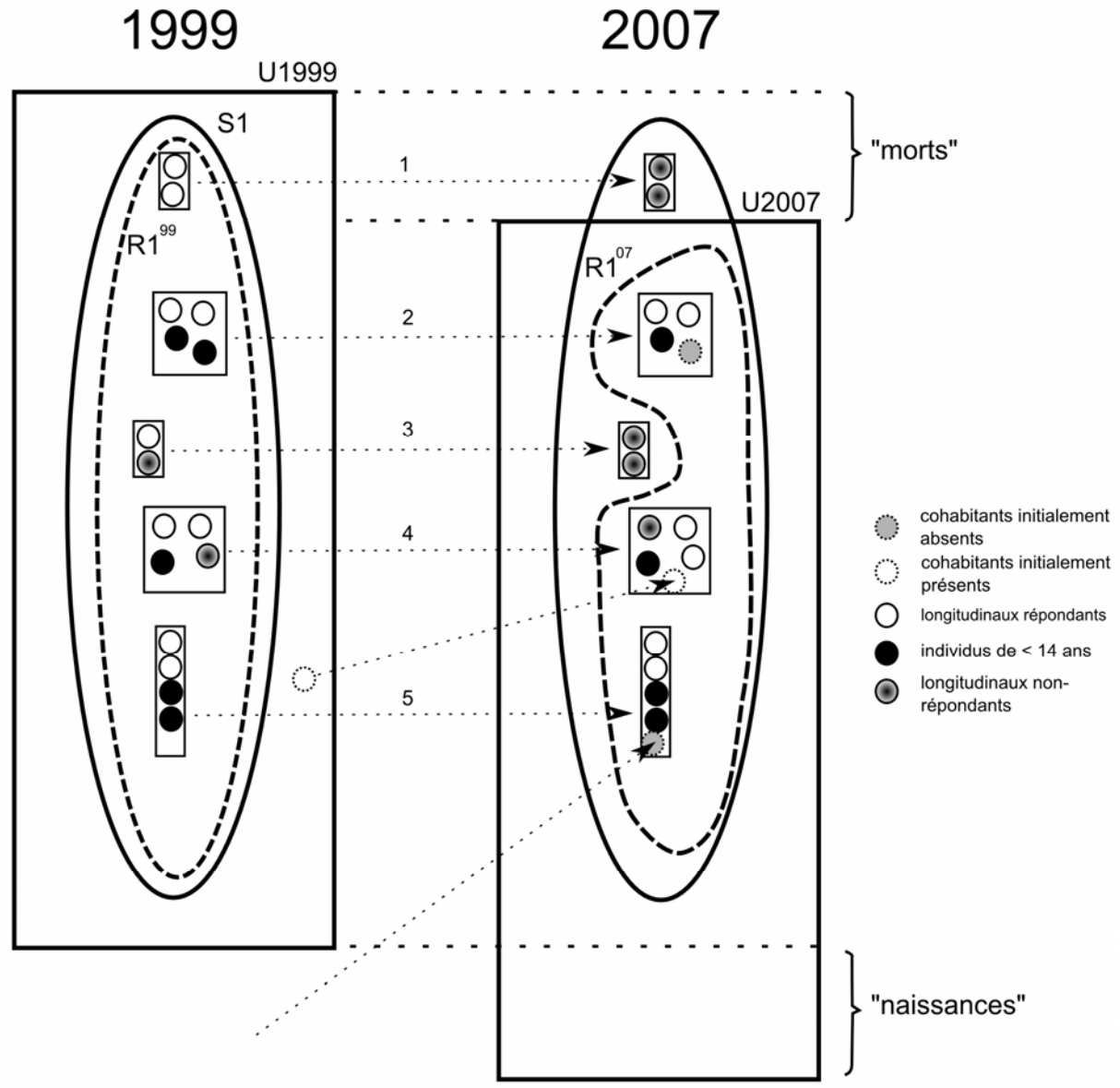
La figure 17 permet de visualiser les définitions et concepts listés dans ce chapitre. Les cercles représentent des individus et les rectangles les ménages les abritant. On a considéré 5 cas de figure courants (flèches numérotées dans la figure) de transitions entre deux années consécutives que nous commentons ci-dessous :

1. Tous les longitudinaux d'un ménage quittent la population transversale cible de l'année courante. Ce sont des « morts » au sens large car ces individus ne

sont pas forcément décédés. Ces individus gardent un poids longitudinal en 2006, mais ils n'auront pas de poids transversal.

2. Un des enfants du ménage devient éligible pour l'interview individuelle. Strictement cet enfant doit être traité comme un individu initialement absent car il ne faisait pas partie de la population cible en 1999 pour le PSM_I, en 2004 s'il s'agit du PSM_II. Dans ce cas, à partir de 2006, s'il répond, cet individu est éligible pour la pondération transversale mais pas pour la pondération longitudinale. Cependant, comme on l'a noté précédemment, depuis la vague 5, on a décidé de considérer cet enfant de longitudinaux comme un longitudinal lui aussi dès qu'il entre dans l'âge d'être interviewé. Il est alors éligible pour les pondérations longitudinale et transversale. Cette stratégie n'est théoriquement pas tout à fait correcte, elle risque d'inclure un biais dans les analyses, mais elle permet de continuer de suivre des ménages dont l'enfant (en 1999 pour le PSM _I, 2004 pour le PSM_II) resterait le seul individu répondant du ménage à la vague courante.
3. Les individus d'un ménage répondant en vague 1 ne répondent plus aux vagues subséquentes mais ils restent dans la population transversale cible. Dans ce cas ils ne sont plus éligibles pour aucune des pondérations.
4. Un individu longitudinal non-répondant en vague 1 se met à répondre à une vague subséquente. Cet individu aura un poids transversal et longitudinal.
   Un autre individu longitudinal, répondant en vague 1, ne répond pas à une vague subséquente mais il reste dans la population transversale cible. Cet individu n'aura ni poids transversal, ni poids longitudinal.
   Un individu qui ne faisait pas partie de l'échantillon S1, mais qui faisait partie de la population transversale cible en vague 1, se joint au ménage. C'est donc un cohabitant initialement présent, il aurait pu être sélectionné pour faire partie de l'enquête en vague 1. Cet individu aura un poids transversal (s'il répond) mais pas de poids longitudinal.
5. Un individu qui ne faisait pas partie de la population transversale cible en vague 1 se joint au ménage. Cet individu aura un poids transversal (seulement s'il répond) mais pas de poids longitudinal.

Figure 17 Illustration des concepts de base des pondérations longitudinale et transversale. On a imaginé cet exemple dans le cadre du PSM_I. La situation pour le PSM_II est exactement pareille, il suffit de remplacer 1999 par 2004.



U1999 Population transversale cible du PSM_I en 1999 = population longitudinale cible du PSM_I en 2006

U2007 Population transversale cible du PSM_I en 2007.

S1 Échantillon brut du PSM_I

R1^{99} Échantillon des personnes dans des ménages ayant répondu au moins au questionnaire grille en 1999, PSM_I (= échantillon net).

R1^{07} Échantillon des personnes dans des ménages ayant répondu au moins au questionnaire grille en 2007, PSM_I (= échantillon net).

‘**’Morts’’** Individus ayant quitté la population transversale cible de l’année courante (individus décédés, partis à l’étranger, institutionnalisés,…)

‘**’Naissances’’** Individus qui ne faisaient pas partie de la population cible l’année précédente (bébés nouveaux né, immigrants venus s’établir en Suisse,…)

5 Étapes de la pondération

5.1 Introduction

Ce paragraphe décrit le principe général; les détails de chaque étape sont expliqués au paragraphe 6.
Pour toutes les pondérations, les individus sont d'abord classés selon leur statut et leurs réponses aux différents questionnaires du PSM. On procède ensuite à un ajustement pour la non-réponse à la grille.

Pour la *pondération transversale individuelle*, un partage des poids est d'abord effectué, puis les poids des longitudinaux et des cohabitants[10] sont ajustés pour la non-réponse au questionnaire individuel. Enfin, un calage sur marge ajuste les poids pour que les totaux des effectifs estimés correspondent à ceux de la population de la Suisse en 2007.

Pour la *pondération transversale ménage*, un partage des poids est fait, puis les poids sont ajustés pour la non-réponse au questionnaire ménage. Enfin, un calage intégré ajuste les poids pour que le nombre total de personnes soit équivalent à l'effectif de la population de la Suisse en 2007, tout en contraignant les poids des personnes d'un même ménage à rester égaux entre eux.

Dans tous les ajustements pour la non-réponse on utilise la méthode de l'analyse par segmentation. Précisons que pour la pondération individuelle aucun ajustement n'est effectué pour la non-réponse au questionnaire ménage. L'ajustement pour la non-réponse au questionnaire individuel est donc fait sur la base de tous les individus des ménages ayant répondu à la grille. Enfin, pour la pondération transversale individuelle, les longitudinaux et les cohabitants ont dû être modélisés ensemble. Les raisons de ces choix sont expliquées en détails dans les sections correspondantes du document. Les distributions de la pondération transversale ménage du PSM_I vague 9, des pondérations transversales ménage et individu des PSM_I & PSM_II combinés, ont été winsorisées aux 1er et 99e percentiles à cause de la présence de quelques valeurs extrêmes et non réalistes et potentiellement trop dérangeantes pour des analyses.

Les calages sur marges ont été effectués à l'aide d'une macro SAS développée par Jean-François Naud et Caroline Cauchon de Statistique Canada pour le Panel Suisse de Ménage. Plus précisément, la méthode utilisée est un calage par régression généralisée qui se base sur une fonction de distance G_k définie par (voir [10], [19]):

$$G_k(d_k, w_k) = \frac{(w_k - d_k)^2}{2\, d_k q_k}$$

où w_k est le poids après calage pour l'unité k, d_k le poids avant calage, et q_k est un facteur d'ajustement indépendant des d_k. On pose le plus souvent $q_k = 1$ pour toutes les observations. Notons que $G_k(d_k, w_k)$ peut produire des poids négatifs. On peut obtenir les mêmes résultats avec la macro CALMAR développée par l'INSEE en appliquant la méthode « linéaire ». La plupart des enquêtes de l'OFS utilisent (CALMAR avec) la méthode « raking ratio », justement pour éliminer tout risque

[10] Voir le chapitre 4 : «Concepts relatifs à la pondération» pour la définition des cohabitants.

d'obtenir des poids négatifs. Cette dernière méthode présente par contre d'autres désavantages, on a par exemple constaté empiriquement que pour enquêtes SILC et PSM, elle produisait plus de poids dans l'intervalle]0,1] ainsi que des poids très élevés.

Les deux méthodes ont des avantages et des inconvénients. Nous avons choisi de garder la méthode linéaire afin de garantir la comparabilité et la compatibilité avec les pondérations produites pour les vagues précédentes. De plus, au niveau ménage, la macro développée par Statistique Canada permet d'effectuer facilement un calage intégré.

Pour la pondération transversale combinant l'échantillon du PSM_I vague 9 avec celui du PSM_II vague 4, il y a deux manières de procéder:

1. On peut combiner les panels en une seule fois après l'ajustement à la grille[11], c'est-à-dire à partir des échantillons longitudinaux. C'est ainsi que les méthodologues de Statistique Canada ont procédé pour combiner les enquêtes PSM _I vague 6, PSM_II vague 1 réalisées en 2004.
2. L'alternative est de combiner de manière transversale (à partir des échantillons transversaux), soit après l'ajustement pour la non-réponse au questionnaire individuel pour la pondération transversale individuelle et après l'ajustement pour la non-réponse au questionnaire ménage pour la pondération transversale ménage.

Comme pour l'année passée, c'est la deuxième manière de faire qui est appliquée et décrite dans ce document. Notons qu'elle est similaire à celle utilisée dans l'enquête SILC dans d'autres pays (voir [15]).
Précisons encore que les enquêtes sont combinées au niveau des grandes régions, les facteurs d'allocation des panels[12] variant d'une région à l'autre. On calcule donc sept différents facteurs d'allocation (un par région) au lieu d'en calculer un seul au niveau de la Suisse.
Pour une discussion plus détaillée des deux alternatives pour combiner les panels, on peut se rapporter aux références [12],[13] et [15].
Le contenu des fichiers de données utilisés pour construire les pondérations sont décrits dans l'annexe A.

5.2 Types de pondérations produites

Pour la vague 9 du PSM_I et la vague 4 du PSM_II, trois pondérations ont été produites :

1 – poids transversal ménage
- *représentant la population des ménages en 2007 (vagues 9 et 4)*

2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2007 (vagues 9 et 4)*

3 – poids longitudinal individuel
- *représentant la population des individus de 14+ ans en 1999 pour le PSM_I vague 9, en 2004 pour le PSM_II vague 4.*

Pour les échantillons combinés PSM_I vague 9 & PSM_II vague 4, les trois mêmes types de pondérations ont été produites. La pondération longitudinale du PSM_I est

[11] Statistique Canada se base sur [14] pour faire la combinaison avec les longitudinaux seulement.
[12] Voir le paragraphe 10.1 pour plus de précisions concernant les facteurs d'allocation des panels.

utilisée pour des analyses portant sur plus de 5 ans. Pour les autres exploitations c'est en principe les pondérations combinées qui sont utilisées puisqu'elles permettent d'utiliser toutes les données récoltées en 2007 et donc bénéficier d'une taille d'échantillon plus grande, ce qui implique une meilleure précision.

1 – poids transversal ménage

- représentant la population des ménages en 2007

2 – poids transversal individuel

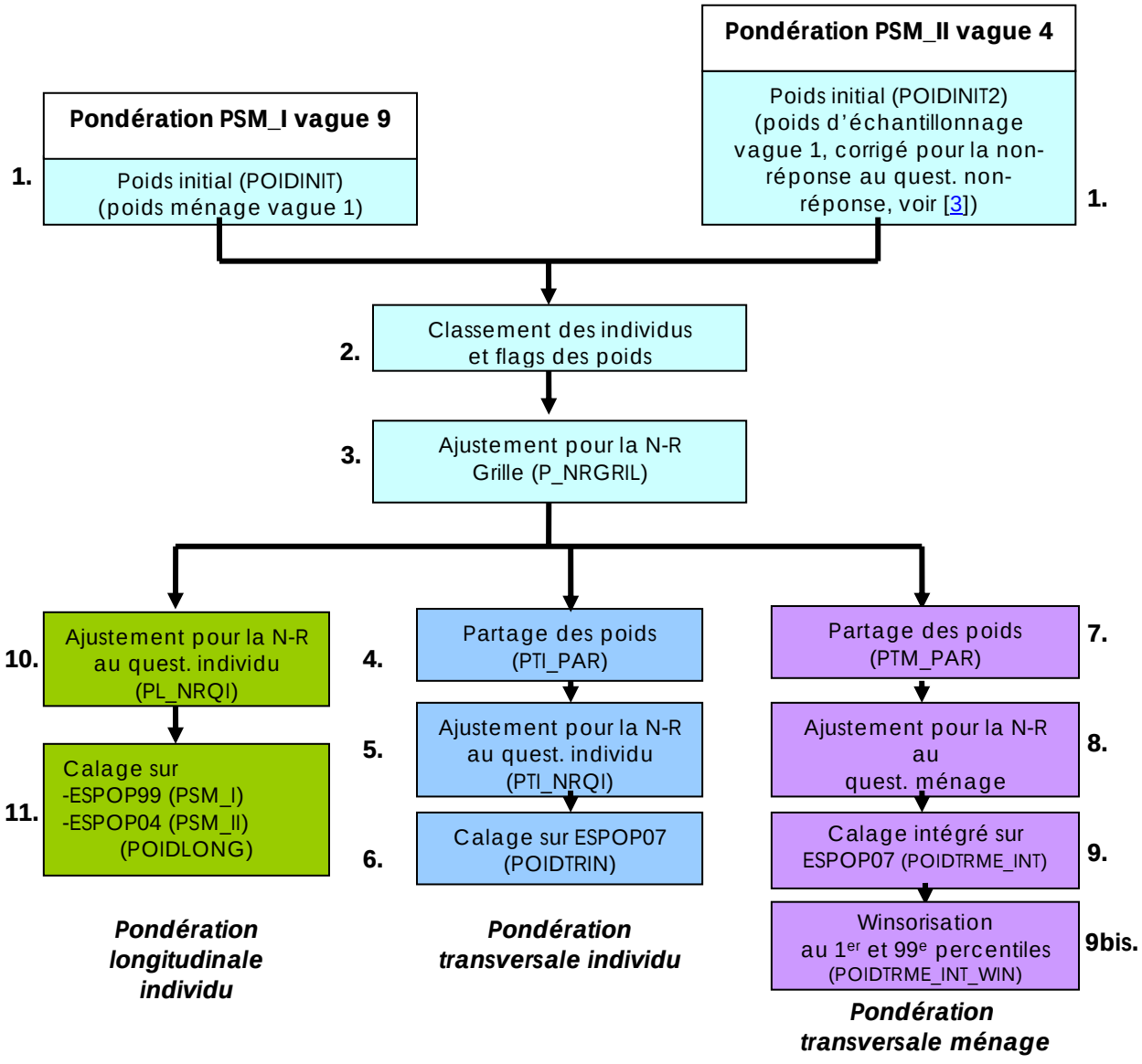
- représentant la population des individus de 14+ ans en 2007

3 – poids longitudinal individuel

- représentant la population des individus de 14+ ans en 2004.

Les trois premières étapes de la pondération sont communes aux trois types de poids, tandis que les suivantes sont spécifiques à chacun. La figure 18 résume le processus de la pondération de la vague 9 du PSM_I et de la vague 4 du PSM_II, alors que la figure 19 traite du processus de pondération des échantillons combinés.

Figure 18 Étapes de la pondération du PSM_I vague 9 et du PSM_II vague 4. Entre parenthèses, dans les cases, on a toujours le nom du jeu de poids créé à l'étape considérée. L'étape 9bis n'a été réalisée que pour la pondération transversale ménage du PSM_I seul.



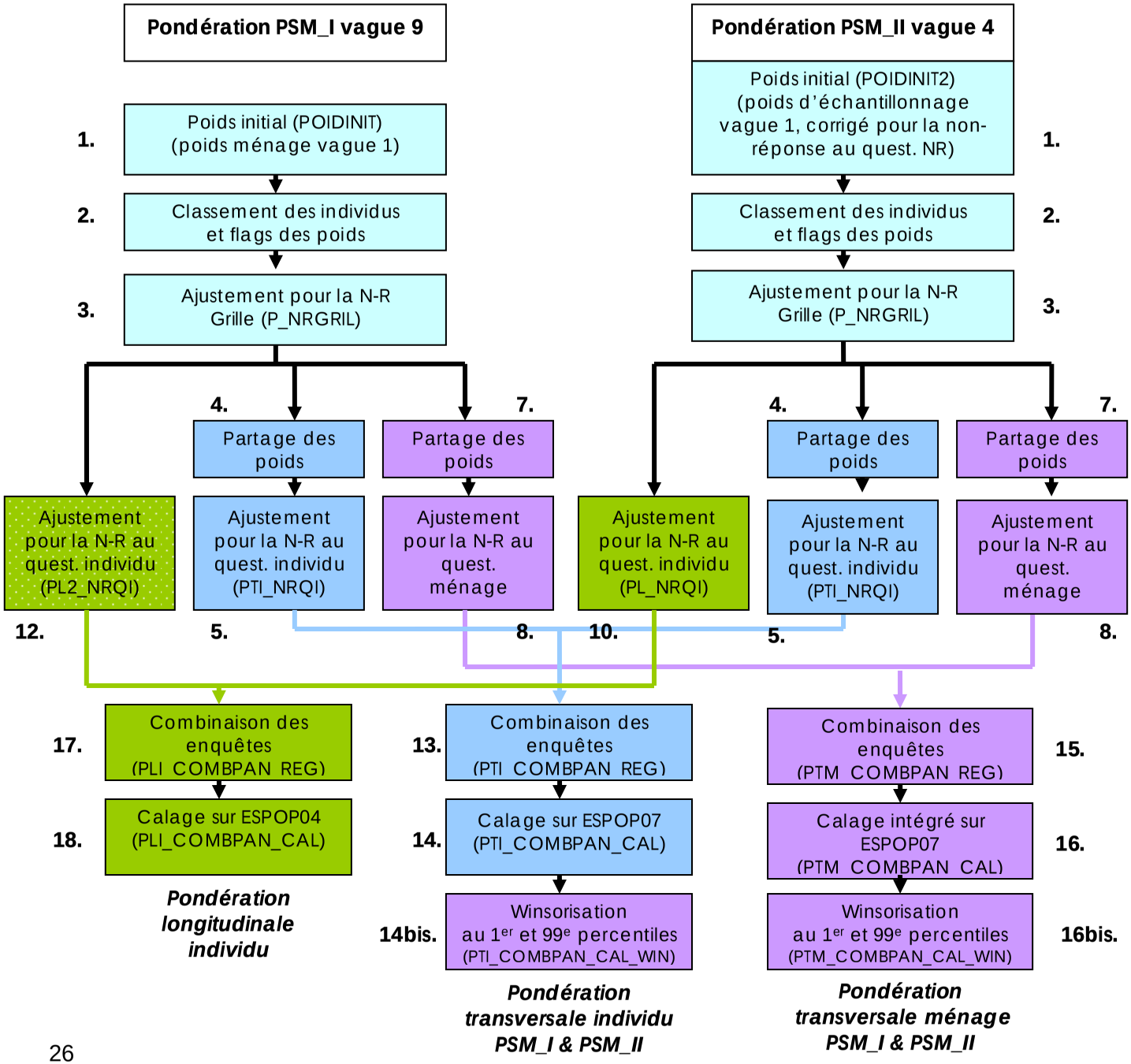


Figure 19 Étapes de la pondération permettant d'agréger l'échantillon du PSM_I vague 9 avec celui du PSM_II vague 4. Il y a une étape supplémentaire comparativement à la pondération séparée par enquête. Il s'agit de la combinaison des échantillons. Il faut combiner les échantillons pour chaque pondération (ménage et individu). Pour les pondérations transversales, les étapes allant du calcul des poids initiaux jusqu'à l'ajustement pour la non réponse au niveau des individus (ou des ménages) sont les mêmes que celles calculées lors de la pondération séparée des deux enquêtes. Pour la pondération longitudinale combinée, il faut refaire un nouvel ajustement pour la non-réponse au questionnaire individuel pour les individus du PSM_I (voir le chapitre 12 pour plus de détails). Entre parenthèses, dans les cases, on a toujours le nom du jeu de poids créé à l'étape considérée.

6 Étapes communes pour tous les types de poids

Tout en se rapportant aux figures 18 et 19, cette section décrit les étapes 1 à 3 communes tant à la production des poids aux niveaux individu que ménage, transversaux que longitudinaux. On a procédé de la même manière pour pondérer la vague 9 du PSM_I et la vague 4 du PSM_II à la différence près que le poids initial du PSM_II fut obtenu après un ajustement supplémentaire pour la non réponse au questionnaire non-réponse (voir [3]). Après une brève introduction à la méthode de segmentation, on traite du calcul du poids initial, puis de la façon dont les individus sont classés selon leur statut et leur réponse et, enfin, de l'ajustement pour la non-réponse à la grille.

6.1 Modélisation de la non réponse par la méthode de segmentation

6.1.2 La méthode de segmentation

Une fois les poids initiaux calculés, if s'agit de calculer un facteur d'ajustement pour la non-réponse (totale) de chaque unité (ménage ou individu selon les cas) considérée comme répondante. Le facteur d'ajustement pour la non-réponse appliqué à chaque unité répondante est défini comme étant l'inverse du taux de réponse du groupe homogène de réponse (GHR) dans lequel l'unité se situe. Les GHR sont formés en regroupant ensemble des unités dont les caractéristiques expliquant la non-réponse sont similaires. Si les GHR sont définis de telle sorte que la non-réponse soit complètement aléatoire à l'intérieur d'un groupe, alors le biais dû à la non-réponse est négligeable (voir [20], [21]). Pour trouver un ajustement adéquat pour amplifier les poids des unités répondantes afin que ces dernières soient représentatives de l'ensemble des unités (répondantes et des non-répondantes), on procède à la modélisation de la non-réponse.

La première étape de la modélisation de la non-réponse consiste à déterminer quelles sont les variables explicatives qui doivent être considérées dans le modèle afin de bien prédire la variable dichotomique de réponse. Le choix des variables est souvent assez restreint car on doit détenir l'information tant pour les unités répondantes que pour les non-répondantes. Il s'agit de dresser une liste des variables disponibles pouvant décrire le processus de non-réponse. La nature de cette liste varie selon les thèmes abordés par l'enquête, le mode d'interview, le plan d'échantillonnage, etc. Tout facteur susceptible d'expliquer la non-réponse doit être pris en compte dans la mesure du possible. Les variables peuvent êtres des grandeurs mesurées dans l'enquête elle-même ou des informations auxiliaires qu'on peut obtenir par d'autres biais (par exemple des registres) sur les unités répondantes et non-répondantes.
Dans une enquête panel comme celle qui nous intéresse ici, on a dès la vague deux, un choix de variables beaucoup plus grand pour modéliser la non-réponse aux questionnaires ménage (voir paragraphe 8.2) ou individuel (voir paragraphe 7.2) car on peut alors se reposer sur des informations récoltée en vague 1.
Notons encore que les variables catégorielles dont le nombre de modalités (m) est supérieur à 2 doivent être recodées en m variables dichotomiques.

La méthode utilisée pour la modélisation de la non-réponse est l'analyse par segmentation (voir [8]). Cette méthode est utilisée pour tous les ajustements nécessaires dus à la non réponse et présentés dans ce document.
La modélisation par segmentation utilise un processus itératif pour partitionner le fichier de données et ainsi former l'arbre de décision. Le premier nœud de l'arbre de décision est tout simplement l'ensemble complet des données. La première itération définit le premier embranchement de l'arbre. Ce dernier est créé en déterminant la variable qui influence le plus significativement la variable réponse parmi les variables retenues pour la modélisation. Une mesure statistique reflétant l'importance de la variable pour discriminer la non-réponse est calculée pour chaque variable; d'après cette mesure, la variable qui est la plus importante significativement est choisie. Le premier embranchement est ainsi formé, créant un nœud à chaque extrémité. Pour chaque nœud formé, une nouvelle itération est effectuée en répétant le même mécanisme. De nouveaux embranchements et de nouveaux nœuds sont ainsi créés. Si pour un nœud donné, aucune variable n'est significative pour expliquer la non-réponse, ou qu'un des critères d'arrêt de l'algorithme est rencontré, aucun autre embranchement n'est créé.

Plusieurs méthodes peuvent être employées pour déterminer la variable qui influence le plus la réponse (voir [22]). Nous utilisons ici l'algorithme de CHAID[13], qui sélectionne la variable avec la valeur statistique du chi-deux de Pearson la plus élevée.

6.1.2 Segmentation et régression logistique

Une alternative pour modéliser la non-réponse serait d'utiliser une régression logistique. Comme on a voulu rester le plus compatible possible avec les pondérations des années passées, on n'a pas remis en cause cette méthode utilisée depuis 9 ans déjà.
Par ailleurs, mentionnons que les auteurs de [7] ont montré sur deux études[14] empiriques similaires au PSM que la méthode de segmentation était meilleure que la régression logistique pour créer des groupes homogènes de réponse (GHR) et par conséquent pour réduire le biais. Effectivement, la méthode de segmentation est plus flexible que la régression logistique. La régression logistique force le modèle à être symétrique : une fois que les variables significatives sont déterminées, les GHR sont formés en prenant toutes les combinaisons possibles de l'intersection entre ces variables. En forçant le modèle à être symétrique, de petites cellules ou des cellules qui ne sont pas significatives peuvent être formées, ce qui peut affecter la fiabilité des estimations. La modélisation par segmentation quant à elle détermine à chaque embranchement de l'arbre de décision quelle variable est la plus significative, ce qui garantit que chaque GHR formé est significatif et résulte habituellement en un modèle asymétrique. Cela présente aussi l'avantage de pouvoir utiliser un plus grand nombre de variables pour modéliser la non-réponse.

