

2018–
2023

00

Bases
statistiques
et généralités

Neuchâtel 2024

Enquêtes thématiques 2018–2023

Plans de sondage, tirages, pondération et estimation de variance

Rapports de méthodes de la section Méthodes statistiques

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.

Enquêtes thématiques 2018–2023

Plans de sondage, tirages, pondération et estimation de variance

Rédaction Djalel-Eddine Meskaldji, OFS
Christian Panchard, OFS
Lionel Qualité, OFS
Jacques Saliba, OFS

Éditeur Office fédéral de la statistique (OFS)

Neuchâtel 2024

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Djalel-Eddine Meskaldji, tél. +41 58 463 02 38
djalel-eddine.meskaldji@bfs.admin.ch

Rédaction: Djalel-Eddine Meskaldji, OFS
Christian Panchard, OFS
Lionel Qualité, OFS
Jacques Saliba, OFS

Contenu: Section méthodes statistiques, OFS

Série: Statistique de la Suisse

Domaine: 00 Bases statistiques et généralités

Langue du texte original: français

Mise en page: Publishing et diffusion PUB, OFS
Ce document a été produit automatiquement à partir de banques de données. Il ne répond donc pas aux normes typographiques des publications de l'OFS.

En ligne: www.statistique.ch

Imprimés: www.statistique.ch
Office fédéral de la statistique, CH-2010 Neuchâtel
order@bfs.admin.ch, tél. +41 58 463 60 60
Impression réalisée en Suisse

Copyright: OFS, Neuchâtel 2024
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.

Numéro OFS: 338-0090

ISBN: 978-3-303-00762-4

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Enquêtes thématiques	8
2.1 Microrecensement Formation de Base et Formation Continue - MZB	8
2.2 Enquête suisse sur la Santé - ESS	8
2.3 Enquête sur les Familles et les Générations - EFG	8
2.4 Enquête sur la Langue, la Religion et la Culture - ELRC	8
2.5 Microrecensement Mobilité et Transports - MRMT	9
3 Plan et tirage	10
3.1 Définition de la population cible	10
3.2 Tailles nettes, stratification et allocation de l'échantillon OFS	10
3.3 Densification de l'échantillon OFS	11
3.4 Tailles brutes et estimation des taux de réponse	11
3.5 Réserves et fractionnement en plusieurs vagues	11
3.6 Tirage coordonné et tailles d'échantillon effectives	12
3.7 Plan et tirage MZB 2021	12
3.7.1 Définition de la population cible	12
3.7.2 Tailles nettes, stratification et allocation de l'échantillon OFS	12
3.7.3 Densification de l'échantillon OFS	12
3.7.4 Tailles brutes et estimation des taux de réponse	13
3.7.5 Réserves et fractionnement en plusieurs vagues	13
3.7.6 Tirage coordonné et tailles d'échantillon effectives	13
3.7.7 Table récapitulative	14
3.8 Plan et tirage ESS 2022	15
3.8.1 Définition de la population cible	15
3.8.2 Tailles nettes, stratification et allocation de l'échantillon OFS	15
3.8.3 Densification de l'échantillon OFS	15
3.8.4 Tailles brutes et estimation des taux de réponse	16
3.8.5 Réserves et fractionnement en plusieurs vagues	17
3.8.6 Tirage coordonné et tailles d'échantillon effectives	17
3.8.7 Tables récapitulatives	18
3.9 Plan et tirage EFG 2023	20
3.9.1 Définition de la population cible	20
3.9.2 Tailles nettes, stratification et allocation de l'échantillon OFS	20
3.9.3 Densification de l'échantillon OFS	21
3.9.4 Tailles brutes et estimation des taux de réponse	21
3.9.5 Réserves et fractionnement en plusieurs vagues	21
3.9.6 Tirage coordonné et tailles d'échantillon effectives	21
3.9.7 Table récapitulative	22
3.10 Plan et tirage ELRC 2019	23
3.10.1 Définition de la population cible	23
3.10.2 Tailles nettes, stratification et allocation de l'échantillon OFS	23
3.10.3 Densification de l'échantillon OFS	23
3.10.4 Tailles brutes et estimation des taux de réponse	23
3.10.5 Réserves et fractionnement en plusieurs vagues	24
3.10.6 Tirage coordonné et tailles d'échantillon effectives	24
3.10.7 Tables récapitulatives	24
3.11 Plan et tirage MRMT 2021	26
3.11.1 Définition de la population cible	26
3.11.2 Tailles nettes, stratification et allocation de l'échantillon OFS	26
3.11.3 Densification de l'échantillon OFS	26
3.11.4 Tailles brutes et estimation des taux de réponse	26

3.11.5 Réserves et fractionnement en plusieurs vagues	26
3.11.6 Tirage coordonné et tailles d'échantillon effectives	27
3.11.7 Table récapitulative	27
4 Pondération	28
4.1 Pondération d'une enquête en une seule vague	28
4.2 Estimation des probabilités de réponse	29
4.3 Deuxième phase d'enquête	29
4.4 Discussion sur le calage	30
4.5 Pondération d'une enquête sélectionnée en plusieurs vagues	30
4.5.1 Coefficients de combinaison des vagues	31
4.5.2 Estimation des probabilités de réponse	32
4.5.3 Calage séparé pour chaque vague et/ou calage des poids finaux	32
4.5.4 Étapes de pondération pour une enquête en plusieurs vagues	32
5 Pondération des enquêtes thématiques	34
5.1 Pondération MZB 2021	34
5.2 Pondération ESS 2022	34
5.3 Pondérations EFG 2018 et 2023	36
5.4 Pondération ELRC 2019	39
5.5 Pondération MRMT 2021	40
6 Estimation de variance	42
6.1 Précision d'estimation d'un total pour une enquête en plusieurs vagues	42
6.2 Linéarisation pour l'estimation d'un total/une moyenne global(e) ou sur un domaine	44
6.3 Prise en considération du calage	44
6.4 Précision pour les estimateurs de l'enquête MRMT 2021	46
7 Conclusions	48
Annexe	49
A Notations	49
B Tables récapitulatives des deux échantillons supplémentaires de l'OFSP pour l'enquête ESS 2022	51
C Distribution des poids des pondérations des enquêtes thématiques	55
D Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage	60
E Variance d'un estimateur calé	64
E.1 Linéarisation d'un estimateur calé	64
E.2 Application aux fonctions de calage usuelles	64
Bibliographie	66
Rapports de méthodes de la section méthodes statistiques de l'OFS	67

Préambule

Les enquêtes thématiques sont l'un des composants du système de recensement de la population en Suisse introduit en 2010 par l'Office fédéral de la statistique (OFS). Le recensement traditionnel est depuis lors remplacé par une utilisation des registres complétée par l'enquête annuelle du relevé structurel, les enquêtes thématiques sur lesquelles portent le présent rapport et des enquêtes omnibus.

