Rapport de méthodes

Pondérations du PSM

PSM_I vague 8, PSM_II vague 3, 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, 2008

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	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	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 PSM

PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés

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

Editeur **Office fédéral de la statistique**

Neuchâtel, 2008

Préambule

L'enquête PSM est une enquête réalisée auprès des ménages et des personnes résidants 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 la nouvelle Fondation suisse pour la recherche en sciences sociales FORS, hébergée par l'Université de Lausanne. L'OFS a reçut le mandat de réaliser les pondérations des enquêtes PSM réalisées en 2007-2008 (vague 8 du PSM_I et vague 3 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 résulte d'une collaboration fructueuse avec Statistique Canada, en particulier avec Caroline Cauchon et Michel Latouche. Que ces personnes soient tout particulièrement remerciées ici.

Toute notre gratitude va aussi à Paul-André Salamin (METH) pour ses remarques avisées après une lecture approfondie de ce rapport.

Résumé

La structure de ce 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 ainsi que les effets des incitatifs, 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. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite séparément les pondérations au niveau ménage de celles au niveau individuel. On relève les spécificités des pondérations transversales par rapport aux pondérations longitudinales. Les dernières sections concernent les pondérations combinées du PSM_I vague 8 et PSM_II vague 3. Finalement, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes qui ne sont 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: 338-0054
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 2008
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.
ISBN: 978-3-303-00412-8

Tables des matières

1 INTRODUCTION : LE PANEL SUISSE DE MÉNAGES (PSM) 7

1.1 INTRODUCTION 7
1.2 LE PROBLÈME DE L'ATTRITION 8
1.3 STRUCTURE DE L'ENQUÊTE PSM 10

2 PLAN D'ÉCHANTILLONNAGE ET DÉROULEMENT DE L'ENQUÊTE 11

2.1 PLAN D'ÉCHANTILLONNAGE ET TAILLES BRUTES DES PSM_I ET PSM_II 11
2.2 UNIVERS DE L'ENQUÊTE 12

3 TAUX DE RÉPONSES ATTEINTS EN 2006 13

3.1 TAUX DE RÉPONSE EN REGARD DES INCITATIFS 13
3.2 EFFETS DES INCITATIFS 14
3.3 RÈGLES DE SUIVI 15
3.3.1 Règles de contact pour les ménages *18*
3.3.2 Règles de suivi pour les individus *18*
3.3.3 Particularités dès 2006 (vagues 8 et 3) *18*
3.4 DÉROULEMENT DES ENQUÊTES PSM_I VAGUE 8 ET PSM_II VAGUE 3 19

4 CONCEPTS RELATIFS À LA PONDÉRATION 20

4.1 CONCEPTS RELATIFS À LA PRODUCTION DE POIDS LONGITUDINAUX 20
4.1.1 Population longitudinale cible et observée *20*
4.1.2 Unité longitudinale *20*
4.1.3 Cohabitants *21*
4.1.4 Éligibilité longitudinale *21*
4.2 CONCEPTS RELATIFS À LA PRODUCTION DES POIDS TRANSVERSAUX 21
4.2.1 Population transversale cible et observée *21*
4.2.2 Unité transversale *22*
4.2.3 Unité dans le champ de l'enquête transversalement *22*
4.2.4 Éligibilité transversale *22*
4.2.5 Unités répondantes *22*
4.3 ILLUSTRATION DES CONCEPTS PAR UN EXEMPLE 22

5 ÉTAPES DE LA PONDÉRATION 25

5.1 INTRODUCTION 25
5.2 TYPES DE PONDÉRATIONS PRODUITES 26

6 ÉTAPES COMMUNES POUR TOUS LES TYPES DE POIDS 29

6.1 MODÉLISATION DE LA NON RÉPONSE PAR LA MÉTHODE DE SEGMENTATION 29
6.1.2 La méthode de segmentation *29*
6.1.2 Segmentation et régression logistique *30*
6.1.3 Application dans le cadre du PSM *31*
6.2 POIDS INITIAL DU PSM_I ET POIDS INITIAL DU PSM_II 32
6.3 CLASSEMENT DES INDIVIDUS: ÉTAPE 2 33
6.4 AJUSTEMENT POUR LA NON-RÉPONSE – GRILLE: ÉTAPE 3 35

7 PONDÉRATION TRANSVERSALE INDIVIDUELLE 37
7.1 PARTAGE DES POIDS: ÉTAPE 4 37
7.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE INDIVIDUEL: ÉTAPE 5 38
7.3 CALAGE: ÉTAPE 6 39
8 PONDÉRATION TRANSVERSALE MÉNAGE 41
8.1 PARTAGE DES POIDS: ÉTAPE 7 41
8.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE MÉNAGE: ÉTAPE 8 42
8.3 CALAGE: ÉTAPE 9 43
8.4 WINSORISATION DES POIDS APRÈS CALAGE: ÉTAPE 9BIS 44
9 PONDÉRATION LONGITUDINALE INDIVIDUELLE 46
9.1 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE INDIVIDUEL: ÉTAPE 10 46
9.2 CALAGE: ÉTAPE 11 47
10 PONDÉRATION TRANSVERSALE INDIVIDUELLE COMBINÉE PSM_I ET PSM_II 49
10.1 COMBINAISON TRANSVERSALE DE PANELS À PARTIR DES ÉCHANTILLONS TRANSVERSAUX DE CHACUN DES DEUX PANELS 49
10.2 COMBINAISON DES DEUX PANELS: ÉTAPE 13 51
10.3 CALAGE: ÉTAPE 14 52
11 PONDÉRATION TRANSVERSALE COMBINÉE MÉNAGE PSM_I ET PSM_II 54
11.1 COMBINAISON DES DEUX PANELS: ÉTAPE 15 54
11.2 CALAGE: ÉTAPE 16 55
12 PONDÉRATION LONGITUDINALE INDIVIDUELLE COMBINÉE DU PSM_I & PSM_II 57
12.1 AJUSTEMENT POUR LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL POUR LE PSM_I, ANNÉE DE RÉFÉRENCE 2004 57
12.2 COMBINAISON DES PANELS: ÉTAPE 17 59
12.3 CALAGE: ÉTAPE 18 60
13 CONCLUSIONS 62
13.1 LIGNES DIRECTRICES CONCERNANT L'APPLICATION DES PONDÉRATIONS PRODUITES 62
13.2 RECOMMANDATIONS CONCERNANT LES ESTIMATIONS DE LA PRÉCISION 63
14 BIBLIOGRAPHIE 64
ANNEXE A – FICHIERS DE DONNÉES POUR PSM_I ET PSM_II 66
ANNEXE B - DÉRIVATION DES VARIABLES POUR LE CLASSEMENT DES INDIVIDUS 68
ANNEXE C – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE GRILLE 71
ANNEXE D – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION LONGITUDINALE INDIVIDUELLE 77
ANNEXE E – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION TRANSVERSALE INDIVIDUELLE 81
ANNEXE F – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE MÉNAGE, PONDÉRATION TRANSVERSALE MÉNAGE 86

ANNEXE G – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION LONGITUDINALE INDIVIDUELLE DU PSM_I ET PSM_II COMBINÉS **90**

ANNEXE H – CARACTÉRISTIQUES DES DISTRIBUTIONS DES PONDÉRATIONS PRODUITES .. **93**

ANNEXE I – TAUX DE RÉPONSE **100**

ANNEXE J – ALGORITHME D'APPARTENANCE À LA POPULATION CIBLE **102**

1 Introduction : le Panel Suisse de Ménages (PSM)

1.1 Introduction

L'enquête PSM est une enquête réalisée auprès des ménages et des personnes résidants 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 début 2008, le projet est financé par le Fonds National Suisse de la Recherche Scientifique et est mené dans le cadre de la nouvelle Fondation suisse pour la recherche en sciences sociales FORS, hébergée par 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 (SILC). 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 comme le SILC, il n'y a donc pas d'injections régulières de nouvelles unités à enquêter, ni d'abandons volontaires de certaines autres unités. L'enquête PSM voit évoluer deux échantillons en parallèle au fil du temps, nous les noterons dans la suite par :

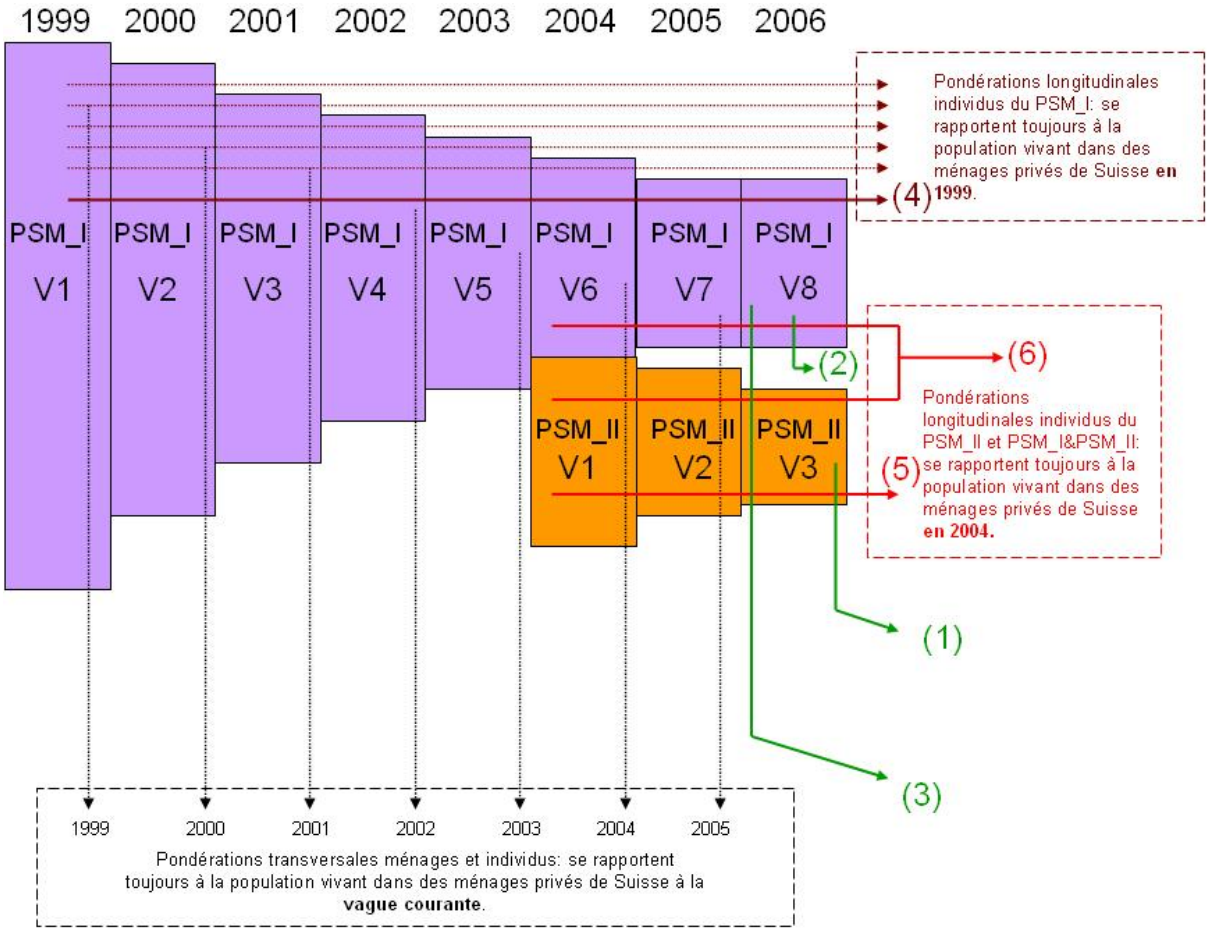
- ***PSM_I vague 8*** : il s'agit de l'échantillon des ménages et personnes sélectionnés en 1999 et interrogés pour la première fois cette année-là. Ce document traite de la pondération de la 8^{e} vague d'interviews de ces unités,
- ***PSM_II vague 3*** : il s'agit de l'échantillon des ménages et personnes sélectionnés et interrogés pour la première fois en 2004. Ce document traite de la pondération de la 3^{e} 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 c'est l'auteur qui a mené à bien cette tâche faisant ainsi l'objet de ce rapport.

Le document dont il est question ici décrit de façon détaillée le processus de production des poids pour l'enquête PSM de 2006[1], à savoir, celles du PSM_I vague 8, du PSM_II vague 3 ainsi que la pondération combinée des échantillons PSM_I et PSM_II pour 2006. La figure 1 donne un aperçu des neuf différents jeux de poids produits.

[1] Pour alléger le texte, nous dirons « réalisé en 2006 » pour l'enquête qui s'est déroulée de septembre 2006 à février 2007.

Figure 1 Schéma des neuf différents jeux de poids produits en 2006 : 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 qui furent effectuées par Statistique Canada pour les années précédentes et parfaitement compatibles avec ces dernières.

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 subit depuis son commencement. Le problème de l'attrition (ou érosion) a été extrêmement préoccupant. En effet, jusqu'à 2004, au cours des 6 premières vagues, le panel perdait 11% d'individus et de ménage par année en moyenne. Ces pertes étaient doublement trop grandes par rapport à ce qu'expérimentaient d'autres enquêtes similaires comme le BHPS en Angleterre ou le GSOEP en Allemagne.

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, réinjection 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 3.3, et pour les taux de réponses atteints en fin de période de terrain, voir le chapitre 3.

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

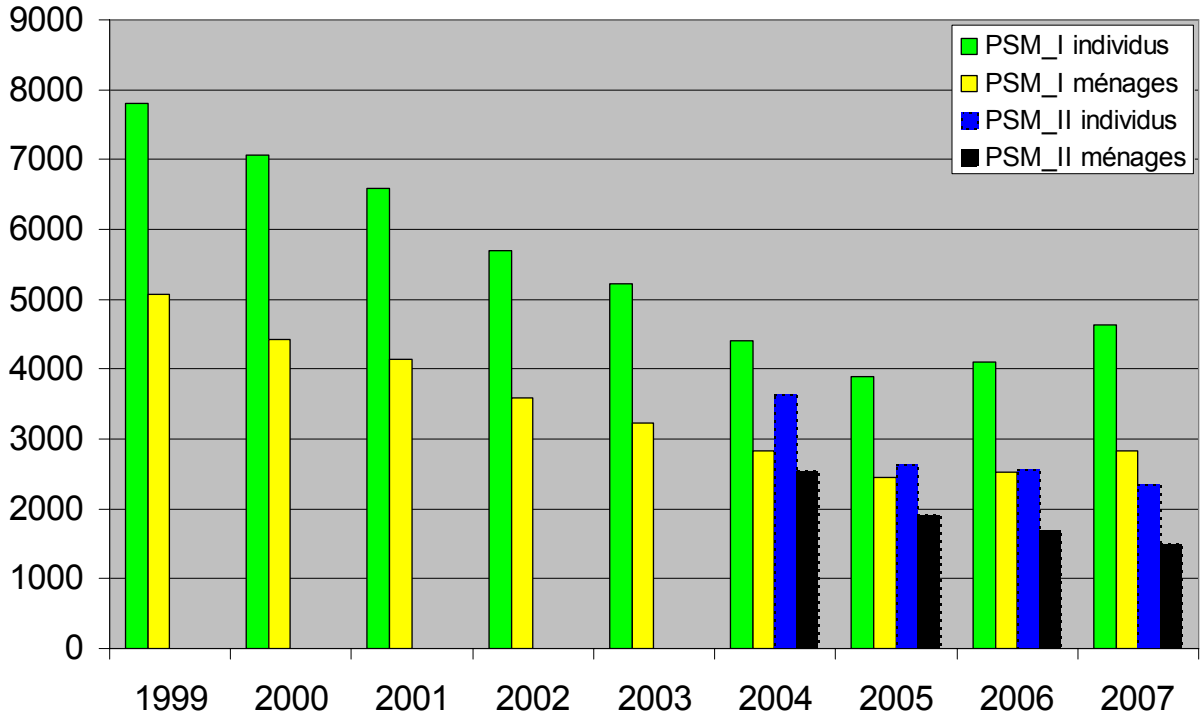


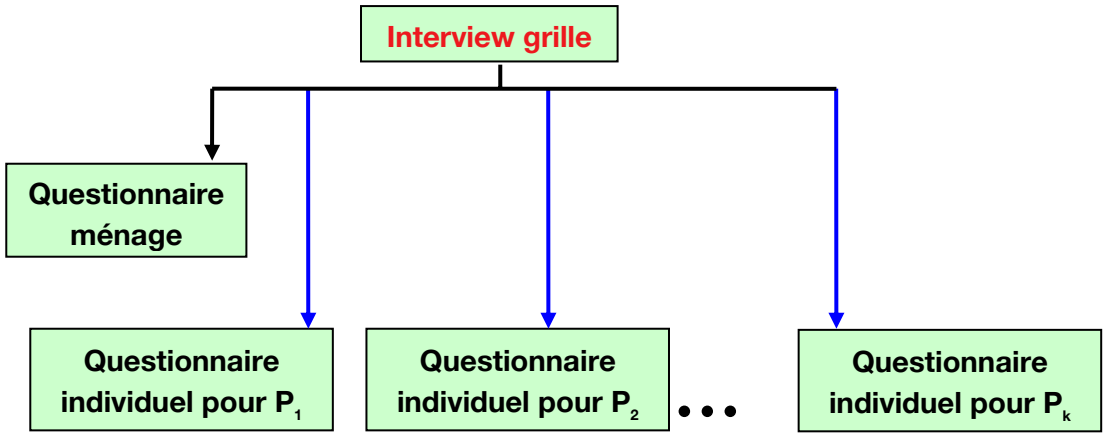
Tableau 3 Nombre de questionnaires individuels/ménages complétés dans l'enquête PSM. L'attrition relative est la proportion du nombre de cas interrogés l'année n sur le nombre de cas l'année (n-1). On a ici également déjà les résultats pour 2007 (vagues 9 et 4) à titre indicatif.

	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*[2] 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 8 du PSM_I et de la vague 3 du PSM_II ainsi que leur combinaison se trouve dans l'annexe A.