Un inconvénient de la modélisation par segmentation est directement relié à l'arbre de décision lui-même. Lorsqu'un arbre de décision est formé, il peut devenir assez lourd : dès qu'il comporte plusieurs embranchements, il devient plus difficile à interpréter. Les arbres de décisions sont aussi assez instables : l'ajout, le retrait ou le

[13] CHAID: Chi-Square Automatic Interaction Detection, algorithme inventé par Kass (voir [8]).
[14] SLID (*Survey of Labour and Income Dynamics*), un panel rotatif, et NLSCY (*National Longitudinal Survey of Children and Youth*) un panel simple.

changement d'une variable ou d'une unité peut avoir un effet important sur l'arbre lui-même.

Le tableau 20 résume quelques caractéristiques comparatives entre les méthodes de modélisation de la non-réponse par segmentation et par régression logistique.

Tableau 20 Comparaison entre la régression logistique et la méthode de segmentation dans le cadre de la modélisation de la non-réponse

Régression logistique	Segmentation
➢ Procédure pas à pas (stepwise) pour obtenir la liste des variables explicatives les plus significatives de la prédisposition à répondre, ➢ Les GHR sont créés sur la base de 2^k combinaisons des k prédicateurs sélectionnés, ➢ Approche qu'on peut qualifier de symétrique, ➢ On peut ajouter des contraintes pour les GHR (typiquement n>30 et taux de réponse >50% par GHR), ➢ Le nombre de variables pouvant être utilisées pour modéliser la non-réponse est limité.	➢ Utilise l'algorithme de CHAID qui divise l'échantillon en plus petits sous-groupes d'après leur taux de réponse jusqu'à ce que plus aucun prédicateur significatif ne soit trouvé. Les sous-groupes finaux forment les GHR. ➢ Approche qu'on peut qualifier de non symétrique. ➢ Les contraintes doivent être satisfaites pour chaque GHR (par exemple, n>30 et taux de réponse >50%: le choix de ces contraintes a une influence directe sur le nombre final de GHR), ➢ Peut utiliser un éventail plus large de variables pour modéliser la non-réponse. Mais l'arbre de segmentation peut devenir instable si trop de variables explicatives sont sélectionnées ou si certaines sont trop corrélées entre elles.

6.1.3 Application dans le cadre du PSM

On utilise une macro SAS qui fut développée par Jean-François Naud de Statistique Canada lorsque ce dernier reçut le mandat d'effectuer les pondérations du Panel Suisse de Ménages pour la première fois en 2001. La macro permet de modéliser la non-réponse pour plusieurs types d'ajustements (se reporter aux figures 18 et 19 pour la définition des étapes mentionnées):

- ajustement pour la non-réponse au questionnaire grille (paramètre Type_aj=1), étape 3,
- ajustement pour la non-réponse au questionnaire individuel dans le cadre de la pondération longitudinale (paramètre Type_aj=2), étape 10
- ajustement pour la non-réponse au questionnaire individuel dans le cadre de la pondération transversale (paramètre Type_aj=3), étape 5,

- ajustement pour la non-réponse au questionnaire ménage dans le cadre de la pondération transversale (paramètre Type_aj=4), étape 8.

Les unités répondantes de chaque GHR voient leur poids avant ajustement multiplié par un facteur correspondant à l'inverse du taux de réponse pondéré de leur GHR.

Les exigences minimales (critères d'arrêt de l'algorithme dans sa détermination d'embranchements supplémentaires, outre le fait de ne plus trouver de variable expliquant la variable réponse de manière significative), pour chaque GHR, ont été, suivant les cas, soit 50% de taux de réponse minimal et 30 unités au minimum, soit 30% de taux de réponse et 50 unités au minimum.

6.2 Poids initial du PSM_I et poids initial du PSM_II

Étape 1

Pour le PSM_I, comme cela a été fait les années précédentes, le poids initial est le poids ménage calculé à la vague 1 (voir [2]). Il s'agit du poids d'échantillonnage corrigé pour la non-réponse à la grille et au questionnaire ménage en 1999. Les facteurs de corrections furent alors calculés à l'aide de modèle de régressions logistiques.

$$POIDINIT_k^{PSM_I} = \frac{1}{\pi_k \cdot p_k^1 \cdot p_k^2 \cdot p_k^3 \cdot f}$$

Où : π_k est le poids d'échantillonnage du ménage k (vague 1, 1999)

p_k^1 est la probabilité de réponse à la grille en vague 1 (sous l'hypothèse de non-réponse uniforme) étant donné le plan d'échantillonnage

p_k^2 est la probabilité de réponse au questionnaire ménage en vague 1 (obtenue par un modèle de non-réponse logistique) étant donné la répartition à la grille

p_k^3 est la probabilité qu'au moins un questionnaire individuel ait été complété en vague 1 (obtenue par un modèle de non-réponse logistique), i.e. que le ménage soit complet, étant donné la répartition à la grille et au questionnaire ménage

f est un facteur de correction qui fut appliqué en vague 1 pour que les poids ménage multipliés par la taille du ménage somment à la taille de la population résidente en Suisse d'après l'ESPOP 1999.

Rappelons qu'en 2000, lors de la vague 2 du PSM_I, on tenta de recontacter seulement les ménages complets en vague 1. Tous les ménages pour lesquels uniquement le questionnaire grille ou le questionnaire grille plus le questionnaire ménage avaient été complétés furent abandonnés. Ainsi, il est naturel de prendre le poids ménage *POIDINIT* décrit plus haut comme poids initial.

Pour le PSM_II, le poids initial est le POIDINIT2 calculé comme étant le poids d'échantillonnage en vague 1 (2004) multiplié par un facteur d'ajustement corrigeant pour la non-réponse au questionnaire non-réponse de vague 1. Ce

facteur d'ajustement a été calculé à partir d'un modèle par segmentation décrit dans [3].

$$POIDINIT2_k^{PSM_II} = \frac{1}{\pi_k \cdot f_k^{QNR}}$$

Où : π_k est le poids d'échantillonnage du ménage k (vague 1, 2004)

f_k^{QNR} est un facteur de correction obtenu par la méthode de segmentation pour la non-réponse du ménage k au questionnaire non-réponse.

Rappelons qu'en 2005, lors de la vague 2 du PSM_II, on tenta de recontacter tous les ménages qui avaient au moins complété un questionnaire grille en vague 1. Les ménages ayant répondu au questionnaire non-réponse ont été définitivement abandonnés[15]. Ainsi, il est naturel de prendre le poids ménage *POIDINIT2* décrit ici comme poids initial.

Des statistiques descriptives des poids initiaux du PSM_I et PSM_II se trouvent dans l'annexe H.

6.3 Classement des individus: étape 2

Cette étape a pour but de créer un fichier de base pour la pondération contenant tous les individus caractérisés par des variables de classification qui servent lors des étapes subséquentes.

Le fichier contient tous les individus des ménages ayant rempli au moins le questionnaire grille en vague 1 et en vague courante[16]. Pour le PSM_I, il on trouve donc tous les individus de INDIVIDUS 1999, INDIVIDUS 2007 et LONGOUT 2007, pour le PSM_II, on trouve tous ceux de INDIVIDUS 2004, INDIVIDUS 2007 et LONGOUT 2007 (ces fichiers sont décrits dans l'annexe A). Les variables contenues dans le fichier « classement » sont les suivantes:

[15] Les ménages répondants à l'enquête non-réponse effectuée en 2004 n'ont donc pas de poids (poids final = 0) contrairement à ce qui fut fait en vague 1 par Statistique Canada (voir [4]). Dans les étapes ultérieures de la pondération, ces ménages ne sont plus considérés du tout. L'information qu'ils ont délivrée est « contenue » dans les poids initiaux (POIDINIT2).

[16] A ce stade, le PSM_I est traité séparément du PSM_II. Il y a donc un classement distinct pour ces deux enquêtes.

Nom de la variable	Contenue dans le fichier classement PSM_I v9	Contenue dans le fichier classement PSM_II v4	Libellé
IDPERS	(x)	(x)	identificateur de la personne
idhous_o	(=IDHOUS99)	(=IDHOUS04)	identificateur du ménage à la vague originale (=vague 1)
idhous_c	(=IDHOUS07)	(=IDHOUS07)	identificateur du ménage à la vague courante (2007) (= vague 9 pour PSM_I, vague 4 pour PSM_II)
f_plong	(x)	(x)	indicateur de poids longitudinal
f_ptrin	(x)	(x)	indicateur de poids transversal individuel
f_ptrme	(x)	(x)	indicateur de poids transversal ménage
agecorro	(x)	(x)	âge de l'individu à la vague originale
agecorrc	(x)	(x)	âge de l'individu à la vague courante
poidinit	(x)		poids initial ménage vague 1, PSM_I (étape 1)
poidinit2		(x)	poids initial corrigé pour la NR au QNR (étape 1), PSM_II
longst	(x)	(x)	statut de l'individu (longitudinal / cohabitant)
coderep	(x)	(x)	code de réponse à la vague courante
cr_vx	cr_v2 à cr_v8	cr_v1 à cr_v3	code de réponse à la vague x
	13752 obs.	6813 obs.	

On peut trouver les détails sur la manière dont ces variables ont été construites dans l'annexe B.

Les tableaux 21 et 22 donnent les distributions des individus selon leur statut, leur code de réponse et s'ils doivent ou non avoir un poids non-nul.

Tableau 21 Distribution des individus du PSM_I et du PSM_II selon leur statut et leur code de réponse[17]. Le nombre de longitudinaux adultes et enfants ne change en principe pas d'une vague à l'autre (pour PSM_I) car on fait toujours référence aux statuts qu'avaient les individus en l'année de sélection de l'échantillon panélisé. Suite à une replausibilisation de l'âge deux longitudinaux enfants du PSM_II sont passé en adultes par rapport à l'année précédente. Pour le PSM_II toujours, les nombres sont donnés en excluant les individus des ménages ayant répondu au questionnaire non-réponse en 2004. Suite à des plausibilisations des données, on a ici 8 (4) questionnaires individuels réalisés de moins pour le PSM_I (PSM_II) par rapport aux nombres donnés dans le tableau 8.

	Statut de l'individu – **LONGST**									
	1 – Longitudinal adulte		2 – Longitudinal enfant		3 – Cohabitant[18] initialement présent		4 – Cohabitant initialement absent		**TOTAL**	
Code de réponse – **CODEREP**	PSM_I	PSM_II	PSM_I	PSM_II	PSM_I	PSM_II	PSM_I	PSM_II	PSM_I	PSM_II
0 – < 14 ans à la vague courante	0	0	1051	796	30	16	488	100	1569	912
1 – Hors-cible transversalement	189	41	5	1	0	0	0	0	194	42
2 – Non-répondant total	5382	2610	679	112	0	0	0	0	6021	2722
3 – Répondant grille seulement	134	117	18	5	7	8	1	0	160	130
4 – Répondant grille et questionnaire ménage seulement, dans un ménage où tous les individus sont non-répondants au questionnaire individuel	72	68	16	5	4	3	2	0	94	76
5 – Répondant grille et questionnaire ménage seulement, dans un ménage où au moins un individu est répondant au questionnaire individuel	844	524	201	33	152	56	19	8	1216	621
6 – Répondant grille, questionnaire ménage et questionnaire individuel	3768	2136	607	112	218	89	28	1	4621	2338
7 – Répondant grille et questionnaire individuel seulement	2	8	3	1	0	0	0	0	5	9
Total	**10351**	**5504**	**2580**	**1065**	**411**	**172**	**538**	**109**	**13880**	**6850**

[17] Les enfants de moins de 14 ans sont tous classés hors-cible transversalement.
[18] Voir paragraphe 4.1.3 pour les définitions des cohabitants.

Tableau 22 Distribution des individus du PSM_I et du PSM_II selon leur indicateur de poids longitudinal, transversal ménage[19] et transversal individuel. Pour le PSM_II, les nombres sont donnés en excluant les individus des ménages ayant répondu au questionnaire non-réponse en 2004.

Indicateur de poids	PSM_I V7	PSM_I V8	PSM_I V9	PSM_II V2	PSM_II V3	PSM_II V4
F_PLONG = 0 – Poids longitudinal = 0	10166	10135	9921	4137	4360	4665
F_PLONG = 1 – Poids longitudinal > 0	3495	3617	3959	2589	2453	2185
Total	**13661**	**13752**	**13880**	**6726**	**6813**	**6850**
F_PTRME = 0 – Poids transversal ménage = 0	8469	9664	7949	4078	4245	3815
F_PTRME = 1 – Poids transversal ménage > 0	5192	4088	5931	2648	2568	3035
Total	**13661**	**13752**	**13880**	**6726**	**6813**	**6850**
F_PTRIN = 0 – Poids transversal individuel = 0	9776	8380	9254	3026	3374	4503
F_PTRIN = 1 – Poids transversal individuel > 0	2885	5372	4626	3700	3439	2347
Total	**13661**	**13752**	**13880**	**6726**	**6813**	**6850**

6.4 Ajustement pour la non-réponse – Grille: étape 3

Tous les ménages choisis pour être réactivés pour l'enquête en 2007 et dans lesquels vivaient des longitudinaux étaient susceptibles de répondre à la grille.
Toutefois, comme l'unité longitudinale est l'individu et non le ménage, la probabilité de réponse à la grille a été modélisée au niveau de l'individu. On modélise donc la probabilité qu'un individu fasse partie d'un ménage ayant répondu à la grille. Les longitudinaux adultes et enfants sont utilisés. Ils sont cependant divisés selon qu'ils soient ou non des enfants dès la première étape de modélisation. La somme des poids longitudinaux avant et après l'ajustement pour la non-réponse à la grille reste la même pour les enfants et pour les adultes. La modélisation est faite à partir des données ménage et individuelles de la vague 1.
On crée un fichier ne contenant que les longitudinaux adultes (LONGST=1) et enfants (LONGST=2). On définit la variable :

REPGRIL = 0 Non-répondant à la grille.
(CODEREP=0, 2).
1 Répondant à la grille.
(CODEREP=3, 4, 5, 6 ou 7).

Tableau 23 Distribution des longitudinaux selon la réponse à la grille. On a 11686 et 5731 individus au total car on a écarté les cohabitants et les individus hors cible transversalement, ainis que les < 14 ans à la vague courante (voir tableaux 21 et 22)

Réponse à la grille (REPGRIL)		Nombre de longitudinaux (Proportion %)		
		Adultes	Enfants	Total
0 – Non-répondant à la grille	PSM_I	5342	679	**6021 (51.5)**
	PSM_II	2610	112	**2722 (47.5)**
1 – Répondant à la grille	PSM_I	4820	845	**5665 (48.5)**
	PSM_II	2853	156	**3009 (52.5)**
Total	**PSM_I**	**10162**	**1524**	**11686**
	PSM_II	**5463**	**268**	**5731**

[19] Les enfants de moins de 14 ans ont F_PTRME=0 même si leur ménage a POIDTRME > 0.

Pour le PSM_I, il a fallu relaxer les contraintes du modèle de segmentation par rapport aux premières années: un taux de réponse pondéré d'au moins 30 % et un effectif minimal de 50 individus ont été exigés dans chaque GHR. Pour le PSM_II les contraintes étaient un taux de réponse pondéré d'au moins 50 % et un effectif minimal de 30 individus pour chaque GHR
Aucun ajustement pour la non-réponse n'a été fait pour les individus hors-cible transversalement (CODEREP=1). Leur poids ajusté pour la non-réponse sera identique au poids initial.

Le modèle de non-réponse créé avec l'analyse par segmentation a conduit à la création de 64 GHR pour le PSM_I, respectivement 6 pour le PSM_II. Le tableau 24 donne quelques informations sur le facteur d'ajustement ainsi calculé et appliqué aux poids initiaux pour obtenir les poids P_NRGRIL.

Tableau 24 Statistiques descriptives du facteur d'ajustement, f_aj_nrgril , pour les 5665 et 3009 individus longitudinaux dans les ménages répondants à la grille (ceux pour qui REPGRIL=1, voir tableau 23). Les 194 et 45 individus hors-cible transversalement ont aussi un poids P_NRGRIL>0 mais ne subissent pas d'ajustement pour la non-réponse à la grille (i.e. P_NRGRIL=poids initial, f_aj_nrgril=1).
PSM_I: P_NRGRIL = f_aj_nrgril * POIDINIT
PSM_II: P_NRGRIL = f_aj_nrgril * POIDINIT2

	PSM_I	PSM_II		PSM_I	PSM_II
Moyenne	2.06	1.92	Min	1.18	1.45
Médiane	1.87	1.98	Max	3.33	1.98
Écart-type	0.61	0.15	Étendue	2.15	0.53

Les détails sur cette modélisation de la non-réponse se trouvent à l'annexe C.

7 Pondération transversale individuelle

Dans ce chapitre on décrit les étapes 4, 5 et 6 de la figure 18. La procédure est la même pour le PSM_I vague 9 et le PSM_II vague 4. Le poids de départ de la pondération transversale individuelle est le poids ajusté pour la non-réponse à la grille des longitudinaux âgés de 14 ans ou plus à la vague courante. L'étape du partage des poids permet d'attribuer un poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante. On procède ensuite à un ajustement pour la non-réponse au questionnaire individuel et, finalement, à un calage sur marge en utilisant des totaux de contrôle pour la population résidant dans des ménages privés de Suisse en 2007.

7.1 Partage des poids: étape 4

Après l'étape de l'ajustement pour la non-réponse à la grille, seuls les longitudinaux des ménages répondants à la grille ont un poids P_NRGRIL non-nul. L'étape du partage des poids consiste à attribuer une partie de ces poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante. Le partage des poids est effectué séparément pour chacun des deux panels PSM_I et PSM_II.
Pour les considérations et validations théoriques quant à la méthode du partage des poids, voir [9].
Notons que les enfants longitudinaux qui ont atteint 14 ans à la vague courante sont traités comme des longitudinaux, et non pas comme des cohabitants initialement absents. Ainsi on tient compte de la possibilité qu'un enfant longitudinal de plus de 14 ans soit le seul individu longitudinal d'un ménage.

Processus

On utilise tous les individus faisant partie de la population cible transversale et répondants à la grille (variable CODEREP = 3, 4, 5, 6 ou 7). Le poids de base (avant partage) est P_NRGRIL. Il est nul pour tous les cohabitants.
Le poids après le partage est PTI_PAR. **Ce poids peut être différent pour les individus d'un même ménage.**
Le poids partagé se calcule de la manière suivante :
i : individu
L : nombre de longitudinaux dans le ménage
P : nombre de cohabitants initialement présents dans le ménage

- *si le ménage comporte des cohabitants initialement présents*, le poids partagé est le même pour tous et est égal à la somme des poids des longitudinaux divisée par le nombre de longitudinaux et de cohabitants initialement présents dans le ménage.

$$PTI_PAR_i = \frac{\sum_{i=1}^{L} P_NRGRIL_i}{L+P}$$

- *s'il n'y a aucun cohabitant initialement présent dans le ménage*, alors les poids des longitudinaux demeurent les mêmes et ceux des cohabitants initialement absents sont égaux à la moyenne des poids des longitudinaux.

$$PTI_PAR_i = \begin{cases} P_NRGRIL_i & \textit{pour i longitudinaux} \\ \dfrac{\sum_{i=1}^{L} P_NRGRIL_i}{L} & \textit{pour i cohabitants initialement absent} \end{cases}$$

On utilise la macro SAS %partage() développée par Jean-François Naud de Statistique Canada pour effectuer le partage des poids. Pour plus de précisions quant à l'utilisation de cette macro se reporter à son code source.

7.2 Ajustement pour la non-réponse – questionnaire individuel: étape 5

Une fois que le ménage a rempli la grille, les individus peuvent répondre au questionnaire individuel, peu importe que le questionnaire ménage ait été rempli ou non. C'est pourquoi, pour être cohérent avec le processus de collecte et parce que peu d'information du questionnaire ménage s'avérerait utile pour la modélisation, nous faisons l'ajustement pour la non-réponse au questionnaire individuel directement à partir du poids partagé qui représente les individus dans les ménages ayant répondu à la grille.

Processus

On crée un fichier contenant les individus (longitudinaux adultes et cohabitants) dont les ménages ont répondu à la grille et faisant partie de la population cible transversale, i.e. tous ceux pour qui PTI_PAR > 0.

On définit : REPQI = 0 Non-répondant au questionnaire individuel (CODEREP=3, 4 ou 5)
1 Répondant au questionnaire individuel (CODEREP=6 ou 7).

Tableau 25 Distribution des adultes faisant partie de la population cible transversale et appartenant à des ménages répondants à la grille selon la réponse au questionnaire individuel.

	PSM_I *Vague 9*	*PSM_II* *Vague 4*
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1470 (24.1)	827 (26.1)
1 – Répondant au Q.I.	4626 (75.9)	2347 (73.9)
Total	**6096**	**3174**

Tableau 26 Statistiques descriptives du facteur d'ajustement, f_aj_nrtiqi, pour les 4626 et 2347 individus répondants au questionnaire individuel (ceux pour qui REPQI=1, voir tableau 25). PTI_NRQI = P_NRGRIL * f_aj_nrtqi.
Dans les deux enquêtes on a des individus pour qui f_aj_nrtqi=1 car, comme on peut le voir dans l'annexe E, certains groupes homogènes de réponse formés par l'algorithme de segmentation ont atteint 100% de taux de réponse, donc pas d'ajustement pour ces GHR-là.

	PSM_I	PSM_II		PSM_I	PSM_II
Moyenne	1.32	1.36	Min	1.00	1.00
Médiane	1.24	1.27	Max	3.29	1.96
Écart-type	0.31	0.28	Étendue	2.29	1.00

L'ajustement se fait au niveau des individus. Étant donné que le nombre de cohabitants est très faible (voir tableau 21), il a été nécessaire de faire la modélisation des longitudinaux et des cohabitants ensemble. Les données provenant de la grille de la **vague courante** ont été utilisées pour la modélisation.
Pour le PSM_I vague 9, le modèle obtenu avec l'analyse par segmentation comporte 39 GHR et 23 GHR pour le PSM_II vague 4 (voir annexe E). Comme pour l'ajustement pour la non-réponse à la grille, un taux de réponse pondéré d'au moins 50% et un effectif minimal de 30 individus étaient exigés dans chaque GHR pour le PSM_II, tandis que pour le PSM_I, les contraintes étaient un taux de réponse de 30% et 50 unités pour chaque GHR.

Le poids ajusté pour la non-réponse (questionnaire individuel) est PTI_NRQI.

7.3 Calage: étape 6

Le poids transversal ajusté pour la non-réponse au questionnaire individuel (PTI_NRQI) représente la population de la Suisse à la vague courante. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2007 selon l'ESPOP (de décembre 2007). Cette étape est réalisée pour chacune des deux enquêtes. Les totaux se trouvent dans le tableau 28.

Poids avant calage : PTI_NRQI

Totaux de contrôle
1. Sexe x (14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)
4. État civil (marié; pas marié)

Poids final : POIDTRIN[20]

Le tableau 27 donne quelques nombres caractérisant les calages effectués.

[20] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal individuel, donc POIDTRIN, est renommée en:
- Wp07TP1P pour le PSM_I vague 9
- Wp07TP2P pour le PSM_II vague 4

Tableau 27 Caractéristiques du calage sur marge, pondération transversale individuelle. On a respectivement 4626 et 2347 observations avec un POIDTRIN>0 pour PSM_I et PSM_II.

	facteur d'ajustement du calage = POIDTRIN/PTI_NRQI		*POIDTRIN*	
	PSM_I	*PSM_II*	*PSM_I*	*PSM_II*
Moyenne	0.92	0.88	1358	2676
Médiane	0.84	0.77	1033	2232
Écart-type	0.31	0.40	1043	1508
Min	0.61	0.49	166	257
Max	2.40	2.59	10736	12306
Étendue	1.79	2.11	10569	12050

Tableau 28 Effectifs de population estimés avant et après calage pour le PSM_I. Poids transversal individuel. Notons que ces totaux reflètent la population de 14 ans et plus. Le total des pti_nrqi du PSM_II vague 4 est plus élevé à cause des nombreuses corrections pour la non-réponse (questionnaire non-réponse, questionnaire grille, questionnaire individuel). En effet, les poids corrigés pour la non-réponse somment généralement à un nombre plus élevé que la somme des poids après calage. Le phénomène est d'autant plus accentué que l'attrition est grande et sélective.

	Estimation avec le poids avant calage (PTI_NRQI)		*Population (et estimation avec le poids après calage POIDTRIN)*
Variable	*PSM_I vague 9*	*PSM_II vague 4*	
Homme 14-24 ans	631598	741323	485575
Homme 25-34 ans	339159	324648	485027
Homme 35-44 ans	569641	659394	610375
Homme 45-54 ans	549138	719048	550107
Homme 55 ans et plus	1065601	1022071	944765
Femme 14-24 ans	559773	545506	461919
Femme 25-34 ans	329447	395063	490511
Femme 35-44 ans	646153	752947	608983
Femme 45-54 ans	633059	820173	543721
Femme 55 ans et plus	1231258	1270238	1099785
Suisse	5832564	6576356	4955440
Étranger	722261	674055	1325328
Lémanique, Centrale et Tessin	2300250	2286437	1999361
Mitteland, N-O, Zürich, Orientale.	4254575	4963974	4281408
Non marié	3028509	3210740	2867978
Marié	3526316	4039671	3412790
Total	**6554824**	**7158158**	**6280768**

8 Pondération transversale ménage

Dans ce chapitre on décrit les étapes 7, 8 et 9 de la figure 18. La procédure est la même pour le PSM_I vague 9 et le PSM_II vague 4. Le poids de départ de la pondération transversale ménage est aussi le poids ajusté pour la non-réponse à la grille des longitudinaux adultes. L'étape du partage des poids permet d'attribuer un poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante et d'obtenir un poids au niveau du ménage. Ensuite, on procède à un ajustement pour la non-réponse au questionnaire ménage et, finalement, à un calage sur marges en utilisant des totaux de contrôle pour la population résidant dans des ménages privés de Suisse en 2007.

8.1 Partage des poids: étape 7

Après l'étape de l'ajustement pour la non-réponse à la grille, tous les longitudinaux des ménages répondant à la grille ont un poids P_NRGRIL > 0. L'étape du partage des poids consiste à attribuer une partie de ces poids aux cohabitants qui se sont joints aux ménages des longitudinaux depuis la vague 1.

Comme pour la pondération transversale individuelle, notons que les enfants longitudinaux qui ont atteint 14 ans sont traités comme des longitudinaux, et non pas comme des cohabitants initialement absents comme on pourrait aussi le faire, ceci pour tenir compte de la possibilité qu'a un enfant longitudinal de plus de 14 ans d'être le seul individu longitudinal d'un ménage.

Processus

On utilise tous les individus faisant partie de la population cible transversale et répondant à la grille (variable CODEREP = 3, 4, 5, 6 ou 7). Le poids de base (avant le partage) est P_NRGRIL. Le poids après le partage est PTM_PAR. **Ce poids est le même pour tous les individus d'un ménage.**

Le poids ménage partagé est égal à la somme des poids des longitudinaux divisée par le nombre de longitudinaux et de cohabitants initialement présents dans le ménage.

$$PTI_PAR_i = \frac{\sum_{i=1}^{L} P_NRGRIL_i}{L + P}$$

i : individu
L : nombre de longitudinaux dans le ménage
P : nombre de cohabitants initialement présents dans le ménage

Le fichier résultant est au niveau du ménage (identificateur IDHOUS07).
Comme pour la pondération transversale individus (voir paragraphe 7.1), on utilise la macro `%partage()` pour effectuer le partage des poids.

8.2 Ajustement pour la non-réponse – questionnaire ménage: étape 8

Après l'étape du partage des poids, tous les ménages ayant répondu à la grille de la vague courante ont un poids transversal. Comme seuls les ménages qui ont répondu au questionnaire ménage auront un poids transversal ménage final, il faut procéder à un autre ajustement pour la non-réponse, cette fois au niveau du ménage. Seules des données de la grille (**vague courante**) sont utilisées pour la modélisation. La modélisation est faite pour chacune des deux enquêtes (séparément).