Avec des échantillons de 10 000 à 40 000 personnes, les enquêtes thématiques permettent, chaque année, d'approfondir l'un des thèmes du relevé structurel ou de relever des informations sur l'état de santé de la population.

La section des méthodes statistiques (METH) de l'OFS prépare les plans de sondage des enquêtes thématiques, en collaboration avec les sections responsables de celles-ci. Elle sélectionne les échantillons et calcule les poids d'extrapolations une fois l'enquête réalisée. Elle fournit enfin les outils permettant d'estimer la précision des résultats obtenus.

Ce rapport décrit l'ensemble de ces travaux pour les années 2018 à 2023. Au cours de cette période, les enquêtes thématiques ont connu des évolutions notables : certaines sont passées de tirages trimestriels à un seul tirage annuel, d'autres sont passées d'une collecte par entretien téléphonique à un questionnaire électronique auto-administré. Enfin, des progrès qualitatifs ont été réalisés en utilisant, lors de la pondération, des données administratives nouvellement disponibles, telles les revenus assurés enregistrés par les caisses de compensation AVS.

Les auteurs remercient les relecteurs de ce document qui ont grandement contribué à l'améliorer : D. Assoulin, D. Kilchmann, J.-P. Renfer (METH), E. Von Erlach (BILD-S), L. Audergon, F. Helfer, A. Mosimann, F. Müller (DEM), M. Storni (GESB) et J.-L. Muralti (MOBIL).

Résumé

Ce rapport présente les méthodes statistiques appliquées par l'Office fédéral de la statistique (OFS) pour les cinq enquêtes thématiques menées entre 2018 et 2023 auprès des personnes et des ménages dans le cadre du système de recensement de la population suisse : l'Enquête sur les Familles et les Générations (EFG 2018 et 2023), l'Enquête sur la Langue, la Religion et la Culture (ELRC 2019), le Microrecensement Formation de Base et Formation Continue (MZB 2021), le Microrecensement Mobilité et Transports (MRMT 2021) et l'Enquête Suisse sur la Santé (ESS 2022).

Le rapport détaille les plans de sondage, la sélection et la préparation des échantillons, ainsi que le calcul des poids d'extrapolation. Pour ces derniers, il décrit les différentes étapes, comme la modélisation de la non-réponse et le calage, en prenant en compte les spécificités propres à chaque enquête. Une attention particulière est accordée à l'estimation de la variance, notamment pour les enquêtes en plusieurs vagues.

Les enquêtes thématiques, bien que largement homogènes, évoluent régulièrement grâce à l'intégration de nouvelles techniques, comme les questionnaires électroniques et à des améliorations méthodologiques, telles que l'utilisation de nouvelles variables auxiliaires (par exemple, les données de l'AVS) pour améliorer la précision des estimations. Ce rapport fournit ainsi une base des méthodes statistiques visant à standardiser dans la mesure du possible ces dernières pour les enquêtes thématiques et dont pourraient bénéficier d'autres enquêtes de l'OFS.

1 Introduction

Ce rapport présente les aspects méthodologiques de la sélection des échantillons et de leur préparation pour produire les résultats des cinq enquêtes thématiques menées par l'Office fédéral de la statistique (OFS) dans le cadre du système de recensement de la population suisse : le Microrecensement Formation de Base et Formation Continue (MZB), l'Enquête Suisse sur la Santé (ESS), l'Enquête sur les Familles et les Générations (EFG), l'Enquête sur la Langue, la Religion et la Culture (ELRC) et le Microrecensement Mobilité et Transports (MRMT). Le calcul de la précision des estimations basées sur ces enquêtes est également abordé. Ces cinq enquêtes sont réalisées en alternance, au rythme d'une enquête par an, et permettent d'approfondir les résultats du recensement de la population sur les thèmes concernés.

La section 2 donne une présentation succincte des différentes enquêtes thématiques. La section 3 décrit de manière générale le processus d'élaboration du plan de sondage des échantillons des enquêtes thématiques, puis précise les spécificités de chaque enquête.

Dans la section 4, le principe général utilisé pour le calcul des poids d'extrapolation des enquêtes thématiques est décrit. Y sont abordés entre autres la modélisation de la non-réponse et le calage des poids, avec le choix des variables auxiliaires utilisées pour ces étapes, mais aussi la combinaison des différents sous-échantillons pour des enquêtes sélectionnées en plusieurs vagues. La section 5 précise les options retenues pour chacune des enquêtes thématiques, en lien avec leurs spécificités.

La section 6 se concentre sur l'estimation de la variance, pour un estimateur calé et pour une enquête en plusieurs vagues, adaptée aux enquêtes thématiques. Une table des notations figure dans l'annexe A.

2 Enquêtes thématiques

Chaque année l'OFS réalise l'une des cinq enquêtes thématiques auprès des personnes et des ménages en alternant les thèmes d'enquête.

2.1 Microrecensement Formation de Base et Formation Continue - MZB

«Le microrecensement formation de base et formation continue fournit des informations sur les comportements de la population suisse dans le domaine de la formation. Les données collectées servent de base à l'observation statistique du secteur de la formation, à la préparation des décisions politiques dans ce domaine et à l'élaboration d'analyses approfondies.» ([Fiche signalétique, MZB](#)). Cette enquête a lieu tous les cinq ans depuis 2011. En 2011 et 2016, l'enquête était réalisée sous la forme d'entretiens téléphoniques assistés par ordinateur (CATI). Depuis 2021, le mode de collecte de données est mixte. Il consiste en des entretiens en ligne (CAWI) et des entretiens téléphoniques assistés par ordinateur.