[2] 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 trableau 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 est le SRH[3] 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**

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

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

Tableau 7 Tailles S_h de l'échantillon brut par strates. Les S_h représentent le nombre d'adresses injectées dans 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 encouru 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 2006

Le tableau 8 résume les taux de réponse obtenus par rapport à l'échantillon brut, i.e. par rapport au nombre d'adresses injectées dans les enquêtes indépendamment de leur statut final (validité).

Tableau 8 Taux de réponse aux différents niveaux de collecte des données pour le PSM_I en 1999 (vague 1) et 2006 (vague 8), ainsi que pour le PSM_II en 2004 (vague 1) et 2006 (vague 3) (source: statistiques finales M.I.S-Trend SA).

		PSM_I vague 1, 1999				PSM_I vague 8, 2006			
Niveau ménages	Echantillon brut (nombre d'adresses injectées)	14174	100%			3207	100%		
	Nombre de grilles réalisées	7063	50%	100%		2579	80%	100%	
	Grilles + quest. Ménage réalisés	6001	42%	85%	100%	2534	79%	98%	100%
	Nombre de ménages complets[4]	5074	36%	72%	85%	2486	78%	96%	98%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	10293	100%			5331	100%		
	Nombre de quest. Individuel réalisés	7799	76%			4093	77%		

		PSM_II vague 1, 2004				PSM_II vague 3, 2006			
Niveau ménages	Echantillon brut (nombre d'adresses injectées)	5000	100%			2555	100%		
	Nombre de grilles réalisées	2704	54%	100%		1752	69%	100%	
	Grilles + quest. Ménage réalisés	2535	51%	94%	100%	1682	66%	96%	100%
	Nombre de ménages complets	2335	47%	86%	92%	1611	63%	92%	96%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	5384	100%			3499	100%		
	Nombre de quest. Individuel réalisés	3652	68%			2570	73%		

3.1 Taux de réponse en regard des incitatifs

Afin de lutter contre l'attrition, le PSM a décidé pour 2006 de tenter l'expérience de l'utilisation d'incitatifs pour gagner des répondants. Les adresses du PSM_I et PSM_II furent ainsi partagées aléatoirement en 4 groupes de tailles égales.

1. Le premier groupe ne reçu aucun incitatif comme ce fut le cas pour les années précédentes.

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

2. Au deuxième on a joint un carnet de timbres courrier A, personnalisés avec le logo du Panel Suisse de Ménages, à la lettre annonçant les vagues 8 et 3 de l'enquête.
3. Pour le troisième groupe d'adresses on a annoncé aux ménages concernés afin de les remercier pour leurs participations passées et futures on les ferait participer à une loterie avec trois gains de 5000, 3000 et 2000 francs suisses à la clé.
4. Enfin, les individus du quatrième groupe d'adresses avaient la possibilité, à la fin de leur interview individuelle, de faire un don de CHF 10.-- à certaines associations caritatives.

Tableau 9 Taux de réponse aux différents niveaux de collecte des données pour le PSM_I et le PSM_II pour les différents groupes d'adresses par rapport aux incitatifs. En gras on a les meilleurs résultats parmi les quatre groupes. Les ménages de vague 5 (voir paragraphe 3.3.3) ainsi que les ménages qui avaient refusé de participer aux vague 7 et 2 ne faisaient pas partie de l'expérience des incitatifs. En effet ces ménages-là ont été injectés plus tard dans l'enquête et directement traités par les interviewers chargés de faire les rappels de refus.

	Hors expérience	1. Pas d'incitatif		2. Timbres		3. Dons		4. Loterie	
	PSM_I	PSM_I		PSM_I		PSM_I		PSM_I	
Echantillon brut (nombre d'adresses injectées) Total = 3207	*636*	654	100%	643	100%	655	100%	619	100%
Nombre de grilles réalisées	*224*	597	91%	593	**92%**	597	91%	568	**92%**
Grilles + quest. Ménage réalisés	*209*	592	**91%**	584	**91%**	589	90%	560	90%
Nbr quest.indiv. Réalisés || Nbr. moyen d'indiv. Interrogés par ménage	*259*	937	1.57	974	1.64	1011	**1.69**	912	1.61
	PSM_II	PSM_II		PSM_II		PSM_II		PSM_II	
Echantillon brut (nombre d'adresses injectées) Total = 2555	*588*	497	100%	509	100%	459	100%	502	100%
Nombre de grilles réalisées	*151*	411	**83%**	424	**83%**	357	78%	409	81%
Grilles + quest. Ménage réalisés	*140*	397	**80%**	409	**80%**	339	74%	397	79%
Nbr quest.indiv. Réalisés || Nbr. moyen d'indiv. Interrogés par ménage	*178*	641	**1.56**	645	1.52	523	1.46	583	1.43

3.2 Effets des incitatifs

Comme on peut le voir dans le tableau 9, sur le plan général, on n'a pas observé d'effet très significatif suite à l'utilisation d'incitatifs, du moins pas en fin de période de récolte des données.
Les résultats n'ont pas été les mêmes sur le PSM_I que sur le PSM_II tout comme ils étaient différents au niveau ménage (questionnaires grille et ménage) et au niveau des individus.

Comme les graphiques 10 à 15 le montrent, les incitatifs ont eu plus d'influence sur le PSM_II que sur le PSM_I. Ceci n'est pas surprenant puisque les participants au PSM_I ont été fidélisés depuis plus d'années que ceux du PSM_II. On constate aussi que les timbres ont directement aidé au niveau des ménages: ils ont contribués à compléter plus rapidement les questionnaires grilles (et ménages). Ce qui a indirectement aidé à accélérer la collecte des questionnaires individuel. La courbe montrant l'évolution du taux de réponse au fil des jours travaillé a donc été plus « saine », même si les résultats finaux ne sont pas forcément significativement différents. Par ailleurs, les dons qui étaient une mesure incitative au niveau individuel ont contribués à augmenter le nombre d'individus interrogés par ménage pour le PSM_I.

3.3 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 suivis pour les ménages déterminant quels ménages recontacter par rapport à l'années précédente, et des règles de suivi pour les personnes pour savoir qui interroger réellement.

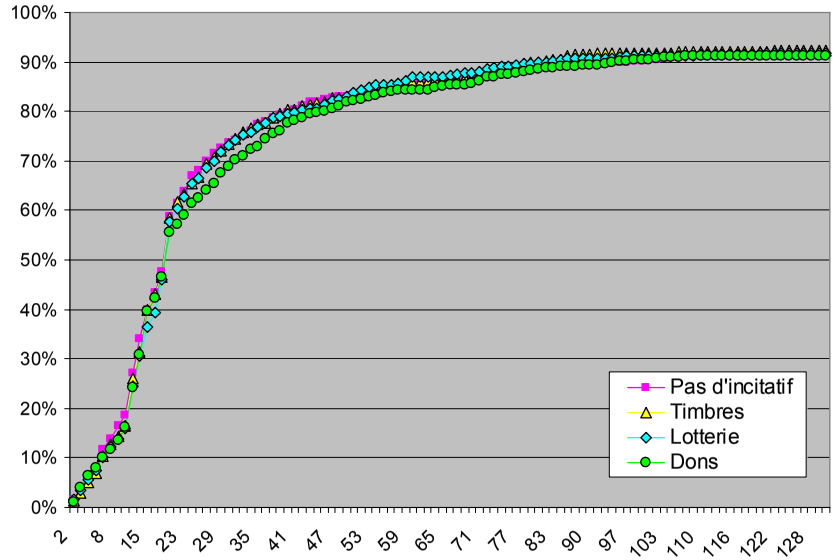
Pour le PSM_II, on a libéré les mêmes adresses que lors de la vague précédente, soit tous les ménages complets de la vague 1 ainsi qu'une partie des ménages dont on n'avait obtenu que la grille.

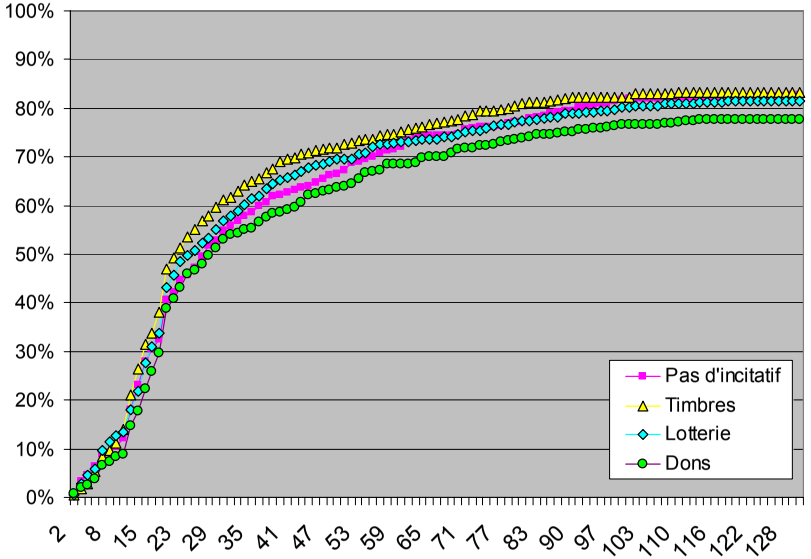
Tous les refus écrits reçus de chaque année, les décès ainsi que les adresses non-retrouvées à la suite des retours du courrier préliminaire n'ont par contre jamais été réinjectées.

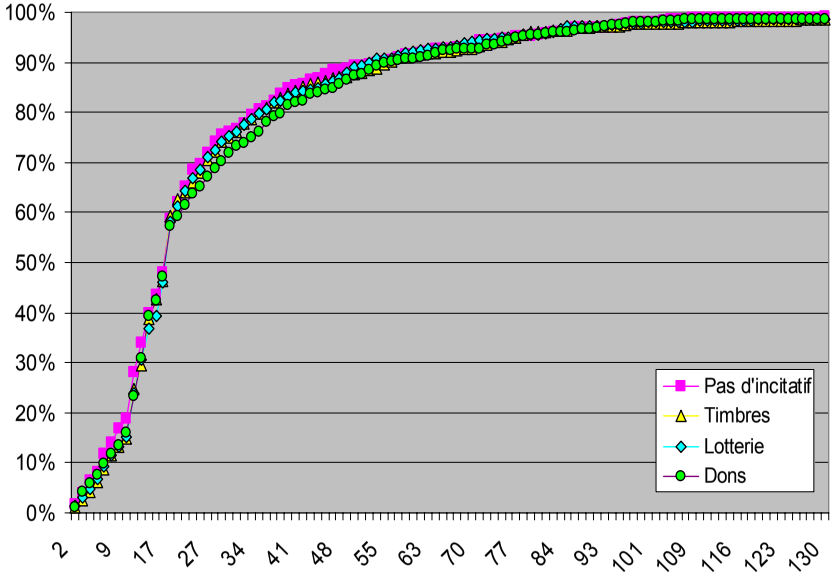
En définitive, les nombres d'adresses suivants ont été injectés:

PSM I		2672 adresses
- Ménages vague 5 réinjectés		535 adresses (voir paragraphe 3.3.3)
PSM II		2555 adresses

	Total	**5762 adresses**



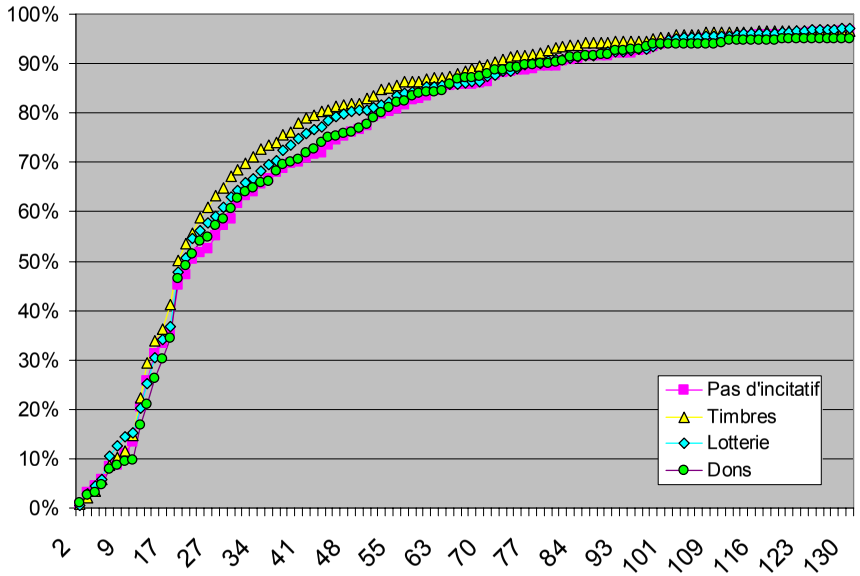


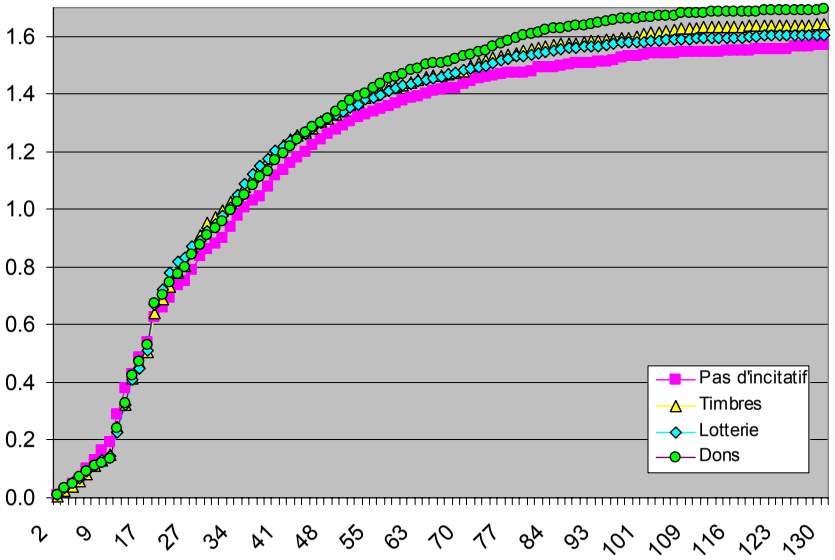


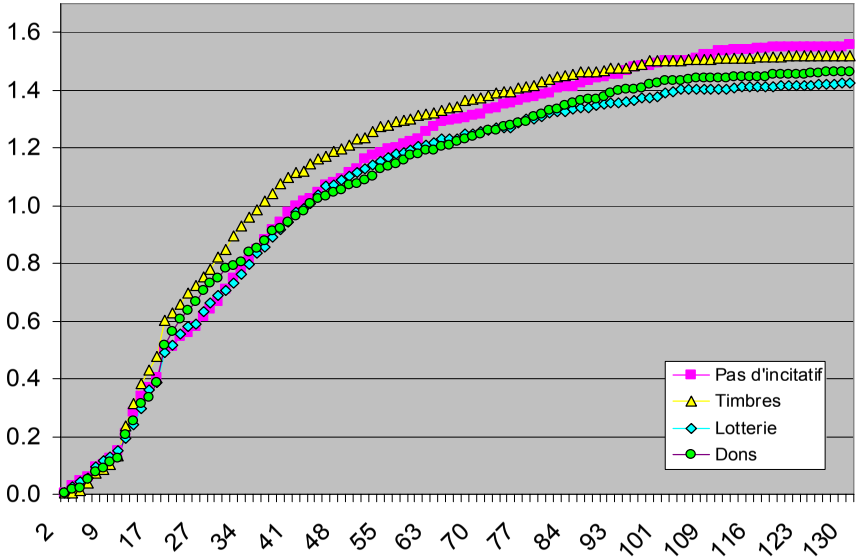
Graphique 10 En-haut à gauche: PSM_I vague 8, questionnaires grille complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre d'adresses attribués à chaque groupe.

Graphique 11 En-haut à droite: PSM_II vague 3, questionnaires grille complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre d'adresses attribués à chaque groupe.

Graphique 12 En bas à gauche: PSM_I vague 8, questionnaires ménages complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre de questionnaires grille réalisés.







Graphique 13 En-haut à gauche: PSM_II vague 3, questionnaires ménages complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre de questionnaires grille réalisés.

Graphique 14 En-haut à droite: PSM_I vague 8, questionnaires individuels complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre de questionnaires grille réalisés.

Graphique 15 En bas à gauche: PSM_II vague 3, questionnaires individuels complétés en fonction du nombre de jours d'interviews. Les taux de réponse sont calculés par rapport au nombre de questionnaires grille réalisés.

3.3.1 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.

3.3.2 Règles de suivi pour les individus

4. Les répondants longitudinaux (OSM[5]) et les enfants issus de ceux-ci sont suivis en tous temps, tandis que les cohabitants[6] 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.

3.3.3 Particularités dès 2006 (vagues 8 et 3)

Il a été décidé de libérer, pour la vague 8, les adresses de tous les ménages complets interrogés en vague 5 (2004), 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 retenté les ménages qui avaient été abandonné suite à un refus écrit. 535 ménages ont ainsi été «réintégrés» à l'échantillon brut, 35% d'entre eux ont accepté de participer à nouveau à l'enquête. Cette opération a donc été un succès.

De plus, 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 pour la première fois eu recours aux services de la société *Dcl.data care* spécialisée dans le traçage des déménagements et autres changements d'adresses.

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

3.4 Déroulement des enquêtes PSM_I vague 8 et PSM_II vague 3

La période de terrain a duré du 4 septembre 2006 au 17 février 2007, soit 5 mois et demi. Comme lors de la vague précédente, le fichier d'adresses à injecter 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). Quatre lettres ont été rédigées dans chaque langue, soit une pour chaque type de remerciement (incitatifs): timbres, loterie, dons ou… rien du tout. C'est la première fois dans l'histoire du PSM qu'on a utilisé des incitatifs. Un courrier spécial a en outre été envoyé préalablement aux rappels de refus à tous les ménages ou personnes ayant refusé de participer à la vague courante ainsi qu'aux ménages complets de vague 5 perdus depuis pour cause de double refus.