Processus

On crée un fichier contenant les ménages qui ont répondu à la grille et qui font partie de la population cible transversale, i.e. tous ceux pour qui PTM_PAR > 0 et qui ont le bon code de réponse (CODEREP=3, 4, 5, 6 ou 7).

On définit: REPQM= 0 Non-répondant au questionnaire ménage
(CODEREP = 3 ou 7 pour tous les individus)
1 Répondant au questionnaire ménage
(CODEREP = 4, 5 ou 6 pour tous les individus).

Tableau 29 Distribution des ménages faisant partie de la population cible transversale et répondant à la grille selon la réponse au questionnaire ménage.

	PSM_I *Vague 9*	*PSM_II* *Vague 4*
Réponse au QM (REPQM)	Nombre de ménages (proportion en %)	Nombre de ménages (proportion en %)
0 – Non-répondant au questionnaire ménage	76 (2.6)	54 (3.5)
1 – Répondant au questionnaire ménage	2817 (97.4)	1489 (96.5)
Total	**2893**	**1543**

La modélisation est basée sur la grille de la vague courante soit celle de la vague 9 pour le PSM_I et celle de la vague 4 pour le PSM_II. Le modèle obtenu avec l'analyse par segmentation comporte 10 GHR pour le PSM_I et 7 pour le PSM_II. Un taux de réponse pondéré d'au moins 30% et un effectif minimal de 50 individus ont été exigé dans chaque GHR pour le PSM_I, pour le PSM_II les restrictions étaient 50% et 30 individus par GHR.

Le poids ajusté pour la non-réponse (questionnaire ménage) est PTM_NRQM. Le tableau 30 donne quelques caractéristiques numériques de cet ajustement.

Tableau 30 Statistiques descriptives du facteur d'ajustement, f_aj_nrtm, pour les respectivement 2817 et 1489 ménages répondants (voir tableau 29) des PSM_I et PSM_II. PTM_NRQM = f_aj_tm_nrqm * PTM_PAR. Comme on peut le voir dans l'annexe F, certaine GHR formées ont un taux de réponse de 100%, d'où le minimum égal à un du facteur d'ajustement.

	f_aj_tm_nrqm	
	PSM_I vague 9	PSM_II vague 4
Moyenne	1.03	1.04
Médiane	1.00	1.00
Écart-type	0.09	0.11
Min	1.00	1.00
Max	1.44	1.45
Étendue	0.44	0.45

Pour plus de détails sur la modélisation, se rapporter à l'annexe F.

8.3 Calage: étape 9

Comme pour la pondération transversale individuelle, on veut que le nombre de personnes dans les ménages représente le total de la population résidente en Suisse. Il s'agit ici de faire un calage sur marges *intégré*[21] pour le poids ménage en utilisant les comptes de population Suisse en 2007 (vague courante), d'après l'ESPOP de décembre 2007.

Le poids transversal ajusté pour la non-réponse au questionnaire ménage (PTM_NRQM) représente la population résidante en Suisse à la vague courante. L'étape du calage ajustera ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population en 2007. Cette étape est réalisée pour chacune des deux enquêtes. Les totaux se trouvent dans le tableau 32.

Poids avant calage : PTM_NRQM

Totaux de contrôle

- Sexe x (< 14 ans; 14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
- Nationalité (suisse; étranger)
- Région (région lémanique & Suisse centrale & Tessin; espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse orientale)
- État civil (marié; pas marié)

[21] Dans un calage intégré, on veut satisfaire les totaux de chaque marge au niveau individuel, tout en forçant les poids finaux (après calage) des individus d'un même ménage à rester égaux entre eux.

Pour le PSM_I, l'attrition importante que l'échantillon a subie, s'est faite ressentir à cette étape du calage. En effet, comme l'année précédente, un unique ménage s'est vu attribuer un poids après calage négatif, ce qui n'est pas acceptable. Comme les méthodologues de Statistique Canada le proposent (voir [10]), une manière de régler ce problème est d'effectuer un deuxième calage après avoir préalablement réattribué le poids avant premier calage (soit PTM_NRQM) aux observations qui posent problème.
Pour PSM_II un seul calage a suffit.
Les tableaux 31 et 32 donnent quelques nombres caractérisant les calages effectués pour PSM_I et PSM_II.

Poids calé: POIDTRME_INT

Tableau 31 Caractéristiques du calage sur marge, pondération transversale ménage. On a respectivement 7034 individus dans 2817 ménages, et 3617 individus dans 1489 ménages avec un POIDTRME>0 pour PSM_I et PSM_II. Les caractéristiques des facteurs d'ajustement sont données au niveau des individus, alors que celles des poids après calage le sont au niveau des ménages.

	facteur d'ajustement du calage = poids avant / poids après				*Poids calés*	
	PSM_I 1er calage	*PSM_I 2e calage*	*PSM_II*	*PSM_I Poids après 1er calage*	*PSM_I POIDTRME_INT 2e calage*	*PSM_II POIDTRME_INT*
	Niveau individus dans les ménages répondants au questionnaire ménage			*Niveau ménages répondants au questionnaire ménage*		
Moyenne	0.92	1.00	0.88	1045	1048	2037
Médiane	0.80	1.00	0.79	798	796	1756
Écart-type	0.45	0.01	0.44	803	802	1101
Min	**-0.32**	0.93	0.14	**-353**	100	47
Max	3.72	1.01	2.77	7206	7287	7947
Étendue	4.05	0.08	2.63	7558	7187	7900

8.4 Winsorisation des poids après calage: étape 9bis

Pour le PSM_I, le rapport poids maximal sur poids minimal vaut 73, c'est moins que l'année précédente (227.8), mais pour rester cohérent d'une année à l'autre, nous avons décidé de tout de même winsoriser les poids calés aux 1er et 99e percentiles. De même pour le PSM_II où ledit rapport vaut 169. Ainsi le rapport max/min est passé à 20 pour le PSM_I, à 13 pour le PSM_II. Le calage intégré (poids égal pour tous les individus d'un même ménage) représente, en plus des marges à respecter une contrainte forte pour le problème d'optimisation qu'il reflète. C'est pourquoi, avec une attrition comme celle que le PSM_I subit, les poids finaux comportent certaines valeurs extrêmes qui ne sont pas forcément réalistes.
Un autre moyen d'éviter les poids extrêmes, outre la winsorisation, serait de, par exemple, renoncer à caler sur l'état civil. Néanmoins, nous n'avons pas choisi cette option car elle implique des différences plus importantes dans la distribution des

poids 2006 par rapport à celles des années précédentes que si l'on tronque la distribution comme cela a été fait.

Poids final: POIDTRME_INT_WIN[22]

Des statistiques descriptives des poids winsorisés se trouvent dans l'annexe H.

Tableau 32 Effectifs de population estimés avant et après calage pour PSM_I et PSM_II. Poids transversal ménage. Notons que ces totaux reflètent toute la population (tous âges, y compris les moins de 14 ans). Les totaux après winsorisation obtenus avec les POIDTRME_INT_WIN ne diffèrent que très peu de ceux-ci:

	Estimation avec le poids avant calage (PTM_NRQM)		*Population (et estimation avec le poids après calage) POIDTRME_INT*	*POIDTRME_INT_WIN*	
Variable	*PSM_I vague 9 (avant 1er calage)*	*PSM_II vague 4*		*PSM_I vague 9*	*PSM_II vague 4*
Homme < 14 ans	580389	711161	559468	562083	557453
Homme 14-24 ans	597207	712552	485575	485499	488611
Homme 25-34 ans	327991	349557	485027	478379	485480
Homme 35-44 ans	566482	645053	610375	607834	608230
Homme 45-54 ans	564549	685990	550107	550191	552705
Homme 55+ ans	1089160	1094235	944765	949319	946982
Femme < 14 ans	554507	651840	529354	530410	524566
Femme 14-24 ans	552785	559461	461919	462600	457426
Femme 25-34 ans	322006	399606	490511	486688	488868
Femme 35-44 ans	623909	728327	608983	610126	610781
Femme 45-54 ans	655691	753529	543721	542034	546768
Femme 55+ ans	1269935	1296985	1099785	1104428	1101721
Suisse	6908202	7681608	5781145	5824299	5806118
Étranger	796410	906687	1588445	1545291	1563472
Lémanique, Centrale et Tessin	2719061	2731192	2364422	2364839	2373875
Mitteland, N-O, Zürich, Orientale.	4985551	5857103	5005168	5004751	4995715
Non marié	4120575	4586378	3956800	3960945	3955896
Marié	3584037	4001917	3412790	3408645	3413694
Total	**7704612**	**8588296**	**7369590**	**7369590**	**7369590**

[22] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal ménage, donc POIDTRME, a été renommée en:

- Wh07TP1P pour le PSM_I vague 9
- Wh07TP2P pour le PSM_II vague 4

9 Pondération longitudinale individuelle

Une fois que les poids des longitudinaux adultes ont été ajustés pour la non-réponse à la grille, la pondération longitudinale se poursuit avec des étapes distinctes de celles de la pondération transversale. Les poids des longitudinaux adultes ajustés pour la non-réponse à la grille sont ajustés pour la non-réponse au questionnaire individuel. Puis, un calage sur marges permet que les effectifs de population estimés correspondent à certains totaux de contrôle de la population suisse à la vague 1 (on utilise les nombres de l'ESPOP 1999 pour le PSM_I et ceux de l'ESPOP 2004 pour le PSM_II). Dans ce chapitre, on traite des étapes 10 et 11 de la figure 19.

9.1 Ajustement pour la non-réponse – questionnaire individuel: étape 10

Une fois que le ménage a rempli la grille, les individus peuvent répondre au questionnaire individuel, peu importe que le questionnaire ménage ait été rempli ou non. C'est pourquoi, pour être cohérent avec le processus de collecte et parce que peu d'information du questionnaire ménage s'avérerait utile pour la modélisation, nous faisons l'ajustement pour la non-réponse au questionnaire individuel directement à partir du poids ajusté pour la non-réponse à la grille.

Processus

On crée un fichier contenant les individus longitudinaux adultes (LONGST=1) qui sont dans des ménages ayant répondu à la grille ou qui sont hors cible transversalement (CODEREP=1,3,4,5,6 ou 7), i.e. tous ceux pour qui P_NRGRIL > 0.

On définit : REPQI = 0 Non-répondant au questionnaire individuel (CODEREP=3, 4 ou 5)
1 Répondant au questionnaire individuel (CODEREP=6 ou 7).

Les individus hors-cible transversalement (CODEREP=1) ne subissent pas d'ajustement pour la non-réponse, leur facteur d'ajustement est donc égal à 1 (poids avant ajustement = poids après ajustement) et la macro SAS pour la segmentation `%aj_nr(…)` leur attribue un GHR=. (manquant).

Tableau 33 Distribution des longitudinaux adultes faisant partie de la population cible transversale et appartenant à des ménages répondants à la grille selon la réponse au questionnaire individuel. Notons qu'on a 189 et 41 longitudinaux hors-cible transversalement respectivement pour le PSM_I et le PSM_II. Ces derniers conserveront leur P_NRGRIL. Notons qu'on peut retrouver ces nombres dans le tableau 21.

	PSM_I Vague 9	*PSM_II Vague 4*
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1050 (21.8)	709 (24.9)
1 – Répondant au Q.I.	3770 (78.2)	2144 (75.1)
Total	**4820**	**2853**

Tableau 34 Statistiques descriptives du facteur d'ajustement, f_aj_nrlqi, pour les 3770 et 2144 individus répondants au questionnaire individuel (ceux pour qui REPQI=1, voir tableau 33). PL_NRQI = P_NRGRIL * f_aj_nrlqi. Les GHR formées où le taux de réponse atteint 100% conduisent à un facteur d'ajustement de 1.00 (i.e. pas d'ajustement).
Pour les individus hors-cible transversalement (non-compris dans les statistiques du tableau ci-dessous), on a f_aj_nrlqi=1.

	PSM_I	PSM_II		PSM_I	PSM_II
Moyenne	1.28	1.34	Min	1.00	1.00
Médiane	1.20	1.28	Max	2.38	1.95
Écart-type	0.26	0.28	Étendue	1.38	0.95

L'ajustement se fait au niveau des individus. Les données provenant de la grille de la **vague courante** ont été utilisées pour la modélisation.
Pour le PSM_I vague 9, le modèle obtenu avec l'analyse par segmentation comporte 26 GHR et 20 GHR pour le PSM_II vague 4 (voir annexe D). Comme pour l'ajustement pour la non-réponse à la grille, un taux de réponse pondéré d'au moins 50% et un effectif minimal de 30 individus étaient exigés dans chaque GHR pour le PSM_II, tandis que pour le PSM_I, les contraintes étaient un taux de réponse de 30% et 50 unités pour chaque GHR.

Le poids ajusté pour la non-réponse (questionnaire individuel) est PL_NRQI.

Les respectivement (3770+189=) 3959 et (2144+41=) 2185 des PSM_I et PSM_II ayant un poids PL_NRQI non-nul sont ceux qui auront un poids longitudinal final non-nul (F_PLONG=1).

9.2 Calage: étape 11

Le poids longitudinal ajusté pour la non-réponse au questionnaire individuel (PL_NRQI) représente la population de la Suisse au moment de la sélection de l'échantillon du PSM à la vague 1. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 1999 selon l'ESPOP pour le PSM_I, en 2004 pour le PSM_II. Cette étape est réalisée pour chacune des deux enquêtes. Les totaux se trouvent dans le tableau 36.

Poids avant calage : PL_NRQI

Totaux de contrôle[23]

1. Sexe x (14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)

Poids final : POIDLONG[24]

[23] On ne cale pas selon l'état-civil pour la pondération longitudinale pour assurer l'uniformité de cette pondération au fil des années. En effet, pour la pondération transversale, on cale selon l'état civil que depuis 2004, vagues 6 du PSM_I et 1 du PSM_II.
[24] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal individuel, donc POIDLONG, est renommée en:

Le tableau 35 donne quelques nombres caractérisant les calages effectués.

Tableau 35 Caractéristiques du calage sur marge (individus hors-cible transversalement inclus), pondération longitudinale individuelle. On a respectivement 3959 et 2185 observations avec un POIDLONG>0 pour PSM_I et PSM_II.

	facteur d'ajustement du calage = POIDLONG/PL_NRQI		*POIDLONG*	
	PSM_I	*PSM_II*	*PSM_I*	*PSM_II*
Moyenne	1.02	0.88	1515	2783
Médiane	0.93	0.85	1258	2476
Écart-type	0.25	0.31	912	1284
Min	0.66	0.56	374	629
Max	1.96	1.90	8291	11265
Étendue	1.30	1.34	7917	10637

Tableau 36 Effectifs de population estimés avant et après calage pour le PSM_I. Poids transversal individuel. Notons que ces totaux reflètent la population de 14 ans et plus.

	Estimation avec le poids avant calage (PL_NRQI)		*Population (et estimation avec le poids après calage POIDLONG)*	
Variable	PSM_I vague 9	PSM_II vague 4	PSM_I (ESPOP1999)	PSM_II (ESPOP 2004)
Homme 14-24 ans	501917	685307	463320	469973
Homme 25-34 ans	507049	420987	530410	481790
Homme 35-44 ans	590288	730396	585647	611688
Homme 45-54 ans	524310	610018	497434	513613
Homme 55 ans et plus	675158	916782	824282	888469
Femme 14-24 ans	387220	483780	445927	449624
Femme 25-34 ans	569875	463404	540607	491394
Femme 35-44 ans	641663	833054	571212	611586
Femme 45-54 ans	609298	748890	492648	509749
Femme 55 ans et plus	764981	1064097	1046725	1052980
Suisse	5019412	6208605	4861829	4841446
Étranger	752348	748110	1136383	1239420
Lémanique, Centrale et Tessin	1983768	2220163	1900541	1924056
Mitteland, N-O, Zürich, Orientale.	3787991	4736551	4097671	4156810
Total	**5771759**	**6956714**	**5998212**	**6080866**

- Wp07LP1P pour le PSM_I vague 9
- Wp07LP2P pour le PSM_II vague 4

10 Pondération transversale individuelle combinée PSM_I et PSM_II

10.1 Combinaison transversale de panels à partir des échantillons transversaux de chacun des deux panels

La pondération transversale combinée est faite en combinant les deux panels transversalement.
Fixons les notations, soient:

- t année de la vague courante: 2007,
- t_{S1} année de sélection du PSM_I: 1999,
- t_{S2} année de sélection du PSM_II: 2004.

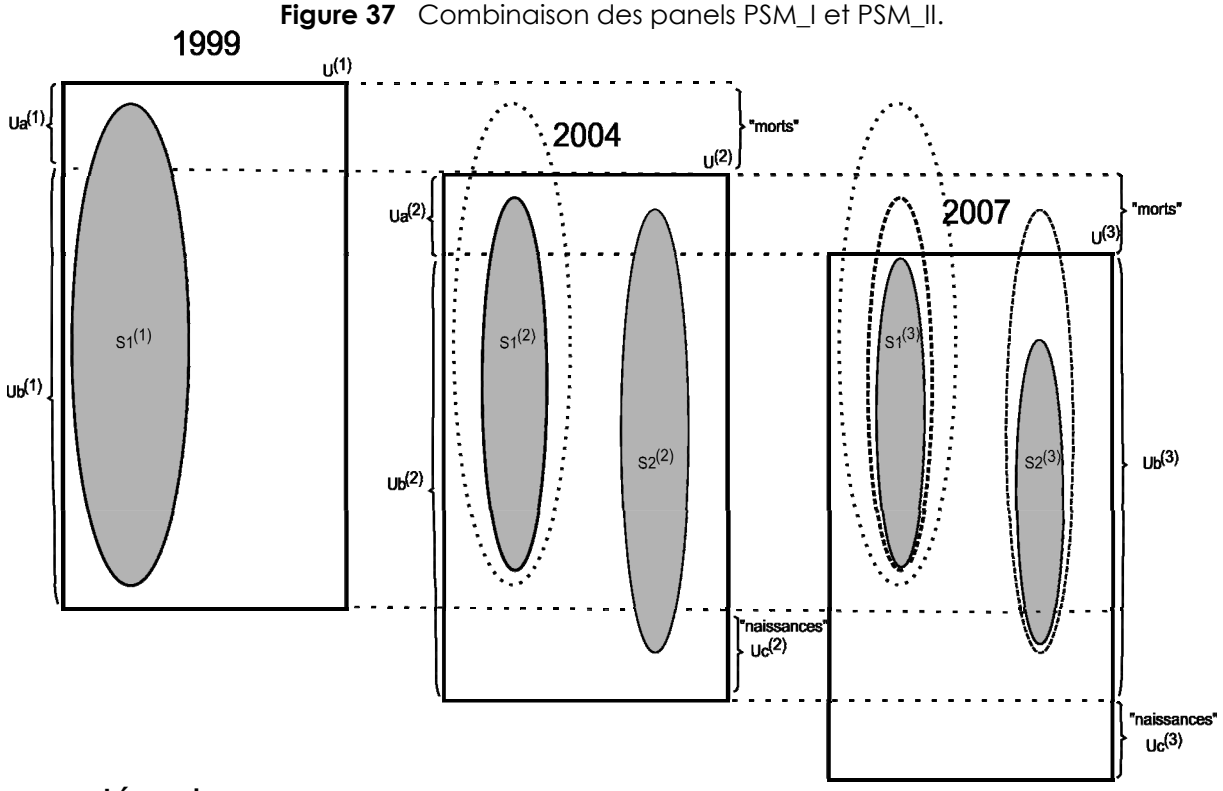
Pour combiner transversalement le PSM_I au PSM_II, on crée d'abord des poids afin que chacun des deux échantillons soit représentatif de la population transversale au temps t. En fait, on récupère les étapes de l'ajustement pour la non-réponse à la grille, du partage des poids pour les cohabitants et de l'ajustement pour la non-réponse au questionnaire individuel déjà faites séparément pour chacune des deux enquêtes. On part pour chacun des deux PSM du poids PTI_NRQI (étape 5, voir figure 19). On obtient donc deux « panels » représentatifs de la population. Ensuite, on combine les deux échantillons transversaux.

Notons que les individus de $U_a^{(1)}$ (voir figure 37) ayant quitté la population cible entre la sélection du PSM_I et le temps t_{S2} de la sélection du PSM_II ne peuvent pas faire partie du PSM _II. De plus les individus de $U_a^{(1)}$ et $U_a^{(2)}$ ayant quitté la population cible entre la sélection du PSM_I et le temps t ne font pas partie de l'échantillon combiné $S_1^{(3)} \cup S_2^{(3)}$.

Les différences entre les populations $U^{(1)}$, $U^{(2)}$ et $U^{(3)}$ sont dues aux personnes qui sont sorties de la population cible entre septembre 1999 et septembre 2004, ainsi qu'entre septembre 2004 et septembre 2007 (personnes décédées, institutionnalisées, parties à l'étranger, etc.) aux les nouveaux arrivés dans la population après septembre 1999, respectivement 2004 (immigrants, naissances). Une partie de l'échantillon du premier panel comprend des personnes des populations sortantes $U_a^{(1)}$ et $U_a^{(2)}$, de même, une partie de l'échantillon du deuxième panel comprend des personnes de la population sortante $U_a^{(2)}$. Les individus de ces populations ne font alors plus partie de la population suisse au moment de la vague courante qui nous intéresse en 2007. Les individus de $U_a^{(1)}$ avaient une probabilité nulle d'être choisis pour le PSM_II en 2004.
D'autre part, une partie du second panel est composée d'individus qui faisaient partie de la population suisse au moment de la sélection de cet échantillon mais qui n'étaient pas présents lors de la sélection du premier panel ($U_c^{(2)}$); ces personnes n'avaient aucune chance (probabilité nulle) d'être choisies au moment t_{S1}.
En résumé, les individus qui ont quitté la population cible (les « morts ») entre la sélection des panels et le temps t sont retirés de l'échantillon. Après le partage des poids le premier panel sera représentatif de la population au temps t moins les ménages composés uniquement d'immigrants s'étant joints à la population entre t_{S1} et t. De même, le deuxième panel sera représentatif de la population au temps t

moins les ménages composés uniquement d'immigrants s'étant joints à la population entre t_{S2} et t.

Figure 37 Combinaison des panels PSM_I et PSM_II.



Légende

$U^{(1)} = U_a^{(1)} \cup U_b^{(1)}$	Population longitudinale cible pour le PSM_I (panel 1)
$U^{(2)} = U_b^{(2)} \cup U_c^{(2)}$	Population longitudinale cible du PSM_II (panel 2). Population longitudinale cible du PSM_I & PSM_II combinés.
$U^{(3)} = U_b^{(3)} \cup U_c^{(3)}$	Population transversale cible en 2007 pour le PSM_I seul (vague 9), pour le PSM_II seul (vague 4) et pour les PSM_I & PSM_II combinés (vague 4).
$S1^{(1)}$	Échantillon des répondants du PSM_I en 1999, vague 1
$S1^{(2)}$	Échantillon des répondants du PSM_I en 2004, vague 6
$S1^{(3)}$	Échantillon des répondants du PSM_I en 2006, vague 9
$S2^{(2)}$	Échantillon des répondants du PSM_II en 2004, vague 1
$S2^{(3)}$	Échantillon des répondants du PSM_II en 2004, vague 4

Afin de développer une méthodologie qui permet de combiner l'information des échantillons transversaux des deux panels, il suffit de considérer les deux panels au moment de l'introduction du deuxième panel. La population que l'on doit considérer est $U^{(2)}$, celle effective lors de l'introduction du deuxième panel. Les échantillons longitudinaux à considérer sont les échantillons originaux choisis à la vague une des panels. L'échantillon du premier panel ayant été réduit à cause des individus qui ont quitté la population et des non-répondants. La méthode présentée ici a été développée par Merkouris (voir [14] et [16]). Soit *Y* un paramètre de la population. Pour estimer *Y* il s'agit d'utiliser l'estimateur combiné suivant (qui donne une importance relative à chacun des panels selon un certain facteur *p*) :

$$\hat{Y} = p \cdot \hat{Y}_1 + (1-p) \cdot \hat{Y}_2$$

où $\hat{Y}$ est l'estimation transversale pour une variable d'intérêt, $\hat{Y}_i$ est l'estimation produite pour cette même variable d'intérêt pour le panel *i*, et *p* est le facteur

d'allocation des panels. **La valeur optimale du facteur d'allocation des panels a été déterminée de façon à ce que la variance de $\hat{Y}$ soit minimale.** Le facteur d'allocation des panels est défini comme suit :

$$p = \frac{n_1}{n_1 + n_2(\frac{d_1}{d_2})}$$

où d_i désigne l'*effet de plan d'échantillonnage*[25] pour le panel *i* et n_i représente le nombre de personnes[26] (ou de ménages au chapitre suivant) du panel *i* répondantes au temps *t* (t étant l'année de référence traitée, 2007 en l'occurrence). Comme les plans de sondage de PSM_I et PSM_II sont identiques, on fait l'hypothèse que les effets de plan sont sensiblement les mêmes ($d_1/d_2=1$). Pour calculer les facteurs d'allocation, on utilise le nombre de personnes (ou de ménages) répondants.

Notons que bien que cette façon d'allouer des facteurs minimise la variance, l'idéal serait de minimiser l'erreur quadratique moyenne (EQM). Mais comme cette dernière n'est pas facilement mesurable dans le cadre du PSM, on a choisi de se restreindre à minimiser la variance. Précisons encore que dans notre cas, vu que le PSM_I est deux fois plus grand en nombre de répondants que le PSM_II, cette méthode donne au panel le plus vieux (PSM_I) une plus grande importance. Or on soupçonne le PSM_I d'être plus biaisé que le PSM_II. On n'aimerait donc pas que les facteurs d'allocation ne s'éloignent trop de ½, ce qui arrive tout de même dû aux différences de nombre d'observations par grande région (tableau 38).
On traite dans ce chapitre les étapes 13 et 14 de la figure 19.

10.2 Combinaison des deux panels: étape 13

On combine les deux panels afin de former un nouveau panel qui inclut les deux échantillons. La combinaison s'effectue en ajustant les poids de chaque unité selon le facteur d'allocation des panels cité plus haut.

Calcul du poids combiné:

- On applique *p* aux individus du PSM_I.
- On applique (*1-p*) aux individus du PSM_II.

Le tableau 38 présente les facteurs d'allocation calculés pour chaque grande région.

Le poids après combinaison s'obtient donc comme suit:

[25] L'effet de plan d'échantillonnage est défini comme étant le rapport entre la variance de l'estimation *sous-pondérée* (i.e. sans calage sur marge pour respecter les totaux de contrôle de la population, terminologie utilisée par StatCan) et la variance qui aurait résulté d'un échantillon aléatoire simple de même taille.

[26] Les méthodologues de Statistique Canada suggèrent de combiner les deux panels avant d'intégrer les cohabitants par la méthode du partage des poids, i.e. que les n_i représentent le nombre de répondants longitudinaux seulement. Omettre les cohabitants dans le calcul des facteurs *p* contribue à réduire l'importance du panel le plus âgé et ainsi minimiser les impacts potentiels dus au biais d'attrition.
Mais comme dans notre enquête le nombre de cohabitants est assez faible proportionnellement à celui des longitudinaux, nous n'avons pas distingué les cohabitants des longitudinaux pour le calcul des n_i qui représentent, dans notre cas, la somme des deux.

$$PTI_COMBPAN_REG = \begin{cases} PTI_NRQI * p_h & \text{pour les individus du PSM_I} \\ PTI_NRQI * (1- p_h) & \text{pour les individus du PSM_II} \end{cases}$$

Tableau 38 Facteurs d'allocation par grande région. Pondération transversale individuelle combinée du PSM_I et le PSM_II.