2.2 Enquête suisse sur la Santé - ESS

«L'enquête suisse sur la santé livre des informations sur l'état de santé de la population et ses facteurs déterminants, sur les conséquences de maladies et sur l'utilisation des services de santé. Sa périodicité permet d'observer les évolutions dans ce domaine et d'étudier les effets des mesures de politique sanitaire.» ([Fiche signalétique, ESS](#)). Elle a lieu tous les cinq ans et est disponible depuis 1992. Les données sont collectées lors d'entretiens téléphoniques suivis par un questionnaire écrit.

2.3 Enquête sur les Familles et les Générations - EFG

«L'enquête sur les familles et les générations (EFG) a pour objectif de fournir des données sur l'état actuel et l'évolution des familles et plus largement sur les relations entre les générations. Les données collectées servent de base pour l'observation des changements et pour des analyses approfondies. Elles permettent de mieux orienter la politique familiale et la politique des générations en Suisse.» ([Fiche signalétique, EFG](#)). Cette enquête thématique auprès des personnes et des ménages a lieu tous les cinq ans depuis 2013. Depuis 2023, il s'agit d'un relevé en mode mixte, par questionnaire en ligne sur internet ou entretien téléphonique. En 2013 et 2018, l'enquête se déroulait par entretien téléphonique suivi d'un questionnaire en ligne ou écrit.

2.4 Enquête sur la Langue, la Religion et la Culture - ELRC

«L'enquête sur la langue, la religion et la culture (ELRC) a pour but de fournir des informations statistiques sur l'état actuel des pratiques linguistiques, religieuses ou spirituelles et culturelles ainsi que sur leurs évolutions en Suisse. Les données collectées servent de base à l'observation des changements et à des analyses approfondies. Elles contribuent à mieux orienter l'action politique en matière de plurilinguisme, de diversité culturelle et d'intégration en Suisse.» ([Fiche signalétique, ELRC](#)). L'enquête a lieu tous les cinq ans depuis 2014. Depuis 2024, il s'agit d'une collecte en mode mixte, en ligne ou par téléphone auprès de deux échantillons séparés. L'un des échantillons reçoit le questionnaire portant sur le thème langue et religion (ELR), l'autre reçoit le questionnaire portant sur le thème de la culture (EC). En 2014 et 2019, l'enquête se déroulait par entretien téléphonique suivi d'un questionnaire en ligne ou écrit.

2.5 Microrecensement Mobilité et Transports - MRMT

«Le microrecensement mobilité et transports fournit des informations sur le comportement de la population en matière de mobilité : possession de véhicules, de permis de conduire et d'abonnements de transports publics, intensité journalière et motifs des déplacements, moyens de transport utilisés. Les résultats fournissent une image détaillée du trafic de personnes en Suisse. Ils servent de référence statistique pour l'élaboration et le contrôle de mesures politiques et constituent des données de base pour des analyses approfondies de l'évolution des transports.» ([Fiche signalétique, MRMT](#)). Cette enquête a lieu tous les cinq ans et ses résultats sont disponibles depuis 1974. Les informations sont relevées au cours d'entretiens téléphoniques.

3 Plan et tirage

Cette partie décrit tout d'abord de manière générale le processus d'élaboration du plan de sondage et le processus de tirage des enquêtes thématiques (sections 3.1 à 3.6), puis précise les spécificités de chacune de ces enquêtes thématiques à partir de la section 3.7.

3.1 Définition de la population cible

La population cible des enquêtes thématiques est constituée des personnes résidentes permanentes de la Suisse, à leur domicile principal et vivant dans un ménage privé, à l'exclusion :

- des demandeurs d'asile ;
- des personnes sous tutelle / curatelle ;
- des fonctionnaires internationaux ;
- des personnes ayant des permis de séjour de courte durée ;
- et des personnes n'ayant dans les registres qu'une adresse de référence hors de Suisse.

D'autre part, selon les besoins des différentes enquêtes thématiques, cette population cible peut encore être restreinte à certaines classes d'âge.

Pour l'élaboration du plan, la population cible prise en considération, de taille N , est généralement celle du dernier cadre de sondage SRPH (Stichprobenrahmen für Personen- und Haushaltserhebungen) disponible, ce cadre étant construit avec les données trimestrielles provenant de registres communaux, cantonaux et fédéraux, qui sont aussi utilisées pour la Statistique de la Population et des Ménages STATPOP.

3.2 Tailles nettes, stratification et allocation de l'échantillon OFS

La taille nette r de l'échantillon OFS d'une enquête thématique est déterminée dans l'annexe à l'Ordonnance du 30 juin 1993 sur les relevés statistiques¹.

La stratification de ces enquêtes est généralement conçue sur la base de dimensions géographiques comme les grandes régions, les cantons, les agglomérations et les villes. D'autres dimensions, comme des caractéristiques de ménage, peuvent aussi y être intégrées.

La taille nette r de l'échantillon OFS est généralement allouée aux strates proportionnellement à la taille de ces dernières. On a alors la taille nette par strate

$$r_h = r \cdot \frac{N_h}{N},$$

où $h = 1, 2, \dots, H$ désigne les strates et N_h les tailles de la population cible de ces strates.

De plus, si les grandes régions font partie de la stratification, la taille de population cible N_h de la grande région du Tessin est ramenée à celle de la plus petite des autres grandes régions, c'est-à-dire la grande région de la Suisse centrale, ceci afin d'assurer que la taille de son échantillon soit aussi grande que la taille de l'échantillon de cette dernière.

1. [RS 431.012.1](#).

3.3 Densification de l'échantillon OFS

Les cantons et les régions officielles peuvent demander une densification de leur échantillon OFS. Les tailles nettes de ces densifications, par strate, sont notées r_h^+ et les tailles nettes de l'échantillon OFS augmenté des densifications $r_h^0 = r_h + r_h^+$.

3.4 Tailles brutes et estimation des taux de réponse

Pour obtenir des tailles brutes n_h^0 à partir des tailles nettes r_h^0 , des taux de réponse u_h^0 sont estimés au niveau des strates.