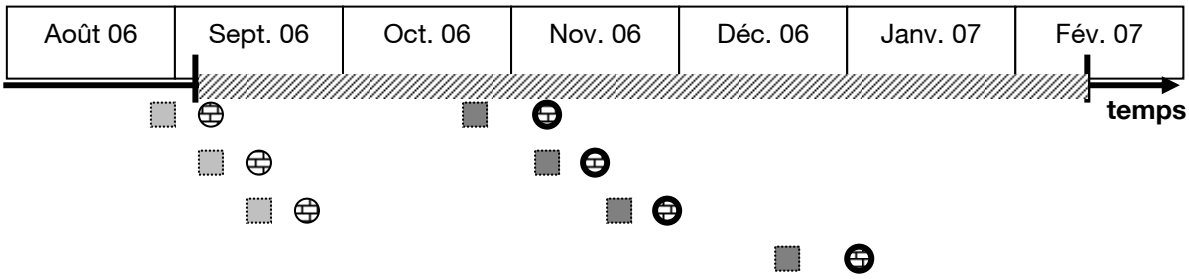
Pour le PSM_I, comme mentionné auparavant, on a réinjecté les adresses de tous les ménages complets réalisés en vague 5. Cependant, les ménages perdus depuis, à la suite de double refus, n'ont pas été injectés dès le début de la prise période de terrain, mais plus tard, lors du démarrage des rappels de refus, et inclus dans l'étude « refus ». On a fait cela parce qu'on a considéré que ces derniers nécessitaient une approche particulière.

Il n'y eut pas d'enquête non réponse en 2006.

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

Tableau 16 Dates d'envois de courrier et d'injections 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 injection d'adresses (pour le première fois en 2006, aussi pour les rappels de refus).

Envois du courrier préliminaire	Injection des échantillons d'adresses	Envois des courriers rappels de refus	Injection des échantillons d'adresses rappels de refus
28 août 2006	4 sept. 2006	27 oct. 2006	9 nov. 2006
4 sept. 2006	12 sept. 2006	7 nov. 2006	16 nov. 2006
13 sept. 2006	21 sept. 2006	20 nov. 2006	23 nov. 2006
		21 déc. 2006	3 janv. 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. 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'à 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 sans égard par rapport à 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[7]. 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 PSM_II, ainsi que certaines personnes qui étaient institutionnalisées ou non sélectionnables pour une autre raison en septembre 2006.

[7] Cependant, comme cela a été fait l'année passée, on a enfreint 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.

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

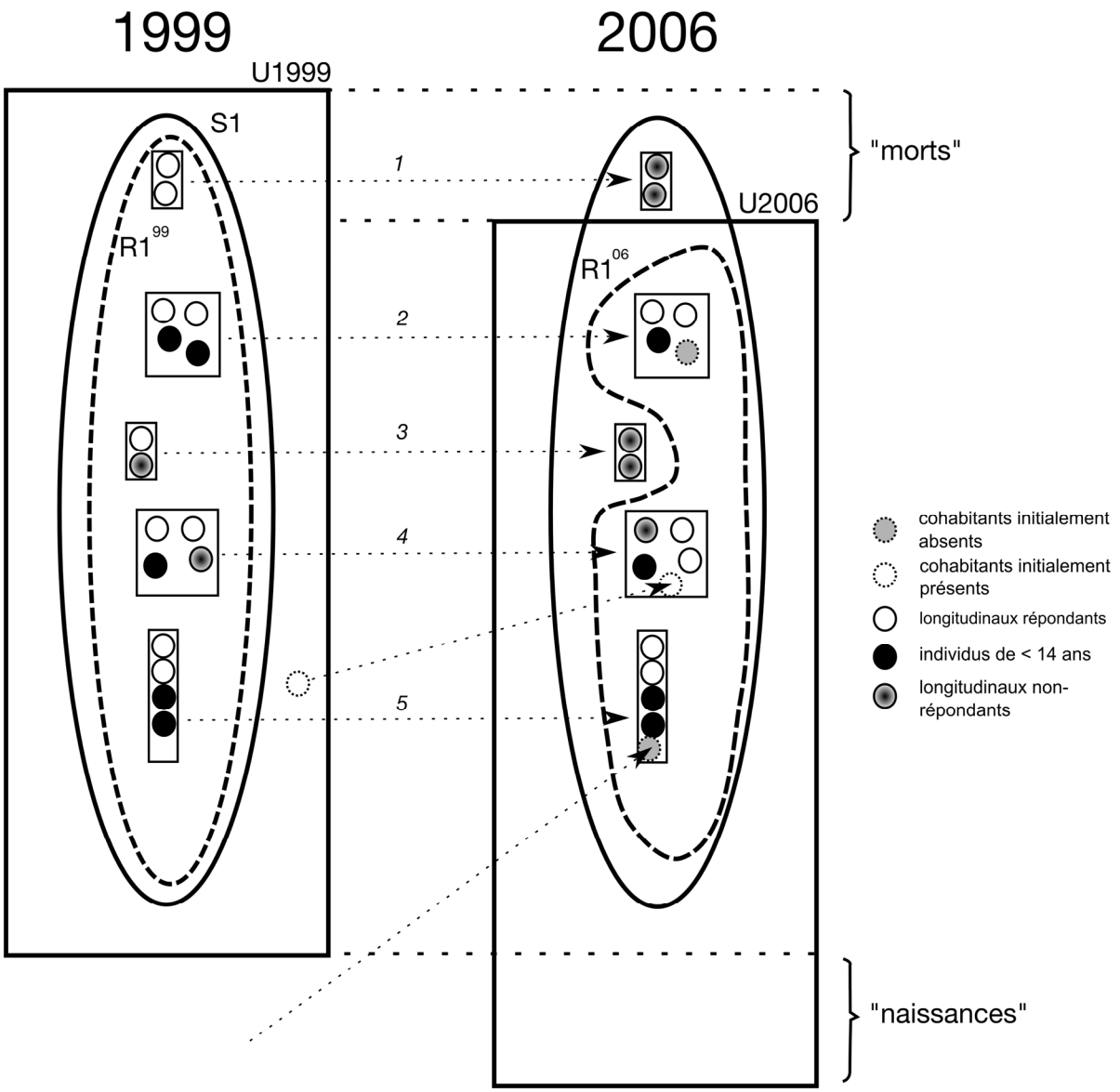
La figure 17 permet de visualiser les définitions et concepts listés dans ce chapitre. Les cercles représentent des individus et les rectangles les ménages les abritant. On a considéré 5 cas de figure courants (flèches numérotées dans la figure) de transitions entre deux années consécutives que nous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

U2006 Population transversale cible du PSM_I en 2006.

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^{06} Échantillon des personnes dans des ménages ayant répondu au moins au questionnaire grille en 2006, 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

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[8] 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 2006.

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 2006, 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. La distribution de la pondération transversale ménage du PSM_I vague 8 a été winsorisée aux 1er et 99^{e} percentiles à cause de la présence de quelques valeurs extrêmes et non réalistes.

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 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.

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

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 8 avec celui du PSM_II vague 3, il y a deux manières de procéder:

1. On peut combiner les panels en une seule fois après l'ajustement à la grille[9], 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[10] 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 8 du PSM_I et la vague 3 du PSM_II, trois pondérations ont été produites :

1 – poids transversal ménage
- *représentant la population des ménages en 2006 (vagues 8 et 3)*

2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2006 (vagues 8 et 3)*

3 – poids longitudinal individuel
- *représentant la population des individus de 14+ ans en 1999 pour le PSM_I vague 8, en 2004 pour le PSM_II vague 3.*

Pour les échantillons combinés PSM_I vague 8 & PSM_II vague 3, les trois mêmes types de pondérations ont été produites. En plus de la pondération longitudinale du PSM_I pour les analyses sur plus de 5 ans, c'est en principe les pondérations combinées qui sont utilisées pour les analyses de données puisqu'elles permettent d'utiliser toutes les données récoltées en 2006 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 2006*

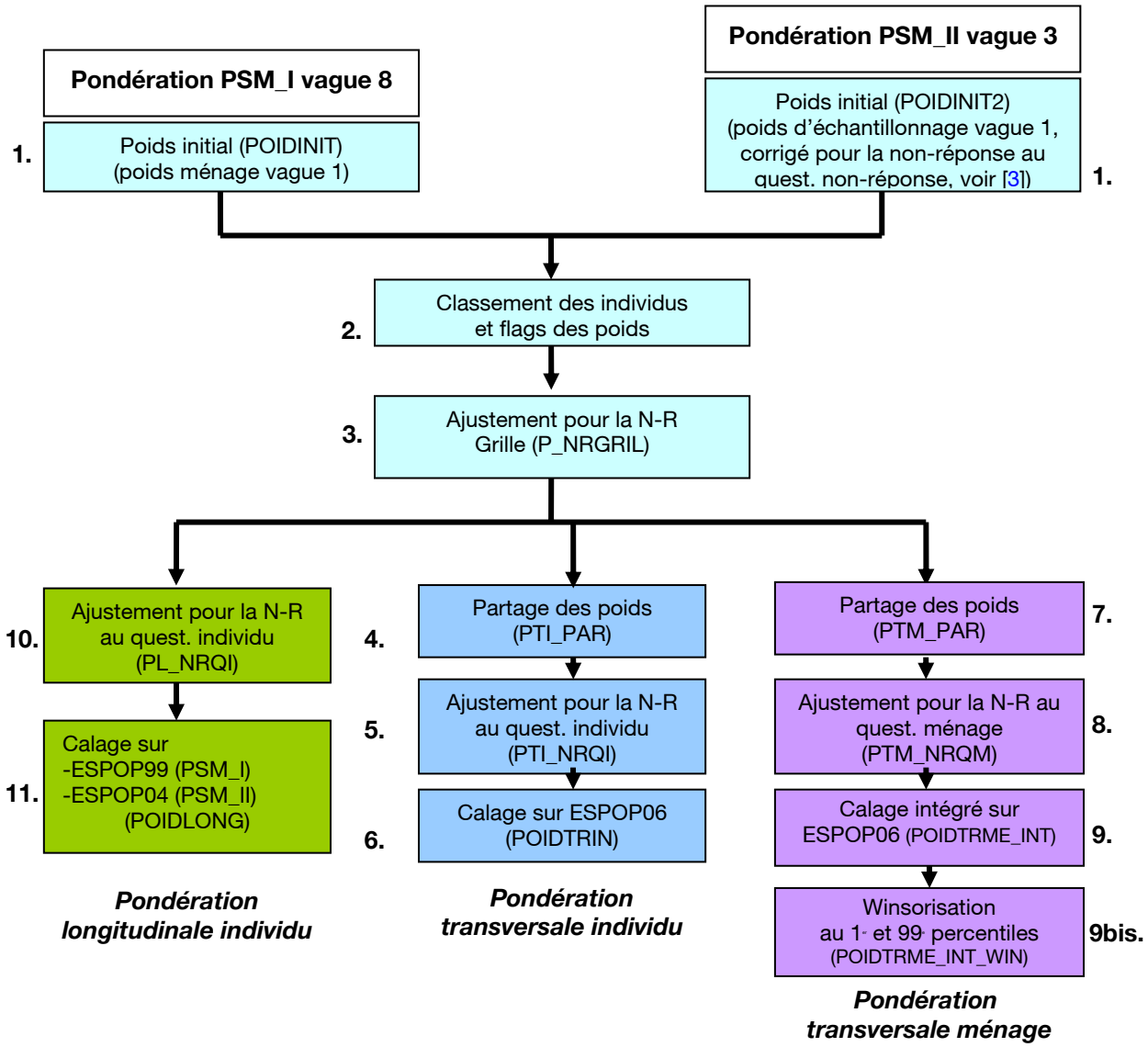
2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2006*

3 – poids longitudinal individuel
- *représentant la population des individus de 14+ ans en 2004.*

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

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 8 du PSM_I et de la vague 3 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 8 et du PSM_II vague 3. 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.



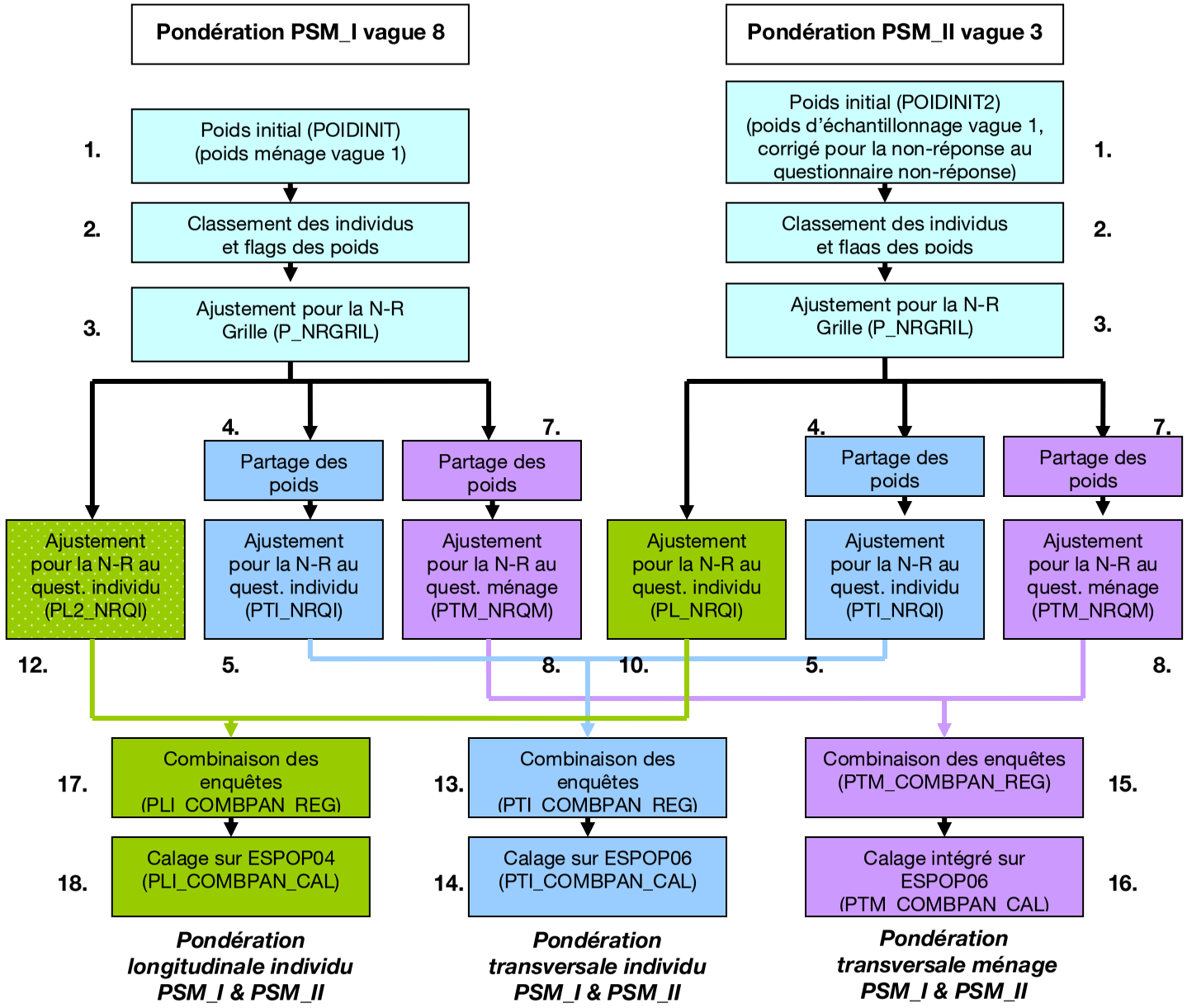


Figure 19 Étapes de la pondération permettant d'agréger l'échantillon du PSM_I vague 8 avec celui du PSM_II vague 3. Il y a une étape supplémentaire comparativement à la pondération séparée par enquête. Il s'agit de la combinaison des échantillons. Il faut combiner les échantillons pour chaque pondération (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 8 du PSM_I et la vague 3 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 gonfler 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 par 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 catégoriques 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 créer 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[11], 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 l'utilisation de cette méthode déjà depuis 9 ans.
Par ailleurs, mentionnons que les auteurs de [7] ont montré sur deux études[12] 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 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.

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

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)

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

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.

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[13]. Ainsi, il est naturel de prendre le poids ménage *POIDINIT2* décrit ici comme poids initial.

[13] 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).

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[14]. Pour le PSM_I, il on trouve donc tous les individus de INDIVIDUS 1999, INDIVIDUS 2006 et LONGOUT 2006, pour le PSM_II, on trouve tous ceux de INDIVIDUS 2004, INDIVIDUS 2006 et LONGOUT 2006 (ces fichiers sont décrits dans l'annexe A). Les variables contenues dans le fichier « classement » sont les suivantes:

Nom de la variable	Contenue dans le fichier classement PSM_I v8	Contenue dans le fichier classement PSM_II v2	Libellé
IDPERS	(x)	(x)	identificateur de la personne
idhous_o	(=IDHOUS99)	(=IDHOUS04)	identificateur du ménage à la vague originale (=vague 1)
idhous_c	(=IDHOUS06)	(=IDHOUS06)	identificateur du ménage à la vague courante (2006) (= vague 8 pour PSM_I, vague 3 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_v7	cr_v1, cr_v2	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.

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

Tableau 21 Distribution des individus du PSM_I et du PSM_II selon leur statut et leur code de réponse[15]. 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. Pourtant une replausibilisation de l'âge des répondants a modifié les comptes pour le PSM_II: 16 personnes considérées comme des longitudinaux adultes en 2005 sont passées en longitudinaux enfants en 2006. 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 5 (2) questionnaires individuels réalisés de moins pour le PSM_I (PSM_II) par rapport aux nombres donnés dans les tableaux 8 et 9.