REGION07 h	*Nombre total de personnes n_h*	*Nombres de personnes PSM_ I: n_{h1}*	*Nombres de personnes PSM_II: n_{h2}*	*Facteur d'allocation des panels: p_h*
1	1234	821	413	0.67
2	1719	1135	584	0.66
3	987	692	295	0.70
4	1243	781	462	0.63
5	869	576	293	0.66
6	669	438	231	0.65
7	252	183	69	0.73
TOTAL	6973	4626	2347	

10.3 Calage: étape 14

Le poids combiné (PTI_COMBPAN_REG) représente la population de la Suisse à la vague courante. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2007. Les totaux se trouvent dans le tableau 40.
Poids avant calage : PTI_COMBPAN_REG.

Les totaux de contrôle sont les mêmes que pour les pondérations transversales individuelles prises pour chaque enquête séparément :

1. Sexe x (14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)
4. État civil (marié; pas marié)

Poids final : POIDTRIN.

Le tableau 39 donne quelques nombres caractérisant le calage effectué.

10.4 Winsorisation des poids après calage: étape 14bis

Le rapport poids maximal sur poids minimal vaut 94, nous avons décidé de winsoriser les poids calés aux 1er et 99e percentiles. Ainsi le rapport max/min est passé à 16. On constate (tableau 40) que la winsorisation affecte le plus les marges « étrangers » et la tranche d'âge des 25 à 34 tant chez les hommes que chez les femmes.

Poids final: POIDTRIN_WIN[27]

Des statistiques descriptives des poids winsorisés se trouvent dans l'annexe H, tableau 39 et tableau 40

Tableau 39 Caractéristiques du calage sur marge, pondération transversale individuelle. On a 6973 individus avec un POIDTRIN>0 pour PSM_I et PSM_II combinés.

	facteur d'ajustement du calage = POIDTRIN/ PTI_COMBPAN_REG	*POIDTRIN*	*POIDTRIN_WIN*
Moyenne	0.85	901	901
Médiane	0.80	695	705
Écart-type	0.27	714	639
Min	0.57	98	268
Max	2.14	9221	4161
Étendue	1.57	9124	3893

Tableau 40 Effectifs de population estimés avant et après calage pour le PSM_I combiné au PSM_II. Poids transversal individuel combiné.

Variable	*Estimation avec le poids avant calage* PTI_COMBPAN_REG	*Population (et estimation avec le poids après calage POIDTRIN) ESPOP 2007*	*Estimation avec le poids winsorisé POIDTRIN_WIN*	*Différences relative par rapport aux nbr. ESPOP pour les estimations POIDTRIN_WIN*
Homme 14-24 ans	688236	485575	487444	-0.38%
Homme 25-34 ans	360004	485027	475576	1.95%
Homme 35-44 ans	638304	610375	604916	0.89%
Homme 45-54 ans	627357	550107	552521	-0.44%
Homme 55 ans et plus	1096250	944765	946982	-0.23%
Femme 14-24 ans	563392	461919	465801	-0.84%
Femme 25-34 ans	384723	490511	480114	2.12%
Femme 35-44 ans	731286	608983	608278	0.12%
Femme 45-54 ans	720359	543721	549681	-1.10%
Femme 55 ans et plus	1291893	1099785	1109456	-0.88%
Suisse	6293386	4955440	5022093	-1.35%
Étranger	808417	1325328	1258676	5.03%
Lémanique, centrale et Tessin	2393124	1999361	1988624	0.54%
Mitteland, N-O, Zürich, orientale.	4708679	4281408	4292145	-0.25%
Pas Marié	3204303	2867978	2874550	-0.23%
Marié	3897500	3412790	3406218	0.19%
Total	**7101803**	**6280768**	**6280768**	

[27] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRIN, est renommée en wp07T1P.

11 Pondération transversale combinée ménage PSM_I et PSM_II

La pondération transversale combinée ménage se fait de la même manière que la pondération transversale combinée individuelle sauf qu'on combine deux échantillons transversaux composés de ménages et non d'individus. On combine donc les panels après l'ajustement pour la non-réponse au questionnaire ménage. Le poids avant combinaison est donc PTM_NRQM (qui a été calculé lors des pondérations transversales ménage de chacune des deux enquêtes, étape 8, voir figure 18). En dernier lieu, on procède à un calage intégré en utilisant les totaux de la population résidant dans des ménages privés de Suisse en 2007.
On développe ici les étapes 15 et 16 de la figure 19. Pour l'introduction théorique nous renvoyons le lecteur au paragraphe 10.1 du chapitre précédent.

11.1 Combinaison des deux panels: étape 15

On combine les deux panels afin de former un nouveau panel basé sur les deux échantillons. La combinaison s'effectue en ajustant les poids de chaque unité selon le facteur d'allocation des panels. Comme dans la pondération transversale individuelle, on applique un facteur d'allocation (voir paragraphe 10.1), sauf qu'ici ledit facteur est calculé au niveau des ménages.

Calcul du poids combiné:
- On applique *p* aux individus du PSM_I.
- On applique (*1-p*) aux individus du PSM_II.

Le tableau 41 présente les facteurs d'allocation calculés pour chaque grande région de Suisse.

Tableau 41 Facteurs d'allocation par région. Pondération transversale ménage du PSM_I combiné au PSM_II.

REGION07 h	*Nombre total de personnes n_h*	*Nombres de personnes PSM_I: n_{h1}*	*Nombres de personnes PSM_II: n_{h2}*	*Facteur d'allocation des panels: p_h*
1	773	505	268	0.65
2	1070	693	377	0.65
3	592	418	174	0.71
4	781	490	291	0.63
5	537	345	192	0.64
6	386	249	137	0.65
7	167	117	50	0.70
TOTAL	4306	2817	1489	

Le poids après combinaison s'obtient comme suit:

$$\text{PTM_COMBPAN_REG} = \begin{cases} \text{PTM_NRQM} * p_h & \text{pour les ménages du PSM_I} \\ \text{PTM_NRQM} * (1- p_h) & \text{pour les ménages du PSM_II} \end{cases}$$

11.2 Calage: étape 16

Comme pour la pondération transversale ménage séparée par panel, on veut que le nombre de personnes dans les ménages représente le total de la population résidant dans des ménages privés de Suisse. On procède donc à nouveau à un calage sur marges intégré en utilisant les comptes de l'ESPOP (de décembre 2007). Le poids transversal combiné (PTM_COMBPAN_REG) représente la population de la Suisse à la vague courante. Les facteurs d'ajustement se trouvent dans le tableau 42 et les totaux dans le tableau 43.

Poids avant calage : P_ PTM_COMBPAN_REG.

Totaux de contrôle

- Sexe x (moins de 14; 14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
- Nationalité (suisse; étranger)
- Région (région lémanique & Suisse centrale & Tessin; espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse orientale)
- État civil (marié; pas marié)

Poids final : POIDTRME

Comme pour la pondération transversale ménage du PSM_I (voir paragraphe 8.3), le calage du poids transversal ménage est effectué deux fois car après le premier calage, il y a un ménage dont le poids est négatif (d'une valeur de -104). Avant de refaire le second calage, on remet les poids négatifs à leur valeur avant le premier calage (soit PTM_COMBPAN_REG). POIDTRME est le poids transversal ménage final de la pondération combinée (donc après le second calage). Le tableau 42 résume les faits.

11.3 Winsorisation des poids après calage: étape 16bis

Le rapport poids maximal sur poids minimal vaut 136, nous avons décidé de winsoriser les poids calés aux 1er et 99e percentiles. Ainsi le rapport max/min est passé à 17. On constate (tableau 43) que, comme pour la pondération transversale individuelle, la winsorisation affecte plus particulièrement les marges « étrangers » et la tranche d'âge des 25 à 34 tant chez les hommes que chez les femmes.

Poids final: POIDTRME_WIN[28]

Des statistiques descriptives des poids winsorisés se trouvent dans l'annexe H, tableau 42 et tableau 43.

[28] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRME, est renommée en wh07T1P.

Tableau 42 Caractéristiques du calage sur marge, pondération transversale ménage combinée PSM_I & PSM_II. On a 10651 individus dans 4306 ménages avec un POIDTRME>0 pour PSM_I combiné au PSM_II. Les caractéristiques des facteurs d'ajustement sont données au niveau des individus, alors que celle des poids après calage le sont au niveau des ménages.

	Facteur d'ajustement du calage ($=p_{avant} / p_{après}$)				
	1er calage	*2e calage*	*Après premier calage*	*POIDTRME*	*Poids finaux POIDTRME_WIN*
	Niveau individus dont le ménage a répondu au questionnaire ménage		*Niveau ménages répondants au questionnaire ménage*		
Moyenne	0.88	1.00	735	735	735
Médiane	0.80	1.00	606	605	612
Écart-type	0.40	0.00	479	480	439
Min	**-0.15**	0.96	**-104**	40	168
Max	3.08	1.01	5494	5447	2853
Étendue	3.23	0.05	5598	5407	2684

Tableau 43 Effectifs de population estimés avant et après calage pour le PSM_I combiné au PSM_II. Poids transversal ménage combiné.

Variable	*Estimation avec le poids avant calage PTM_COMBPAN_REG*	*Population ESPOP 2007 (et estimation avec le poids après calage: POIDTRME)*	*Estimation avec le poids winsorisé POIDTRME_WIN*	*Différences relative par rapport aux nbr. ESPOP pour les estimations POIDTRME_WIN*
Homme moins de 14 ans	642256	559468	557388	0.37%
Homme 14-24 ans	647094	485575	486450	-0.18%
Homme 25-34 ans	349497	485027	478831	1.28%
Homme 35-44 ans	613532	610375	603908	1.06%
Homme 45-54 ans	616815	550107	550776	-0.12%
Homme 55 ans et plus	1097556	944765	954377	-1.02%
Femme moins de 14 ans	598546	529354	526743	0.49%
Femme 14-24 ans	557530	461919	464171	-0.49%
Femme 25-34 ans	365953	490511	481791	1.78%
Femme 35-44 ans	676899	608983	608120	0.14%
Femme 45-54 ans	696063	543721	546869	-0.58%
Femme 55 ans et plus	1299649	1099785	1110167	-0.94%
Suisse	7267376	5781145	5838753	-1.00%
Étranger	894014	1588445	1530837	3.63%
Lémanique, centrale et Tessin	2769851	2364422	2364028	0.02%
Mitteland, N-O, Zürich, orientale	5391539	5005168	5005562	-0.01%
Pas Marié	4375570	3956800	3962520	-0.14%
Marié	3785820	3412790	3407070	0.17%
Total	**8161390**	**7369590**	**7369590**	

12 Pondération longitudinale individuelle combinée du PSM_I & PSM_II

Il est possible de combiner deux panels n'ayant pas été sélectionnés au même moment si on dispose de l'information concernant l'éligibilité des membres de chacun des panels lorsque ceux-ci ont été sélectionnés. Il a été décidé de combiner les PSM_I au PSM_II au niveau longitudinal d'abord pour ensuite produire des poids à partir de ce panel longitudinal combiné. On combine donc les échantillons après l'ajustement pour la non-réponse au questionnaire individuel pour les longitudinaux (étapes 10 et 12, figure 19).

On combine les deux panels afin de former un nouveau panel qui inclut les deux échantillons. *L'année de référence du panel combiné est celle du panel le plus jeune, soit 2004*.
Les individus du premier panel qui ont quitté la population cible au temps t_{S2} (voir figure 37), ceux appartenant à la population $U_a^{(1)}$ sont exclus du panel combinant les deux échantillons.
La combinaison s'effectue en ajustant les poids de chaque unité selon le facteur d'allocation des panels (voir paragraphe 10.1) mais le calcul de ce facteur est un peu plus complexe que dans le cas des pondérations transversales ménages et individus.
Une fois les panels combinés, l'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2004.
Dans ce chapitre on traite des étapes 12, 17 et 18 de la figure 19.

12.1 Ajustement pour la non-réponse au questionnaire individuel pour le PSM_I, année de référence 2004

L'ajustement pour la non-réponse au questionnaire individuel présenté au paragraphe 9.1 (étape 10) fait dans le cadre de la pondération longitudinale du PSM_I seul doit être refait avant qu'on puisse combiner les deux panels. En effet, la pondération longitudinale du PSM_I seul fait référence à l'année 1999 alors que la pondération longitudinale combinée des PSM_I et PSM_II combinés fait référence à l'année 2004. Ceci nécessite la révision du statut des personnes du PSM_I. La variable longst_comb est une mise à jour de la variable longst (voir paragraphe 6.3) et est définie comme suit:

- pour les individus du PSM_II: $longst_comb = longst_{PSM_II}$
- pour les individus du PSM_I:

$$longst_comb = \begin{cases} longst_{PSM_I} & \text{si } longst_{PSM_I} = 1,3 \text{ ou } 4 \\ 1 & \text{si } longst_{PSM_I} = 2 \text{ et âge en } 2004 \geq 14 \text{ ans} \\ 2 & \text{si } longst_{PSM_I} = 2 \text{ et âge en } 2004 < 14 \text{ ans} \end{cases}$$

En résumé, les individus ayant un statut de longitudinal enfant dans le PSM_I, mais qui ont atteint ou dépassé l'âge de 14 ans en 2004, changent leur statut et « deviennent » des longitudinaux adultes pour la pondération combinée des panels. C'est ce qui nécessite de refaire l'ajustement pour la non-réponse au questionnaire individuel en tenant compte de ces nouveaux statuts de longst_comb (pour le PSM_I uniquement).

Processus

On crée un fichier contenant les individus longitudinaux adultes (LONGST_comb=1) qui sont dans des ménages ayant répondu à la grille ou qui sont hors cible transversalement (CODEREP=1,3,4,5,6 ou 7), i.e. tous ceux pour qui P_NRGRIL > 0.

On définit : REPQI = 0 Non-répondant au questionnaire individuel (CODEREP=3, 4 ou 5)

1 Répondant au questionnaire individuel (CODEREP=6 ou 7).

Les individus hors-cible transversalement (CODEREP=1) ne subissent pas d'ajustement pour la non-réponse, leur facteur d'ajustement est donc égal à 1 (poids avant ajustement = poids après ajustement) et la macro SAS pour la segmentation %aj_nr(…) leur attribue un GHR=. (manquant).

Tableau 44 Distribution des longitudinaux adultes faisant partie de la population cible transversale et appartenant à des ménages répondant à la grille selon la réponse au questionnaire individuel. Notons qu'on a 50 longitudinaux hors-cible transversalement pour le PSM_I en prenant l'année de référence 2004. Ces derniers conserveront leur P_NRGRIL.

	PSM_I Vague 9
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1195 (22.4)
1 – Répondant au Q.I.	4129 (77.6)
Total	**5324**

Tableau 45 Statistiques descriptives du facteur d'ajustement, f_aj_nrlqi, pour les 4129 individus répondants au questionnaire individuel (ceux pour qui REPQI=1, voir tableau 44). PL_NRQI = P_NRGRIL * f_aj_nrlqi. Les GHR formées où le taux de réponse atteint 100% conduisent à un facteur d'ajustement de 1.00 (i.e. pas d'ajustement).
Pour les 50 individus hors-cible transversalement, on a f_aj_nrlqi=1 (ces derniers ne sont pas pris en compte dans les statistiques du tableau ci-dessous).

	PSM_I		PSM_I
Moyenne	1.30	Min	1.00
Médiane	1.20	Max	2.63
Écart-type	0.28	Étendue	1.63

L'ajustement se fait au niveau des individus. Les données provenant de la grille de la **vague courante** ont été utilisées pour la modélisation.
Le modèle obtenu avec l'analyse par segmentation comporte 19 GHR pour le PSM_I. Un taux de réponse pondéré d'au moins 30 % et un effectif minimal de 50 individus étaient exigés dans chaque GHR (voir annexe G).

Le poids ajusté pour la non-réponse (questionnaire individuel) est PL2_NRQI.

Les (4129+50=) 4179 individus du PSM_I ayant un poids PL2_NRQI non-nul sont ceux qui auront un poids longitudinal combiné final non-nul (F_PLONG=1) parmi les individus du PSM_I.

12.2 Combinaison des panels: étape 17

On applique donc aux poids des individus longitudinaux un facteur d'allocation p, où $p \leq 1$ et $p = \frac{n_1}{n_1 + n_2}$. Le facteur p est en fait la proportion du nombre de longitudinaux répondants du panel le plus âgé (n_1) sur le nombre total de longitudinaux ($n_1 + n_2$).

On applique le facteur de la manière suivante :

- p aux individus du PSM_I (panel le plus âgé) qui étaient éligibles lors de la sélection du panel le moins âgé (le PSM_II) au moment t_{S2} (voir figure 37).
- $1 - p$ aux individus du PSM_II (le panel le moins âgé) qui faisaient partie de la population cible au moment t_{S1}.
- 1 (i.e. pas de facteur) aux individus du PSM_II qui ne faisaient pas partie de la population cible au moment t_{S1}. Ces individus font partie da la population $U_c^{(2)}$ des « naissances » entre 1999 et 2004 (voir figure 37).

Le nouveau panel obtenu est ainsi représentatif de la même population que le PSM_II, soit la population résidant dans des ménages privés de Suisse en 2004. Ce nouveau panel en est à sa vague 4 en 2007.
Afin de décider comment appliquer le facteur d'allocation des panels (p), il est impératif de connaître l'éligibilité des membres de chacun des panels. Pour ce faire, on a créé deux nouvelles variables:

- PRES_P1: variable indiquant si la personne faisait partie de la population cible au moment de la sélection du panel le plus âgé (le PSM_I, en 1999). Vaut 1 si la personne en faisait partie, 2 si non.
- PRES_P2: variable indiquant si la personne faisait partie de la population cible au moment de la sélection du panel le moins âgé (le PSM_II, en 2004). Vaut 1 si la personne en faisait partie, 2 si non.

Les variables PRES_P1 et PRES_P2 ont déjà dû être construites pour la pondération longitudinale combinée PSM_I & PSM_II 2005 (voir [3]). Malgré cela, un certain nombre d'individus longitudinaux (longst_comb=1) ont répondu à leur questionnaire individuel en 2007 mais ne l'avaient pas fait en 2005. Il y a donc de nouveaux cas pour lesquels il faut déterminer les valeurs de PRES_P1 et PRES_P2. Pour ces cas restants, on a appliqué l'algorithme décrit dans l'annexe J.

Les échantillons sont combinés au niveau des grandes régions. Un facteur d'allocation est donc calculé pour chaque grande région. Le poids avant combinaison est PL2_NRQI (étape 12) pour les individus du PSM_I et PL_NRQI (étape 10) pour ceux du PSM_II.
Calcul du poids combiné

- On applique p aux individus du PSM_I qui ont PRES_P2=1.

- On applique $1-p$ aux individus du PSM_II qui ont PRES_P1=1.
- On n'applique pas de facteur (donc on applique 1) aux individus du PSM_II qui ont PRES_P1=2.

Poids après combinaison: PLI_COMBPAN_REG

Le tableau 46 présente les facteurs d'allocation calculés pour chaque grande région.

Tableau 46 Facteurs d'allocation par région. Pondération longitudinale individuelle du PSM_I combiné au PSM_II.

REGION04 h	*Nombre total de personnes* n_h	*Nombres de personnes PSM_I:* n_{h1}	*Nombres de personnes PSM_II:* n_{h2}	*Facteur d'allocation des panels:* p_h
1	1129	746	383	0.66
2	1580	1035	545	0.66
3	895	629	266	0.70
4	1131	687	444	0.61
5	799	529	270	0.66
6	596	384	212	0.64
7	234	169	65	0.72
TOTAL	**6364**	**4179**	**2185**	

12.3 Calage: étape 18

Le poids longitudinal ajusté pour la non-réponse au questionnaire individuel (PLI_COMBPAN_REG) représente la population de la Suisse au moment de la sélection de l'échantillon du PSM_II à la vague 1. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2004 selon l'ESPOP. Les totaux se trouvent dans le tableau 48.

Poids avant calage : PLI_COMBPAN_REG

Totaux de contrôle
- Sexe*(14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
- Nationalité (suisse; étranger)
- Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)
- Etat civil (marié; non marié)

Poids final : PLI_COMBREG_CAL[29]

Le tableau 47 donne quelques nombres caractérisant le calage effectué.

[29] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final longitudinal individuel, donc PLI_COMBREG_CAL, est renommée en WP07L1P.

Tableau 47 Caractéristiques du calage sur marge (individus hors-cible transversalement inclus), pondération longitudinale individuelle combinée PSM_I & PSM_II. On a 6364 observations avec un PLI_COMBREG_CAL>0.

	facteur d'ajustement du calage = PLI_COMBREG_CAL/PLI_COMBPAN_REG	PLI_COMBREG_CAL
Moyenne	0.87	956
Médiane	0.82	769
Écart-type	0.25	676
Min	0.55	213
Max	1.94	8260
Étendue	1.39	8046

Tableau 48 Effectifs de population estimés avant et après calage. Poids longitudinal individuel combiné PSM_I & PSM_II. Notons que ces totaux reflètent la population de 14 ans et plus.

	*Estimation avec le poids avant calage (*PLI_COMBPAN_REG*)*	*Population* ESPOP 2004 *(et estimation avec le poids après calage* PLI_COMBREG_CAL*)*
Homme 14-24 ans	658891	469973
Homme 25-34 ans	399272	481790
Homme 35-44 ans	684968	611688
Homme 45-54 ans	580162	513613
Homme 55 ans et plus	944735	888469
Femme 14-24 ans	498202	449624
Femme 25-34 ans	431335	491394
Femme 35-44 ans	788455	611586
Femme 45-54 ans	678971	509749
Femme 55 ans et plus	1084101	1052980
Suisse	5960561	4841446
Étranger	788530	1239420
Lémanique, Centrale et Tessin	2272801	1924056
Mitteland, N-O, Zürich, Orientale.	4476290	4156810
Non marié	2947102	2686477
Marié	3801989	3394388
Total	**6749091**	**6080866**

13 Conclusions

Les pondérations décrites dans ce document ont été calquées sur celles de l'année précédente qui elles-mêmes étaient à peu de choses près de réplications de celles produites par des méthodologues de Statistique Canada (voir [3], [11]) pour les données du Panel Suisse de Ménages (PSM) les 7 premières années. On a fait de gros efforts pour étoffer le rapport au maximum par des exemples et des explications concernant les divers aspects théoriques ardus mis en pratique aux diverses étapes de la production des poids.

Cette année, comme pour la précédente, diverses mesures ont été entreprises pour lutter contre l'attrition (particularités de 2007, voir paragraphe 3.1.3) ainsi que pour essayer de prendre ces dernières en compte dans les modèles de non-réponse de la production des pondérations. L'attrition est un problème d'importance croissante au fur et à mesure que le PSM prend de l'âge. En effet, les modèles pour expliquer la non-réponse ont leurs limites et les calages sur marges deviennent difficiles. Ainsi on est forcé de souvent winsoriser les poids pour limiter leur variabilité. Cependant, comme le montrent les figures 2 et 3, en 2007-08, les gros efforts fournis ont contrebalancé l'effet de l'attrition et on a fini, pour le PSM_I, avec plus de d'interviews individuelles et ménages réalisées que l'année précédente. Par ailleurs le PSM_II a continué de s'éroder, mais moins que pour ses premières vagues.

Les données des fichiers livrés ne sont pas pondérées. Il est du ressort des utilisateurs de données d'appliquer les poids appropriés compte tenu des analyses qu'ils veulent produire. On risque de forts biais si l'on n'applique pas les poids.

13.1 Lignes directrices concernant l'application des pondérations produites

De façon générale, plus l'échantillon est petit, plus la variabilité d'échantillonnage est grande. Par conséquent les estimations concernant de petits sous-groupes de population sont tendanciellement moins fiables que celles relatives à des (sous-)groupes importants.
Dans cet ordre d'idée, il y a trois premières mesures à considérer :

(1) Lors d'exploitations transversales (au niveau individuel ou ménage), il convient toujours d'utiliser le maximum d'information à disposition:
   - si une variable est disponible dans les deux enquêtes (PSM_I et PSM_II), il faut exploiter les deux échantillons et utiliser les pondérations combinées. Le fait d'utiliser l'information ne provenant que du PSM_I ou que du PSM_II ferait aboutir à une perte de précision évitable dans les résultats produits.

(2) Lors d'exploitations longitudinales (au niveau individuel), il convient également d'utiliser le maximum d'information à disposition:
   - si l'exploitation ne repose que sur des données récoltées depuis 2004, il convient d'utiliser les pondérations qui permettent de combiner longitudinalement les deux PSM. Notons que dans ce cas la population étudiée est celle résidant dans des ménages privés de Suisse en 2004 (même pour les données du PSM_I).
   - si l'exploitation repose (même que partiellement) sur des données récoltées avant 2004, alors seul le PSM_I peut être utilisé et il convient d'appliquer les poids longitudinaux du PSM_I seul.

(3) Si une estimation repose sur moins d'une trentaine d'observations (nombre non pondéré d'observations), celle-ci ne devrait être diffusée qu'avec réserves, peu importe la valeur de son coefficient de variation. En effet, plus le groupe d'observations étudié est petit, plus le risque de biais élevé est grand.

Les concepts de variance, de coefficient de variation et d'intervalle de confiance fournissent une indication de l'ampleur de la variabilité d'échantillonnage. Cependant, nous rappelons aux utilisateurs que les estimations de ces grandeurs fournies par les procédures classiques de la plupart des logiciels statistiques courants (SAS, SPSS, Stata, MS Excel,...) sont erronées lorsque les données analysées proviennent d'enquêtes complexes comme le PSM. Ces logiciels supposent entre autres qu'on a procédé à un échantillonnage aléatoire simple. C'est donc dire qu'ils ne tiennent pas compte des caractéristiques spéciales du plan de sondage du PSM comme la stratification, ni des corrections apportées pour la non-réponse, ni de la post-stratification (calage sur marge). Même si les procédures classiques de ces logiciels permettent le recours à la pondération dans la production d'estimations, elles ne prennent pas correctement en compte toutes les spécificités de l'enquête PSM lors du calcul des estimations de la variance, un élément essentiel pour la plupart des tests statistiques. En fait, on a tendance à sous-estimer ladite variance et par conséquent les intervalles de confiance liés à des estimations.

13.2 Recommandations concernant les estimations de la précision

Malgré le fait que la variance est souvent sous-estimée à l'aide des logiciels cités au paragraphe précédent, nous recommandons de la calculer tout de même mais en utilisant impérativement la pondération. Effectivement, dans le cas des intervalles de confiance autour d'estimations produites par exemple, ceux-ci auront tendance à être trop étroit. Mais si une différence ou une grandeur s'avère déjà être non significative avec un tel calcul, il ne sera pas nécessaire d'investiguer davantage.
Des travaux récents ont montré (voir [23]) qu'il est possible avec notamment les logiciels SAS et SPSS d'utiliser des procédures permettant de prendre en compte toutes les spécificités du plan de sondage. Il s'agit des procédures SURVEY de SAS et des procédures du module COMPLEX SAMPLE de SPSS. Elles calculent la variance selon la méthode de linéarisation en série de Taylor et donnent des résultats très similaires (du moins pour le PSM) à ceux produit par des méthodes de ré-échantillonnage plus lourdes à implémenter comme le bootstrap ou le jackknife. De ce fait, pour les utilisateurs de SAS ou de SPSS, dans la mesure du possible, nous recommandons d'utiliser ces procédures qui, assurément, fournissent des résultats bien plus corrects que ceux fournis par les procédures standards.

14 Bibliographie

[1] Graf, E (2008). *Pondérations du SILC pilote: SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés*. Office Fédéral de la Statistique, N°338-0051.

[2] Cornali, A. et Vonlanthen, C. (2000). *Description de la pondération du Panel Suisse de Ménages*. Document interne de l'Office Fédéral de la Statistique.

[3] Cauchon, C. et Latouche, M. (2007). *Pondératon du Panel Suisse de Ménages : PSM_I vague 7, PSM_II vague 2, PSM_I et PSM_II combinés ; description détaillée des tâches.* Panel Suisse de Ménages.