À partir de ces taux de réponse, on peut alors calculer :

$$n_h^0 = \frac{r_h^0}{u_h^0}.$$

De manière générale les taux de réponse u_h^0 sont estimés à partir du cadre de sondage SRPH le plus récent, ainsi que d'une ou de plusieurs enquêtes thématiques récemment réalisées et en considérant :

- les taux t^{TEL} des personnes du cadre de sondage pour lesquelles au moins un numéro de téléphone (raccordement fixe) a été trouvé, généralement au niveau des strates ;
- les taux de réponse tr^{TEL} des personnes avec au moins un numéro de téléphone, généralement au niveau des strates ;
- les taux de réponse tr^{ALTEL} des personnes sans numéro de téléphone, généralement au niveau des strates ou au niveau des grandes régions ;

puis en calculant :

$$u_h^0 = [t^{TEL} \cdot tr^{TEL}] + [(1 - t^{TEL}) \cdot tr^{ALTEL}].$$

Pour obtenir des estimations des taux de réponse adéquats, il est tenu compte, dans le choix des enquêtes thématiques considérées, du mode de leur questionnaire.

3.5 Réserves et fractionnement en plusieurs vagues

Afin de pouvoir activer² plus d'observations si les taux de réponse prévus u_h^0 s'avèrent surestimés, des réserves, de taille brute l_h^0 , sont ajoutées aux tailles brutes n_h^0 . Le choix du nombre de réserves et de leurs tailles est spécifique à chaque enquête. Si, lors de la réalisation de l'enquête, il apparaît que les tailles nettes escomptées ne seront probablement pas atteintes, des réserves sont activées. Cette activation peut être effectuée au niveau des strates.

Pour les enquêtes qui se déroulent sur une longue période, comme par exemple une année, les tailles brutes peuvent être fractionnées en plusieurs vagues, de manière à pouvoir effectuer

2. Dans ce document, lorsqu'on parle d'activation d'observations ou d'activer des observations, on fait référence aux observations effectivement tirées dans les différents échantillons, ce qui n'exclut donc pas les cas particuliers comme celui des observations exclues de l'enquête suite à leur participation à une autre enquête de l'OFS, ou celui, dans le cas de fractionnement en plusieurs vague, des observations exclues de l'enquêtes suite à un retraitage de la même observation dans deux vagues différentes.

des tirages dans les cadres de sondage les plus récents par rapport aux différentes périodes d'interview prévues.

3.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage d'un échantillon de taille égale à la totalité des tailles brutes de l'enquête, tailles brutes des réserves comprises, est effectué avec le système de coordination des tirages du SRPH sous un plan de Poisson. Cet échantillon est ensuite réparti, à nouveau sous un plan de Poisson, en plusieurs parties correspondant aux différents sous-échantillons et différentes réserves, selon leur propre taille brute. Pour les enquêtes fractionnées en vagues, les différents tirages des vagues sont coordonnés négativement. Comme le tirage de l'échantillon et sa répartition sont réalisés sous des plans de Poisson, les tailles brutes obtenues ne sont pas exactement égales aux tailles brutes prévues.

Le tirage est effectué dans le SRPH dont la date de référence est la plus récente possible par rapport à la période d'interview. Ce cadre est généralement plus récent que celui qui est pris en considération pour l'élaboration du plan. Dans le cas des enquêtes fractionnées en vagues, c'est le SRPH de tirage de la première vague qui est généralement plus récent que celui qui est pris en considération pour le plan, alors que les SRPH de tirage des vagues suivantes le sont toujours.

3.7 Plan et tirage MZB 2021

3.7.1 Définition de la population cible

La population cible de l'enquête MZB 2021 est restreinte aux personnes âgées de 15 à 74 ans, les âges étant calculés au 31 décembre 2020.

La taille de la population cible prise en considération pour le plan de sondage et qui correspond au SRPH avec la date de référence du 30 septembre 2020, créé le 10 novembre 2020, est alors $N = 6\,392\,973$.

3.7.2 Tailles nettes, stratification et allocation de l'échantillon OFS

La taille nette de l'échantillon OFS est fixée à $r = 10\,000$ personnes.

La stratification est constituée des cantons, en regroupant les demi-cantons d'Appenzell Rhodes-Intérieures (AI) et d'Appenzell Rhodes-Extérieures (AR) et les demi-cantons de Nidwald (NW) et d'Obwald (OW), mais pas les demi-cantons de Bâle-Campagne (BL) et de de Bâle-Ville (BS), ainsi que de la ville de Zurich prise séparément du reste de son canton.

La taille nette de l'échantillon OFS est allouée aux strates selon [3.2](#).

3.7.3 Densification de l'échantillon OFS

Les densifications des cantons et des villes sont données dans la [table 1](#).

La taille nette totale est ainsi $r^0 = \sum_h r_h + \sum_h r_h^+ = 15\,836$.

Les tailles nettes r_h^0 , par strate, sont données dans la [table 2](#).

TABLE 1 Densifications (tailles nettes) des cantons et des villes de l'enquête thématique MZB 2021.

Canton ou Ville	Densification (r_h^+)
BL	700
GE	3 487
LU	800
VS	380
Ville de Zurich	469
Total	5 836

3.7.4 Tailles brutes et estimation des taux de réponse

Les tailles brutes n_h^0 par strate sont estimées à partir :

- des taux t^{TEL} et des taux de réponse tr^{TEL} , par strate, de l'enquête ELRC 2019 ;
- des taux de réponse tr^{ALTEL} , par grande région, également de l'enquête ELRC 2019.

La taille brute totale est ainsi $n^0 = \sum_h n_h^0 = 37\,901$.

Cette taille a été diminuée de 15% pour tenir compte de la différence du mode de collecte entre ELRC 2019 (CATI) et MZB 2021 (CAWI/CATI), en se basant sur ce qui a été observé pour l'enquête Omnibus TIC lors du changement de mode entre 2014 et 2017.

La taille brute totale est alors finalement $n^0 = \sum_h n_h^0 = 32\,217$.

Les tailles brutes n_h^0 , par strate, sont données dans la table 2.

3.7.5 Réserves et fractionnement en plusieurs vagues

L'adaptation de -15% du paragraphe précédent est ensuite également prise en compte dans le calcul de la taille de deux réserves de taille brute égale à 15% de n^0 , par strate.