	Statut de l'individu – **LONGST**									
	1 – Longitudinal adulte		2 – Longitudinal enfant		3 – Cohabitant[16] 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	1250	873	22	16	418	65	1690	954
1 – Hors-cible transversalement	170	34	3	0	0	0	0	0	173	34
2 – Non-répondant total	5780	2156	652	69	0	0	0	0	6432	2225
3 – Répondant grille seulement	60	137	8	2	13	11	1	0	82	150
4 – Répondant grille et questionnaire ménage seulement, dans un ménage où tous les individus sont non-répondants au questionnaire individuel	82	139	16	6	8	10	3	0	109	155
5 – Répondant grille et questionnaire ménage seulement, dans un ménage où au moins un individu est répondant au questionnaire individuel	812	617	208	32	136	73	22	5	1178	727
6 – Répondant grille, questionnaire ménage et questionnaire individuel	3447	2412	440	81	175	64	23	0	4085	2557
7 – Répondant grille et questionnaire individuel seulement	0	7	3	4	0	0	0	0	3	11
Total	**10351**	**5502**	**2580**	**1067**	**354**	**174**	**467**	**70**	**13752**	**6813**

[15] Les enfants de moins de 14 ans sont tous classés hors-cible transversalement.
[16] 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[17] 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 Vague 7	PSM_I Vague 8	PSM_II Vague 2	PSM_II Vague 3
F_PLONG = 0 – Poids longitudinal = 0	10166	10135	4137	4360
F_PLONG = 1 – Poids longitudinal > 0	3495	3617	2589	2453
Total	**13661**	**13752**	**6726**	**6813**
F_PTRME = 0 – Poids transversal ménage = 0	8469	9664	4078	4245
F_PTRME = 1 – Poids transversal ménage > 0	5192	4088	2648	2568
Total	**13661**	**13752**	**6726**	**6813**
F_PTRIN = 0 – Poids transversal individuel = 0	9776	8380	3026	3374
F_PTRIN = 1 – Poids transversal individuel > 0	2885	5372	3700	3439
Total	**13661**	**13752**	**6726**	**6813**

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

Tous les ménages choisis pour être réinjectés dans l'enquête en 2006 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 12758 et 5535 individus au total car on a écarté les cohabitants et les individus hors cible transversalement (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	5780	1902	**7682 (60.2)**
	PSM_II	2156	942	**3098 (47.4)**
1 – Répondant à la grille	PSM_I	4401	675	**5076 (39.8)**
	PSM_II	3312	125	**3437 (52.6)**
Total	**PSM_I**	**10181**	**2577**	**12758**
	PSM_II	**5468**	**1067**	**5535**

[17] 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 53 GHR pour le PSM_I, respectivement 31 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 5076 et 3437 individus longitudinaux dans les ménages répondants à la grille (ceux pour qui REPGRIL=1, voir tableau 24). Les 173 et 34 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).
Pour le PSM_II on a tout de même des individus pour qui f_aj_nrgril=1 parmi ceux dans la champ de l'enquête car, comme on peut le voir dans l'annexe C, un des groupe homogène de réponse formé par l'algorithme de segmentation a atteint 100% de taux de réponse, donc pas d'ajustement pour ce GHR.
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	2.26	1.65	Min	1.30	1.00
Médiane	2.04	1.63	Max	3.32	2.00
Écart-type	0.64	0.28	Étendue	2.02	1.00

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 8 et le PSM_II vague 3. 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 2006.

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 8*	*PSM_II* *Vague 3*
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1369 (25.1)	1032 (28.7)
1 – Répondant au Q.I.	4088 (74.1)	2568 (71.3)
Total	**5457**	**3600**

Tableau 26 Statistiques descriptives du facteur d'ajustement, f_aj_nrtiqi, pour les 4088 et 2568 individus répondants au questionnaire individuel (ceux pour qui REPQI=1, voir tableau 25). PTI_NRQI = P_NRGRIL * f_aj_nrtqi.
Pour le PSM_I on a des individus pour qui f_aj_nrtqi=1 car, comme on peut le voir dans l'annexe E, un des groupe homogène de réponse formé par l'algorithme de segmentation a atteint 100% de taux de réponse, donc pas d'ajustement pour ce GHR.

	PSM_I	PSM_II		PSM_I	PSM_II
Moyenne	1.34	1.40	Min	1.00	1.03
Médiane	1.29	1.35	Max	2.73	1.99
Écart-type	0.29	0.31	Étendue	1.73	0.96

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 8, le modèle obtenu avec l'analyse par segmentation comporte 29 GHR et 27 GHR pour le PSM_II vague 3 (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 2006 selon l'ESPOP (de décembre 2006). 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*(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[18]

Le tableau 27 donne quelques nombres caractérisant les calages effectués.

Tableau 27 Caractéristiques du calage sur marge, pondération transversale individuelle. On a respectivement 4088 et 2568 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	0.93	0.86	1516	2413
Médiane	0.89	0.81	1207	2027
Écart-type	0.29	0.26	1023	1295
Min	0.56	0.51	138	515
Max	2.17	1.82	10999	12966
Étendue	1.62	1.30	10861	12451

[18] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal individuel, donc POIDTRIN, est renommée en:
- Wp06TP1P pour le PSM_I vague 8
- Wp06TP2P pour le PSM_II vague 3

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 3 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 avant correction. 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 8*	*PSM_II vague 3*	
Homme 14-24 ans	618086	761007	479986
Homme 25-34 ans	381372	405358	476752
Homme 35-44 ans	564637	611251	610459
Homme 45-54 ans	502484	601807	535289
Homme 55 ans et plus	984332	957608	925494
Femme 14-24 ans	593827	600127	457634
Femme 25-34 ans	378807	428620	485042
Femme 35-44 ans	644548	745728	610869
Femme 45-54 ans	608574	721665	531036
Femme 55 ans et plus	1185807	1324987	1083787
Suisse	5723900	6312192	4918185
Étranger	738574	845965	1278163
Lémanique, Centrale et Tessin	2332998	2342646	1969700
Mitteland, N-O, Zürich, Orientale.	4129476	4815512	4226648
Non marié	2969539	3352252	2800971
Marié	3492935	3805906	3395378
Total	**6462474**	**7158158**	**6196348**

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 8 et le PSM_II vague 3. 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 2006.

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 répondants 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'à 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 répondants 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 IDHOUS06).
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 répondants à la grille selon la réponse au questionnaire ménage.

	PSM_I *Vague 8*	*PSM_II* *Vague 3*
Réponse au QM (REPQM)	Nombre de ménages (proportion en %)	Nombre de ménages (proportion en %)
0 – Non-répondant au questionnaire ménage	43 (1.7)	70 (4.0)
1 – Répondant au questionnaire ménage	2537 (98.3)	1684 (96.0)
Total	**2580**	**1754**

La modélisation est basée sur la grille de la vague courante soit celle de la vague 8 pour le PSM_I et celle de la vague 3 pour le PSM_II. Le modèle obtenu avec l'analyse par segmentation comporte 7 GHR pour le PSM_I et 9 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 2537 et 1684 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 8	PSM_II vague 3
Moyenne	1.02	1.04
Médiane	1.00	1.00
Écart-type	0.07	0.14
Min	1.00	1.00
Max	1.39	1.75
Étendue	0.39	0.75

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*[19] pour le poids ménage en utilisant les comptes de population Suisse en 2006 (vague courante), d'après l'ESPOP de décembre 2006.

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 2006. 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*(< 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)

Pour le PSM_I, l'attrition importante que l'échantillon a subie, s'est fait ressentir à cette étape du calage. En effet, 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

[19] 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.

Tableau 31 Caractéristiques du calage sur marge, pondération transversale ménage. On a respectivement 6489 individus dans 2537 ménages, et 4087 individus dans 1684 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	0.91	1.00	0.87	1123	1126	1783
Médiane	0.84	1.00	0.83	904	907	1548
Écart-type	0.40	0.01	0.28	801	800	885
Min	**-0.15**	0.93	0.30	**-304**	32	222
Max	3.08	1.02	2.06	7321	7321	6644
Étendue	3.24	0.09	1.76	7625	7289	6422

8.4 Winsorisation des poids après calage: étape 9bis

Pour le PSM_I, le rapport poids maximal sur poids minimal vaut 227.8 dû à quelques poids extrêmes. Cela pourrait rendre certaines analyses instables. Nous avons décidé de winsoriser les poids calés aux 1er et 99e percentiles. Ainsi le rapport max/min est passé à 25.6. 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 choisit 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[20]

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

[20] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal ménage, donc POIDTRME, a été renommée en:
- Wh06TP1P pour le PSM_I vague 8
- Wh06TP2P pour le PSM_II vague 3

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)*
Variable	*PSM_I vague 8 (avant 1er calage)*	*PSM_II vague 3*	
Homme < 14 ans	622077	646178	560649
Homme 14-24 ans	596427	693972	479986
Homme 25-34 ans	348433	422527	476752
Homme 35-44 ans	571083	610753	610459
Homme 45-54 ans	560056	660634	535289
Homme 55 ans et plus	1022690	1040058	925494
Femme < 14 ans	604911	594029	530906
Femme 14-24 ans	566582	585898	457634
Femme 25-34 ans	364176	435083	485042
Femme 35-44 ans	648662	718963	610869
Femme 45-54 ans	655832	731770	531036
Femme 55 ans et plus	1181987	1259741	1083787
Suisse	6894824	7286145	5746587
Étranger	848092	1113461	1541316
Lémanique, Centrale et Tessin	2773628	2732175	2334421
Mitteland, N-O, Zürich, Orientale.	4969288	5667432	4953481
Non marié	4120884	4471916	3892525
Marié	3622033	3927690	3395378
Total	**7742916**	**8399607**	**7287903**

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 170 et 34 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 8	*PSM_II Vague 3*
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	954 (21.7)	893 (27.0)
1 – Répondant au Q.I.	3447 (78.3)	2419 (73.0)
Total	**4401**	**3312**

Tableau 34 Statistiques descriptives du facteur d'ajustement, f_aj_nrlqi, pour les 3447 et 2419 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-desous), on a f_aj_nrlqi=1.

	PSM_I	PSM_II		PSM_I	PSM_II
Moyenne	1.28	1.38	Min	1.00	1.02
Médiane	1.26	1.35	Max	2.59	1.85
Écart-type	0.26	0.26	Étendue	1.59	0.83

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 8, le modèle obtenu avec l'analyse par segmentation comporte 17 GHR et 16 GHR pour le PSM_II vague 3 (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 (3447+170=) 3617 et (2419+34=) 2453 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[21]

1. Sexe*(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[22]

Le tableau 35 donne quelques nombres caractérisant les calages effectués.

[21] 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.
[22] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal individuel, donc POIDLONG, est renommée en:

- Wp06LP1P pour le PSM_I vague 8
- Wp06LP2P pour le PSM_II vague 3

Tableau 35 Caractéristiques du calage sur marge (individus hors-cible transversalement inclus), pondération longitudinale individuelle. On a respectivement 3617 et 2453 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	1658	2479
Médiane	0.93	0.85	1439	2179
Écart-type	0.26	0.28	968	1203
Min	0.67	0.53	381	644
Max	1.99	1.78	10121	11886
Étendue	1.32	1.25	9741	11242

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 8	PSM_II vague 3	PSM_I (ESPOP1999)	PSM_II (ESPOP 2004)
Homme 14-24 ans	522981	645772	463320	469973
Homme 25-34 ans	482245	439395	530410	481790
Homme 35-44 ans	552028	671936	585647	611688
Homme 45-54 ans	498853	568896	497434	513613
Homme 55 ans et plus	675090	893973	824282	888469
Femme 14-24 ans	453897	632237	445927	449624
Femme 25-34 ans	548081	455320	540607	491394
Femme 35-44 ans	643356	780415	571212	611586
Femme 45-54 ans	580965	668664	492648	509749
Femme 55 ans et plus	814262	1196414	1046725	1052980
Suisse	5067028	6162946	4861829	4841446
Étranger	704731	790075	1136383	1239420
Lémanique, Centrale et Tessin	2037774	2276206	1900541	1924056
Mitteland, N-O, Zürich, Orientale.	3733986	4676816	4097671	4156810
Total	**5771759**	**6953021**	**5998212**	**6080866**

10 Pondération transversale individuelle combinée PSM_I et PSM_II

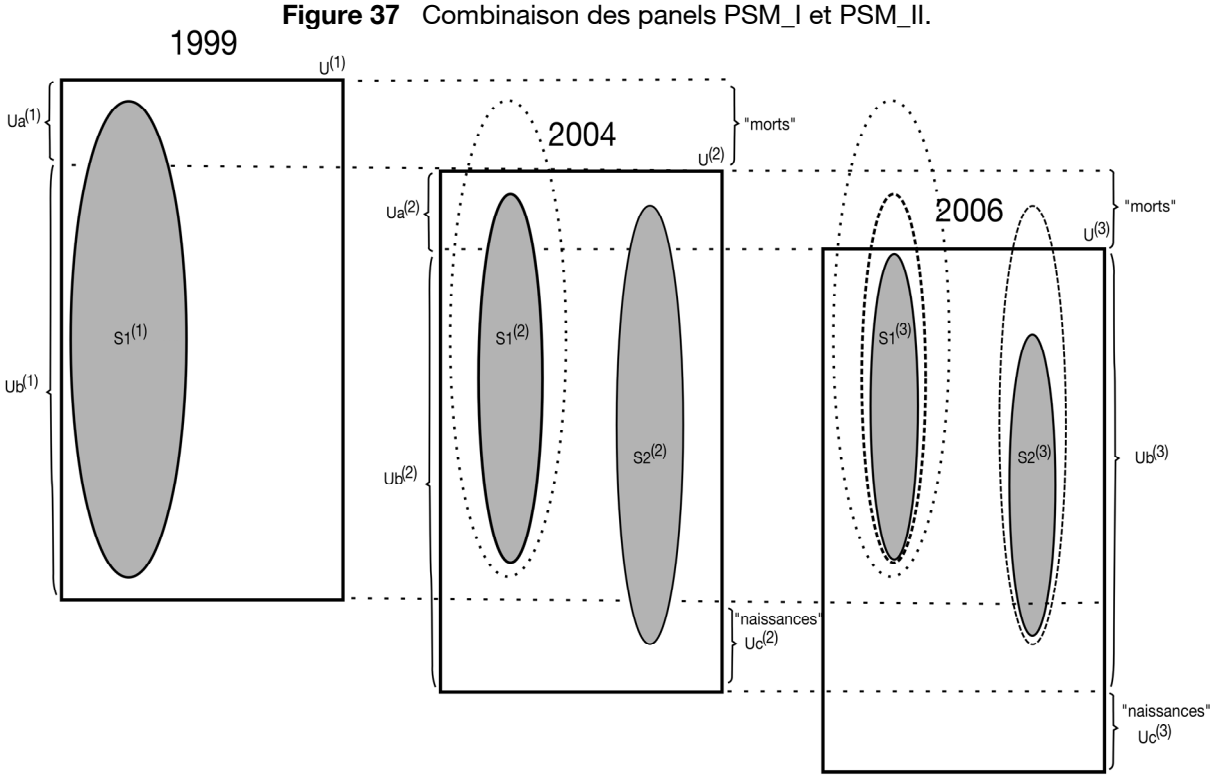
10.1 Combinaison transversale de panels à partir des échantillons transversaux de chacun des deux panels

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

t année de la vague courante: 2006,
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.

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 2006 pour le PSM_I seul (vague 8), pour le PSM_II seul (vague 3) et pour les PSM_I & PSM_II combinés (vague 3).
$S1^{(1)}$	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 8
$S2^{(2)}$	Échantillon des répondants du PSM_II en 2004, vague 1
$S2^{(3)}$	Échantillon des répondants du PSM_II en 2004, vague 3

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 2006 (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 sortantes $U_a^{(2)}$. Les individus de ces populations ne font alors plus partie de la population suisse au moment de la vague courante qui nous intéresse en 2006. 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, les 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.

Afin de développer une méthodologie qui permet de combiner l'information des échantillons transversaux des deux panels, il suffit de considérer les deux panel 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*[23] pour le panel *i* et n_i représente le nombre de personnes[24] (ou de ménages au chapitre suivant) du panel *i* répondantes au temps *t* (t étant l'année de référence traitée, 2006 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 choisit de se restreindre à minimiser la variance. Précisons encore que dans notre cas, vu que le PSM_I est 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. Il faut donc s'assurer que les facteurs d'allocation ne s'éloignent pas trop de ½.
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:

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

[23] 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.
[24] 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.

Tableau 38 Facteurs d'allocation par grande région. Pondération transversale individuelle combinée du PSM_I et le PSM_II. Les facteurs varient entre 0.58 et 0.69.

REGION06 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	1224	748	476	0.61
2	1662	1026	636	0.62
3	919	602	317	0.66
4	1187	692	495	0.58
5	781	456	325	0.58
6	638	396	242	0.62
7	245	168	77	0.69
TOTAL	6656	4088	2568	

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 2006. 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*(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[25].

Le tableau 39 donne quelques nombres caractérisant le calage effectué.

Tableau 39 Caractéristiques du calage sur marge, pondération transversale individuelle. On a 6656 individus avec un POIDTRIN>0 pour PSM_I et PSM_II combinés.

	facteur d'ajustement du calage = POIDTRIN/ PTI_COMBPAN_REG	*POIDTRIN*
Moyenne	0.90	931
Médiane	0.84	761
Écart-type	0.27	575
Min	0.53	85
Max	2.01	6189
Étendue	1.48	6104

[25] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRIN, est renommée en wp06T1P.

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 (P_COMBPAN_IND)*	*Population (et estimation avec le poids après calage POIDTRIN) ESPOP 2006*
Homme 14-24 ans	676520	479986
Homme 25-34 ans	393105	476752
Homme 35-44 ans	584360	610459
Homme 45-54 ans	541293	535289
Homme 55 ans et plus	973467	925494
Femme 14-24 ans	597447	457634
Femme 25-34 ans	400077	485042
Femme 35-44 ans	683496	610869
Femme 45-54 ans	651743	531036
Femme 55 ans et plus	1243228	1083787
Suisse	5963795	4918185
Étranger	780941	1278163
Lémanique, centrale et Tessin	2342589	1969700
Mitteland, N-O, Zürich, orientale.	4402148	4226648
Pas Marié	3123409	2800971
Marié	3621328	3395378
Total	**6744736**	**6196348**

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 2006.
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 précédant.

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. Le facteur varie entre 0.58 et 0.67.