[4] Cauchon, C. et Latouche, M. (2006). *Pondération du SILC, SILC vague 1, PSM II et SILC combinés, PSM I, PSM II et SILC combinés; description détaillée des tâches*. Panel Suisse de Ménage.

[5] Graf, E. et Tillmann, R. (2005). *Comparaison du déroulement des enquêtes PSMI, PSMII et SILC (2004-2005).* Working Paper 4_05, Panel Suisse de Ménage.

[6] Graf, E. et Tillmann, R. (2006). *Comparaison du déroulement des enquêtes PSM et SILC (2005-2006) : l'exemple de la gestion des interviews.* Bulletin de la société suisse de statistiques, 53 (avril 2006), 9-13.

[7] J. Dufour, F. Gagnon, Y. Morin, M. Renaud, C.-E. Särndal (1998). *Measuring the impact of alternative weighting schemes for longitudinal data.* SAA proceedings, Survey Research Methods Section, pp. 552-557.

[8] Kass, G.V. (1980). *An exploratory technique for investigating large quantities of categorical data*. Applied Statistics, 29, 119-127.

[9] Lavallée, P. (2002). Le sondage indirect ou la méthode généralisée du partage des poids.

[10] Latouche, M. (2007). *Calage des poids et estimations de la variance pour le Panel Suisse de ménages.* Panel Suisse de Ménages, document interne.

[11] Cauchon, C. et Latouche, M. (2007). *Pondération du Panel Suisse de Ménages, PSM_I vague 7, PSM II vague 2, PSM_I et PSM_II combinés; description détaillée des tâches*. Panel Suisse de Ménage.

[12] Latouche, M. et Naud, J.-F. (2005). Panel suisse de ménages, stratégie de pondération en présence de plusieurs panels. Document interne du PSM.

[13] Naud, J.-F. (2005). Deux façons de combiner des panels. Document interne du PSM.

[14] Merkouris, T. (2001). *Estimation transversale dans le cas des enquêtes auprès des ménages à panels multiples.* Techniques d'enquête, vol. 27, pp. 189-200.

[15] Ardilly, P. et Lavallée, P. (2003). *La méthode du partage des poids et l'enquête européenne sur le revenu et les conditions de vie.* Recueil du Symposium de

Statistique Canada sur les Défis à la réalisation d'enquêtes pour la prochaine décennie.

[16] Lévesque, I. et Franklin, S. (2000). *Pondérations longitudinale et transversale de l'Enquête sur la Dynamique du travail et du revenu, Année de référence 19*97. Statistique Canada, 75F0002MIF-00004.

[17] Naud, J.-F. (2004). *Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu, 1996-2002*. Statistique Canada, 75F0002MIF2004008.

[18] Latouche, M., Dufour, J. et Merkouris, T. (2000). *Pondération transversale : combinaison de deux panels ou plus*. Statistique Canada, 75F0002MIF-00006.

[19] Deville, J.-C. et Särndal, C. E. (1992). *Calibration Estimators in Survey Sampling.* Journal of the American Statistical Association, Vol87, No. 418 (June 1992), PP. 376-382.

[20] Tambay, J. L., Schiopu-Kratina, I., Mayda, J., Stukel, D. et Nadon, S. (1998). *Traitement de la non-réponse du cycle de deux enquêtes sur la santé de la population.* Techniques d'enquête, vol. 24, no 2, pp.159-169.

[21] Kalton, G. et Kasprzyk, D. (1986). *Le traitement des données d'enquête manquantes.* Techniques d'enquête, 12, pp. 1-17.

[22] Nakache, J.-P. et Confais, J. (2003). *Statistique explicative appliquée : analyse discriminante, modèle logistique, segmentation par arbre.* Editions Technip.

[23] Plaza, S. et Graf, E. (2007). *Recommandations et exemples pratiques concernant l'application des pondérations*. Panel Suisse de Ménages.

[24] Graf, E. (2008). Pondérations du PSM: PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro de commande: 338-0054

[25] M.I.S. Trend SA (2008), Panel Suisse de Ménages 2007-2008, rapport technique.

ANNEXE A – Fichiers de données pour PSM_I et PSM_II

Voici une description sommaire des fichiers de données à partir desquels les pondérations du PSM_I et du PSM_II ont été effectuées. Pour plus de détails, se rapporter aux documents de travail internes produits par l'auteur.

MÉNAGES 2007 ce fichier contient les ménages qui ont répondu à la grille à la vague 9 du PSM_I et à la vague 4 du PSM_II.
Le fichier contient 4436 enregistrements, 2893 pour le PSM_I, 1543 pour le PSM_II, 52 variables. Nom: `MEN07.sas7bdat`, construit par le programme `EGR_PSM_men07_Ind07_20081113.sas`

Tableau A –1 Liste des variables du fichier MEN07 :

IDHOUS07	identifiant du ménage
FILTER07	1= PSMI, 0= PSMII
HLINGU07	langue du ménage, 1=français, 2=allemand, 3=italien
REGION07	région du ménage (7 régions)
COM2_07	Topologie de la commune 2 : de 1999 pour les ménages du PSMI, de 2004 pour ceux du SILC et du PSMII (9 typologies), 1=Centres, 2=Communes suburbaines, 3=Communes à revenu élevé, 4=Communes périurbaines, 5=Communes touristiques, 6=Communes industrielles et tertiaires, 7=Communes pendulaires rurales, 8=Communes agraires-mixtes, 9=Communes agricoles
NBPERS07	nombre de personnes dans le ménage
NBADUL07	nombre d'adultes (18 ans et +)
NBKID07	nombre d'enfants (- de 18 ans)
H07H15	variable originelle, type d'immeuble, 1=immeuble avec plusieurs appartements, 2=maison individuelle, 3=autre type de maison/appartement.
H07H29	1=locataire, 2=propriétaire,3=occupant à titre gratuit
H07H37	frais liés au logement, en francs
H07H20	Nombre de pièces du logement
FRANC	questionnaire ménage rempli en français (1) ou dans une autre langue (0)
ALLEM	questionnaire ménage rempli en allemand (1) ou dans une autre langue (0)
ITAL	questionnaire ménage rempli en italien (1) ou dans une autre langue (0)
LAKE_GE	région du ménage = lac de Genève (1) ou autre (0)
MIDDLE	région du ménage = middleland (1) ou autre (0)
NOREST	région du ménage =nord-est (1) ou autre (0)
ZURICH	région du ménage = zurich (1) ou autre (0)
EST	région du ménage = est de la Suisse (1) ou autre (0)
CENTRAL	région du ménage = suisse centrale (1) ou autre (0)
TESSIN	région du ménage = Tessin (1) ou autre (0)
CENTRES	type de commune: agglomérations (1) ou autre (0)
SUBURB	type de commune: banlieues (1) ou autre (0)
WEALTH	type de commune: commune riche (1) ou autre (0)
PERIPH	type de commune: communes périphériques (1) ou autre (0)
TOURIST	type de commune: communes touristiques (1) ou autre (0)
INDUS	type de commune: communes industrielles ou tertiaires (1) ou autre (0)
RURAL	type de commune: cités dortoirs (1) ou autre (0)
MIXED	type de commune: communes agricoles mixtes (1) ou autre (0)
SEULE	ménage à une seule personne (1) ou à plusieurs personnes (0)

AD_SEUL	ménage à un seul adulte (avec ou sans enfants) (1) vs ménage à plusieurs adultes (0)
NO_KID	ménage sans enfants (1) ou avec enfant(s) (0)
APPART	ménage vivant dans un appartement dans un immeuble (1) ou autre (0)
DETACH	ménage vivant dans une maison indépendante (1) ou autre (0)
PROPRIO	propriétaire (1) ou autre (0)
RAT	ratio, nombre de pièces /nb de personnes
RAT1	1= premier tiers le plus bas pour le ratio nb pièces/nb de personnes (<=1.34), sinon 0
RAT2	1= second tiers le plus bas pour le ratio nb pièces/nb de personnes (>1.34, <=2), sinon 0
RAT3	1= troisième tiers le plus bas pour le ratio nb pièces/nb de personnes (> 2), sinon 0
LOY_1	premier tiers le plus bas pour le coût du logement (>=0, <=1000 fs/mois) (1) vs autre (0)
LOY_2	second tiers le plus bas pour le coût du logement (>1000, <=1500 fs/mois) (1) vs autre (0)
LOY_3	tiers le plus haut pour le coût du logement (>1500 fs/mois et +) (1) vs autre (0)
Q_INDIV	0=aucun questionnaire individuel complété dans le ménage, 1=au moins un questionnaire individuel complété dans le ménage
Nbr_Qind_complete	Nombre de questionnaires individuels complétés dans le ménage
STATHH06	1=questionnaire ménage complété, 0=questionnaire ménage non complété ou que partiellement complété
MAISINDEP	1= ménage vit dans une maison indépendante, sinon 0
MEN_RR	1= ménage est passée dans rappel de refus, sinon 0
NOTEASYADR	1= PSM_I (respectivement PSM_II) ménage avec seulement la grille en W8 (respectivement W3) et au moins le quest. ménage ou 1e quest. individuel en W7 (respectivement W2) ou ménage non-répondant total en W8 (respectivement W3) et au moins le quest. ménage ou 1e quest. individuel en W7 (respectivement W2) ou ménage non-répondant total en W8 (respectivement W3) et seulement la grille en W7 (respectivement W2)
P1ADRFROMW2_3_4	1=PSM 1 ménage complets pour la dernière fois en W2, W3 et W4, sinon 0
P1EASYADR	1= ménage complets en W8 ou seulement le quest. ménage ou seulement 1 quest. individuel et le ménage fait parti du PSM 1, sinon 0
P2EASYADR	1= ménage complets en W3 ou seulement le quest. ménage ou seulement 1 quest. individuel et le ménage fait parti du PSM 2, sinon 0

INDIVIDUS 2007 ce fichier contient tous les individus des ménages qui ont répondu à la grille à la vague 9 du PSM_I et à la vague 4 du PSM_II. Tous font partie de la population cible transversale. Ils peuvent être répondants ou non-répondants au questionnaire individuel.

Tableau A –2 Liste des variables du fichier PERS07 :

IDPERS	Identificateur de personne
IDHOUS07	Identificateur de ménage
STATUS07	Mode de collecte des données 0=répondant questionnaire individuel, 1=proxy, donc non éligible, 2=non

	répondant au Q. individuel
SEX	Sex
AGE07	Age en années à la fin 2007
BIRTHY	Année de naissance
CIVSTA07	Etat civil en 2007 1=célibataire, 2=marié(e), 3=séparé(e), 4= divorcé(e), 5= veuf/veuve
EDUCAT07	Plus haut degré de formation atteint (information du quest. grille et individu) 0 = école obligatoire inachevée 1= école obligatoire, formation professionnelle élémentaire 2 = stage ménager, 1 année d'école commerciale courte 3 = école de formation générale 4 = apprentissage (CFC) 5 = école professionnelle à plein temps 6 = baccalauréat/maturité 7 = formation professionnelle supérieure avec maîtrise, brevet fédéral 8 = école technique ou professionelle 9 = école professionnelle supérieure ETS etc. 10 = université, haute école universitaire, HEP, HES
OCCUPA07	Actual occupation
NAT_1_07	Première nationalité
NAT_2_07	Deuxième nationalité
NAT_3_07	Troisième nationalité
HAB_CH07	Durée d'habitation en Suisse en années (pour les non-suisses)
HOMME	homme (1) ou femme (0)
AGE_30	âge inférieur ou égal à 30 ans (1) ou supérieur à 30 (0)
CELIB	célibataire (1) ou autre (0)
MARIE	marié (1) ou autre (0)
SEPARE	séparé (1) ou autre (0)
DIVORCE	divorcé (1) ou autre (0)
VEUF	Veuf (1) ou autre (0)
NO_FORM	pas de formation spécifique (1) ou autre (0)
FORM_PRO	formation professionnelle (1) ou autre (0)
FORM_SUP	formation de type supérieur (1) ou autre (0)
TRAVAIL	personnes actives professionnellement (1) ou autre (0)
SWISS	nationalité suisse (1) ou autre (0)
AGE_1	classe d'âge 15 à 24 (1) sinon (0)
AGE_2	classe d'âge 25 à 34 (1) sinon (0)
AGE_3	classe d'âge 35 à 44 (1) sinon (0)
AGE_4	classe d'âge 45 à 54 (1) sinon (0)
AGE_5	classe d'âge 55 et plus (1) sinon (0)
FILTER07	**1**= PSMI, **0**= PSMII
PERMIC	Permis de séjour C (1) ou pas (0)
PERMIB	Permis de séjour B (1) ou pas (0)

10993 enregistrements, 7225 pour le PSM_I, 3768 pour le PSM_II, 33 variables. Nom: `PERS07.sas7bdat`, construit par le programme `EGR_PSM_men07_Ind07_20081113.sas`

PARTIS 2007

Ce fichier contient la liste des individus qui, au moment de la vague à pondérer, ne vivaient plus dans le ménage dans lequel ils étaient recensés lors de la **vague précédente**. La variable

OSM indique si la personne ayant quitté le ménage est une OSM[30] ou pas, selon le codage suivant:

1 = OSM
2 = Enfant d'OSM
3 = NON-OSM

Tableau A –3 Liste des variables du fichier PARTIS07 :

FILTER07	**1**= PSMI, **0**= PSMII
IDPERS	Identification of person
IDHOUS07	Identification of household
IDSOUR07	Numéro du ménage d'origine (pour les split-off, il est différent de IDHOUS06)
OSM	Statut de la personne (1=OSM,2=Enfant d'OSM, 3=Non-OSM)
G07RNH	Raison pour laquelle ne vit plus dans le ménage 1=Décès 2=hospitalisation, institutionnalisation 3=autre type d'institutionnalisation 4=divorce, séparation, rupture (y compris enfant ayant suivi un de ses parents) 5=enfant qui a quitté "définitivement" le domicile parental 6=enfant qui a quitté "momentanément" le domicile parental 7=départ à l'étranger (professionnel ou autre) 8=autre raison
SEX	Sexe, information du masterfile 2007
AGE07	Age en 2007, information du masterfile 2007
G07CIV	Etat civil en 2007 1 = célibataire, 2 = marié, 3 = séparé, 4 = divorcé, 5 = veuf
G07EDU	Plus haut degré d'éducation achevée, variable interne 0=AUCUNE formation achevée 1=école obligatoire 2=formation prof. élémentaire 3=apprentissage (niveau CFC) 4=école professionnelle /métier à plein temps 5=maturité professionnelle 6=école de degré diplôme ou de culture générale (école d'adm.) 7=gymnase/collège, école normale, études pédagogiques, baccalauréat 8=école commerciale/formation générale/stage ménager (au pair) 12=formation professionnelle supérieure avec maîtrise, brevet fédéral 13=école technique ou prof. (2 ans pl.tps/3 ans temps partiel) 14=école prof. sup./technicum /ETS (3 ans pl.tps/4 ans temps partiel) 15=université, Hautes-écoles (lic., doct., post-grade) 30=scolarité pas encore commencée 31=pré-scolarité non-obligatoire 32=école spécialisée/école pour handicapés
G07OCC	Occupation actuelle 1 = travail rémunéré plein temps (min 37 heures par semaine) 2 = travail rémunéré à temps partiel (5 à 36 heures par semaine) 3 = travail rémunéré à temps partiel (1 à 4 heures par semaine) 4 = en scolarité, apprentissage 5 = travail dans entreprise familiale 6 = travail dans atelier protégé (pour personnes handicapées) 7 = enfant, femme, home à la maison (âge max 62/65 ans) 8 = à la retraite (personnes âgées) 9 = autres personne rentières (invalidité, etc.)"

[30] OSM = Original Sample Member, i.e. un individu longitudinal.

	10 = sans emploi 11 = autre situation, congé non payé
G07NAA	First nationality
G07NAB	Second nationality
G07NAC	Third nationality
HOMME	homme (1) ou femme (0)
CELIB	célibataire (1) ou autre (0)
MARIE	marié (1) ou autre (0)
SEPARE	séparé (1) ou autre (0)
DIVORCE	divorcé (1) ou autre (0)
NO_FORM	pas de formation spécifique (1) ou autre (0). G07EDU= 0,1,2,8,30,31,32
FORM_PRO	formation professionnelle (1) ou autre (0). G07EDU= 3,4,6
FORM_SUP	formation de type supérieur (1) ou autre (0). G07EDU= 5,7,12,13,14,15
TRAVAIL	personnes actives professionnellement (1) ou autre (0). G07OCC= 1,2,4
SWISS	nationalité suisse (1) ou autre (0)
AGE_1	classe d'âge 15 à 24 (1) sinon (0)
AGE_2	classe d'âge 25 à 34 (1) sinon (0)
AGE_3	classe d'âge 35 à 44 (1) sinon (0)
AGE_4	classe d'âge 45 à 54 (1) sinon (0)
AGE_5	classe d'âge 55 et plus (1) sinon (0)

286 enregistrements au total, 181 pour le PSM_I, 105 pour le PSM_II, 29 variables. Nom: `partis07.sas7bdat`, construit par le programme *`EGR_partis07.sas`*

Tableau A – 4 Répartition de la variable OSM :

PSM_I vague 8	PSM_II vague 3
127OSM	80 OSM
6 Enfant d'OSM	6 Enfant d'OSM
48 non-OSM	19 non-OSM

REVENUS 2007

Ce fichier contient la liste des individus qui ont quitté le ménage lors des vagues précédentes et qui y sont revenus à la vague à pondérer. **Attention:** on ne prend en compte que les ménages dont les autres membres ont continué à répondre ! Il ne s'agit donc pas de ménages ayant complètement disparu une année et qui réapparaissent par la suite. Le fichier contient les mêmes variables que le PARTIS07 (mais avec la variable G07RET[31] à la place de la G07RNH).
25 enregistrements au total, 19 pour le PSM_I, 6 pour le PSM_II.
Nom : `revenus07.sas7bdat`, construit par le programme `EGR_revenus07.sas`

[31] La variable G07RET indique le retour d'anciens membres du ménage (1=oui, 2=non, elle est construite à partir des variables G07RET01à G07RET16 renseignées pour chaque personne). Ces variables ne sont renseignées qu'à partir de la vague 4 ou tout de même déjà en vague 3 dans des cas spéciaux de split-off de ménages.

LONGOUT07

Ce fichier est fabriqué à partir des fichiers PARTIS00, PARTIS01, PARTIS02, PARTIS03, PARTIS04, PARTIS05, PARTIS06, PARTIS07 et des fichiers personnes (les plus récents) de chaque vague. Il contient la liste des individus longitudinaux (OSM) qui ont quitté leur ménage entre la vague 1 et la vague à pondérer et qui ne sont recensés dans aucun autre ménage, donc jamais revenus (individus non présents dans le fichier individuel de la vague à pondérer (=vague 9 pour le PSM_I, vague 4 pour le PSM_II)), avec la raison du départ.
Ce fichier ne contient donc pas que des longitudinaux qui sont hors de la population cible de l'année en cours, car certains des individus peuvent vivre en Suisse dans un ménage qui n'a pas pu être contacté.

Tableau A – 5 Liste des variables du fichier LONGOUT07 :

IDPERS	Numéro d'identification de la personne
RNH	Raison du départ, *(signification des codes, voir fichier PARTIS07)*
WAVE	Vague à laquelle le départ a été enregistré *Vagues 2 à 9 (pour le PSM_II on prend le numéro de la vague à laquelle en est le PSM_I).*
FILTER07	**1**= PSM_I, **0**= PSM_II
OSM	**1**= longitudinal adulte, **2**= longitudinal enfant

1134 enregistrements au total, 873 pour le PSM_I, 261 pour le PSM_II, 5 variables. Nom : `longout07.sas7bdat`, construit par le programme `EGR_longout07.sas`

LISTPERS07

Contient, pour tous les ménages recensés en 2007, la liste de toutes les personnes qui les composaient avec leur numéro d'identification personnel (`IDPERS`) *à lal vague précédente.*
4441 enregistrements, 2893 pour le PSM_I, 1548 pour le PSM_II, 19 variables.
Nom : `listpers07.sas7bdat`, construit par le programme `EGR_listpers07.sas`

OSM07

Contient toutes les personnes de chaque vague avec l'indication quant à leur statut OSM. Ce fichier doit être remis à jour chaque année.
30'0288 enregistrements au total, deux variables : IDPERS et OSM.
Nom : `OSM_2008_10_20.sas7bdat`, construit par le programme `EGR_osm07.sas`

ANNEXE B - Dérivation des variables pour le classement des individus

Les variables produites lors de l'étape 2 (classement des individus) ont été dérivées de la manière suivante :

LONGST = **1** **Longitudinal – adulte**
(si dans INDIVIDUS de la vague 1, AGECORRO >= 14).

2 **Longitudinal – enfant**
(si dans INDIVIDUS de la vague 1, AGECORRO < 14).

3 **Cohabitant initialement présent**
(Si pas dans INDIVIDUS de la vague 1 ET
(HAB_CH07 < 1999 (PSM_I)
< 2004 (PSM_II)
ET AGECORRO>=0)).
Pour les HAB_CH07 < 0 (ne sait pas, pas de réponse), on supposera qu'ils sont initialement présents.

4 **Cohabitant initialement absent**
(Si pas dans INDIVIDUS de la vague 1 ET
(HAB_CH07 >= 1999 (PSM_I)
>= 2004 (PSM_II)
OU AGECORRO<0)).

Note : les enfants (< 14 ans) longitudinaux n'ont pas de poids longitudinal. Cependant, les enfants qui ont 14 ans ou plus à la vague 9 du PSM_I ou à la vague 4 du PSM_II ont été traités comme des longitudinaux lors de la pondération *transversale* et ils ont un poids transversal, même si l'enfant est le seul membre longitudinal du ménage.

CODEREP = **0** **Enfants**, AGECORRC< 14

1 **Hors-cible transversalement**
(LONGST=1 ou 2 et dans LONGOUT et RNH=1, 2, 3 ou 7)

2 **Non-répondant total**
(AGECORRC>= 14 et pas dans INDIVIDUS 2007 et pas dans LONGOUT et avec RNH=1, 2, 3 ou 7)

3 **Répondant grille seulement**
(AGECORRC >= 14 et dans INDIVIDUS 2007 et STATUS06=1 ou 2 et STATHH07 = 0)

4 **Répondant grille et questionnaire ménage seulement, dans un ménage ou tous les individus sont non-répondants au questionnaire individuel**
(AGECORRC >= 14 et dans INDIVIDUS 2007 et STATUS07=1 ou 2 et STATHH07 = 1 et Q_INDIV=0)

5 **Répondant grille et questionnaire ménage seulement, dans un ménage ou au moins un individu est répondant au questionnaire individuel**
(AGECORRC >= 14 et dans INDIVIDUS 2007 et STATUS07=1 ou 2 et STATHH07 = 1 et Q_INDIV=1)

6 **Répondant grille, questionnaire ménage et questionnaire individuel**
(AGECORRC >= 14 et dans INDIVIDUS 2007 et STATUS07=0 et STATHH07=1)

7 **Répondant grille et questionnaire individuel seulement**
(AGECORRC >= 14 et dans INDIVIDUS 2007 et STATUS07=0 et STATHH07=0)

Notons que tous les enfants (i.e. < 14 ans à la vague courante) sont hors-cible transversalement.

Indicateurs de poids

Ces variables indiquent, pour chaque individu, s'il a un poids positif ou nul.

Poids longitudinal (individuel):

F_PLONG = 1 Poids longitudinal positif (LONGST=1, CODEREP=1, 6 ou 7)
0 Poids longitudinal nul (autres)

Poids transversal ménage:

F_PTRME = 1 Poids transversal ménage positif (CODEREP=4, 5 ou 6)
0 Poids transversal ménage nul (autres)

Poids transversal individuel:

F_PTRIN = 1 Poids transversal individu positif (CODEREP=6 ou 7)
0 Poids transversal individu nul (autres)

*Construction des variables d'âge pour **PSM_I vague 9***

IF age99 >= 0 THEN agecorro = age99;
ELSE IF age07 >= 0 THEN agecorro = age07-8;
ELSE IF annee1 >= 1800 THEN agecorro = 1999-annee1;
ELSE IF annee9 >= 1800 THEN agecorro = 1999-annee9;
→ Les bébés nés après la première vague auront un AGECORR0 négatif.
Annee1, annee9 : année de naissance communiquée en vague 1 respectivement 9.

Imputation des âges manquants
Âges imputés cette année et aussi les 2 années précédentes :
IF idpers=2402101 THEN agecorro = 45;
IF idpers=2782101 THEN agecorro = 27;
IF idpers=9090101 THEN agecorro = 70;
Âges imputés cette année mais pas les années précédentes : on a mis l'âge du (supposé) partenaire
IF idpers=6233105 THEN agecorro = 42;

agecorrc= agecorro + 8

*Construction des variables d'âge pour **PSM_II vague 3***

IF age07 >= 0 THEN agecorro = age07-3;
ELSE IF age04 >= 0 THEN agecorro = age04;
ELSE IF annee4 >= 1800 THEN agecorro = 2004-annee4;
ELSE IF annee1 >= 1800 THEN agecorro = 2004-annee1;
→ Les bébés nés après la première vague auront un AGECORR0 négatif.
Annee1, annee4 : année de naissance communiquée en vague 1 respectivement 4.

Imputation des âges manquants

if IDPERS=20046101 then agecorro=57;
if IDPERS=20046102 then agecorro=57;

```
if IDPERS=20873102 then agecorro=19;
if IDPERS=20873103 then agecorro=21;
if IDPERS=21103102 then agecorro=59;
if IDPERS=21811101 then agecorro=27;
if IDPERS=21811102 then agecorro=27;
if IDPERS=22647102 then agecorro=56;
if IDPERS=23562102 then agecorro=44;
if IDPERS=24640102 then agecorro=68;
if IDPERS=20046103 then agecorro=39;
if IDPERS=22604104 then agecorro=22;
if IDPERS=22864111 then agecorro=38;
if IDPERS=23482102 then agecorro=50;
if IDPERS=23784101 then agecorro=44;

agecorrc=  agecorro + 3
```

Note sur l'imputation des âges manquants : on impute un âge seulement aux individus adultes ayant répondu au questionnaire ménage ou au questionnaire individuel. Aux autres on attribue un âge de 9999 ans afin qu'ils soient traités comme des adultes tout en restant facilement identifiables.
Les imputations d'âge manquant se font au cas par cas, manuellement, en étudiant la situation sociale/dans son ménage de la personne considérée.