Deux réserves sont ainsi prévues par strate, d'une taille brute totale $l^0 = \sum_h l_h^0 = 5688$. Aucune de ces réserves n'est activée.

Les tailles brutes des réserves l_h^0 , par strate, sont données dans la table 2.

Suite au passage en mixed-mode en 2021, il n'y a plus de fractionnement de l'échantillon en plusieurs vagues pour les enquêtes MZB.

3.7.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage est effectué dans le SRPH avec la date de référence du 31 décembre 2020, créé le 9 février 2021.

Les tailles de la population cible de tirage N_h , par strate, sont données dans la table 2.

Les tailles brutes obtenues \tilde{n}_h^0 et les taille brutes des réserves obtenues \tilde{l}_h^0 , par strate, sont données dans la table 2.

3.7.7 Table récapitulative

La table récapitulative 2 donne, par strate, les différentes tailles de l'enquête.

TABLE 2 Tailles, par strate, de l'enquête thématique MZB 2021.

Strate (h)	Taille population cible (N_h)	Taille échantillon				
		net prévu (r_h^0)	brut		réserve	
			prévu (n_h^0)	tiré (\tilde{n}_h^0)	prévu (l_h^0)	tiré (\tilde{l}_h^0)
AG	522 072	772	1 622	1 686	572	632
AI + AR	52 831	78	134	168	48	47
BE	766 495	1 134	2 236	2 150	790	777
BL	212 978	1 015	2 105	2 097	744	689
BS	143 637	213	505	481	178	157
FR	241 625	357	700	669	248	264
GE	345 553	3 999	8 333	8 244	2 940	2 963
GL	30 335	45	112	113	40	38
GR	148 862	220	447	481	158	150
JU	542 20	80	140	136	50	50
LU	309 560	1 258	2 440	2 322	862	787
NE	129 687	192	392	389	138	148
NW + OW	61 483	91	162	154	58	57
SG	383 027	568	1 165	1 139	412	392
SH	61 360	91	190	196	66	94
SO	207 394	307	635	659	224	221
SZ	123 037	181	393	426	138	142
TG	212 775	315	587	625	208	227
TI	259 538	912	1 480	1 471	522	526
UR	26 955	40	99	98	36	35
VD	597 174	883	1 706	1 794	602	598
VS	256 408	759	1 651	1 643	582	618
ZG	96 596	142	276	291	98	92
Ville de Zurich	317 699	941	2 216	2 150	782	744
ZH sans ville de Zurich	840 982	1 243	2 491	2 422	880	878
Total	6 402 283	15 836	32 217	32 004	11 376	11 326

3.8 Plan et tirage ESS 2022

3.8.1 Définition de la population cible

La population cible de l'enquête ESS 2022 est restreinte aux personnes âgées de 15 ans ou plus, les âges étant calculés au 31 décembre 2021.

La taille de la population cible prise en considération pour le plan de sondage et qui correspond au SRPH avec la date de référence du 30 juin 2021, créé le 15 août 2021, est alors $N = 7\,191\,991$.

3.8.2 Tailles nettes, stratification et allocation de l'échantillon OFS

La taille nette de l'échantillon OFS est fixée à $r = 10\,000$ personnes.

La stratification est constituée des cantons et demi-cantons, ainsi que de la ville de Zurich prise séparément du reste de son canton.

La taille nette de l'échantillon OFS est allouée aux strates selon 3.2.

3.8.3 Densification de l'échantillon OFS

Les densifications des cantons et des villes sont données dans la table 3.

TABLE 3 Densifications (tailles nettes) des cantons et des villes de l'enquête thématique ESS 2022.

Canton ou Ville	Densification (r_h^+)
AG	400
AR	600
BL	500
BS	790
FR	651
GE	587
GR	776
LU	550
NE	400
NW	400
SG	433
SZ	821
TG	888
TI	600
VD	800
VS	618
ZG	859
Ville de Zurich	539
Total	11 212

La taille nette totale est ainsi $r^0 = \sum_h r_h + \sum_h r_h^+ = 21\,212$.

Les tailles nettes r_h^0 , par strate, sont données dans la table 5.

D'autre part, l'Office Fédéral de la Santé Publique OFSP a demandé deux densifications de l'échantillon OFS. Ces deux densifications, qui seront désignées dans la suite du document sous l'appellation d'échantillons supplémentaires, sont :

1. de taille nette $r^1 = 750$, constitué d'étrangers de nationalité d'Europe de l'est ou du sud-est (ESEE)³, dont la taille de la population cible vaut $N = 450\,354$;
2. de taille nette $r^2 = 250$, constitué d'étrangers de nationalité d'Europe du sud-ouest (ESO)⁴, dont la taille de la population cible vaut $N = 560\,093$.

Les tailles nettes par strate de ces deux échantillons sont allouées proportionnellement à la taille de leur population cible.

Les tailles nettes r_h^1 et r_h^2 , par strate, sont données respectivement dans les tables 20 et 22 en annexe.

3.8.4 Tailles brutes et estimation des taux de réponse

Les tailles brutes n_h^0 par strate sont estimées à partir :

- des taux t^{TEL} , par strate, du SRPH avec la date de référence du 30 juin 2021, créé le 15 août 2021 ;
- des taux de réponse tr^{TEL} , par strate et des taux de réponse tr^{ALTEL} , par grande région, égaux à la moyenne arithmétique des taux de réponse des enquêtes ESS 2017, EFG 2018 et ELRC 2019 ;

alors que les tailles brutes n_h^1 et n_h^2 par strate le sont à partir :

- des taux t^{TEL} , par strate, du SRPH avec la date de référence du 30 juin 2021, créé le 15 août 2021 ;
- des taux de réponse globaux tr^{ESEE} et tr^{ESO} , par grande région, de l'enquête ESS 2017 pour les étrangers de nationalité d'Europe de l'est ou du sud-est et pour les étrangers de nationalité d'Europe du sud-ouest.

Les tailles brutes totales sont ainsi $n^0 = \sum_h n_h^0 = 48\,109$, $n^1 = \sum_h n_h^1 = 2\,539$ et $n^2 = \sum_h n_h^2 = 564$.

Les tailles brutes n_h^0 , par strate, sont données dans la table 5.