REGION06 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	763	460	303	0.60
2	1056	628	428	0.59
3	567	369	198	0.65
4	769	446	323	0.58
5	509	294	215	0.58
6	389	227	162	0.58
7	168	113	55	0.67
TOTAL	4221	2537	1684	

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 2006).
Le poids transversal combiné (PTM_COMBPAN_REG) représente la population de la Suisse à la vague courante. Les totaux se trouvent dans le tableau 43.

Poids avant calage : P_ PTM_COMBPAN_REG.

Totaux de contrôle

- Sexe*(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[26]

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 -101). Avant de refaire le second calage, on remet les poids négatifs à leur valeur avant le premier calage (soit P_ 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.

Tableau 42 Caractéristiques du calage sur marge, pondération transversale ménage combinée. On a 10576 individus dans 4221 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 $= \frac{\textit{poids avant}}{\textit{poids après}}$			
	PSM_I & PSM_II 1er calage	*PSM_I & PSM_II 2e calage*	*PSM_I & PSM_II Après premier calage*	*Poids finaux PSM_I & PSM_II POIDTRME*
	Niveau individus dont le ménage a répondu au questionnaire ménage		*Niveau ménages répondants au questionnaire ménage*	
Moyenne	0.89	1.00	689	689
Médiane	0.84	1.00	568	568
Écart-type	0.35	0.01	420	420
Min	**-0.09**	0.96	**-101**	41
Max	2.67	1.01	3686	3687
Étendue	2.76	0.05	3787	3646

[26] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRME, est renommée en wh06T1P.

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 (et estimation avec le poids après calage POIDTRME) ESPOP 2006*
Homme moins de 14 ans	634021	560649
Homme 14-24 ans	636775	479986
Homme 25-34 ans	379201	476752
Homme 35-44 ans	588907	610459
Homme 45-54 ans	601561	535289
Homme 55 ans et plus	1031667	925494
Femme moins de 14 ans	601685	530906
Femme 14-24 ans	573996	457634
Femme 25-34 ans	392925	485042
Femme 35-44 ans	676503	610869
Femme 45-54 ans	685368	531036
Femme 55 ans et plus	1215344	1083787
Suisse	7064497	5746587
Étranger	953457	1541316
Lémanique, centrale et Tessin	2763969	2334421
Mitteland, N-O, Zürich, orientale.	5253984	4953481
Pas Marié	4264576	3892525
Marié	3753378	3395378
Total	**8017954**	**7287903**

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 mis 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 répondants à la grille selon la réponse au questionnaire individuel. Notons qu'on a 30 longitudinaux hors-cible transversalement respectivement pour le PSM_I en prenant l'année de référence 2004. Ces derniers conserveront leur P_NRGRIL.

	PSM_I Vague 8
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1113 (22.8)
1 – Répondant au Q.I.	3761 (77.2)
Total	**4874**

Tableau 45 Statistiques descriptives du facteur d'ajustement, f_aj_nrlqi, pour les 3761 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 30 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	1.29	Max	2.63
Écart-type	0.27	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 (3761+30=) 3791 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 éligible 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 3 en 2006.
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 2006 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. Le facteur varie entre 0.57 et 0.68.

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	1142	686	456	0.60
2	1564	958	606	0.61
3	850	554	296	0.65
4	1109	635	474	0.57
5	760	442	318	0.58
6	583	356	227	0.61
7	236	160	76	0.68
TOTAL	**6244**	**3791**	**2453**	

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[27]

Le tableau 47 donne quelques nombres caractérisant le calage effectué.

Tableau 47 Caractéristiques du calage sur marge (individus hors-cible transversalement inclus), pondération longitudinale individuelle combinée. On a 6244 observations avec un PLI_COMBREG_CAL>0.

	facteur d'ajustement du calage = PLI_COMBREG_CAL/PLI_COMBPAN_REG	PLI_COMBREG_CAL
	PSM_I & PSM_II	*PSM_I & PSM_II*
Moyenne	0.88	974
Médiane	0.83	805
Écart-type	0.25	641
Min	0.51	205
Max	1.84	10075
Étendue	1.33	9870

[27] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final longitudinal individuel, donc PLI_COMBREG_CAL, est renommée en WP06L1P.

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

	*Estimation avec le poids avant calage (*PLI_COMBPAN_REG*)*	*Population (et estimation avec le poids après calage* PLI_COMBREG_CAL*)*
Variable	PSM_I & PSM_II	PSM_I & PSM_II (ESPOP 2004)
Homme 14-24 ans	646068	469973
Homme 25-34 ans	426263	481790
Homme 35-44 ans	649359	611688
Homme 45-54 ans	553382	513613
Homme 55 ans et plus	900567	888469
Femme 14-24 ans	598699	449624
Femme 25-34 ans	431990	491394
Femme 35-44 ans	756559	611586
Femme 45-54 ans	634215	509749
Femme 55 ans et plus	1177703	1052980
Suisse	5970924	4841446
Étranger	803880	1239420
Lémanique, Centrale et Tessin	2324690	1924056
Mitteland, N-O, Zürich, Orientale.	4450114	4156810
Non marié	3049915	2686477
Marié	3724889	3394388
Total	**6774804**	**6080866**

13 Conclusions

Les pondérations décrites dans ce document ont été calquées sur celles produites par des méthodologues de Statistique Canada (voir [3], [11]) pour les données du Panel Suisse de Ménages (PSM) les années précédentes. On a fait de gros efforts pour étoffer le rapport au maximum par des exemples et des explications supplémentaires –par rapport aux années précédentes– concernant les divers aspects théoriques ardus mis en pratique aux diverses étapes de la production des poids.

Cette année se différencie un peu des précédentes par de nouvelles tentatives tant pour lutter contre l'attrition (incitatifs, voir paragraphe 2.2) que pour essayer de la prendre en compte dans les modèles de non-réponse de la production des pondérations. L'attrition est un problème grandissant pour le PSM, à tel point que son avenir pourrait être remis en question (du moins du point de vue statistique-méthodologique) si l'on ne parvient pas à freiner le phénomène. En effet, les modèles pour expliquer la non-réponse ont leurs limites et les calages sur marges deviennent difficiles (voir paragraphes 8.3 et 11.2). Comme le montrent les figures 2 et 3, en 2007-08, de gros efforts ont à nouveau été faits et l'attrition a été (momentanément) amoindrie. Ceci fera l'objet d'un rapport ultérieur (celui des pondérations des vagues 4 et 9). Cependant le nombre de refus stricts et définitifs, ainsii que celui d'individus et de ménages dont on perd la trace, augmentent de plus en plus. Dès lors, la représentativité des échantillons est de plus en plus remise en cause…
L'attrition grandissante doit inciter les utilisateurs à appliquer les poids dans toutes les situations. 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. L'utilisation des poids est impérative, si l'on n'applique pas les poids appropriés, les estimations effectuées ne peuvent pas être considérées comme représentatives de la population observée. 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 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 de 30 observations (nombre non pondéré d'observations), celle-ci ne devrait pas être diffusée, peu importe la valeur de son coefficient de variation.

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 sera souvent sous-estimée à l'aide des logiciels cités au paragraphe précédant, 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 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 standard.

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.

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 2006 : ce fichier contient les ménages qui ont répondu à la grille à la vague 8 du PSM_I et à la vague 3 du PSM_II.
Le fichier contient 4334 enregistrements, 2580 pour le PSM_I, 1754 pour le PSM_II, 52 variables. Nom: `MEN06.sas7bdat`, construit par le programme `EGR_PSM_men06_Ind06_2007-09-03.sas`

INDIVIDUS 2006 : ce fichier contient tous les individus des ménages qui ont répondu à la grille à la vague 8 du PSM_I et à la vague 3 du PSM_II. Tous font partie de la population cible transversale. Ils peuvent être répondants ou non-répondants au questionnaire individuel.
10863 enregistrements, 6587 pour le PSM_I, 4276 pour le PSM_II, 33 variables. Nom: `PERS06.sas7bdat`, construit par le programme `EGR_PSM_men06_Ind06_2007-09-03.sas`

PARTIS 2006 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[28] ou pas, selon le codage suivant:

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

348 enregistrements au total, 180 pour le PSM_I, 168 pour le PSM_II, 29 variables. Nom: `partis06.sas7bdat`, construit par le programme *`EGR_partis06.sas`*

Tableau A –1 Répartition de la variable OSM :

PSM_I vague 8	PSM_II vague 3
128OSM	139 OSM
5 Enfant d'OSM	5 Enfant d'OSM
47 non-OSM	24 non-OSM

REVENUS 2006 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 PARTIS06 (mais avec la variable G06RET[29] à la place de la G06RNH).

[28] OSM = Original Sample Member, i.e. un individu longitudinal.
[29] La variable G06RET indique le retour d'anciens membres du ménage (elle est construite à partir des variables G06RET01à G06RET16 renseignées pour chaque personne). Ces variables ne sont renseignées qu'à partir de la vague 3 ou tout de même déjà en vague 2 dans des cas spéciaux de split-off de ménages.

20 enregistrements au total, 18 pour le PSM_I, 2 pour le PSM_II.
Nom : `revenus06.sas7bdat`, construit par le programme `EGR_revenus06.sas`

LONGOUT06[30]

Ce fichier est fabriqué à partir des fichiers PARTIS00, PARTIS01, PARTIS02, PARTIS03, PARTIS04, PARTIS05, PARTIS06 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 8 pour le PSM_I, vague 3 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.
996 enregistrements au total, 792 pour le PSM_I, 204 pour le PSM_II, 5 variables. Nom : `longout06.sas7bdat`, construit par le programme `EGR_longout06.sas`

LISTPERS06

Contient, pour tous les ménages recensés en 2006, la liste de toutes les personnes qui les compose avec leur numéro d'identification personnel (`IDPERS`).
4334 enregistrements, 2580 pour le PSM_I, 1754 pour le PSM_II, 19 variables.
Nom : `listpers06.sas7bdat`, construit par le programme `EGR_listpers06.sas`

[30] Attention, on a modifié le programme construisant ce fichier par rapport aux années passées. Effectivement l'ancien programme contenait des erreurs de construction qui devaient être corrigées à chaque fois a posteriori lors du calcul effectif des pondérations.

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_CH06 < 1999 (PSM_I)
  < 2004 (PSM_II)
  ET AGECORRO>=0)).
  Pour les HAB_CH06[31] < 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_CH06 >= 1999 (PSM_I)
  >= 2004 (PSM_II)
  OU AGECORRO<0)).

Note : les enfants (< 14 ans) longitudinaux n'ont pas de poids longitudinal. Les enfants qui ont 14 ans ou plus à la vague 8 du PSM_I ou à la vague 3 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 2006 et pas dans LONGOUT et avec RNH=1, 2, 3 ou 7)
- **3** **Répondant grille seulement**
  (AGECORRC >= 14 et dans INDIVIDUS 2006 et STATUS06=1 ou 2 et STATHH06 = 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 2006 et STATUS06=1 ou 2 et STATHH06 = 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 2006 et STATUS06=1 ou 2 et STATHH06 = 1 et Q_INDIV=1)
- **6** **Répondant grille, questionnaire ménage et questionnaire individuel**
  (AGECORRC >= 14 et dans INDIVIDUS 2006 et STATUS06=0 et STATHH06=1)
- **7** **Répondant grille et questionnaire individuel seulement**
  (AGECORRC >= 14 et dans INDIVIDUS 2006 et STATUS06=0 et STATHH06=0)

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

[31] HAB_CH06 := durée de résidence en Suisse, depuis quelle année.

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 8***

IF age99 >= 0 THEN agecorro = age99;
ELSE IF age06 >= 0 THEN agecorro = age06-7;
ELSE IF annee1 >= 1800 THEN agecorro = 1999-annee1;
ELSE IF annee8 >= 1800 THEN agecorro = 1999-annee8;
→ Les bébés nés après la première vague auront un AGECORR0 négatif.
Annee1, annee8 : année de naissance communiquée en vague 1 respectivement 8.

Imputation des âges manquants
IF IDPERS=2402101 then agecorro=45;
ELSE IF IDPERS=2782101 then agecorro=27;
ELSE IF IDPERS=9090101 then agecorro=70;
ELSE IF IDPERS=683201 then agecorro=18;
ELSE IF IDPERS=7170106 then agecorro=41;
ELSE IF IDPERS=12420103 then agecorro=26;

agecorrc= agecorro + 7

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

IF age06 >= 0 THEN agecorro = age06-2;
ELSE IF age04 >= 0 THEN agecorro = age04;
ELSE IF annee3 >= 1800 THEN agecorro = 2004-annee3;
ELSE IF annee1 >= 1800 THEN agecorro = 2004-annee1;
→ Les bébés nés après la première vague auront un AGECORR0 négatif.
Annee1, annee3 : année de naissance communiquée en vague 1 respectivement 3.

Imputation des âges manquants

IF IDPERS=20873102 then agecorro=19;
ELSE IF IDPERS=20873103 then agecorro=21;
ELSE IF IDPERS=21811101 then agecorro=27;
ELSE IF IDPERS=21811102 then agecorro=27;

```
ELSE IF IDPERS=22647102  then agecorro=56;
ELSE IF IDPERS=23562102  then agecorro=44;
ELSE IF IDPERS=24640102  then agecorro=68;
ELSE IF IDPERS=23482102  then agecorro=50;
ELSE IF IDPERS=20226103  then agecorro=22;
ELSE IF IDPERS=21838102  then agecorro=28;
ELSE IF agecorro=.       then agecorro=9999;

agecorrc=    agecorro + 2
```

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.

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).
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	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
FRANC	NON (voir tableau C-1)	NON (voir tableau C-1)	Questionnaire grille ou non réponse rempli en français
ITAL	OUI	OUI	Questionnaire grille ou non réponse rempli en italien
ALLEM	**OUI**	**OUI**	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	NON (voir tableau C-1)	NON (voir tableau C-1)	Région Lémanique
MIDDLE	**OUI**	**OUI**	Région Mitteland
NOREST	**OUI**	**OUI**	Région Nord-Ouest
EST	**OUI**	OUI	Région Suisse orientale
CENTRAL	**OUI**	**OUI**	Région Suisse centrale
ZURICH	**OUI**	OUI	Région Zurich
MIXED	OUI	**OUI**	Type de commune : agricole mixte
CENTRES	**OUI**	**OUI**	Type de commune : agglomération
SUBURB	**OUI**	OUI	Type de commune : banlieue
WEALTH	OUI	OUI	Type de commune : riche
PERIPH	OUI	OUI	Type de commune : périphérique
TOURIST	OUI	OUI	Type de commune : touristique
INDUS	**OUI**	OUI	Type de commune : industrielle ou tertiaire
RURAL	OUI	**OUI**	Type de commune : cité dortoir
AD_SEUL	NON (voir tableau C-1)	NON (voir tableau C-1)	Ménage avec un seul adulte
NO_KID	**OUI**	**OUI**	Ménage sans enfant
SEULE	OUI	**OUI**	Ménage d'une seule personne
LOY_1	**OUI**	OUI	Coût du logement jusqu'à 1000 Fr./mois
LOY_2	**OUI**	**OUI**	Coût du logement entre 1001-1500 Fr./mois
LOY_3	OUI	**OUI**	Coût du logement 1501 Fr./mois et plus
RAT1	OUI	OUI	nbr. de pièces / nbr. de personnes <= 1.34
RAT2	**OUI**	**OUI**	1.34 < nbr. de pièces / nbr. de personnes < 2
APPART	**OUI**	**OUI**	Ménage vivant en appartement
DETACH	NON (voir tableau C-1)	NON (voir tableau C-1)	Ménage vivant dans une maison indépendante
PROPRIO	**OUI**	**OUI**	Propriétaire
REV_1	**OUI**	NON	Premier tiers pour le revenu du ménage
REV_2	**OUI**	NON	Deuxième tiers pour le revenu du ménage
REV_3	**OUI**	NON	Troisième tiers pour le revenu du ménage

Niveau individu

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
AGEV1_1	**OUI**	OUI	15-24 ans (à la vague 1)
AGEV1_2	**OUI**	OUI	25 à 34 ans (à la vague 1)
AGEV1_3	**OUI**	OUI	35 à 44 ans (à la vague 1)
AGEV1_4	**OUI**	**OUI**	45 à 54 ans (à la vague 1)
AGEV1_5	**OUI**	OUI	55+ ans (à la vague 1)
CELIB	OUI	**OUI**	Célibataire
DIVORCE	OUI	**OUI**	Divorcé(e)
SEPARE	OUI	OUI	Séparé(e)
MARIE	**OUI**	**OUI**	Marié(e)
FORM_PRO	OUI	OUI	Formation professionnelle
FORM_SUP	**OUI**	**OUI**	Formation supérieure
NO_FORM	**OUI**	**OUI**	Pas de formation
HOMME	**OUI**	OUI	Homme
SWISS	**OUI**	OUI	Nationalité suisse.
TRAVAIL	**OUI**	OUI	Personne active professionnellement

Tableau C – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.36 en valeur absolue. On a décidé de supprimer LAKE_GE, FRANC, DETACH. On a omit de supprimer PROPRIO.
AD_SEUL et SEULE corrèlent à 0.89 pour le PSM_I et 0.90 pour le PSM_II: on a décidé de supprimer AD_SEUL

	FRANC	LAKE_GE
ALLEM	0.88 (PSM_I) 0.89 (PSM_II)	0.58 (PSM_I) 0.67 (PSM_II)
FRANC	1	0.66 (PSM_I) 0.75 (PSM_II)

	PROPRIO	DETACH
APPART	0.60 (PSM_I) 0.64 (PSM_II)	0.92 (PSM_I) 0.80 (PSM_II)
PROPRIO	1	0.61 (PSM_I) 0.51(PSM_II)

Modélisation de la non-réponse – analyse par segmentation

PSM_I vague 8
Le modèle de non-réponse créé avec l'analyse par segmentation a utilisé 29 variables différentes et conduit à la création de 53 groupes homogènes de réponse (GHR), 45 pour les longitudinaux adultes (figures C-2 et C-3), 8 pour les longitudinaux enfants (figure C-4).