ANNEXE C – Modélisation de la non-réponse au questionnaire grille

Modélisation de la non-réponse – variables utilisées

Les variables servant à la modélisation sont des variables dichotomiques qui caractérisent les individus et les ménages et qui proviennent des données (questionnaires grille) de la vague 1 (1999 pour le PSM_I, 2004 pour le PSM_II). Leur description et signification se trouve dans l'ANNEXE A.
Voici une liste des variables qui pouvaient potentiellement faire partie du modèle de non-réponse (en **gras** celles qui ont été retenues pour le modèle :

Niveau ménage

Variable (valeur en vague 1)	Utilisée pour PSM_I en 06	**Utilisée pour PSM_I en 07**	Utilisée pour PSM_II en 06	**Utilisée pour PSM_II en 07**
FRANC	NON	NON (voir figure C-1)	NON	NON (voir figure C-1)
ITAL	OUI	**OUI**	OUI	OUI
ALLEM	**OUI**	**OUI**	**OUI**	OUI
LAKE_GE	NON	NON (voir figure C-1)	NON	NON (voir figure C-1)
MIDDLE	**OUI**	**OUI**	**OUI**	**OUI**
NOREST	**OUI**	**OUI**	**OUI**	OUI
EST	**OUI**	**OUI**	OUI	OUI
CENTRAL	**OUI**	OUI	**OUI**	OUI
ZURICH	**OUI**	OUI	OUI	OUI
MIXED	OUI	**OUI**	**OUI**	OUI
CENTRES	**OUI**	**OUI**	**OUI**	OUI
SUBURB	**OUI**	**OUI**	OUI	OUI
WEALTH	OUI	OUI	OUI	OUI
PERIPH	OUI	**OUI**	OUI	OUI
TOURIST	OUI	OUI	OUI	OUI
INDUS	**OUI**	OUI	OUI	OUI
RURAL	OUI	**OUI**	**OUI**	OUI
AD_SEUL	NON	NON (voir figure C-1)	NON	NON (voir figure C-1)
NO_KID	**OUI**	**OUI**	**OUI**	OUI
SEULE	OUI	**OUI**	**OUI**	OUI
LOY_1	**OUI**	**OUI**	OUI	OUI
LOY_2	**OUI**	**OUI**	**OUI**	**OUI**
LOY_3	OUI	**OUI**	**OUI**	**OUI**
RAT1	OUI	NON (voir figure C-1)	OUI	NON (voir figure C-1)
RAT2	**OUI**	**OUI**	**OUI**	OUI
APPART	**OUI**	NON (voir figure C-1)	**OUI**	NON (voir figure C-1)
DETACH	NON	NON (voir figure C-1)	NON	NON (voir figure C-1)
PROPRIO	**OUI**	**OUI**	**OUI**	OUI
REV_1[32]	**OUI**	**OUI**	NON	NON
REV_2	**OUI**	**OUI**	NON	NON
REV_3	**OUI**	**OUI**	NON	NON

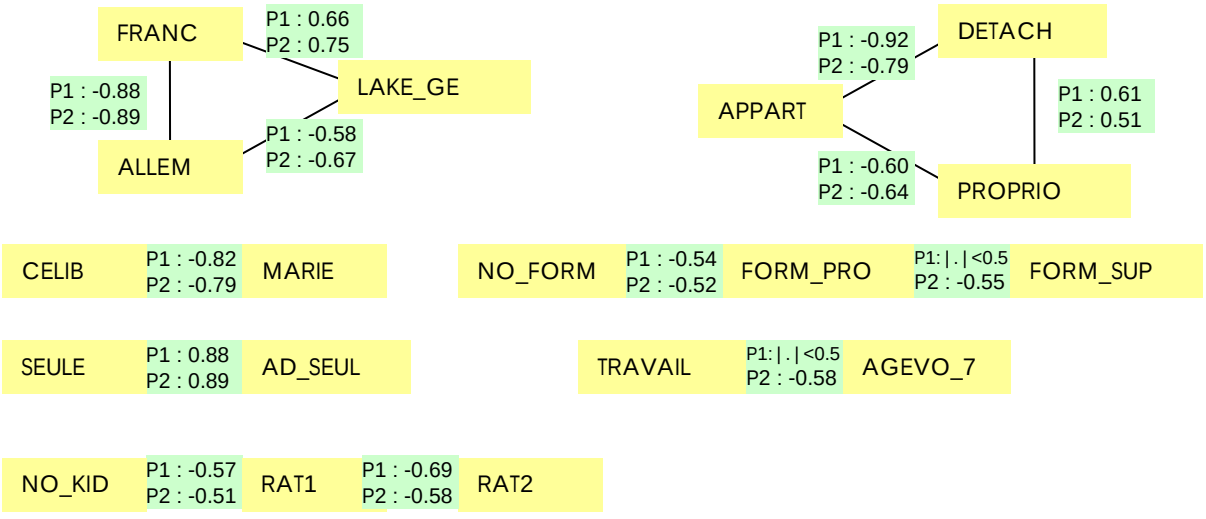
[32] REV_1 à REV_3 classifient le revenus des ménages répondants à la vague 1 de 1999 en trois groupes (premier, deuxième et troisième tiers). Ces variables ne sont disponibles que pour la vague 1 du PSM_I.

Niveau individu

On a introduit une nouvelle définition des classes d'âge en 2007, en séparant les 15 à 19 ans, des 20 à 24 ans. En effet on a observé que ces deux sous-groupes de population se comportaient de manière opposée en matière de taux de participation. On a également rajouté la classe d'âge des 55 à 64 ans.

Variable (valeur en vague 1)	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**	**Description**
AGEV1_1	**OUI**	OUI	OUI	OUI	15-19 ans (à la vague 1)
AGEV1_2		OUI		OUI	20-24 ans (à la vague 1)
AGEV1_3	**OUI**	**OUI**	OUI	OUI	25 à 34 ans (à la vague 1)
AGEV1_4	**OUI**	**OUI**	OUI	OUI	35 à 44 ans (à la vague 1)
AGEV1_5	**OUI**	**OUI**	OUI	OUI	45 à 54 ans (à la vague 1)
AGEV1_6	**OUI**	**OUI**	**OUI**	**OUI**	55 à 64 ans (à la vague 1)
AGEV1_7		**OUI**		OUI	65+ ans (à la vague 1)
CELIB	OUI	NON (voir figure C-1)	**OUI**	NON (voir figure C-1)	Voir annexe A.
DIVORCE	OUI	OUI	**OUI**	OUI	
SEPARE	OUI	OUI	OUI	OUI	
MARIE	**OUI**	**OUI**	**OUI**	OUI	
FORM_PRO	OUI	OUI	OUI	NON (voir figure C-1)	
FORM_SUP	**OUI**	**OUI**	**OUI**	OUI	
NO_FORM	**OUI**	**OUI**	**OUI**	OUI	
HOMME	**OUI**	**OUI**	OUI	OUI	
SWISS	**OUI**	**OUI**	OUI	OUI	
TRAVAIL	**OUI**	**OUI**	OUI	NON (voir figure C-1)	

Figure C – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.5 en valeur absolue. On donne ici les corrélations observées.



Modélisation de la non-réponse – analyse par segmentation

PSM_I vague 9

Le modèle de non-réponse créé avec l'analyse par segmentation a utilisé 31 variables différentes (dont 25 en commun avec la modélisation pour 2006) et conduit à la création de 64 groupes homogènes de réponse (GHR), 54 pour les longitudinaux adultes, 10 pour les longitudinaux enfants (figure C-2).

Critères d'arrêt pour la segmentation : 30% de taux de réponse et 50 observations par GHR au minimum.

PSM_II vague 4

Le modèle de non-réponse créé avec l'analyse par segmentation a utilisé 4 variables différentes (toutes aussi choisies pour la modélisation en 2006) et conduit à la création de 6 GHR pour les longitudinaux adultes 4 pour les longitudinaux adultes, 2 pour les longitudinaux enfants (figure C-3).

Critères d'arrêt pour la segmentation : 50% de taux de réponse et 30 observations par GHR au minimum.

Figure C-2: Modèle complet de non-réponse à la grille. Arbre pour les longitudinaux adultes et enfants, PSM_I vague 9

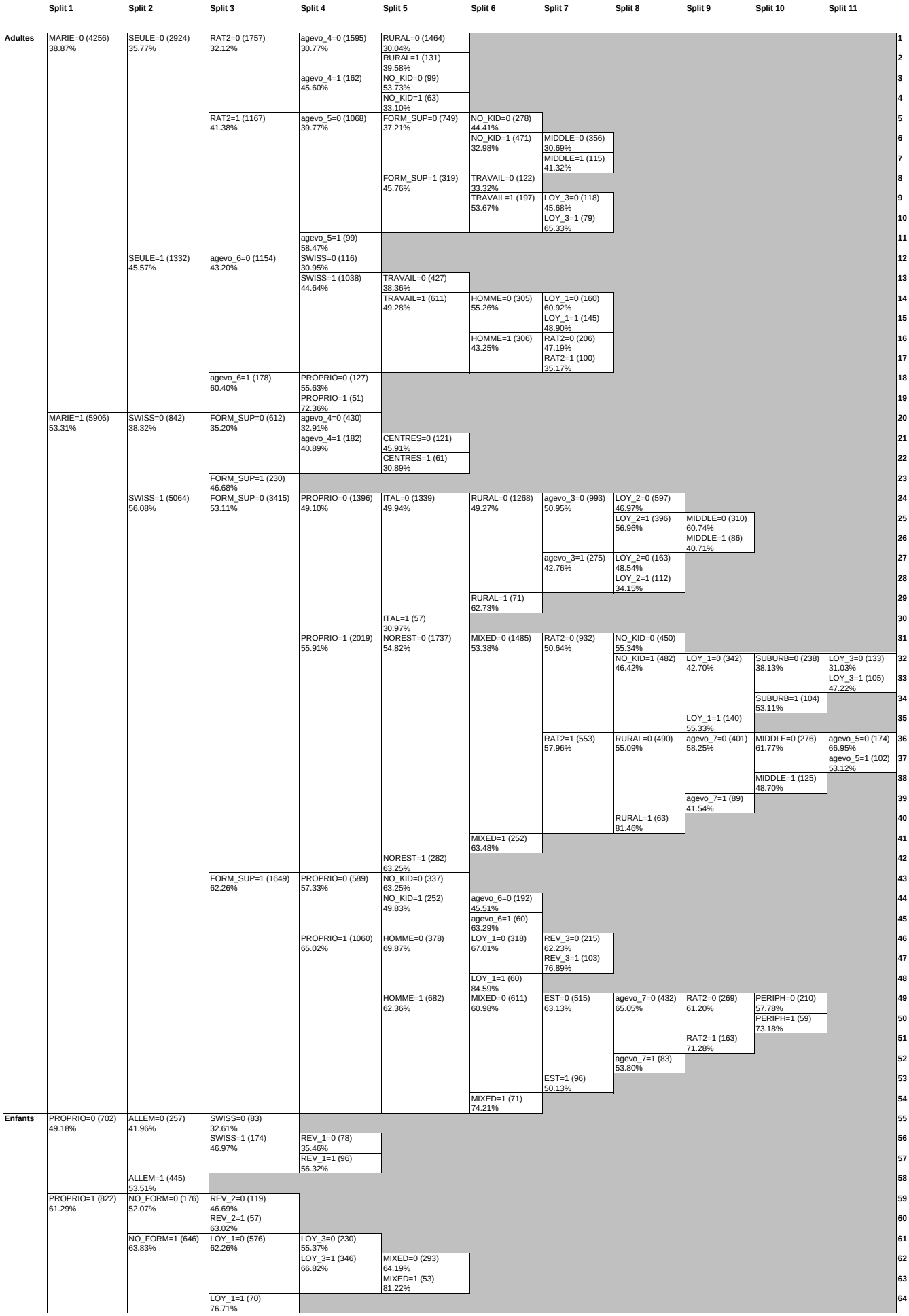


Figure C—3 Arbre pour les longitudinaux, PSM_II vague 4. L'arbre obtenu est très petit en raison des critères d'arrêt (50% et 30 observations au minimum par GHR). Le taux de réponse global du PSM_II étant à peine supérieur à 50%, la segmentation s'arrête rapidement. On a renoncé pour cette année encore à relaxer les contraintes pour mettre, comme au PSM_I 30% et 50 observations car l'arbre devenait alors, à l'inverse d'ici, trop grand.

	Split 1	Split 2	Split 3	rhg
Adultes	agevo_6=0 (4744) 50.51%			**1**
	agevo_6=1 (719) 59.45%	MIDDLE=0 (526) 56.45%	loy_3=0 (359) 52.04%	**2**
			loy_3=1 (167) 65.92%	**3**
		MIDDLE=1 (193) 68.14%		**4**
Enfants	loy_2=0 (197) 53.60%			**5**
	loy_2=1 (71) 69.04%			**6**

ANNEXE D – Modélisation de la non-réponse au questionnaire individuel, pondération longitudinale individuelle

Les variables dichotomiques utilisées pour la modélisation de la non-réponse proviennent de la grille de la vague courante. Pour la description des variables, se référer à l'annexe A.
Les variables ADRFROMW5, REFW2W7CONV, REFW3W8CONV, INC_STAMPS, INC_LOTT, INC_GIFT présentent en 2006 n'était plus disponibles en 2007. Par contre les variables MAISINDEP, MEN_RR, NOTEASYADR, P1ADRFROMW2_3_4, P1EASYADR, P2EASYADR, HH1, HH2, HH3, HH4+, APPART, DETACH, PROPRIO, LOY_1, LOY_2, LOY_3, RAT1, RAT2, RAT3 ont été ajoutées.

Niveau ménage

Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**
FRANC	NON	NON (voir figure D-1)	NON	NON (voir figure D-1)
ITAL	OUI	OUI	OUI	OUI
ALLEM	**OUI**	**OUI**	OUI	OUI
LAKE_GE	NON	NON (voir figure D-1)	NON	NON (voir figure D-1)
MIDDLE	OUI	OUI	OUI	OUI
NOREST	OUI	OUI	**OUI**	OUI
EST	**OUI**	**OUI**	OUI	OUI
CENTRAL	**OUI**	OUI	OUI	OUI
ZURICH	OUI	OUI	OUI	OUI
MIXED	OUI	OUI	OUI	OUI
CENTRES	OUI	OUI	OUI	**OUI**
SUBURB	OUI	**OUI**	OUI	OUI
WEALTH	OUI	OUI	OUI	OUI
PERIPH	OUI	OUI	OUI	OUI
TOURIST	OUI	OUI	OUI	OUI
INDUS	OUI	OUI	OUI	OUI
RURAL	OUI	OUI	OUI	OUI
AD_SEUL	NON	NON (voir figure D-1)	NON	NON (voir figure D-1)
NO_KID	**OUI**	**OUI**	OUI	NON (voir figure D-1)
SEULE	**OUI**	**OUI**	**OUI**	**OUI**
ADRFROMW5	**OUI**	**--**	**--**	**--**
REFW2W7CONV	OUI	**--**	**OUI**	**--**
REFW3W8CONV	OUI	**--**	OUI	**--**
INC_STAMPS	**OUI**	**--**	OUI	**--**
INC_LOTT	OUI	**--**	**OUI**	**--**
INC_GIFT	OUI	**--**	OUI	**--**
MAISINDEP	**--**	OUI	**--**	OUI
MEN_RR	**--**	**OUI**	**--**	**OUI**
NOTEASYADR	**--**	**OUI**	**--**	NON (voir figure D-1)
P1ADRFROMW2_3_4	**--**	NON (voir figure D-1)	**--**	--
P1EASYADR	**--**	NON (voir figure D-1)	**--**	--
HH2	**--**	NON (voir figure D-1)	**--**	OUI

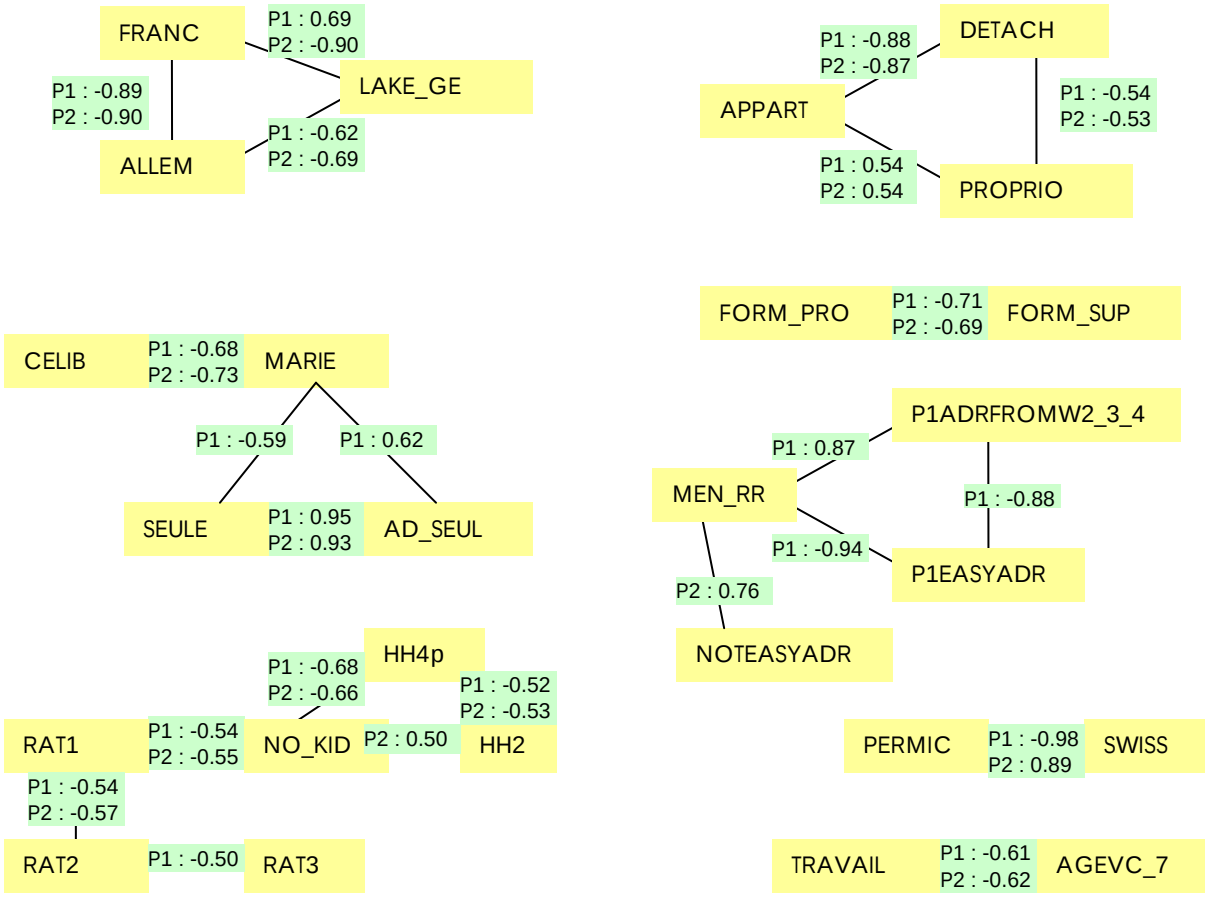
HH3	--	OUI	--	OUI
HH4+	--	OUI	--	NON (voir figure D-1)
APPART		NON (voir figure D-1)		NON (voir figure D-1)
DETACH		NON (voir figure D-1)		NON (voir figure D-1)
PROPRIO		**OUI**		**OUI**
LOY_1		**OUI**		**OUI**
LOY_2		OUI		**OUI**
LOY_3		**OUI**		**OUI**
RAT1		NON (voir figure D-1)		NON (voir figure D-1)
RAT2		NON (voir figure D-1)		**OUI**
RAT3		**OUI**		OUI

Niveau individu

On a introduit une nouvelle définition des classes d'âge en 2007, en séparant les 15 à 19 ans, des 20 à 24 ans. En effet on a observé que ces deux sous-groupes de population se comportaient de manière opposée en matière de taux de participation. On a également rajouté la classe d'âge des 55 à 64 ans.

Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**
AGEVC_1	**OUI**	**OUI**	OUI	OUI
AGEVC_2		OUI		OUI
AGEVC_3	OUI	OUI	OUI	OUI
AGEVC_4	OUI	OUI	**OUI**	OUI
AGEVC_5	OUI	OUI	**OUI**	OUI
AGEVC_6	OUI	OUI	OUI	OUI
AGEVC_7		NON (voir figure D-1)		**OUI**
CELIB	NON	OUI	NON	OUI
DIVORCE	OUI	OUI	OUI	OUI
SEPARE	OUI	OUI	OUI	OUI
MARIE	OUI	NON (voir figure D-1)	**OUI**	OUI
VEUF	OUI	OUI	OUI	OUI
FORM_PRO	NON	NON (voir figure D-1)	NON	NON (voir figure D-1)
FORM_SUP	**OUI**	**OUI**	**OUI**	**OUI**
NO_FORM	**OUI**	**OUI**	**OUI**	**OUI**
HOMME	**OUI**	**OUI**	**OUI**	**OUI**
SWISS	OUI	**OUI**	**OUI**	**OUI**
TRAVAIL	OUI	**OUI**	OUI	NON (voir figure D-1)
PERMIB	OUI	NON (trop peu d'obs.)	OUI	OUI
PERMIC	NON	NON (voir figure D-1)	NON	NON (voir figure D-1)

Figure D – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue. P1=PSM_I, P2=PSM_II.



Synthèse

11 variables différentes (8 communes avec la modélisation pour la vague précédente) ont été choisies pour le PSM_I, 13 pour le PSM_II (5 communes avec la modélisation pour la vague précédente), dont 9 se retrouvent dans les deux modèles.

Le modèle obtenu avec l'analyse par segmentation comporte 26 GHR pour le PSM_I et 20 GHR pour le PSM_II. Un taux de réponse pondéré d'au moins 30% et un effectif minimal de 50 individus ont été exigé dans chaque GHR pour le PSM_I, pour le PSM_II les restrictions étaient 50% et 30 individus par GHR.

Figure D – 2 Modèle complet de non-réponse au q.i: pondération longitudinale individuelle du PSM_I.

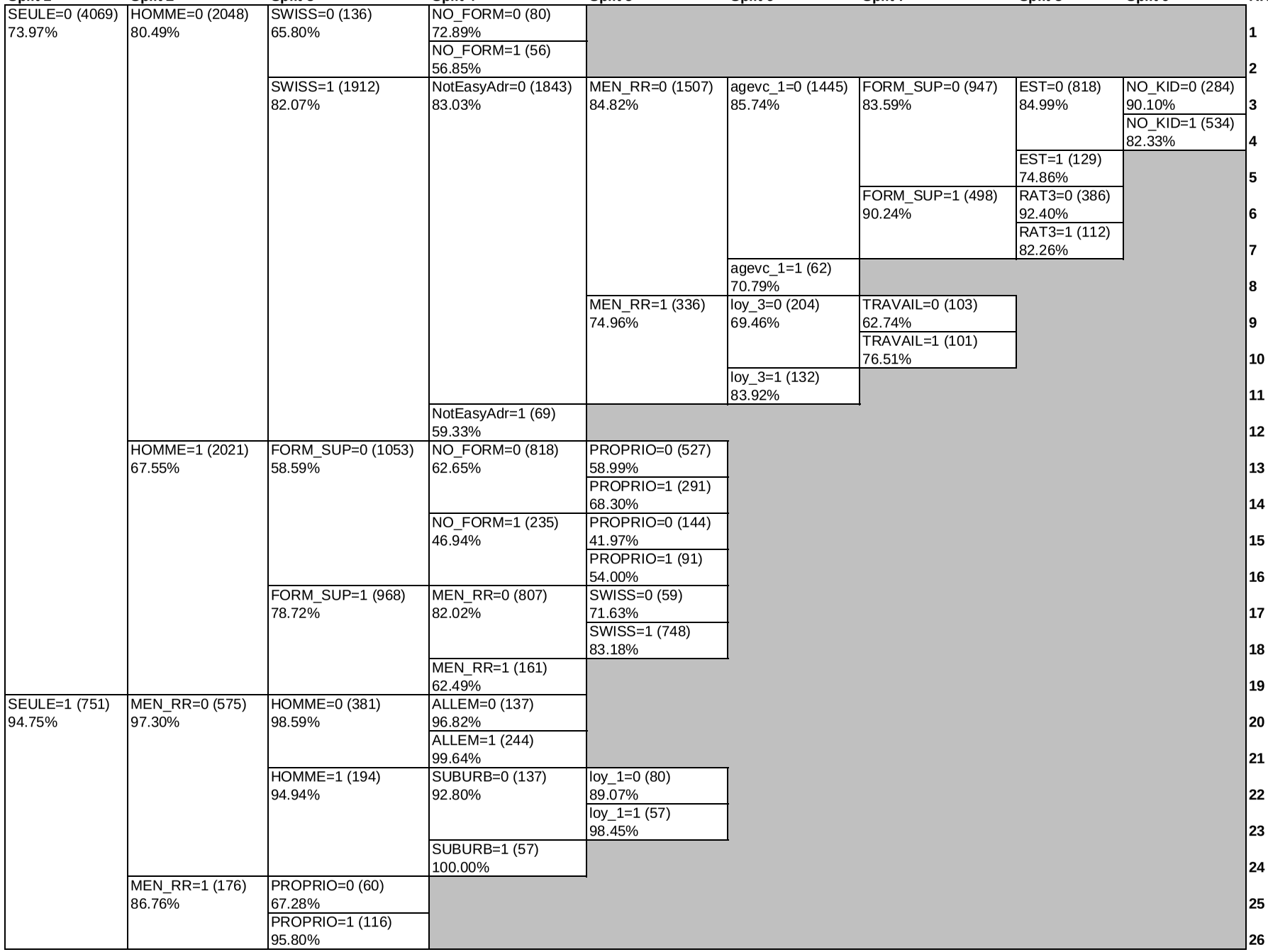
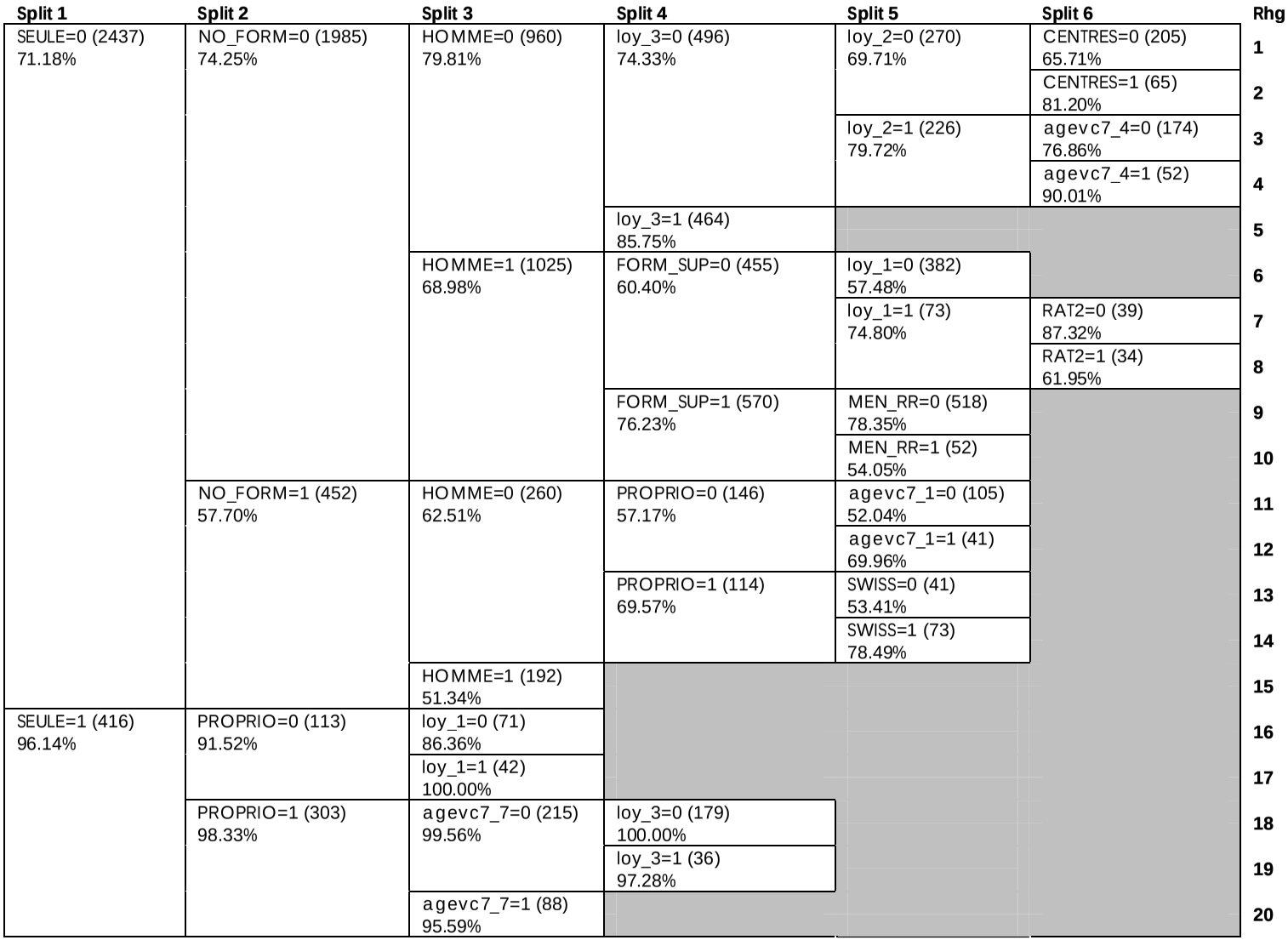


Figure D – 3 Modèle complet de non-réponse au questionnaire individuel : pondération longitudinale individuelle du PSM_II. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.