Les tailles brutes n_h^1 et n_h^2 , par strate, sont données respectivement dans les tables 20 et 22 en annexe.

3. Nationalités ESEE : Albanie, Bulgarie, Grèce, Yougoslavie, Pologne, Roumanie, Union des Républiques Socialistes Soviétiques, Tchécoslovaquie, Turquie, Hongrie, Chypre, Slovaquie, République tchèque, Serbie, Serbie-et-Monténégro, Croatie, Slovénie, Bosnie et Herzégovine, Serbie centrale, Monténégro, Macédoine, Kosovo, Province de Voïvodine, Estonie, Lettonie, Lituanie, Moldava, Russie, Ukraine et Bélarus

4. Nationalités ESO : Andorre, Gibraltar, Italie, Malte, Portugal, Saint-Marin, Espagne et Cité du Vatican

3.8.5 Réserves et fractionnement en plusieurs vagues

Pour chacun des échantillons, soit l'échantillon OFS augmenté des densifications et les deux échantillons supplémentaires de l'OFSP, une réserve est prévue par strate, de tailles brutes totales respectives $l^0 = \sum_h l_h^0 = 4811$, $l^1 = \sum_h l_h^1 = 255$ et $l^2 = \sum_h l_h^2 = 55$.

Les tailles brutes sont fractionnées en quatre vagues $i = 1, 2, 3, 4$, avec un quart de ces tailles pour chaque vague.

Les tailles brutes n_h^0 , par vague, sont données la table 4.

Les tailles brutes n_h^1 et n_h^2 , par vague, sont données respectivement dans les tables 19 et 21 en annexe.

Au vu des taux de réponse particulièrement bas de cette enquête, une réserve supplémentaire est tirée pour la deuxième vague, quatre pour la troisième vague et deux pour la quatrième vague.

Les tailles brutes des réserves l_i^0 , par vague, sont données dans la table 4 et l_h^0 , par strate, dans la table 5.

Les tailles brutes des réserves l_i^1, l_i^2 , par vague, sont données respectivement dans les tables 19 et 21 en annexe et l_h^0 et l_h^2 , par strate, respectivement dans les tables 20 et 22 en annexe.

3.8.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage de la première vague est effectué dans le SRPH avec la date de référence du 30 septembre 2021, créé le 11 novembre 2021.

Le tirage de la deuxième vague est effectué dans le SRPH avec la date de référence du 31 décembre 2021, créé le 9 février 2022.

Le tirage de la troisième vague est effectué dans le SRPH avec la date de référence du 31 mars 2022, créé le 10 mai 2022.

Le tirage de la quatrième vague est effectué dans le SRPH avec la date de référence du 30 juin 2022, créé le 11 août 2022.

Les tailles brutes obtenues \tilde{n}_i^0 , les taille brutes des réserves obtenues \tilde{l}_i^0 et les tailles des réserves activées \bar{l}_i^0 , par vague, sont données dans la table récapitulative 4.

Les tailles brutes obtenues \tilde{n}_h^0 , les taille brutes des réserves obtenues \tilde{l}_h^0 et les tailles des réserves activées \bar{l}_h^0 , par strate, sont données dans la table récapitulative 5.

Les tailles brutes obtenues \tilde{n}_i^1 et \tilde{n}_i^2 , les taille brutes des réserves obtenues \tilde{l}_i^1 et \tilde{l}_i^2 et les tailles des réserves activées \bar{l}_i^1 et \bar{l}_i^2 , par vague, sont données respectivement dans les tables récapitulatives 19, 21 en annexe.

Les tailles brutes obtenues \tilde{n}_h^1 et \tilde{n}_h^2 , les taille brutes des réserves obtenues \tilde{l}_h^1 et \tilde{l}_h^2 et les tailles

des réserves activées \bar{l}_h^1 et \bar{l}_h^2 , par strate, sont données respectivement dans les tables récapitulatives 20, 22 en annexe.

3.8.7 Tables récapitulatives

Les tables récapitulatives 4 et 5 donnent, par vague et par strate, les différentes tailles de l'enquête, pour l'échantillon OFS augmenté des densifications des cantons et des villes.

Les tables récapitulatives 19, 20, 21 et 22 de l'annexe B donnent, par vague et par strate, les différentes tailles de l'enquête, pour les échantillons supplémentaires de l'OFSP.

TABLE 4 Tailles, par vague, de l'enquête thématique ESS 2022, pour l'échantillon OFS augmenté des densifications des cantons et des villes.

Vague (i)	Taille population cible (N_i)	Taille échantillon					
		brut		réserve			total
		prévu (n_i^0)	tiré (\tilde{n}_i^0)	prévu (l_i^0)	tiré (\tilde{l}_i^0)	activé (\bar{l}_i^0)	activé ($\tilde{n}_i^0 + \bar{l}_i^0$)
1	7 186 228	12 026	11 871	1 203	1 178	1 178	13 049
2	7 182 252	12 026	12 022	2 406	2 375	2 375	14 397
3	7 179 559	12 026	12 100	6 015	6 177	3 695	15 795
4	7 172 329	12 026	11 869	3 609	3 543	1 807	13 676
Total	-	48 104	47 862	13 233	13 273	9 055	56 917

TABLE 5 Tailles, par strate, de l'enquête thématique ESS 2022, pour l'échantillon OFS augmenté des densifications des cantons et des villes.