PSM_II vague 3
Le modèle de non-réponse créé avec l'analyse par segmentation a utilisé 20 variables différentes et conduit à la création de 31 GHR pour les longitudinaux adultes (figure C-4). Il n'y a pas eu de modélisation pour les longitudinaux enfants en raison de leur trop petit nombre.

Figure C—2 Arbre pour les longitudinaux adultes (longst=1), PSM_I vague 8: **Partie 1**

split1	split2	split3	split4	split5	split6	rhg
agevo_1=0 (8387) 45.9%	FORM_SUP=0 (5652) 42.8%	SWISS=0 (762) 33.0%	APPART=0 (157) 39.6%			1
			APPART=1 (605) 31.4%			2
		SWISS=1 (4890) 44.5%	MARIE=0 (1491) 38.1%	REV_1=0 (1050) 41.3%	agevo_2=0 (665) 46.0%	3 4 5
					agevo_2=1 (385) 32.8%	6 7
				REV_1=1 (441) 30.7%		8
			MARIE=1 (3399) 47.3%	NO_KID=0 (1534) 51.5%	LOY_1=0 (1281) 49.6%	9 10 11 12 13 14 15
					LOY_1=1 (253) 61.1%	16
				NO_KID=1 (1865) 44.1%	agevo_5=0 (770) 39.5%	17 18 19 20 21 22
					agevo_5=1 (1095) 47.1%	23 24 25 26 27 28 29
	FORM_SUP=1 (2735) 52.3%	NO_KID=0 (1156) 57.8%	SWISS=0 (147) 39.7%			30
			SWISS=1 (1009) 60.7%	CENTRAL=0 (926) 59.0%	MARIE=0 (88) 44.6%	31
					MARIE=1 (838) 60.5%	32 33 34
				CENTRAL=1 (83) 76.9%		35
		NO_KID=1 (1579) 48.4%	agevo_2=0 (1142) 51.9%	TRAVAIL=0 (389) 46.3%	LOY_1=0 (257) 49.9%	36 37 38
					LOY_1=1 (132) 39.3%	39
				TRAVAIL=1 (753) 54.9%	INDUS=0 (699) 56.3%	40 41 42
					INDUS=1 (54) 36.6%	43
			agevo_2=1 (437) 39.0%			44
agevo_1=1 (1794) 30.2%						45

Figure C—3 Arbre pour les longitudinaux adultes (longst=1), PSM_I vague 8: **Partie 2**

split7	split8	split9	split10	rhg
				1
				2
TRAVAIL=0 (217) 39.2%				3
TRAVAIL=1 (448) 49.4%	EST=0 (396) 51.3%			4
	EST=1 (52) 36.3%			5
MIDDLE=0 (288) 30.1%				6
MIDDLE=1 (97) 42.1%				7
				8
ZURICH=0 (1098) 47.9%	agevo_3=0 (514) 52.6%	REV_2=0 (388) 49.0%		9
		REV_2=1 (126) 63.5%	APPART=0 (74) 75.3%	10
			APPART=1 (52) 46.9%	11
	agevo_3=1 (584) 43.7%	TRAVAIL=0 (165) 50.3%		12
		TRAVAIL=1 (419) 41.1%	NOREST=0 (350) 38.8%	13
			NOREST=1 (69) 53.5%	14
ZURICH=1 (183) 59.4%				15
				16
CENTRES=0 (608) 37.2%	PROPRIO=0 (195) 30.2%			17
	PROPRIO=1 (413) 40.6%	NOREST=0 (341) 37.9%	ALLEM=0 (114) 46.6%	18
			ALLEM=1 (227) 33.5%	19
		NOREST=1 (72) 54.4%		20
CENTRES=1 (162) 48.0%	agevo_4=0 (54) 35.1%			21
	agevo_4=1 (108) 54.2%			22
LOY_2=0 (821) 43.6%	NO_FORM=0 (533) 46.6%	REV_1=0 (386) 49.8%	SUBURB=0 (266) 45.6%	23
			SUBURB=1 (120) 59.0%	24
		REV_1=1 (147) 38.0%		25
	NO_FORM=1 (288) 38.1%			26
LOY_2=1 (274) 57.8%	ALLEM=0 (65) 46.1%			27
	ALLEM=1 (209) 61.3%	REV_2=0 (153) 57.2%		28
		REV_2=1 (56) 72.8%		29
				30
				31
EST=0 (700) 62.7%	agevo_2=0 (570) 60.9%			32
	agevo_2=1 (130) 70.5%			33
EST=1 (138) 50.25				34
				35
MARIE=0 (72) 39.9%				36
MARIE=1 (185) 53.7%	REV_3=0 (121) 47.7%			37
	REV_3=1 (64) 65.4%			38
				39
CENTRES=0 (449) 60.1%				40
CENTRES=1 (250) 49.4%	HOMME=0 (95) 58.0%			41
	HOMME=1 (155) 44.1%			42
				43
				44
				45

Figure C—4 Arbre pour les longitudinaux enfants (longst=2), PSM_I vague 8

split1	split2	split3	split4	split5	rhg
SWISS=0 (141) 32.3%					46
SWISS=1 (1186) 53.1%	ZURICH=0 (1022) 51.0%	REV_2=0 (739) 48.1%	LOY_2=0 (532) 50.5%	RAT2=0 (417) 53.1%	47
				RAT2=1 (115) 41.3%	48
			LOY_2=1 (207) 41.8%	ALLEM=0 (72) 31.7%	49
				ALLEM=1 (135) 47.2%	50
		REV_2=1 (283) 58.5%			51
	ZURICH=1 (164) 66.4%	APPART=0 (104) 72.4%			52
		APPART=1 (60) 56.2%			53

Figure C—5 Arbre pour les longitudinaux adultes (longst=1), PSM_II vague 3

split1	split2	split3	split4	split5	split6	split7	rhg
FORM_SUP=0 (3779) 57.7%	loy_3=0 (2613) 55.1%	loy_2=0 (1549) 51.7%	MIDDLE=0 (1146) 50.1%				1
			MIDDLE=1 (403) 56.4%	RAT2=0 (295) 50.8%			2
				RAT2=1 (108) 71.4%	CENTRES=0 (71) 63.7%		3
					CENTRES=1 (37) 84.5%		4
		loy_2=1 (1064) 60.3%	PROPRIO=0 (330) 67.0%	ALLEM=0 (89) 56.8%			5
				ALLEM=1 (241) 70.9%	RURAL=0 (208) 73.7%	NO_FORM=0 (159) 77.4%	6
						NO_FORM=1 (49) 62.5%	7
					RURAL=1 (33) 51.4%		8
			PROPRIO=1 (734) 57.2%	MIDDLE=0 (599) 55.0%	DIVORCE=0 (543) 53.7%		9
					DIVORCE=1 (56) 68.2%		10
				MIDDLE=1 (135) 67.6%	ALLEM=0 (52) 56.2%		11
					ALLEM=1 (83) 74.4%	RAT2=0 (51) 83.1%	12
						RAT2=1 (32) 62.5%	13
	loy_3=1 (1166) 63.5%	MARIE=0 (551) 56.9%	NO_KID=0 (282) 62.6%	APPART=0 (155) 54.6%			14
				APPART=1 (127) 71.9%	CENTRAL=0 (94) 66.0%		15
					CENTRAL=1 (33) 89.2%		16
			NO_KID=1 (269) 51.4%				17
		MARIE=1 (615) 69.7%	CENTRES=0 (506) 72.5%	MIXED=0 (473) 71.3%			18
				MIXED=1 (33) 91.3%			19
			CENTRES=1 (109) 57.8%				20
FORM_SUP=1 (1689) 66.3%	CELIB=0 (1208) 69.3%	ALLEM=0 (363) 61.1%	loy_3=0 (211) 56.6%				21
			loy_3=1 (152) 67.2%				22
		ALLEM=1 ((845) 72.9%	NOREST=0 (684) 75.6%	APPART=0 (338) 70.5%	RAT2=0 (216) 64.8%		23
					RAT2=1 (122) 80.5%		24
				APPART=1 (346) 80.5%	NO_KID=0 (136) 90.6%	agevo_4=0 (102) 87.4%	25
						agevo_4=1 (34) 100.0%	26
					NO_KID=1 (210) 74.3%		27
			NOREST=1 (161) 61.5%				28
	CELIB=1 (481) 58.9%	SEULE=0 (297) 54.2%	loy_2=0 (216) 50.1%				29
			loy_2=1 (81) 65.0%				30
		SEULE=1 (184) 66.6%					31

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.

Niveau ménage

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
FRANC	NON (voir tableau D-1)	NON (voir tableau D-1)	Questionnaire grille ou non réponse rempli en français
ITAL	OUI	OUI	Questionnaire grille ou non réponse rempli en italien
ALLEM	**OUI**	OUI	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	NON (voir tableau D-1)	NON (voir tableau D-1)	Région Lémanique
MIDDLE	OUI	OUI	Région Mitteland
NOREST	OUI	**OUI**	Région Nord-Ouest
EST	**OUI**	OUI	Région Suisse orientale
CENTRAL	**OUI**	OUI	Région Suisse centrale
ZURICH	OUI	OUI	Région Zurich
MIXED	OUI	OUI	Type de commune : agricole mixte
CENTRES	OUI	OUI	Type de commune : agglomération
SUBURB	OUI	OUI	Type de commune : banlieue
WEALTH	OUI	OUI	Type de commune : riche
PERIPH	OUI	OUI	Type de commune : périphérique
TOURIST	OUI	OUI	Type de commune : touristique
INDUS	OUI	OUI	Type de commune : industrielle ou tertiaire
RURAL	OUI	OUI	Type de commune : cité dortoir
AD_SEUL	NON (voir tableau D-1)	NON (voir tableau D-1)	Ménage avec un seul adulte
NO_KID	**OUI**	OUI	Ménage sans enfant
SEULE	**OUI**	**OUI**	Ménage d'une seule personne
ADRFROMW5	**OUI**		Ménages réinjectés dans l'enquête, vus pour la dernière fois en vague 5 (1) ou autres (0)
REFW2W7CONV	OUI	**OUI**	Ménages non répondants de vague 2 PSM_II et vague 7 PSM_I convertis (1) ou autre (0)
REFW3W8CONV	OUI	OUI	Ménages non répondants des vagues en cours, vague 3 PSM_II et vague 8 PSM_I convertis par rappels de refus (1) ou autre (0)
INC_STAMPS	**OUI**	OUI	Ménages du lot ayant reçu un carnet de timbre comme *incitatif* (1) ou autre (0)
INC_LOTT	OUI	**OUI**	Ménages du lot ayant participé à la loterie comme *incitatif* (1) ou autre (0)
INC_GIFT	OUI	OUI	Ménages du lot pouvant faire un don comme *incitatif* (1) ou autre (0)

Niveau individu

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
AGEVC_1	**OUI**	OUI	15-24 ans (à la vague courante)
AGEVC_2	OUI	OUI	25 à 34 ans (à la vague courante)
AGEVC_3	OUI	**OUI**	35 à 44 ans (à la vague courante)
AGEVC_4	OUI	**OUI**	45 à 54 ans (à la vague courante)
AGEVC_5	OUI	OUI	55+ ans (à la vague 1)
CELIB	NON (voir tableau D-1)	NON (voir tableau D-1)	Célibataire
DIVORCE	OUI	OUI	Divorcé(e)
SEPARE	OUI	OUI	Séparé(e)
MARIE	OUI	**OUI**	Marié(e)
VEUF	OUI	OUI	Veuf / veuve
FORM_PRO	NON (voir tableau D-1)	NON (voir tableau D-1)	Formation professionnelle
FORM_SUP	**OUI**	**OUI**	Formation supérieure
NO_FORM	**OUI**	**OUI**	Pas de formation
HOMME	**OUI**	**OUI**	Homme
SWISS	OUI	**OUI**	Nationalité suisse.
TRAVAIL	OUI	OUI	Personne active professionnellement
PERMIB	OUI	OUI	Résident en Suisse avec permis B
PERMIC	NON (voir tableau D-1)	NON (voir tableau D-1)	Résident en Suisse avec permis C

Tableau D – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.36 en valeur absolue. On a décidé de supprimer LAKE_GE et FRANC qui corrélaient fortement avec ALLEM :

	FRANC	LAKE_GE
ALLEM	**0.89** (PSM_I) **0.90** (PSM_II)	**0.60** (PSM_I) **0.68** (PSM_II)
FRANC	1	**0.68** (PSM_I) **0.76** (PSM_II)

En outre les variables corrélant fortement entre elles, on a pris les décisions suivantes:

	PSM_I	PSM_II	décision
Corr(CELIB, MARIE)	**0.70**	**0.74**	supprimer CELIB
Corr(FORM_PRO, FORM_SUP)	**0.73**	**0.66**	supprimer FORM_PRO
Corr(PERMIC, SWISS)	**0.97**	**0.88**	supprimer PERMIC
Corr(SEULE, AD_SEUL)	**0.94**	**0.92**	supprimer AD_SEUL

Synthèse

11 variables différentes ont été choisies tant pour le PSM_I que pour le PSM_II dont 4 se retrouvent dans les deux modèles. Le modèle obtenu avec l'analyse par segmentation comporte 17 GHR pour le PSM_I et 16 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 questionnaire individuel: pondération longitudinale individuelle du PSM_I. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

<table>
<tr><th>split1</th><th>split2</th><th>split3</th><th>split4</th><th>split5</th><th>split6</th><th>split7</th><th>split8</th><th>rhg</th></tr>
<tr><td rowspan="17">SEULE=0
(3761) 73.9%</td><td rowspan="8">HOMME=0
(1897) 80.3%</td><td rowspan="6">agevc_1=0
(1777) 82.2%</td><td rowspan="4">AdrFROM5W5=0
(1647) 83.3%</td><td rowspan="3">NO_FORM=0
(1328) 85.3%</td><td rowspan="2">NO_KID=0
(600) 88.9%</td><td rowspan="2">CENTRAL=0
(547) 90.1%</td><td>EST=0 (471) 91.5%</td><td>1</td></tr>
<tr><td>EST=1 (76) 82.4%</td><td>2</td></tr>
<tr><td></td><td>CENTRAL=1
(53) 79.3%</td><td></td><td>3</td></tr>
<tr><td></td><td>NO_KID=1
(728) 82.5%</td><td></td><td></td><td>4</td></tr>
<tr><td></td><td>NO_FORM=1
(319) 76.4%</td><td></td><td></td><td></td><td>5</td></tr>
<tr><td>AdrFROMW5=1 (130)
67.0%</td><td></td><td></td><td></td><td></td><td></td><td>6</td></tr>
<tr><td rowspan="2">agevc_1=1
(120) 62.0%</td><td>FORM_SUP=0 (64) 48.5%</td><td></td><td></td><td></td><td></td><td></td><td>7</td></tr>
<tr><td>FORM_SUP=1(56) 77.4%</td><td></td><td></td><td></td><td></td><td></td><td>8</td></tr>
<tr><td rowspan="5">HOMME=1
(1864) 67.4%</td><td rowspan="4">FORM_SUP=0
(956) 59.5%</td><td rowspan="3">AdrFROMW5=0
(866) 61.7%</td><td>NO_FORM=0
(691) 65.5%</td><td></td><td></td><td></td><td></td><td>9</td></tr>
<tr><td rowspan="2">NO_FORM=1
(175) 49.5%</td><td>ALLEM=0
(62) 59%</td><td></td><td></td><td></td><td>10</td></tr>
<tr><td>ALLEM=1
(113) 43.8%</td><td></td><td></td><td></td><td>11</td></tr>
<tr><td>AdrFROMW5=1
(90) 38.6%</td><td></td><td></td><td></td><td></td><td>12</td></tr>
<tr><td>FORM_SUP=1 (908)
77.2%</td><td></td><td></td><td></td><td></td><td></td><td>13</td></tr>
<tr><td rowspan="4">SEULE=1
(640) 96.4%</td><td rowspan="3">INC_STAMPS=0
(496) 95.4%</td><td rowspan="2">EST=0
(443) 96.3%</td><td>NO_FORM=0
(361) 97.5%</td><td></td><td></td><td></td><td></td><td>14</td></tr>
<tr><td>NO_FORM=1 (82) 91.7%</td><td></td><td></td><td></td><td></td><td>15</td></tr>
<tr><td>EST=1 (53) 88.9%</td><td></td><td></td><td></td><td></td><td></td><td>16</td></tr>
<tr><td>INC_STAMPS=1
(144) 100.0%</td><td></td><td></td><td></td><td></td><td></td><td></td><td>17</td></tr>
</table>

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.

split1	split2	split3	split4	split5	split6	split7	rhg
SEULE=0 (2834) 68,5%	NO_FORM=0 (2236) 72.6%	HOMME=0 (1069) 77.8%	agevc_3=0 (773) 75.4%	agevc_4=0 (510) 72.8%	REFW2W7CONV=0 (459) 74.1%		1
					REFW2W7CONV=1 (51) 61.2%		2
				agevc_4=1 (263) 80.9%	MARIE=0 (49) 91.2%		3
					MARIE=1 (214) 78.3%	INC_LOTT=0 (172) 82.1%	4
						INC_LOTT=1 (42) 63.0%	5
			agevc_3=1 (296) 84.9%	NOREST=0 (255) 82.5%			6
				NOREST=1 (41) 97.8%			7
		HOMME=1 (1167) 67.8%	FORM_SUP=0 (560) 61.1%	INC_LOTT=0 (426) 63.4%			8
				INC_LOTT=1 (134) 54.1%			9
			FORM_SUP=1 (607) 75%	REFW2W7CONV=0 (566) 76.0%	MARIE=0 (128) 68.8%		10
					MARIE=1 (438) 78.7%		11
				REFW2W7CONV=1 (41) 60.8%			12
	NO_FORM=1 (598) 54.3%						13
SEULE=1 (478) 98.1%	REFW2W7CONV=0 (443) 94.7%	SWISS=0 (42) 84.4%					14
		SWISS=1 (401) 95.8%					15
	REFW2W7CONV=1 (35) 71.9%						16

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.