ANNEXE E – Modélisation de la non-réponse au questionnaire individuel, pondération transversale individuelle

Les variables dichotomiques utilisées pour la modélisation de la non-réponse proviennent de la grille de la vague courante. Pour la description des variables, se référer à l'annexe A.

Niveau ménage

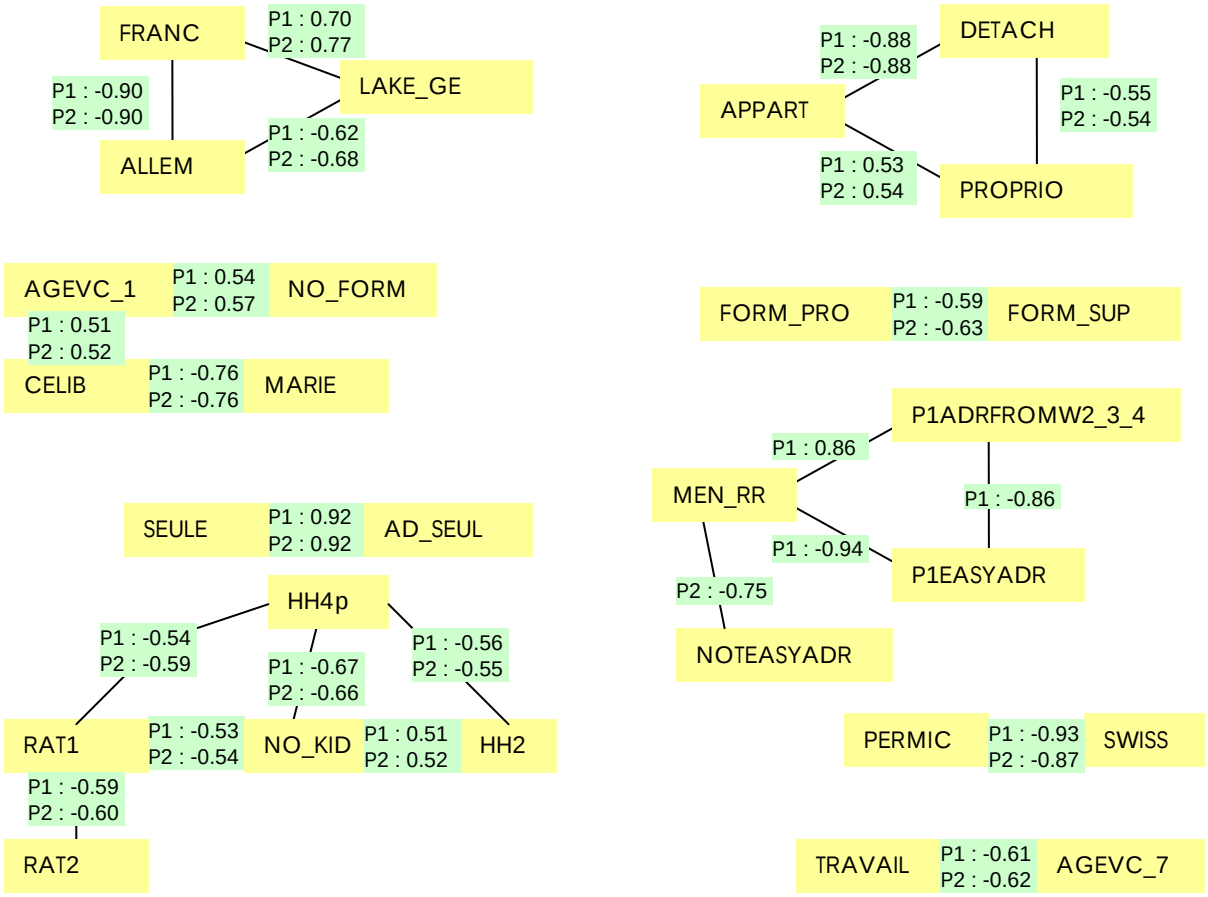
Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**
FRANC	NON	NON (voir figure E-1)	NON	NON (voir figure E-1)
ITAL	OUI	OUI	OUI	OUI
ALLEM	**OUI**	**OUI**	OUI	OUI
LAKE_GE	NON	NON (voir figure E-1)	NON	NON (voir figure E-1)
MIDDLE	OUI	**OUI**	OUI	OUI
NOREST	OUI	**OUI**	OUI	OUI
EST	**OUI**	OUI	OUI	OUI
CENTRAL	OUI	OUI	OUI	OUI
ZURICH	OUI	**OUI**	OUI	OUI
MIXED	**OUI**	OUI	OUI	OUI
CENTRES	**OUI**	OUI	**OUI**	**OUI**
SUBURB	OUI	**OUI**	OUI	OUI
WEALTH	OUI	OUI	OUI	OUI
PERIPH	OUI	OUI	**OUI**	OUI
TOURIST	OUI	OUI	OUI	OUI
INDUS	OUI	OUI	OUI	OUI
RURAL	OUI	OUI	OUI	OUI
AD_SEUL	OUI	NON (voir figure E-1)	**OUI**	NON (voir figure E-1)
NO_KID	**OUI**	**OUI**	OUI	NON (voir figure E-1)
SEULE	**OUI**	**OUI**	OUI	**OUI**
ADRFROMW5	**OUI**			
REFW2W7CONV	OUI		**OUI**	
REFW3W8CONV	OUI		**OUI**	
INC_STAMPS	**OUI**		**OUI**	
INC_LOTT	OUI		**OUI**	
INC_GIFT	OUI		**OUI**	
Loy_1		**OUI**		**OUI**
Loy_2		**OUI**		**OUI**
Loy_3		**OUI**		**OUI**
RAT_1		NON (voir figure E-1		NON (voir figure E-1)
RAT_2		**OUI**		**OUI**
RAT_3		**OUI**		OUI
PROPRIO		**OUI**		**OUI**
MAISINDEP		OUI		OUI
APPART		NON (voir figure E-1)		NON (voir figure E-1)
DETACH		NON (voir figure E-1)		NON (voir figure E-1)
MEN_RR		**OUI**		OUI
NotEasyAdr		**OUI**		NON (voir figure E-1)

P1AdrFromW2_3_4		NON (voir figure E-1)		
P1EasyAdr		NON (voir figure E-1)		
P2EasyAdr				NON (=-MEN_RR)
HH2				OUI
HH3				OUI
HH4p				NON (voir figure E-1)

Niveau individu

Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**
AGEVC_1	**OUI**	**OUI**	OUI	**OUI**
AGEVC_2		**OUI**		OUI
AGEVC_3	**OUI**	OUI	OUI	OUI
AGEVC_4	OUI	**OUI**	**OUI**	**OUI**
AGEVC_5	OUI	OUI	**OUI**	**OUI**
AGEVC_6	**OUI**	**OUI**	OUI	OUI
AGEVC_7		**OUI**		**OUI**
CELIB	NON	NON (voir figure E-1)	NON	NON (voir figure E-1)
DIVORCE	OUI	OUI	**OUI**	OUI
SEPARE	OUI	OUI	OUI	OUI
MARIE	**OUI**	OUI	**OUI**	**OUI**
VEUF	OUI	OUI	OUI	OUI
FORM_PRO	NON	NON (voir figure E-1)	NON	NON (voir figure E-1)
FORM_SUP	**OUI**	**OUI**	**OUI**	**OUI**
NO_FORM	**OUI**	**OUI**	**OUI**	OUI
HOMME	**OUI**	**OUI**	**OUI**	**OUI**
SWISS	**OUI**	**OUI**	**OUI**	**OUI**
TRAVAIL	OUI	NON (voir figure E-1)	OUI	NON (voir figure E-1)
PERMIB	OUI	NON (voir figure E-1)	OUI	OUI
PERMIC	NON	NON (voir figure E-1)	NON	NON (voir figure E-1)

Figure E – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue.



Synthèse

24 et 15 variables différentes ont été choisies pour respectivement le PSM_I et le PSM_II dont 12 se retrouvent dans les deux modèles. Le modèle obtenu avec l'analyse par segmentation comporte 39 GHR pour le PSM_I et 23 GHR pour le PSM_II. Un taux de réponse pondéré d'au moins 30% et un effectif minimal de 50 individus ont été exigés dans chaque GHR pour le PSM_I, pour le PSM_II les restrictions étaient 50% et 30 individus par GHR.

Figure E-2 Modèle complet de non-réponse au questionnaire individuel: pondération transversale individuelle du PSM_I

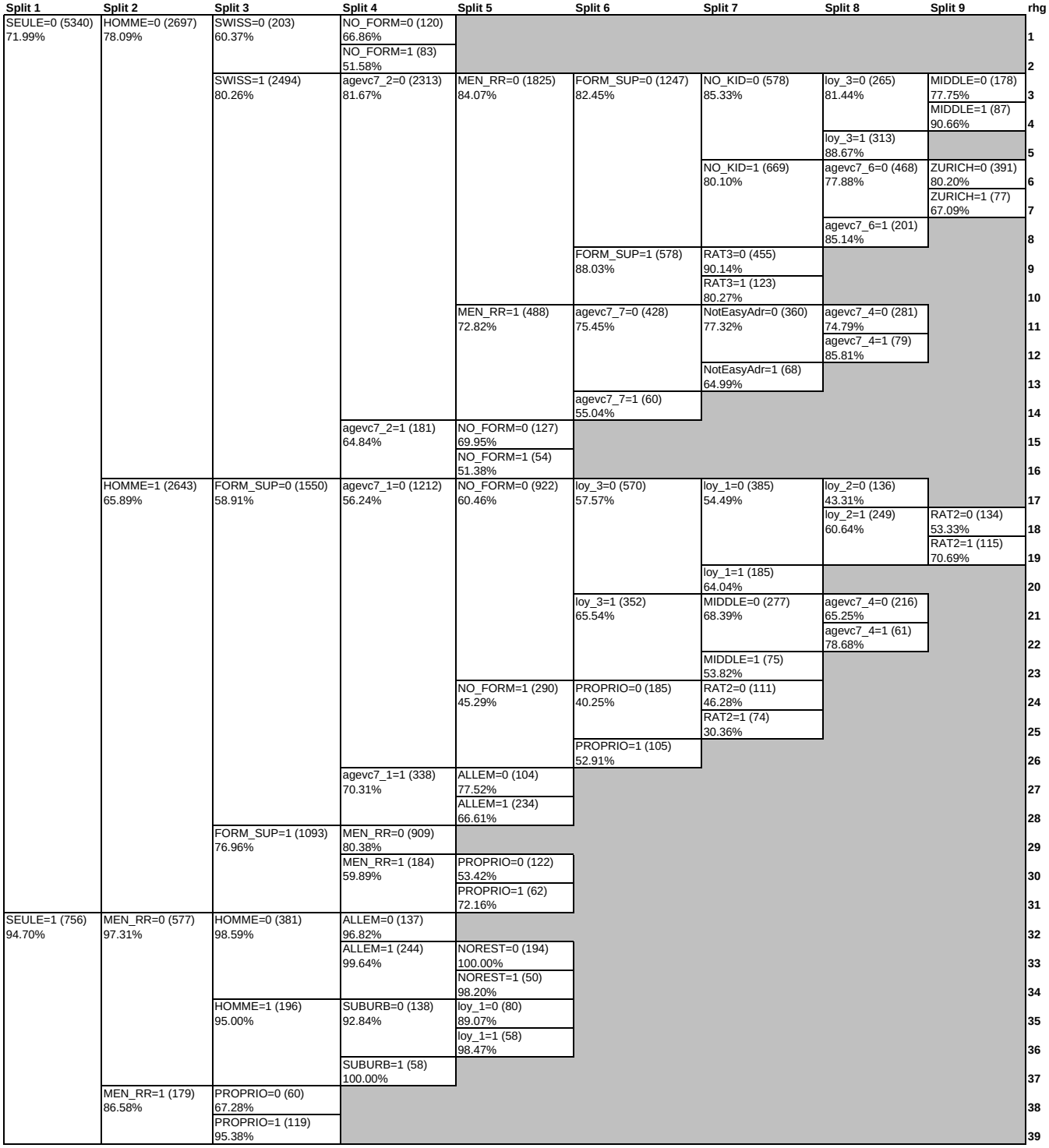


Figure E—3 Modèle complet de non-réponse au questionnaire individuel: pondération transversale individuelle du PSM_ll

Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Rhg
SEULE=0 (2758) 70.22%	FORM_SUP=0 (1716) 65.77%	HOMME=0 (951) 71.45%	loy_3=0 (560) 66.58%	loy_2=0 (320) 60.61%	loy_1=0 (167) 52.67%		1
					loy_1=1 (153) 69.16%		2
				loy_2=1 (240) 74.33%	SWISS=0 (36) 54.20%		3
					SWISS=1 (204) 77.78%	agevc7_4=0 (166) 74.88%	4
						agevc7_4=1 (38) 91.14%	5
			loy_3=1 (391) 78.59%				6
		HOMME=1 (765) 58.81%	loy_1=0 (662) 56.83%	agevc7_1=0 (502) 53.23%	loy_2=0 (338) 50.12%		7
					loy_2=1 (164) 59.45%		8
				agevc7_1=1 (160) 68.94%	CENTRES=0 (128) 73.75%	loy_3=0 (55) 64.15%	9
						loy_3=1 (73) 81.27%	10
					CENTRES=1 (32) 50.62%		11
			loy_1=1 (103) 70.70%				12
	FORM_SUP=1 (1042) 77.96%	loy_3=0 (467) 72.65%	agevc7_5=0 (377) 75.24%				13
			agevc7_5=1 (90) 62.72%	HOMME=0 (41) 75.77%			14
				HOMME=1 (49) 51.86%			15
		loy_3=1 (575) 82.13%	MARIE=0 (181) 76.18%	RAT2=0 (81) 67.55%			16
				RAT2=1 (100) 83.82%			17
			MARIE=1 (394) 84.52%				18
SEULE=1 (416) 96.14%	PROPRIO=0 (113) 91.52%	loy_1=0 (71) 86.36%					19
		loy_1=1 (42) 100.00%					20
	PROPRIO=1 (303) 98.33%	agevc7_7=0 (215) 99.56%	loy_3=0 (179) 100.00%				21
			loy_3=1 (36) 97.28%				22
		agevc7_7=1 (88) 95.59%					23

ANNEXE F – Modélisation de la non-réponse au questionnaire ménage, pondération transversale ménage

Les variables servant à la modélisation sont des variables dichotomiques qui caractérisent les ménages et les individus y vivant et proviennent des données (questionnaires grille) de la vague 1 (1999) pour le PSM_I, respectivement de la vague 1 (2004) pour le PSM_II. Voici une liste des variables qui pouvaient potentiellement faire partie du modèle de non-réponse. Pour la description desdites variables, se référer à l'annexe A.

Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**	Utilisée pour PSM_II 06	**Utilisée pour PSM_II 07**
FRANC	NON	NON (voir figure F-1)	NON	NON (voir figure F-1)
ITAL	OUI	OUI	OUI	OUI
ALLEM	OUI	OUI	OUI	OUI
LAKE_GE	NON	NON (voir figure F-1)	NON	NON (voir figure F-1)
MIDDLE	OUI	OUI	OUI	OUI
NOREST	OUI	OUI	OUI	OUI
EST	OUI	OUI	OUI	OUI
CENTRAL	OUI	OUI	OUI	OUI
ZURICH	OUI	OUI	OUI	OUI
MIXED	OUI	OUI	OUI	OUI
CENTRES	OUI	OUI	OUI	OUI
SUBURB	OUI	OUI	OUI	OUI
WEALTH	OUI	OUI	OUI	OUI
PERIPH	OUI	OUI	OUI	OUI
TOURIST	OUI	OUI	OUI	OUI
INDUS	OUI	OUI	OUI	OUI
RURAL	OUI	OUI	OUI	OUI
AD_SEUL	NON	OUI	NON	NON (voir figure F-1)
NO_KID	OUI	OUI	OUI	OUI
SEULE	OUI	OUI	OUI	OUI
ADRFROMW5	OUI			
REFW2W7CONV	OUI		OUI	
REFW3W8CONV	OUI		OUI	
INC_STAMPS	OUI		OUI	
INC_LOTT	OUI		OUI	
INC_GIFT	OUI		OUI	
APPART	OUI	NON (voir figure F-1)	**OUI**	NON (voir figure F-1)
DETACH	NON	NON (voir figure F-1)	NON	NON (voir figure F-1)
PROPRIO	**OUI**	**OUI**	**OUI**	**OUI**
RAT1	OUI	NON (voir figure F-1)	OUI	NON (voir figure F-1)
RAT2	OUI	OUI	OUI	OUI
RAT3	**OUI**	NON (voir figure F-1)	OUI	NON (voir figure F-1)
LOY_1	**OUI**	**OUI**	**OUI**	**OUI**
LOY_2	**OUI**	**OUI**	**OUI**	**OUI**
LOY_3	**OUI**	**OUI**	**OUI**	**OUI**
tous_hommes	OUI	OUI	OUI	OUI
tous_femmes	NON	NON (voir figure F-1)	NON	OUI

un_marie	NON	NON (voir figure F-1)	NON	NON (voir figure F-1)
max_no_educ	OUI	OUI	OUI	OUI
max_educ_sup	OUI	**OUI**	OUI	OUI
max_educ_pro	NON	NON (voir figure F-1)	NON	OUI
tous_suis	OUI	OUI	**OUI**	OUI
tous_etrang	NON	NON (voir figure F-1)	NON	NON (voir figure F-1)
agemax_1	OUI	OUI	OUI	OUI
agemax_2	OUI	OUI	OUI	OUI
agemax_3	OUI	OUI	OUI	OUI
agemax_4	OUI	OUI	OUI	OUI
agemax_5	**OUI**	NON (voir figure F-1)	**OUI**	OUI
MEN_RR		**OUI**		**OUI**
NotEasyAdr		**OUI**		NON (voir figure F-1)
P1AdrFromW2_3_4		NON (voir figure F-1)		
P1EasyAdr		NON (voir figure F-1)		
P2EasyAdr				NON (voir figure F-1)
HH2		OUI		OUI
HH3		OUI		OUI
HH4p		NON (voir figure F-1)		NON (voir figure F-1)

Synthèse

7 et 5 variables différentes ont été choisies pour respectivement le PSM_I et le PSM_II dont 5 se retrouvent dans les deux modèles. Le modèle obtenu avec l'analyse par segmentation comporte 10 GHR pour le PSM_I et 7 GHR pour le PSM_II. Un taux de réponse pondéré d'au moins 30% et un effectif minimal de 50 individus ont été exigés dans chaque GHR pour le PSM_I, pour le PSM_II les restrictions étaient 50% et 30 individus par GHR.

Figure F – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire ménage ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue.

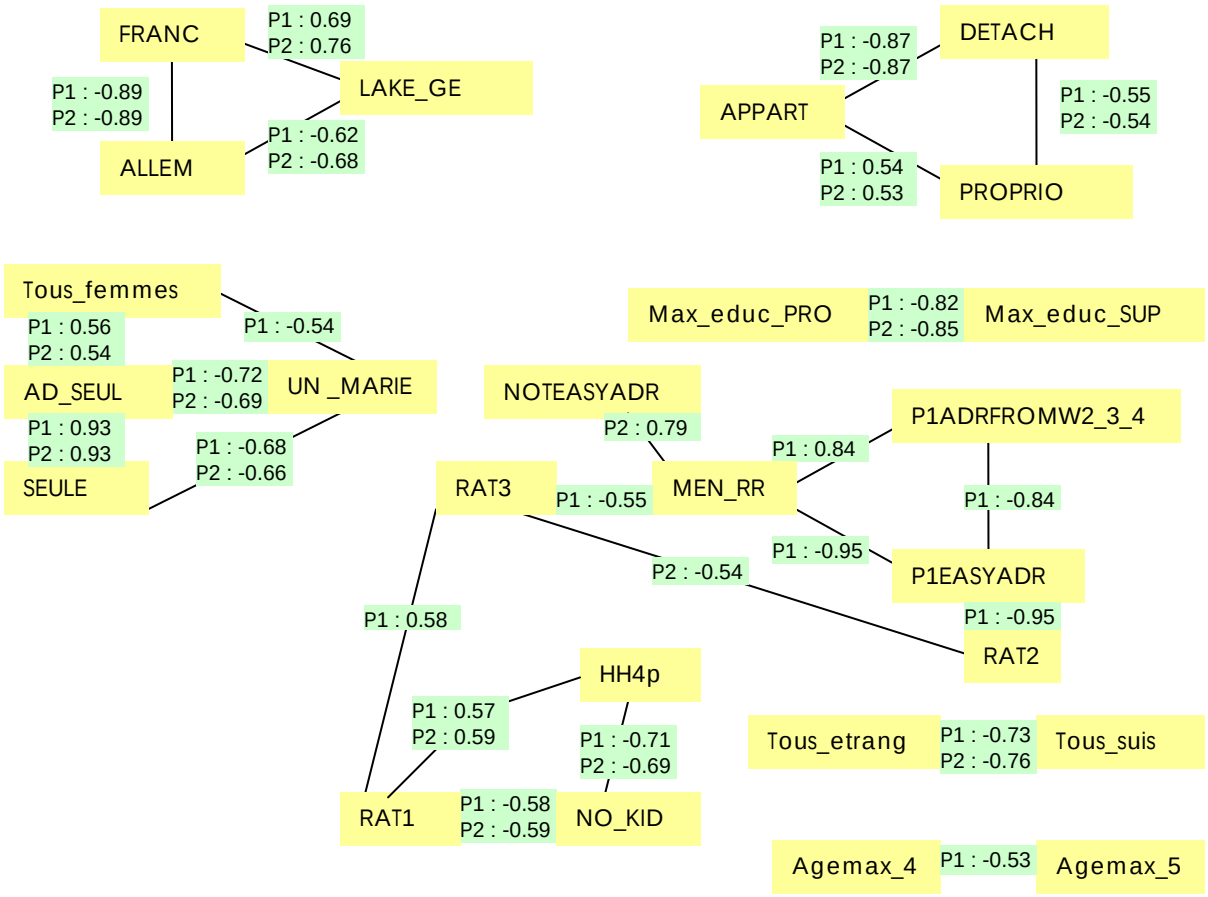


Figure F—2 Modèle complet de non-réponse au questionnaire ménage: pondération transversale ménage du PSM_I. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

Split 1	Split 2	Split 3	Split 4	Split 5	Split 6	Split 7	Rhg
PROPRIO=0 (1576) 94.31%	NotEasyAdr=0 (1514) 95.29%	MEN_RR=0 (1230) 97.16%	loy_3=0 (726) 95.36%	loy_1=0 (441) 92.34%	loy_2=0 (228) 85.46%	max_educ_sup=0 (111) 80.73%	1
						max_educ_sup=1 (117) 90.26%	2
					loy_2=1 (213) 100.00%		3
				loy_1=1 (285) 100.00%			4
			loy_3=1 (504) 100.00%				5
		MEN_RR=1 (284) 87.28%	loy_3=0 (196) 82.38%	loy_1=0 (132) 74.30%			6
				loy_1=1 (64) 100.00%			7
			loy_3=1 (88) 100.00%				8
	NotEasyAdr=1 (62) 69.39%						9
PROPRIO=1 (1317) 100.00%							10

Figure F—3 Modèle complet de non-réponse au questionnaire ménage: pondération transversale ménage du PSM_II. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

<table>
<tr><th>Split 1</th><th>Split 2</th><th>Split 3</th><th>Split 4</th><th>Split 5</th><th></th></tr>
<tr><td rowspan="7">MEN_RR=0
(1374)
97.99%</td><td rowspan="4">PROPRIO=0
(684)
96.05%</td><td rowspan="3">loy_3=0
(401)
93.35%</td><td rowspan="2">loy_1=0
(271)
89.99%</td><td>loy_2=0
(143)
80.11%</td><td>1</td></tr>
<tr><td>loy_2=1
(128)
100.00%</td><td>2</td></tr>
<tr><td>loy_1=1
(130)
100.00%</td><td rowspan="2"></td><td>3</td></tr>
<tr><td>loy_3=1
(283)
100.00%</td><td>4</td></tr>
<tr><td>PROPRIO=1
(690)
100.00%</td><td colspan="3" rowspan="3"></td><td>5</td></tr>
<tr><td rowspan="2">MEN_RR=1 (169)
83.38%</td><td>PROPRIO=0 (85)
68.80%</td><td>6</td></tr>
<tr><td>PROPRIO=1 (84)
100.00%</td><td>7</td></tr>
</table>

ANNEXE G – Modélisation de la non-réponse au questionnaire individuel, pondération longitudinale individuelle du PSM_I et PSM_II combinés

Les variables dichotomiques utilisées pour la modélisation de la non-réponse proviennent de la grille de la vague courante. Cette modélisation ne concerne ici que les individus longitudinaux par rapport à l'année 2004 du PSM_I. En effet on doit refaire la modélisation car pour la pondération longitudinale combinée, l'année de référence est 2004 et non plus 1999, ce qui amène à redéfinir les longitudinaux par rapport à l'année 2004 (voir paragraphe 12.1). Pour la description des variables, se référer à l'annexe A.