Strate (h)	Taille échantillon						
	net prévu (r_h^0)	brut		réserve			total activé ($\tilde{n}_h^0 + \bar{l}_h^0$)
		prévu (n_h^0)	tiré (\tilde{n}_h^0)	prévu (l_h^0)	tiré (\tilde{l}_h^0)	activé (\bar{l}_h^0)	
AG	1 169	2 816	2 857	770	747	482	3 339
AI	18	36	33	11	9	4	37
AR	660	1 364	1 359	374	386	323	1 682
BE	1 149	2 524	2 486	693	694	473	2 959
BL	824	1 944	1 928	539	518	311	2 239
BS	1 005	2 752	2 757	759	787	484	3 241
FR	1 002	2 256	2 193	627	629	396	2 589
GE	1 098	2 452	2 447	671	689	571	3 018
GL	45	112	120	33	34	29	149
GR	999	2 400	2 371	660	629	399	2 770
JU	81	160	140	44	40	28	168
LU	1 006	2 200	2 211	605	564	354	2 565
NE	593	1 288	1 300	352	339	208	1 508
NW	449	968	922	264	270	229	1 151
OW	42	92	93	22	29	22	115
SG	999	2 340	2 413	649	644	421	2 834
SH	92	212	196	55	59	37	233
SO	308	716	734	198	201	127	861
SZ	1 001	2 360	2 298	649	666	436	2 734
TG	1 200	2 612	2 621	715	757	615	3 236
TI	1 508	2 712	2 634	748	710	571	3 205
UR	40	92	94	22	17	9	103
VD	1 680	3 624	3 661	1 001	1 056	650	4 311
VS	1 000	2 396	2 346	660	649	420	2 766
ZG	1 000	2 292	2 261	627	650	396	2 657
Ville de Zurich	996	2 516	2 459	693	673	547	3 006
ZH sans ville de Zurich	1 244	2 868	2 928	792	827	513	3 441
Total	21 212	48 104	47 862	13 233	13 273	9 055	56 917

3.9 Plan et tirage EFG 2023

3.9.1 Définition de la population cible

La population cible de l'enquête EFG 2023 est restreinte aux personnes âgées de 15 à 79 ans, les âges étant calculés au 31 décembre 2022.

La taille de la population cible prise en considération pour le plan de sondage et qui correspond au SRPH avec la date de référence du 30 septembre 2022, créé le 11 novembre 2022, est alors $N = 6\,817\,742$.

3.9.2 Tailles nettes, stratification et allocation de l'échantillon OFS

La taille nette de l'échantillon OFS est fixée à $r = 10\,000$ personnes.

La stratification est constituée :

- des cantons qui densifient leur échantillon et de la ville de Zurich qui densifie également son échantillon ;
- des cantons avec une population cible supérieure ou égale à 200 000 personnes ;
- des grandes régions restantes ;

croisés avec des combinaisons des âges et des compositions de ménage :

1. les personnes âgées de 15 à 24 ans ;
2. les personnes âgées de 25 ans ou plus, dont le ménage comprend au moins un enfant de 13 ans ou moins ;
3. les personnes âgées de 25 ans ou plus, dont le ménage ne comprend pas d'enfant de 13 ans ou moins.

La taille nette de l'échantillon OFS est allouée aux cantons et aux grandes régions restantes de la stratification selon 3.2, puis, après adjonction des tailles de densification demandées (voir ci-dessous), selon les combinaisons des âges et des compositions de ménage, avec les proportions suivantes :

1. 7.5% pour les personnes âgées de 15 à 24 ans ;
2. 40.0% pour les personnes âgées de 25 ans ou plus, dont le ménage comprend au moins un enfant de 13 ans ou moins ;
3. 52.5% pour les personnes âgées de 25 ans ou plus, dont le ménage ne comprend pas d'enfant de 13 ans ou moins ;

tout en fixant une taille nette minimale par strate égale à 30.

Le but de la combinaison, dans la stratification, des grandes régions et des cantons avec les âges et les compositions de ménage est d'augmenter la proportion de familles avec au moins un enfant de 13 ans ou moins dans l'échantillon. Ceci permet d'améliorer la précision des estimations qui portent sur cette population.

3.9.3 Densification de l'échantillon OFS

Les densifications des cantons et des villes sont données dans la table 6.

TABLE 6 Densifications (tailles nettes) des cantons et des villes de l'enquête thématique EFG 2023.

Canton ou Ville	Densification (r_h^+)
GE	2 040
VD	884
JU	900
BL	957
TG	1 256
LU	800
Ville de Zurich	1 395
Total	8 232

La taille nette totale est ainsi $r^0 = \sum_h r_h + \sum_h r_h^+ = 18\,273$.

Les tailles nettes r_g^0 , par canton et grande région restante, sont données dans la table 7.

3.9.4 Tailles brutes et estimation des taux de réponse

Les tailles brutes n_h^0 par strate sont estimées à partir :

- des taux t^{TEL} , par canton et grande région restante, du SRPH avec la date de référence du 30 septembre 2022, créé le 11 novembre 2022 ;
- des taux de réponse tr^{TEL} et tr^{ALTEL} , par grande région, de l'enquête MZB 2021.

La taille brute totale est ainsi $n^0 = \sum_h n_h^0 = 32\,124$.

Les tailles brutes n_g^0 , par canton et grande région restante, sont données dans la table 7.

3.9.5 Réserves et fractionnement en plusieurs vagues

Deux réserves de 10% sont prévues par strate, d'une taille brute totale $l^0 = \sum_h l_h^0 = 3\,213$.

Les tailles brutes des réserves l_g^0 , par canton et grande région restante, sont données dans la table 7.

Suite au passage en mixed-mode en 2023 il n'y a plus de fractionnement de l'échantillon en plusieurs vagues pour les enquêtes EFG.

3.9.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage est effectué dans le SRPH avec la date de référence du 31 décembre 2022, créé le 9 février 2023.

Les tailles de la population cible de tirage N_g , par canton et grande région restante, sont données dans la table 7.

Les tailles brutes obtenues \tilde{n}_g^0 , les taille brutes des réserves obtenues \tilde{l}_g^0 et les tailles des réserves activées \bar{l}_g^0 , par canton et grande région restante, sont données dans la table 7.

3.9.7 Table récapitulative

La table récapitulative 7 donne, par canton et grande région restante, les différentes tailles de l'enquête thématique EFG 2023.

TABLE 7 Tailles, par canton et grande région restante, de l'enquête thématique EFG 2023.