Niveau ménage

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
FRANC	NON (voir tableau E-1)	NON (voir tableau E-1)	Questionnaire grille ou non réponse rempli en français
ITAL	OUI	OUI	Questionnaire grille ou non réponse rempli en italien
ALLEM	**OUI**	OUI	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	NON (voir tableau E-1)	NON (voir tableau E-1)	Région Lémanique
MIDDLE	OUI	OUI	Région Mitteland
NOREST	OUI	OUI	Région Nord-Ouest
EST	**OUI**	OUI	Région Suisse orientale
CENTRAL	OUI	OUI	Région Suisse centrale
ZURICH	OUI	OUI	Région Zurich
MIXED	**OUI**	OUI	Type de commune : agricole mixte
CENTRES	**OUI**	**OUI**	Type de commune : agglomération
SUBURB	OUI	OUI	Type de commune : banlieue
WEALTH	OUI	OUI	Type de commune : riche
PERIPH	OUI	**OUI**	Type de commune : périphérique
TOURIST	OUI	OUI	Type de commune : touristique
INDUS	OUI	OUI	Type de commune : industrielle ou tertiaire
RURAL	OUI	OUI	Type de commune : cité dortoir
AD_SEUL	OUI	**OUI**	Ménage avec un seul adulte [32]
NO_KID	**OUI**	OUI	Ménage sans enfant
SEULE	**OUI**	OUI	Ménage d'une seule personne
ADRFROMW5	**OUI**		Ménages réinjectés dans l'enquête, vus pour la dernière fois en vague 5 (1) ou autres (0)
REFW2W7CONV	OUI	**OUI**	Ménages non répondants de vague 2 PSM_II et vague 7 PSM_I convertis (1) ou autre (0)
REFW3W8CONV	OUI	**OUI**	Ménages non répondants des vagues en cours, vague 3 PSM_II et vague 8 PSM_I convertis par rappels de refus (1) ou autre (0)
INC_STAMPS	**OUI**	**OUI**	Ménages du lot ayant reçu un carnet de timbre comme *incitatif* (1) ou autre (0)
INC_LOTT	OUI	**OUI**	Ménages du lot ayant participé à la loterie comme *incitatif* (1) ou autre (0)
INC_GIFT	OUI	**OUI**	Ménages du lot pouvant faire un don comme *incitatif* (1) ou autre (0)

[32] AD_SEUL corrèle fortement avec SEULE (voir tableau E-1), mais on a oublié de retirer AD_SEUL de la liste de variables proposée à l'algorithme de segmentation. C'est pourquoi AD_SEUL a été choisie en dépit de SEULE pour le premier embranchement de l'arbre (contrairement aux autres arbres de modélisations de la non-réponse au niveau individuel qui commencent toujours par l'embranchement SEULE).

Niveau individu

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
AGEVC_1	**OUI**	**OUI**	15-24 ans (à la vague courante)
AGEVC_2	**OUI**	OUI	25 à 34 ans (à la vague courante)
AGEVC_3	OUI	**OUI**	35 à 44 ans (à la vague courante)
AGEVC_4	OUI	**OUI**	45 à 54 ans (à la vague courante)
AGEVC_5	**OUI**	OUI	55+ ans (à la vague courante)
CELIB	NON (voir tableau E-1)	NON (voir tableau E-1)	Célibataire
DIVORCE	OUI	**OUI**	Divorcé(e)
SEPARE	OUI	OUI	Séparé(e)
MARIE	**OUI**	**OUI**	Marié(e)
VEUF	OUI	OUI	Veuf / veuve
FORM_PRO	NON (voir tableau E-1)	NON (voir tableau E-1)	Formation professionnelle
FORM_SUP	**OUI**	**OUI**	Formation supérieure
NO_FORM	**OUI**	**OUI**	Pas de formation
HOMME	**OUI**	**OUI**	Homme
SWISS	**OUI**	**OUI**	Nationalité suisse.
TRAVAIL	OUI	OUI	Personne active professionnellement
PERMIB	OUI	OUI	Résident en Suisse avec permis B
PERMIC	NON (voir tableau E-1)	NON (voir tableau E-1)	Résident en Suisse avec permis C

Tableau 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. On a décidé de supprimer LAKE_GE, FRANC, qui corrélaient fortement avec ALLEM, de même, on a supprimé CELIB :

	FRANC	LAKE_GE		MARIE	AGE_VC1
ALLEM	**0.90** (PSM_I) **0.90** (PSM_II)	**0.61** (PSM_I) **0.68** (PSM_II)	**CELIB**	**0.77** (PSM_I) **0.76** (PSM_II)	**0.69** (PSM_I) **0.65** (PSM_II)
FRANC	1	**0.68** (PSM_I) **0.76** (PSM_II)	**MARIE**	1	**0.53** (PSM_I) **0.49** (PSM_II)

En outre les variables corrélant fortement entre elles, on a pris les décisions suivantes :

	PSM_I	PSM_II	décision
Corr(FORM_PRO, FORM_SUP)	**0.60**	**0.62**	supprimer FORM_PRO
Corr(PERMIC, SWISS)	**0.92**	**0.86**	supprimer PERMIC
Corr(SEULE, AD_SEUL)	**0.91**	**0.91**	supprimer AD_SEUL

Synthèse

16 et 17 variables différentes ont été choisies pour respectivement le PSM_I et le PSM_II dont 8 se retrouvent dans les deux modèles. Le modèle obtenu avec l'analyse par segmentation comporte 29 GHR pour le PSM_I et 27 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 E—2 Modèle complet de non-réponse au questionnaire individuel: pondération transversale individuelle du PSM_I, **partie 1**. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

<table>
<tr><th>split1</th><th>split2</th><th>split3</th><th>split4</th><th>split5</th><th>rhg</th></tr>
<tr><td rowspan="15">SEULE=0
(4812)
71.1%</td><td rowspan="9">HOMME=0
(2437)
77.1%</td><td rowspan="6">agevc_1=0
(1942)
80.2%</td><td rowspan="2">SWISS=0
(174)
65.6%</td><td>NO_KID=0
(81) 73.5%</td><td>1</td></tr>
<tr><td>NO_KID=1
(93) 58.5%</td><td>2</td></tr>
<tr><td rowspan="2">SWISS=1
(1768) 82.2%</td><td rowspan="2">AdrFROMW5=0
(1642)
83.3%</td><td>3
4
5
6
7
8
9
10</td></tr>
<tr><td>AdrFROMW5=1
(126) 67.7%</td><td>11</td></tr>
<tr><td rowspan="3">agevc_1=1
(495)
65.4%</td><td rowspan="2">MIXED=0
(442) 62.8%</td><td>FORM_SUP=0
(352)
59.8%</td><td>12
13</td></tr>
<tr><td>FORM_SUP=1
(90) 72.2%</td><td>14</td></tr>
<tr><td>MIXED=1 (53)
87.5%</td><td></td><td>15</td></tr>
<tr><td rowspan="6">HOMME=1
(2375)
65.0%</td><td rowspan="3">FORM_SUP=0
(1365)
58.6%</td><td rowspan="2">AdrFROMW5=0
(1246) 60.7%</td><td>CENTRES=0
(976)
62.8%</td><td>16
17
18</td></tr>
<tr><td>CENTRES=1
(270)
53.3%</td><td>19
20</td></tr>
<tr><td>AdrFROMW5=1
(119) 36.7%</td><td></td><td>21</td></tr>
<tr><td rowspan="2">FORM_SUP=1
(1010)
74.9%</td><td>MARIE=0
(259) 69.2%</td><td></td><td>22</td></tr>
<tr><td>MARIE=1
(751) 77.3%</td><td></td><td>23</td></tr>
<tr><td rowspan="6">SEULE=1
(645)
96.4%</td><td rowspan="4">INC_STAMPS=O
(499)
95.4%</td><td rowspan="3">EST=0
(446)
96.3%</td><td rowspan="2">NO_FORM=0
(363) 97.5%</td><td>DIVORCE=0
(365)
98.4%</td><td>24
25</td></tr>
<tr><td>DIVORCE=1
(98) 94.9%</td><td>26</td></tr>
<tr><td>NO_FORM=1
(83) 91.8%</td><td></td><td>27</td></tr>
<tr><td>EST=1
(53) 88.9%</td><td></td><td></td><td>28</td></tr>
<tr><td colspan="2">INC_STAMPS=1
(146) 100.0%</td><td></td><td></td><td></td><td>29</td></tr>
</table>

Figure E—3 Modèle complet de non-réponse au questionnaire individuel: pondération transversale individuelle du PSM_I, **partie 2**. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

split6	split7	split8	split9	split10	rhg
					1
					2
NO_KID=0 (664) 88.1%	FORM_SUP=0 (401) 86.0%	agevc_2=0 (340) 84.5%			3
		agevc_2=1 (61) 94.0%			4
	FORM_SUP=1 (263) 91.7%				5
NO_KID=1 (978) 80.4%	agevc_2=0 (862) 81.8%	MIXED=0 (802) 80.9%	FORM_SUP=0 (570) 79.2%	INC_STAMPS=0 (434) 77.1%	6
				INC_STAMPS=1 (136) 85.5%	7
			FORM_SUP=1 (232) 86.0%		8
		MIXED=1 (60) 93.3%			9
	agevc_2=1 (116) 70.5%				10
					11
NO_KID=0 (245) 65.4%					12
NO_KID=1 (107) 49.4%					13
					14
					15
ALLEM=0 (258) 71.9%	INC_STAMPS=0 (187) 66.0%				16
	INC_STAMPS=1 (71) 87.7%				17
ALLEM=1 (718) 59.3%					18
NO_FORM=0 (184) 58.7%					19
NO_FORM=1 (86) 43.2%					20
					21
					22
					23
agevc_5=0 (150) 97.2%					24
agevc_5=1 (115) 100.0%					25
					26
					27
					28
					29

split1	split2	split3	split4	split5	split6	split7	split8	split9	split10	rhg
AD_SEUL=0 (3035) 66.9%	NO_FORM=0 (2323) 70.6%	SWISS=0 (274) 56.6%	INC_STAMPS=0 (206) 52.8%							1
			INC_STAMPS=1 (68) 68.5%							2
		SWISS=1 (2049) 72.4%	MARIE=0 (583) 65.3%	agevc_4=0 (514) 63.0%	FORM_SUP=0 (276) 57.8%	HOMME=0 (133) 64.1%				3
						HOMME=1 (143) 52.0%				4
					FORM_SUP=1 (238) 70.1%	INC_GIFT=0 (187) 74.3%				5
						INC_GIFT=1 (51) 55.0%				6
				agevc_4=1 (69) 81.4%	DIVORCE=0 (35) 70.3%					7
					DIVORCE=1 (34) 93.4%					8
			MARIE=1 (1466) 75.3%	REFW3W8CONV=0 (1414) 76.2%	HOMME=0 (667) 80.4%	INC_GIFT=0 (529) 78.2%	agevc_3=0 (359) 74.7%	FORM_SUP=0 (238) 70.9%		9
								FORM_SUP=1 (121) 83.4%		10
							agevc_3=1 (170) 86.3%			11
						INC_GIFT=1 (138) 89.0%				12
					HOMME=1 (747) 72.3%	REFW2W7CONV=0 (690) 74.2%	FORM_SUP=0 (306) 67.5%	INC_LOTT=0 (226) 71.9%	PERIPH=0 (193) 69.3%	13
									PERIPH=1 (33) 89.0%	14
								INC_LOTT=1 (80) 54.9%		15
							FORM_SUP=1 (384) 80.8%	CENTRES=0 (278) 84.0%		16
								CENTRES=1 (106) 73.6%		17
						REFW2W7CONV=1 (57) 50.8%				18
				REFW3W8CONV=1 (52) 51.9%						19
	NO_FORM=1 (712) 55.8%	agevc_1=0 (319) 50.3%								20
		agevc_1=1 (393) 60.6%	HOMME=0 (167) 67.3%							21
			HOMME=1 (226) 55.7%							22
AD_SEUL=1 (565) 91.4%	REFW2W7CONV=0 (519) 93.5%	agevc_1=0 (482) 94.5%	SWISS=0 (50) 83.0%							23
			SWISS=1 (432) 95.8%	HOMME=0 (287) 97.1%						24
				HOMME=1 (145) 93.2%						25
		agevc_1=1 (37) 79.1%								26
	REFW2W7CONV=1 (46) 66.6%									27

Figure E – 4 Modèle complet de non-réponse au questionnaire individuel: pondération transversale individuelle du PSM_II. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

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:

Variable	Utilisée pour PSM_I	Utilisée pour PSM_II	Description
FRANC	NON (voir tableau F-1)	NON (voir tableau F-1)	Questionnaire grille ou non réponse rempli en français
ITAL	OUI	OUI	Questionnaire grille ou non réponse rempli en italien
ALLEM	OUI	OUI	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	NON (voir tableau F-1)	NON (voir tableau F-1)	Région Lémanique
MIDDLE	OUI	OUI	Région Mitteland
NOREST	OUI	OUI	Région Nord-Ouest
EST	OUI	OUI	Région Suisse orientale
CENTRAL	OUI	OUI	Région Suisse centrale
ZURICH	OUI	OUI	Région Zurich
MIXED	OUI	OUI	Type de commune : agricole mixte
CENTRES	OUI	OUI	Type de commune : agglomération
SUBURB	OUI	OUI	Type de commune : banlieue
WEALTH	OUI	OUI	Type de commune : riche
PERIPH	OUI	OUI	Type de commune : périphérique
TOURIST	OUI	OUI	Type de commune : touristique
INDUS	OUI	OUI	Type de commune : industrielle ou tertiaire
RURAL	OUI	OUI	Type de commune : cité dortoir
AD_SEUL	NON (voir tableau F-1)	NON (voir tableau F-1)	Ménage avec un seul adulte
NO_KID	OUI	OUI	Ménage sans enfant
SEULE	OUI	OUI	Ménage d'une seule personne
ADRFROMW5	OUI		Ménages réinjectés dans l'enquête, vus pour la dernière fois en vague 5 (1) ou autres (0)
REFW2W7CONV	OUI	OUI	Ménages non répondants de vague 2 PSM_II et vague 7 PSM_I convertis (1) ou autre (0)
REFW3W8CONV	OUI	OUI	Ménages non répondants des vagues en cours, vague 3 PSM_II et vague 8 PSM_I convertis par rappels de refus (1) ou autre (0)
INC_STAMPS	OUI	OUI	Ménages du lot ayant reçu un carnet de timbre comme *incitatif* (1) ou autre (0)
INC_LOTT	OUI	OUI	Ménages du lot ayant participé à la loterie comme *incitatif* (1) ou autre (0)
INC_GIFT	OUI	OUI	Ménages du lot pouvant faire un don comme *incitatif* (1) ou autre (0)
APPART	OUI	**OUI**	Ménage vivant en appartement
DETACH	NON (voir tableau F-1)	NON (voir tableau F-1)	Ménage vivant dans une maison indépendante
PROPRIO	**OUI**	**OUI**	Propriétaire

RAT1	OUI	OUI	nbr. de pièces / nbr. de personnes <= 1.34
RAT2	OUI	OUI	1.34 < nbr. de pièces / nbr. de personnes < 2
RAT3	**OUI**	OUI	nbr. de pièces / nbr. de personnes > 2
LOY_1	**OUI**	**OUI**	Coût du logement jusqu'à 1000 Fr./mois
LOY_2	**OUI**	**OUI**	Coût du logement entre 1001-1500 Fr./mois
LOY_3	**OUI**	**OUI**	Coût du logement 1501 Fr./mois et plus
tous_hommes	OUI	OUI	Ménage composé uniquement de personnes du sexe masculin
tous_femmes	NON (voir tableau F-1)	NON (voir tableau F-1)	Ménage composé uniquement de personnes du sexe féminin
un_marie	NON (voir tableau F-1)	NON (voir tableau F-1)	Au moins une personne mariée dans le ménage
max_no_educ	OUI	OUI	La personne la plus éduquée du ménage est sans formation
max_educ_sup	OUI	OUI	La personne la plus éduquée du ménage a une formation supérieure
max_educ_pro	NON (voir tableau F-1)	NON (voir tableau F-1)	La personne la plus éduquée du ménage a une formation professionnelle
tous_suis	OUI	**OUI**	Toutes les personnes du ménage sont de nationalité suisse
tous_etrang	NON (voir tableau F-1)	NON (voir tableau F-1)	Toutes les personnes du ménage sont de nationalité étrangère
agemax_1	OUI	OUI	La personne la plus âgée du ménage est dans la classe d'âge 1, i.e. 15-24 ans (à la vague courante)
agemax_2	OUI	OUI	La personne la plus âgée du ménage est dans la classe d'âge 2, i.e. 25 à 34 ans (à la vague courante)
agemax_3	OUI	OUI	La personne la plus âgée du ménage est dans la classe d'âge 3, i.e. 35 à 44 ans (à la vague courante)
agemax_4	OUI	OUI	La personne la plus âgée du ménage est dans la classe d'âge 4, i.e. 45 à 54 ans (à la vague courante)
agemax_5	**OUI**	**OUI**	La personne la plus âgée du ménage est dans la classe d'âge 5, i.e. 55+ ans (à la vague 1)

Synthèse

6 et 7 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 7 GHR pour le PSM_I et 9 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.