Niveau ménage

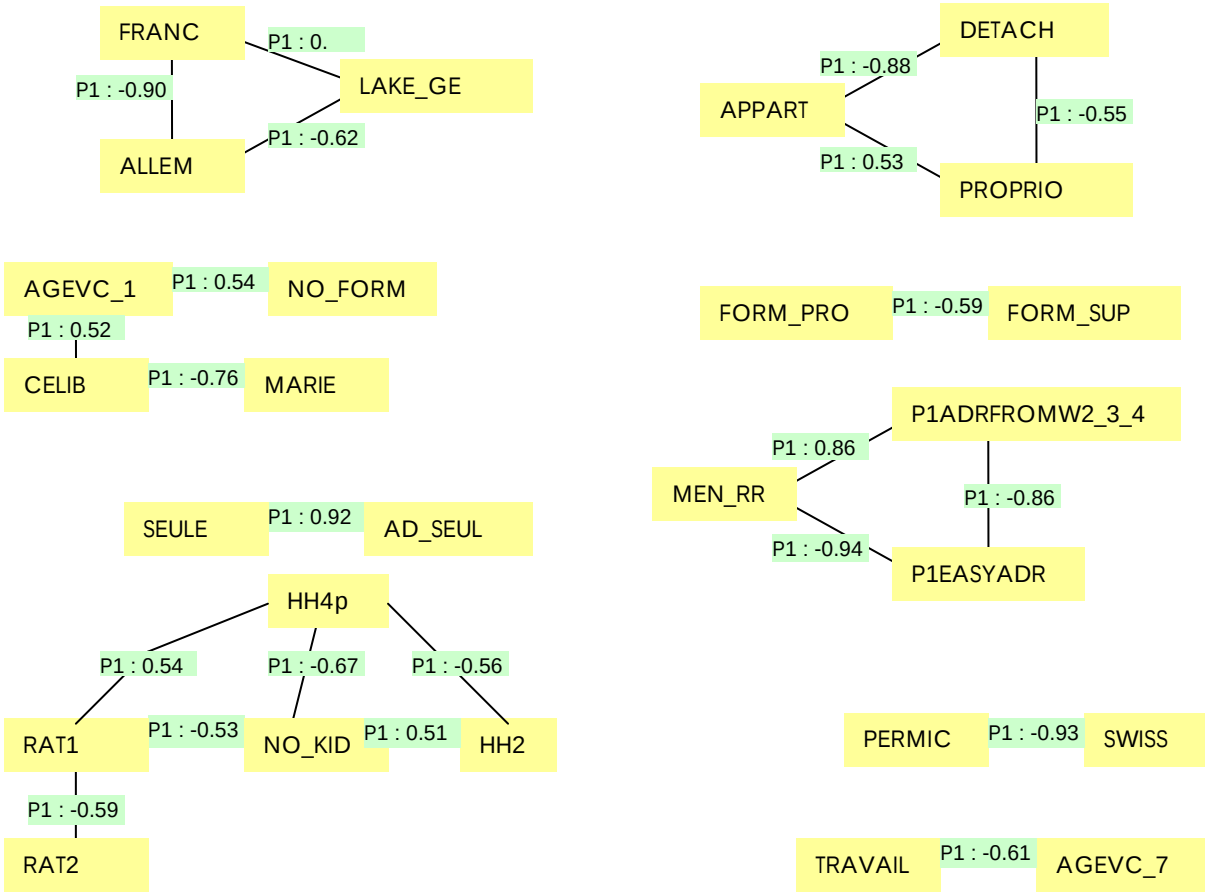
Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**
FRANC	NON	NON (voir figure G-1)
ITAL	OUI	OUI
ALLEM	**OUI**	**OUI**
LAKE_GE	NON	NON (voir figure G-1)
MIDDLE	OUI	OUI
NOREST	OUI	OUI
EST	**OUI**	**OUI**
CENTRAL	**OUI**	OUI
ZURICH	OUI	OUI
MIXED	OUI	OUI
CENTRES	**OUI**	**OUI**
SUBURB	OUI	**OUI**
WEALTH	OUI	OUI
PERIPH	OUI	OUI
TOURIST	OUI	OUI
INDUS	OUI	OUI
RURAL	OUI	OUI
AD_SEUL	NON	NON (voir figure G-1)
NO_KID	**OUI**	NON (voir figure G-1)
SEULE	**OUI**	**OUI**
ADRFROMW5	**OUI**	
REFW2W7CONV	OUI	
REFW3W8CONV	OUI	
INC_STAMPS	**OUI**	
INC_LOTT	OUI	
INC_GIFT	OUI	
Loy_1		**OUI**
Loy_2		**OUI**
Loy_3		**OUI**
RAT_1		OUI
RAT_2		NON (voir figure G-1)
RAT_3		**OUI**
PROPRIO		**OUI**
MAISINDEP		OUI
APPART		NON (voir

Niveau individu

Variable	Utilisée pour PSM_I 06	**Utilisée pour PSM_I 07**
AGEVC_1	**OUI**	OUI
AGEVC_2		**OUI**
AGEVC_3	OUI	**OUI**
AGEVC_4	OUI	OUI
AGEVC_5	OUI	OUI
AGEVC_6	OUI	OUI
AGEVC_7		OUI
CELIB	NON	NON (voir figure G-1)
DIVORCE	OUI	OUI
SEPARE	OUI	OUI
MARIE	OUI	OUI
VEUF	OUI	OUI
FORM_PRO	NON	NON (voir figure G-1)
FORM_SUP	**OUI**	**OUI**
NO_FORM	**OUI**	NON (voir figure G-1)
HOMME	**OUI**	**OUI**
SWISS	OUI	**OUI**
TRAVAIL	OUI	NON (voir figure G-1)
PERMIB	NON	NON (voir figure G-1)
PERMIC	NON	NON (voir figure G-1)

		figure G-1)
DETACH		NON (voir figure G-1)
MEN_RR		**OUI**
NotEasyAdr		OUI
P1AdrFromW2_3_4		NON (voir figure G-1)
P1EasyAdr		NON (voir figure G-1)
P2EasyAdr		NON (voir figure G-1)
HH2		OUI
HH3		OUI
HH4p		NON (voir figure G-1)

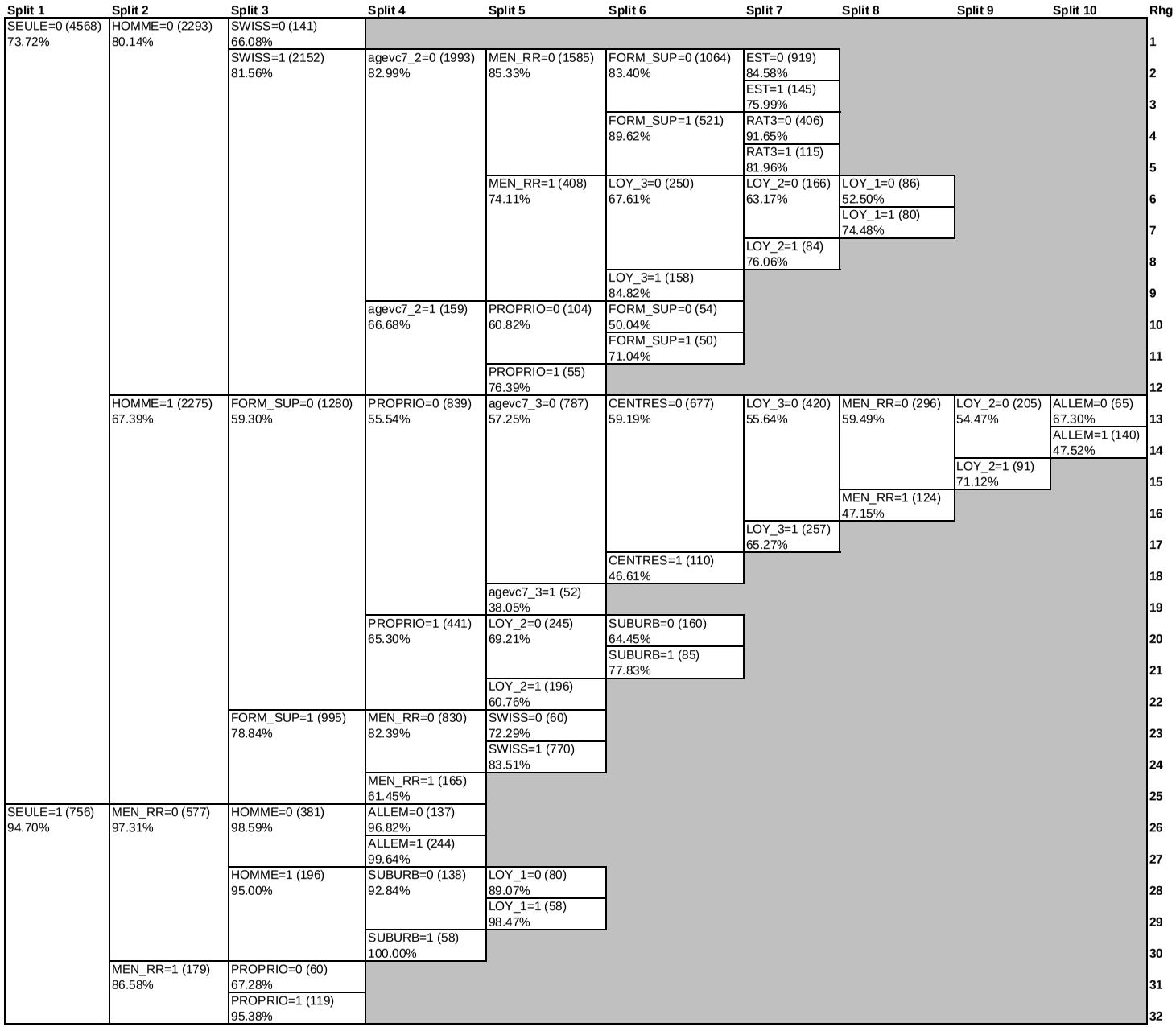
Tableau G – 1 Variables dichotomiques créées pour modéliser la non-réponse ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue.



Modélisation de la non-réponse – analyse par segmentation

Le modèle obtenu avec l'analyse par segmentation comporte 19 GHR pour le PSM_I. On a utilisé 16 variables dont 7 communes avec l'année précédente. Un taux de réponse pondéré d'au moins 30 % et un effectif minimal de 50 individus étaient exigés dans chaque GHR.

Figure G-2 Modèle complet de non-réponse au q. i. pour les longitudinaux du PSM_I (pondération long. ind. du PSM_I PSM_II combinés).



ANNEXE H – Caractéristiques des distributions des pondérations produites

Tableau H – 1 Pondération individuelle longitudinale PSM_I vague 9 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.
PTI_NRQI et PL_NRQI: poids corrigé pour la non réponse au questionnaire individuel
POIDTRIN: poids transversal individuel final
POIDLONG: poids longitudinal individuel final

Nom du jeu de poids	PTI_NRQI 06	**PTI_NRQI 07**	POIDTRIN 06	**POIDTRIN 07**	PL_NRQI 06	**PL_NRQI 07**	POIDLONG 06	**POIDLONG 07**
N	4088	4626	4088	4626	3617	3959	3617	3959
Somme des poids	6462474	6554824	6196348	6554824	5771759	5771759	5998212	5998212
Moyenne	1581	1417	1516	1417	1595	1458	1658	1515
Écart-type	678	651	1023	651	699	660	968	912
Variance	459568	424409	1047053	424409	488682	435770	937888	831666
Entendue	5993	5971	10861	5971	5416	4869	9741	7917
Intervalle interquartile	751	644	716	644	839	726	1034	956
Rapport Max/Min	32	24	80	24	13	12	27	22
Rapport P99 / P1	6	8	11	8	8	8	11	10
Rapport P95 / P5	4	4	5	4	4	4	5	5
Quantiles								
100% Max	6185	6231	10999	10736	5875	5329	10121	8291
99%	3840	3974	5705	6339	4128	3790	5214	5020
95%	2926	2688	3620	3380	2867	2741	3580	3282
90%	2482	2227	2521	2345	2484	2318	2713	2578
75% Q3	1863	1639	1664	1432	1945	1732	2020	1841
50% Médiane	1408	1255	1207	1034	1428	1298	1439	1259
25% Q1	1112	996	948	817	1106	1006	986	885
10%	943	827	790	681	929	827	795	728
5%	836	715	679	599	713	675	694	670
1%	599	514	500	432	501	501	490	502
0% Min	192	260	138	167	460	460	381	374

Tableau H – 2 Pondération transversale ménage PSM_I vague 9 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

P_NRGRIL: poids ajusté pour la non-réponse au questionnaire grille
PTM_NRQM: poids ajusté pour la non-réponse au questionnaire ménage
POIDTRME_INT1: poids après premier calage
POIDTRME_INT: poids après deuxième calage
POIDTRME_INT_WIN: poids transversal ménage final (winsorisé au 1er et au 99e percentiles). La winsorisation s'est fait sur le jeu de poids POIDTRME_INT sur le fichier au niveau individus, c'est pourquoi les nombre donnés ici, calculés au niveau des ménages, ne correspondent pas exactement si l'on compare la somme, le P1, P99 des poids _INT avec les poids _INT_WIN.

Nom du jeu de poids	P_NRGRIL 06	**P_NRGRIL 07**	PTM_NRQM 06	**PTM_NRQM 07**	POIDTRME_ INT1 06	**POIDTRME_ INT1 07**	POIDTRME_ INT 06	**POIDTRME_ INT 07**	POIDTRME_ INT_WIN 06	**POIDTRME_ INT_WIN 07**
N	2028	2301	2537	2817	2537	2817	2537	2817	2537	2817
Somme des poids	2587667	2666800	3140124	3201321	3056981	3103928	3053584	3103183	3064134	3109426
Moyenne	1276	1159	1238	1136	1205	1102	1204	1102	1208	1104
Écart-type	399	383	395	390	730	751	730	751	708	725
Variance	159500	146983	155922	152384	532354	563588	532847	564514	500612	526284
Entendue	2110	2300	2819	2720	7625	7558	7286	7187	4088	4240
Interv. interquartile	625	481	573	459	631	589	633	584	637	588
Rapport Max/Min	4	5	17	13			223	73	26	20
Rapport P99/ P1	3	4	5	5	15	15	15	15	15	15
Rapport P95/ P5	3	3	3	3	5	6	5	6	5	6
Quantiles										
100% Max	2742	2862	2995	2955	7321	7206	7318	7287	4254	4463
99%	2301	2207	2269	2215	4079	4199	4107	4202	4138	4231
95%	2025	1946	1975	1892	2596	2615	2593	2620	2613	2638
90%	1884	1759	1807	1692	1994	1917	1985	1907	2000	1919
75% Q3	1576	1351	1511	1335	1390	1264	1390	1259	1400	1267
50% Médiane	1168	1056	1166	1058	1018	862	1013	862	1021	868
25% Q1	951	870	938	876	759	675	757	674	763	679
10%	826	771	797	730	583	525	580	522	584	526
5%	780	704	719	596	474	443	472	443	476	446
1%	704	616	478	404	266	281	268	280	270	282
0% Min	632	562	176	234	**-304**	**-353**	32	100	166	223

Tableau H – 3 Pondération individuelle longitudinale PSM_II vague 4 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

PTI_NRQI et PL_NRQI: poids corrigés pour la non-réponse au questionnaire individuel
POIDTRIN: poids transversal individuel final
POIDLONG: poids longitudinal individuel final

Nom du jeu de poids	PTI_NRQI 06	**PTI_NRQI 07**	POIDTRIN 06	**POIDTRIN 07**	PL_NRQI 06	**PL_NRQI 07**	POIDLONG 06	**POIDLONG 07**
N	2568	2347	2568	2347	2453	2185	2453	2185
Somme des poids	7158158	7250411	6196348	6280768	6953021	6956714	6080866	6080866
Moyenne	2787	3089	2413	2676	2834	3184	2479	2783
Écart-type	1125	1039	1295	1508	1062	973	1203	1284
Variance	1265490	1080017	1677224	2274016	1127007	947179	1448394	1648812
Entendue	7794	8058	12451	12050	6894	8287	11242	10637
Intervalle interquartile	1429	1211	1331	1346	1427	1284	1370	1309
Rapport Max/Min	12	21	25	48	8	9	18	18
Rapport P99/ P1	6	5	9	10	5	5	7	6
Rapport P95/ P5	3	3	5	4	3	3	4	4
Quantiles								
100% Max	8503	8464	12966	12307	7874	9267	11886	11265
99%	6420	6010	6917	8642	6101	6090	6537	7289
95%	4915	5061	5017	5741	4795	4993	4979	5332
90%	4250	4595	4143	4390	4261	4488	4052	4392
75% Q3	3387	3634	2877	3078	3456	3802	2990	3248
50% Médiane	2565	2879	2027	2232	2612	2926	2179	2476
25% Q1	1958	2423	1546	1732	2029	2517	1620	1939
10%	1596	1987	1230	1468	1703	2175	1305	1569
5%	1434	1596	1104	1283	1566	1980	1191	1451
1%	1028	1181	792	835	1260	1203	958	1239
0% Min	710	406	515	257	980	980	644	629

Tableau H – 4 Pondération transversale ménage PSM_II vague 4 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

POIDINIT2 : poids initial corrigé en vague 1 pour la non-réponse au questionnaire grille
P_NRGRIL: poids ajusté pour la non-réponse au questionnaire grille
PTM_NRQM: poids ajusté pour la non-réponse au questionnaire ménage
POIDTRME_INT: poids transversal ménage final

Nom du jeu de poids	POIDINIT2 06	**POIDINIT2 07**	P_NRGRIL 06	**P_NRGRIL 07**	PTM_NRQM 06	**PTM_NRQM 07**	POIDTRME_INT 06	**POIDTRME_INT 07**	**POIDTRME_INT_WIN 07**
N	1660	1439	1431	1246	1684	1489	1684	1489	1489
Somme des poids	2094703	1814983	3023814	3030473	3543600	3574007	3183038	3199947	3208510
Moyenne	1262	1261	2113	2432	2104	2400	1890	2149	2155
Écart-type	278	277	638	570	749	700	856	982	963
Variance	77299	76856	407126	324461	560613	490027	732642	965169	928199
Entendue	1466	1466	3745	3401	6118	6667	6422	7900	5290
Intervalle interquartile	252	249	775	550	817	647	1007	1165	1171
Rapport Max/Min	3	3	5	3	13	20	30	168	13
Rapport P99/ P1	2	2	3	3	6	5	7	8	8
Rapport P95/ P5	2	2	2	2	3	3	4	4	4
Quantiles									
100% Max	2436	2436	4740	4823	6627	7010	6644	7947	5723
99%	2147	2147	4078	4242	4375	4480	4735	5409	5434
95%	1832	1815	3441	3560	3555	3662	3500	4188	4208
90%	1749	1749	3030	3307	3141	3332	2990	3370	3385
75% Q3	1339	1336	2406	2608	2443	2665	2281	2643	2655
50% Médiane	1173	1173	1991	2292	1971	2279	1713	1961	1970
25% Q1	1087	1087	1631	2057	1626	2018	1275	1478	1484
10%	1015	1019	1447	1944	1352	1756	1038	1168	1174
5%	1000	1000	1347	1707	1122	1188	890	958	963
1%	980	980	1205	1531	749	990	635	638	641
0% Min	970	970	995	1422	510	343	222	47	433

Tableau H – 5 Pondération individuelle longitudinale PSM_I et PSM_II combinés vague 4 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

PL2_NRQI: poids corrigé pour la non réponse au questionnaire individuel
PLI_COMBPAN_REG: poids individuel ajusté pour la non-réponse, combiné au niveau des grandes régions,
PLI_COMBPAN_CAL: poids longitudinal individuel final

Nom du jeu de poids	PL2_NRQI 06	**PL2_NRQI 07**	PLI_COMBPAN_REG 06	**PLI_COMBPAN_REG 07**	PLI_COMBPAN_CAL 06	**PLI_COMBPAN_CAL 07**
N	6244	6364	6244	6364	6244	6364
Somme des poids	13'146'565	13'137'147	6774804	6749091	6080866	6080866
Moyenne	2105	2064	1085	1061	974	956
Écart-type	1022	1116	520	525	641	676
Variance	1044813	1245897	270481	275689	411087	456374
Entendue	7399	8791	5839	6060	9870	8046
Intervalle interquartile	1236	1529	538	496	513	490
Rapport Max/Min	16	19	21	20	49	39
Rapport P99/ P1	7	8	7	8	10	12
Rapport P95/ P5	4	5	3	3	5	5
Quantiles						
100% Max	7874	9267	6130	6372	10075	6364
99%	5417	5324	3095	3214	3417	6080866
95%	4120	4283	1975	1917	2148	956
90%	3553	3660	1660	1601	1609	676
75% Q3	2579	2697	1283	1232	1110	456374
50% Médiane	1873	1756	963	936	805	8046
25% Q1	1343	1168	745	737	597	490
10%	1045	947	624	613	486	39
5%	972	852	568	552	428	12
1%	768	669	440	411	341	5
0% Min	476	476	291	311	205	6364

Tableau H – 6 Pondération individuelle transversale PSM_I et PSM_II combinés vague 4 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

PTI_NRQI_COMB: poids corrigé pour la non-réponse au questionnaire individuel
PTI_COMBPAN_REG: poids individuel ajusté pour la non-réponse, combiné au niveau des grandes régions
PTI_COMBREG_CAL: poids transversal individuel après calage
PTI_COMBREG_CAL _WIN: poids transversal individuel final (winsorisé au 1er et au 99e percentiles)

Nom du jeu de poids	PTI_NRQI _COMB 06	**PTI_NRQI _COMB 07**	PTI_COMBPAN _REG 06	**PTI_COMBPAN _REG 07**	PTI_COMBREG _CAL 06	**PTI_COMBREG _CAL 07**	**PTI_COMBREG _CAL_WIN 07**
N	6656	6973	6656	6973	6656	6973	6973
Somme des poids	13'620'632	13'805'235	6744736	7101803	6196349	6280768	6280768
Moyenne	2046	1980	1013	1018	931	901	901
Écart-type	1056	1127	430	533	575	714	639
Variance	1115407	1269486	184788	283894	331128	509698	407930
Entendue	8311	8204	3487	7168	6104	9124	3893
Intervalle interquartile	1292	1513	512	455	479	441	447
Rapport Max/Min	44	33	30	53	73	94	16
Rapport P99/ P1	8	10	6	9	10	16	16
Rapport P95/ P5	5	5	4	4	5	6	6
Quantiles							
100% Max	8503	8464	3606	7306	6189	9222	4161
99%	5550	5297	2427	3172	3152	4104	4161
95%	4116	4283	1844	1894	2085	2138	2168
90%	3491	3567	1590	1583	1613	1514	1535
75% Q3	2556	2631	1216	1161	1059	978	991
50% Médiane	1761	1595	907	889	761	695	705
25% Q1	1264	1118	703	706	579	537	544
10%	1014	898	597	571	479	433	439
5%	896	783	526	488	420	383	388
1%	681	552	386	351	319	264	268
0% Min	192	260	119	138	85	98	268

Tableau H – 7 Pondération transversale ménage PSM_I et PSM_II combinés vague 4 (2007). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0. On compare également avec les mêmes statistiques pour les pondérations de l'année précédente.

PTM_NRQM_COMB: poids ajusté pour la non-réponse au questionnaire ménage
PTM_COMBPAN_REG: poids transversal ménage ajusté pour la non-réponse, combiné au niveau des grandes régions
PTM_COMBREG_CAL2: poids transversal ménage après 2e calage (=poids final en 2006)
PTM_COMBREG_CAL2_WIN: poids transversal ménage final (winsorisé au 1er et au 99e percentiles). La winsorisation s'est fait sur le jeu de poids PTM_COMBREG_CAL2 sur le fichier au niveau individus, c'est pourquoi les nombre donnés ici, calculés au niveau des ménages, ne correspondent pas exactement si l'on compare la somme, le P1, P99 des poids _CAL2 avec les poids _CAL2_WIN.

Nom du jeu de poids	PTM_NRQM _COMB 06	**PTM_NRQM _COMB 07**	PTM_COMBPAN _REG 06	**PTM_COMBPAN _REG 07**	PTM_COMBREG _CAL2 06	**PTM_COMBREG _CAL 07**	**PTM_COMBREG _CAL2 07**	**PTM_COMBREG _CAL2_WIN 07**
N	4221	4306	4221	4306	4221	10651	4306	4306
Somme poids	6683724	6775328	3305120	3411552	3111109	7369590	3163047	3166716
Moyenne	1583	1573	783	792	737	692	735	735
Écart-type	705	794	267	313	389	504	480	439
Variance	497383	630480	71404	97965	151108	254094	230136	192994
Entendue	6451	6776	2529	4122	3646	5598	5407	2684
Interv. interquart.	862	1124	326	302	399	384	389	393
Rapport Max/Min	38	30	26	35	90		138	17
Rapport P99/ P1	7	9	5	7	10	18	13	13
Rapport P95/ P5	4	5	3	3	5	7	6	6
Quantiles								
100% Max	6627	7010	2632	4242	3687	5494	5447	2853
99%	3945	4017	1588	1980	2135	2838	2605	2633
95%	2964	3075	1283	1297	1500	1694	1635	1652
90%	2458	2608	1140	1158	1214	1167	1199	1212
75% Q3	1927	2090	926	909	882	799	852	861
50% Médiane	1475	1307	738	738	640	548	605	612
25% Q1	1065	965	600	608	483	415	463	468
10%	841	791	494	497	377	316	352	356
5%	778	674	441	389	312	258	297	300
1%	545	447	298	284	222	154	200	202
0% Min	176	234	103	121	41	**-104**	40	168

ANNEXE I – Taux de réponse

Figure I – 1 PSM_I vague 9.

Echantillon brut	**3976**	**100%**				
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	68	2%				
Hors échantillon (numéro invalide, modem, fax, décédé)	414	10%				
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	18	0%				
Adresses éligibles (statut d'éligibilité connu)	3475	88%		**100%**		
	Rendez-vous grille en cours		95	3%		
	Refus, problèmes niveau grille		483	14%		
	Grilles réalisées		**2898**	**83%**		**100%**
			Refus, problèmes niveau ménage		76	2%
			Grille + quest. ménage réalisés		**2822**	**98%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel		5974	100%
	Rendez-vous individuel en cours	8	0%
	Refus, problèmes niveau individuel	1332	22%
	Questionnaires individuel complétés	**4634**	**78%**

Figure I – 2 PSM_II vague 4.

Echantillon brut	**2037**	**100%**
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	34	2%
Hors échantillon (numéro invalide, modem, fax, décédé)	98	5%
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	5	0%
Adresses éligibles (statut d'éligibilité connu)	1900	93%

		100%
Rendez-vous grille en cours	53	3%
Refus, problèmes niveau grille	298	16%
Grilles réalisées	**1549**	**81%**

		100%
Refus, problèmes niveau ménage	53	3%
Grille + quest. ménage réalisés	**1495**	**97%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel	3108	100%

Rendez-vous individuel en cours	11	0%
Refus, problèmes niveau individuel	746	24%
Questionnaires individuel complétés	**2351**	**76%**

ANNEXE J – Algorithme d'appartenance à la population cible

Algorithme auquel il est fait référence au paragraphe 12.2 :

1)…pour les individus du PSM_I
   - si OSM[33]=1,2 alors PRES_P1=1
   - si OSM=1,2 et CR_V6[34]≠1 alors PRES_P2=1
   - si OSM=1,2 et CR_V6=1 (hors-cible transversalement en 2004) alors PRES_P2=2

2)…pour les individus du PSM_II
   - si OSM=1,2 alors PRES_P2=1
   - si OSM=1,2 et CR_V1=1 (hors-cible transversalement en 2004) alors PRES_P2=2

Pour les cas où, après 1) et 2), PRES_P1 ou PRES_P2 sont toujours manquants (appliqué tant aux individus du PSM_I qu'à ceux du PSM_II):

- si HAB_CH07[35]=-5 alors PRES_P1=1 (ont toujours habité en Suisse)
- si 0 < HAB_CH07 < 1999 alors PRES_P1=1 (habitaient déjà en Suisse en 1999)
- si HAB_CH07 > 1999 alors PRES_P1=2 (n'habitaient pas en Suisse en 1999)
- si HAB_CH06=-5 alors PRES_P2=1 (ont toujours habité en Suisse)
- si 0 < HAB_CH07 < 2004 alors PRES_P2=1 (habitaient déjà en Suisse en 2004)
- si HAB_CH07 > 2004 alors PRES_P2=2 (n'habitaient pas en Suisse en 20004)
- si année de naissance > 2004 alors PRES_P2=2

Pour les cas subsistants non résolus, on a attribué le statut d'éligibilité du ménage[36]. Finalement, on a admis que les 20 cas non-couverts restants étaient éligibles.

[33] Original Sample Member, voir paragraphe 3.3.2.
[34] Code réponse à la vague 6, voir paragraphe 6.3.
[35] HAB_CH07: variable du questionnaire individuel indiquant depuis quand la personne vit en Suisse si ce n'a pas toujours été le cas.
[36] Un ménage est dit éligible si au moins un membre du ménage est éligible.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodology reports published by the SFSO's Statistical Methods Unit

Graf, E. (2009). Pondérations du Panel Suisse de Ménages: PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro de commande: 338-0057

Qualité, L., Tillé, Y. (2009). Estimation de la précision des évolutions dans l'enquête sur la valeur ajoutée. Numéro de commande: 338-0056

Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro de commande: 338-0055

Graf, E. (2008). Pondérations du PSM: PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro de commande: 338-0054

Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium".
Numéro de commande: 338-0053

Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro de commande: 338-0052

Graf, E. (2008). Pondérations du SILC pilote: SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro de commande: 338-0051

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050

Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049

Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048

Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze.Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046

Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045

Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044

Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043

Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041

Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer:
338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer:
338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch→Aktuell→Publikationen.

La liste des publications, mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch →Actualités→Publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodology Reports by the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte, werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

La structure du document est la suivante: la façon dont les données ont été collectées est brièvement décrite pour chacun des panels, puis, après avoir présenté les taux de réponse atteints, on définit les différents concepts liés à la pondération. Ensuite, les étapes de la pondération sont présentées, puis décrites une à une pour chacun des panels également. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite aussi séparément les pondérations au niveau ménage de celles au niveau individuel. On relève aussi les spécificités des pondérations transversales par rapport aux pondérations longitudinales. Les dernières sections font l'objet des pondérations combinées du PSM_I vague 9 et PSM_II vague 4. Finalement, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes mentionnées que brièvement dans le texte principal.

N° de commande: 338-0057-05
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 978-3-303-00424-1