Tableau F – 1 Variables dichotomiques créées pour modéliser la non-réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue. On a décidé de supprimer LAKE_GE, FRANC, qui corrélaient fortement avec ALLEM, de même, on a supprimé DETACH, UN_MARIE et TOUS_FEMMES:

	FRANC	LAKE_GE		PROPRIO	DETACH
ALLEM	**0.89** (PSM_I) **0.90** (PSM_II)	**0.61** (PSM_I) **0.67** (PSM_II)	**APPART**	**0.56** (PSM_I) **0.54** (PSM_II)	**0.87** (PSM_I) **0.86** (PSM_II)
FRANC	1	**0.67** (PSM_I) **0.75** (PSM_II)	**PROPRIO**	1	**0.58** (PSM_I) **0.54** (PSM_II)

En outre les variables corrélant fortement entre elles, on a pris les décisions suivantes:

	AD_SEUL	UN_MARIE	TOUS_FEMMES
SEULE	**0.93** (PSM_I) **0.92** (PSM_II)	**0.67** (PSM_I) **0.67** (PSM_II)	**0.49** (PSM_I) **0.51** (PSM_II)
AD_SEUL	1	**0.72** (PSM_I) **0.72** (PSM_II)	**0.55** (PSM_I) **0.57** (PSM_II)
UN_MARIE		1	**0.52** (PSM_I) **0.49** (PSM_II)

	PSM_I	PSM_II	décision
Corr(MAX_EDUC_SUP, MAX_EDUC_PRO)	**0.83**	**0.84**	supprimer MAX_EDUC_PRO
Corr(TOUS_SUIS, TOUS_ETRANG)	**0.74**	**0.77**	supprimer TOUS_ETRANG

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.

split1	split2	split3	split4	split5	split6	rhg
PROPRIO=0 (1356) 96.8%	loy_3=0 (846) 95.1%	loy_1=0 (518) 91.9%	loy_2=0 (268) 84.8%	agemax_5=0 (113) 71.7%		1
				agemax_5=1 (155) 93.0%	RAT3=0 (69) 86.3%	2
					RAT3=1 (86) 98.3%	3
			loy_2=1 (250) 100.0%			4
		loy_1=1 (328) 100.0%				5
	loy_3=1 (510) 100.0%					6
PROPRIO=1 (1224) 100.0%						7

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>split1</th><th>split2</th><th>split3</th><th>split4</th><th>split5</th><th>split6</th><th>split7</th><th>rhg</th></tr>
<tr><td rowspan="8">PROPRIO=0
(852) 92.0%</td><td rowspan="2">tous_suis=0
(96) 72.0%</td><td>loy_3=0
(61) 57.1%</td><td></td><td></td><td></td><td></td><td>1</td></tr>
<tr><td>loy_3=1
(35) 100.0%</td><td></td><td></td><td></td><td></td><td>2</td></tr>
<tr><td rowspan="6">tous_suis=1
(756) 94.4%</td><td rowspan="5">loy_3=0
(481) 91.6%</td><td rowspan="4">loy_1=0
(333)
87.5%</td><td rowspan="3">loy_2=0
(187) 77.7%</td><td rowspan="2">APPART=0
(124) 84.7%</td><td>agemax_5=0
(62) 77.4%</td><td>3</td></tr>
<tr><td>agemax_5=1
(62) 91.6%</td><td>4</td></tr>
<tr><td>APPART=1
(63) 63.2%</td><td></td><td>5</td></tr>
<tr><td>loy_2=1
(146)
100.0%</td><td></td><td></td><td>6</td></tr>
<tr><td>loy_1=1
(148)
100.0%</td><td></td><td></td><td></td><td>7</td></tr>
<tr><td>loy_3=1
(275)
100.0%</td><td></td><td></td><td></td><td></td><td>8</td></tr>
<tr><td>PROPRIO=1
(902) 100.0%</td><td></td><td></td><td></td><td></td><td></td><td></td><td>9</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).

Niveau ménage

Variable	Utilisée pour PSM_I	Description
FRANC	NON (voir tableau G-1)	Questionnaire grille ou non réponse rempli en français
ITAL	OUI	Questionnaire grille ou non réponse rempli en italien
ALLEM	OUI	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	NON (voir tableau G-1)	Région Lémanique
MIDDLE	OUI	Région Mitteland
NOREST	OUI	Région Nord-Ouest
EST	OUI	Région Suisse orientale
CENTRAL	OUI	Région Suisse centrale
ZURICH	OUI	Région Zurich
MIXED	OUI	Type de commune : agricole mixte
CENTRES	OUI	Type de commune : agglomération
SUBURB	OUI	Type de commune : banlieue
WEALTH	OUI	Type de commune : riche
PERIPH	OUI	Type de commune : périphérique
TOURIST	OUI	Type de commune : touristique
INDUS	OUI	Type de commune : industrielle ou tertiaire
RURAL	OUI	Type de commune : cité dortoir
AD_SEUL	NON (voir tableau G-1)	Ménage avec un seul adulte
NO_KID	OUI	Ménage sans enfant
SEULE	OUI	Ménage d'une seule personne
ADRFROMW5	OUI	Ménages réinjectés dans l'enquête, vus pour la dernière fois en vague 5 (1) ou autres (0)
REFW2W7CONV	OUI	Ménages non répondants de vague 2 PSM_II et vague 7 PSM_I convertis (1) ou autre (0)
REFW3W8CONV	OUI	Ménages non répondants des vagues en cours, vague 3 PSM_II et vague 8 PSM_I convertis par rappels de refus (1) ou autre (0)
INC_STAMPS	OUI	Ménages du lot ayant reçu un carnet de timbre comme *incitatif* (1) ou autre (0)
INC_LOTT	OUI	Ménages du lot ayant participé à la loterie comme *incitatif* (1) ou autre (0)
INC_GIFT	OUI	Ménages du lot pouvant faire un don comme *incitatif* (1) ou autre (0)

Niveau individu

Variable	Utilisée pour PSM_I	Description
AGEVC_1	OUI	15-24 ans (à la vague courante)
AGEVC_2	OUI	25 à 34 ans (à la vague courante)
AGEVC_3	OUI	35 à 44 ans (à la vague courante)
AGEVC_4	OUI	45 à 54 ans (à la vague courante)
AGEVC_5	OUI	55+ ans (à la vague courante)
CELIB	NON (voir tableau G-1)	Célibataire
DIVORCE	OUI	Divorcé(e)
SEPARE	OUI	Séparé(e)
MARIE	OUI	Marié(e)
VEUF	OUI	Veuf / veuve
FORM_PRO	NON (voir tableau G-1)	Formation professionnelle
FORM_SUP	OUI	Formation supérieure
NO_FORM	OUI	Pas de formation
HOMME	OUI	Homme
SWISS	OUI	Nationalité suisse.
TRAVAIL	OUI	Personne active professionnellement
PERMIB	NON, trop peu de cas (9)	Résident en Suisse avec permis B
PERMIC	NON (voir tableau G-1)	Résident en Suisse avec permis C

Tableau G – 1 Variables dichotomiques créées pour modéliser la non-réponse ayant des coefficients de corrélations de Pearson > 0.50 en valeur absolue. On a décidé de supprimer LAKE_GE et FRANC qui corrélaient fortement avec ALLEM:

	FRANC	LAKE_GE
ALLEM	**0.89** (PSM_I)	**0.61** (PSM_I)
FRANC	1	**0.68** (PSM_I)

En outre les variables corrélant fortement entre elles, on a pris les décisions suivantes :

	PSM_I	décision
Corr(CELIB, MARIE)	**0.76**	supprimer CELIB
Corr(CELIB, AGE_VC1)	**0.67**	supprimer CELIB
Corr(FORM_PRO, FORM_SUP)	**0.63**	supprimer FORM_PRO
Corr(PERMIC, SWISS)	**0.97**	supprimer PERMIC
Corr(SEULE, AD_SEUL)	**0.92**	supprimer AD_SEUL

Modélisation de la non-réponse – analyse par segmentation

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.

Figure G—2 Modèle complet de non-réponse au questionnaire individuel pour les longitudinaux du PSM_I (pondération longitudinale individuelle du PSM_I & PSM_II combinés). Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

split1	split2	split3	split4	split5	split6	split7	split8	rhg
SEULE=0 (4229) 73.3%	HOMME=0 (2130) 79.5%	agevc_1=0 (1776) 82.3%	AdrFROMW5=0 (1646) 83.5%	NO_FORM=0 (1327) 85.4%	NO_KID=0 (600) 88.9%	CENTRAL=0 (547) 90.1%	EST=0 (471) 91.5%	1
							EST=1 (76) 82.4%	2
						CENTRAL=1 (53) 79.3%		3
					NO_KID=1 (727) 82.7%			4
				NO_FORM=1 (319) 76.4%				5
			AdrFROMW5=1 (130) 67.0%					6
		agevc_1=1 (354) 66.8%	FORM_SUP=0 (269) 62.9%	NO_KID=0 (168) 70.5%				7
				NO_KID=1 (101) 53.8%				8
			FORM_SUP=1 (85) 76.4%					9
	HOMME=1 (2099) 66.9%	FORM_SUP=0 (1175) 59.7%	AdrFROMW5=0 (1070) 61.9%	NO_FORM=0 (705) 65.5%				10
				NO_FORM=1 (365) 55.0%	CENTRES=0 (295) 58.9%	ALLEM=0 (83) 72.8%		11
						ALLEM=1 (212) 52.7%		12
					CENTRES=1 (70) 40.1%			13
			AdrFROMW5=1 (105) 38.0%					14
		FORM_SUP=1 (924) 77.3%						15
SEULE=1 (645) 96.4%	INC_STAMPS=0 (499) 95.4%	EST=0 (446) 96.3%	NO_FORM=0 (363) 97.5%					16
			NO_FORM=1 (83) 91.8%					17
		EST=1 (53) 88.9%						18
	INC_STAMPS=1 (146) 100.0%							19

ANNEXE H – Caractéristiques des distributions des pondérations produites

Tableau H – 1 Pondération individuelle longitudinale PSM_I vague 8 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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	POIDTRIN	PL_NRQI	POIDLONG
N	4088	4088	3617	3617
Somme des poids	6462474	6196348	5771759	5998212
Moyenne	1581	1516	1595	1658
Écart-type	678	1023	699	968
Variance	459568	1047053	488682	937888
Entendue	5993	10861	5416	9741
Intervalle interquartile	751	716	839	1034
Rapport Max/Min	32	80	13	27
Quantiles				
100% Max	6185	10999	5875	10121
99%	3840	5705	4128	5214
95%	2926	3620	2867	3580
90%	2482	2521	2484	2713
75% Q3	1863	1664	1945	2020
50% Médiane	1408	1207	1428	1439
25% Q1	1112	948	1106	986
10%	943	790	929	795
5%	836	679	713	694
1%	599	500	501	490
0% Min	192	138	460	381

Tableau H – 2 Pondération transversale ménage PSM_I vague 8 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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 99^{e} percentiles)

Nom du jeu de poids	P_NRGRIL	PTM_NRQM	POIDTRME_INT1	POIDTRME_INT	POIDTRME_INT_WIN
N	2028	2537	2537	2537	2537
Somme des poids	2587667	3140124	3056981	3053584	3064134
Moyenne	1276	1238	1205	1204	1208
Écart-type	399	395	730	730	708
Variance	159500	155922	532354	532847	500612
Entendue	2110	2819	7625	7286	4088
Intervalle interquartile	625	573	631	633	637
Rapport Max/Min	4	17		223	26
Quantiles					
100% Max	2742	2995	7321	7318	4254
99%	2301	2269	4079	4107	4138
95%	2025	1975	2596	2593	2613
90%	1884	1807	1994	1985	2000
75% Q3	1576	1511	1390	1390	1400
50% Médiane	1168	1166	1018	1013	1021
25% Q1	951	938	759	757	763
10%	826	797	583	580	584
5%	780	719	474	472	476
1%	704	478	266	268	270
0% Min	632	176	**-304**	32	166

Tableau H – 3 Pondération individuelle longitudinale PSM_II vague 3 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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	POIDTRIN	PL_NRQI	POIDLONG
N	2568	2568	2453	2453
Somme des poids	7158158	6196348	6953021	6080866
Moyenne	2787	2413	2834	2479
Écart-type	1125	1295	1062	1203
Variance	1265490	1677224	1127007	1448394
Entendue	7794	12451	6894	11242
Intervalle interquartile	1429	1331	1427	1370
Rapport Max/Min	12	25	8	18
Quantiles				
100% Max	8503	12966	7874	11886
99%	6420	6917	6101	6537
95%	4915	5017	4795	4979
90%	4250	4143	4261	4052
75% Q3	3387	2877	3456	2990
50% Médiane	2565	2027	2612	2179
25% Q1	1958	1546	2029	1620
10%	1596	1230	1703	1305
5%	1434	1104	1566	1191
1%	1028	792	1260	958
0% Min	710	515	980	644

Tableau H – 4 Pondération transversale ménage PSM_II vague 3 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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	P_NRGRIL	PTM_NRQM	POIDTRME_INT
N	1660	1431	1684	1684
Somme des poids	2094703	3023814	3543600	3183038
Moyenne	1262	2113	2104	1890
Écart-type	278	638	749	856
Variance	77299	407126	560613	732642
Entendue	1466	3745	6118	6422
Intervalle interquartile	252	775	817	1007
Rapport Max/Min	3	5	13	30
Quantiles				
100% Max	2436	4740	6627	6644
99%	2147	4078	4375	4735
95%	1832	3441	3555	3500
90%	1749	3030	3141	2990
75% Q3	1339	2406	2443	2281
50% Médiane	1173	1991	1971	1713
25% Q1	1087	1631	1626	1275
10%	1015	1447	1352	1038
5%	1000	1347	1122	890
1%	980	1205	749	635
0% Min	970	995	510	222

Tableau H – 5 Pondération individuelle longitudinale PSM_I et PSM_II combinés vague 3 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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	PLI_COMBPAN_REG	PLI_COMBPAN_CAL
N	6244	6244	6244
Somme des poids	13'146'565	6774804	6080866
Moyenne	2105	1085	974
Écart-type	1022	520	641
Variance	1044813	270481	411087
Entendue	7399	5839	9870
Intervalle interquartile	1236	538	513
Rapport Max/Min	16	21	49
Quantiles			
100% Max	7874	6130	10075
99%	5417	3095	3417
95%	4120	1975	2148
90%	3553	1660	1609
75% Q3	2579	1283	1110
50% Médiane	1873	963	805
25% Q1	1343	745	597
10%	1045	624	486
5%	972	568	428
1%	768	440	341
0% Min	476	291	205

Tableau H – 6 Pondération individuelle transversale PSM_I et PSM_II combinés vague 3 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
PTI_NRQI_COMB: poids corrigé pour la non-réponse au quest. individuel
PTI_COMBPAN_REG: poids individuel ajusté pour la non-réponse, combiné au niveau des grandes régions
PTI_COMBREG_CAL: poids transversal individuel final

Nom du jeu de poids	PTI_NRQI_COMB	PTI_COMBPAN_REG	PTI_COMBREG_CAL
N	6656	6656	6656
Somme des poids	13'620'632	6744736	6196349
Moyenne	2046	1013	931
Écart-type	1056	430	575
Variance	1115407	184788	331128
Entendue	8311	3487	6104
Intervalle interquartile	1292	512	479
Rapport Max/Min	44	30	73
Quantiles			
100% Max	8503	3606	6189
99%	5550	2427	3152
95%	4116	1844	2085
90%	3491	1590	1613
75% Q3	2556	1216	1059
50% Médiane	1761	907	761
25% Q1	1264	703	579
10%	1014	597	479
5%	896	526	420
1%	681	386	319
0% Min	192	119	85

Tableau H – 7 Pondération transversale ménage PSM_I et PSM_II combinés vague 3 (2006). Pour chaque jeu de poids, on a calculé toutes les statistiques sur le sous-ensemble des poids > 0.
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 final

Nom du jeu de poids	PTM_NRQM_COMB	PTM_COMBPAN_REG	PTM_COMBREG_CAL2
N	4221	4221	4221
Somme des poids	6683724	3305120	3111109
Moyenne	1583	783	737
Écart-type	705	267	389
Variance	497383	71404	151108
Entendue	6451	2529	3646
Intervalle interquartile	862	326	399
Rapport Max/Min	38	26	90
Quantiles			
100% Max	6627	2632	3687
99%	3945	1588	2135
95%	2964	1283	1500
90%	2458	1140	1214
75% Q3	1927	926	882
50% Médiane	1475	738	640
25% Q1	1065	600	483
10%	841	494	377
5%	778	441	312
1%	545	298	222
0% Min	176	103	41

ANNEXE I – Taux de réponse

Figure I – 1 PSM_I vague 8.

Echantillon brut	**3207**	**100%**				
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	33	1%				
Hors échantillon (numéro invalide, modem, fax, décédé)	134	4%				
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	5	0%				
Adresses éligibles (statut d'éligibilité connu)	3035	95%	**100%**			
	Rendez-vous grille en cours		66	2%		
	Refus, problèmes niveau grille		390	13%		
	Grilles réalisées		**2579**	**85%**	**100%**	
			Refus, problèmes niveau ménage		45	2%
			Grille + quest. ménage réalisés		**2534**	**98%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel		5331	100%
	Rendez-vous individuel en cours	20	0%
	Refus, problèmes niveau individuel	1218	23%
	Questionnaires individuel complétés	**4093**	**77%**

Figure I – 2 PSM_II vague 3.

Echantillon brut	**2555**	**100%**				
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	40	2%				
Hors échantillon (numéro invalide, modem, fax, décédé)	130	5%				
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	6	0%				
Adresses éligibles (statut d'éligibilité connu)	2379	93%		**100%**		
	Rendez-vous grille en cours		117	5%		
	Refus, problèmes niveau grille		510	21%		
	Grilles réalisées		**1752**	**74%**		**100%**
			Refus, problèmes niveau ménage		70	4%
			Grille + quest. ménage réalisés		**1682**	**96%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel		3499	100%
	Rendez-vous individuel en cours	23	1%
	Refus, problèmes niveau individuel	906	26%
	Questionnaires individuel complétés	**2570**	**73%**

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_CH06[35]=-5 alors PRES_P1=1 (ont toujours habité en Suisse)
- si 0 < HAB_CH06 < 1999 alors PRES_P1=1 (habitaient déjà en Suisse en 1999)
- si HAB_CH06 > 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_CH06 < 2004 alors PRES_P2=1 (habitaient déjà en Suisse en 2004)
- si HAB_CH06 > 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] Oridinal Sample Member, voir paragraphe 3.3.2.
[34] Code réponse à la vague 6, voir paragraphe 6.3.
[35] HAB_CH06: 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. (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 de ce 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 ainsi que les effets des incitatifs, 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. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite séparément les pondérations au niveau ménage de celles au niveau individuel. On relève les spécificités des pondérations transversales par rapport aux pondérations longitudinales. Les dernières sections concernent les pondérations combinées du PSM_I vague 8 et PSM_II vague 3. Finalement, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes qui ne sont mentionnées que brièvement dans le texte principal.

N° de commande: 338-0054
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 978-3-303-00412-8