Canton / Grande région (g)	Taille population cible (N_g)	Taille échantillon						total activé ($\tilde{n}_g^0 + \bar{l}_g^0$)
		net prévu (r_g^0)	brut prévu tiré (n_g^0) (\tilde{n}_g^0)		réserve prévu tiré (l_g^0) (\tilde{l}_g^0)		activé (\bar{l}_g^0)	
VS	276 930	385	680	652	136	117	0	652
GE	364 253	2 545	4 322	4 243	864	830	830	5 073
VD	635 578	1 767	3 048	3 061	610	613	0	3 061
JU	57 267	980	1 627	1 752	324	333	0	1 752
NE	137 024	206	345	315	70	84	0	315
FR	258 500	362	610	608	122	124	0	608
SO	220 988	314	524	492	106	119	0	492
BE	816 397	1 134	1 893	1 825	378	368	0	1 825
BS	152 700	226	437	429	88	97	0	429
BL	227 599	1 273	2 457	2 463	490	467	0	2 463
AG	559 179	776	1 494	1 486	298	338	0	1 486
Ville de Zurich	334 998	1 860	3 417	3 499	684	698	355	3 854
ZH sans ville de Zurich	895 957	1 246	2 270	2 194	454	458	0	2 194
TG	227 242	1 571	2 905	2 928	582	583	0	2 928
SG	411 053	571	1 063	1 071	214	201	0	1 071
Ostschweiz sans TG et SG	313 403	435	811	821	162	162	0	821
LU	331 140	1 259	2 141	2 113	428	447	0	2 113
Zentralschweiz sans LU	328 699	454	777	756	156	144	0	756
TI	277 569	909	1 303	1 266	260	266	138	1 404
Total	6 826 476	18 273	32 124	31 974	6 426	6 449	1 323	33 297

3.10 Plan et tirage ELRC 2019

3.10.1 Définition de la population cible

La population cible de l'enquête ELRC 2019 est restreinte aux personnes âgées de 15 ans ou plus, les âges étant calculés au 31 décembre 2018.

La taille de la population cible prise en considération pour le plan de sondage et qui correspond au SRPH avec la date de référence du 30 septembre 2018, créé le 13 novembre 2018, est alors $N = 7\,026\,399$.

3.10.2 Tailles nettes, stratification et allocation de l'échantillon OFS

La taille nette de l'échantillon OFS est fixée à $r = 10\,000$ personnes.

La stratification est constituée des cantons, en regroupant les demi-cantons d'Appenzell Rhodes-Intérieures (AI) et d'Appenzell Rhodes-Extérieures (AR), les demi-cantons de Nidwald (NW) et d'Obwald (OW), mais pas les demi-cantons de Bâle-Campagne (BL) et de de Bâle-Ville (BS), ainsi que de la ville de Zurich prise séparément du reste de son canton.

La taille nette de l'échantillon OFS est allouée aux strates selon [3.2](#).

3.10.3 Densification de l'échantillon OFS

Les densifications des cantons et des villes sont données dans la [table 8](#).

TABLE 8 Densifications (tailles nettes) des cantons et des villes de l'enquête thématique ELRC 2019.

Canton ou Ville	Densification (r_h^+)
GE	1 091
VD	1 091
VS	385
Ville de Zurich	453
Total	3 020

La taille nette totale est ainsi $r^0 = \sum_h r_h + \sum_h r_h^+ = 13\,020$.

Les tailles nettes r_h^0 , par strate, sont données dans la [table 10](#).

3.10.4 Tailles brutes et estimation des taux de réponse

Les tailles brutes n_h^0 par strate sont estimées à partir :

- des taux t^{TEL} , par strate, du SRPH avec la date de référence du 30 septembre 2018, créé le 13 novembre 2018 ;
- des taux de réponse tr^{TEL} , par strate, de l'enquête EFG 2018 ;
- des taux de réponse tr^{ALTEL} , par grande région, de l'enquête EFG 2018.

La taille brute totale est ainsi $n^0 = \sum_h n_h^0 = 28\,483$.

Les tailles brutes n_h^0 , par strate, sont données dans la table 10.

3.10.5 Réserves et fractionnement en plusieurs vagues

Une réserve de 10% est prévue pour chaque strate, d'une taille brute totale $l^0 = \sum_h l_h^0 = 2854$. Lors du tirage, cette réserve est séparée en trois parties égales.

Les tailles brutes sont fractionnées en trois vagues $i, i = 1, 2, 3$, avec un tiers de ces tailles pour chaque vague.

Au vu des taux de réponse particulièrement bas des première et deuxième vagues, le plan brut est adapté pour la troisième vague, en tenant compte des taux de réponse observés lors des vagues précédentes. À cette troisième vague, une réserve supplémentaire, de la même taille que les trois autres parties de la réserve initialement prévue, est ajoutée.

Les tailles brutes n_h^0 , par strate, sont données dans la table 10 et les tailles brutes n_i^0 , par vague, sont données dans la table 9.

Les tailles brutes des réserves l_h^0 , par strate, sont données dans la table 10 et les tailles brutes l_i^0 , par vague, sont données dans la table 9.

3.10.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage de la première vague est effectué dans le SRPH avec la date de référence du 30 septembre 2018, créé le 13 novembre 2018.

Le tirage de la deuxième vague est effectué dans le SRPH avec la date de référence du 31 décembre 2018, créé le 10 février 2019.

Le tirage de la troisième vague est effectué dans le SRPH avec la date de référence du 31 mars 2019, créé le 13 mai 2019.

Les tailles des populations cibles de tirage N_i , par vague, sont données dans la table 9.

Les tailles brutes obtenues \tilde{n}_h^0 , les tailles brutes des réserves obtenues \tilde{l}_h^0 et les tailles des réserves activées \bar{l}_h^0 , par strate, sont données dans la table récapitulative 10.

Les tailles brutes obtenues \tilde{n}_i^0 , les tailles brutes des réserves obtenues \tilde{l}_i^0 et les tailles des réserves activées \bar{l}_i^0 , par vague, sont données dans la table récapitulative 9.

3.10.7 Tables récapitulatives

La table récapitulative 9 donne, par vague, les différentes tailles de l'enquête et la table récapitulative 9 donne ces mêmes tailles par strate.

TABLE 9 Tailles, par vague, de l'enquête thématique ELRC 2019.

Vague (i)	Taille population cible (N _i)	Taille échantillon					
		brut		réserve			total
		prévu (n _i ⁰)	tiré (\tilde{n}_i^0)	prévu (l _i ⁰)	tiré (\tilde{l}_i^0)	activé (\bar{l}_i^0)	activé ($\tilde{n}_i^0 + \bar{l}_i^0$)
1	7 026 399	9 494	9 626	960	968	0	9 626
2	7 017 473	9 494	9 436	960	949	0	9 436
3	7 007 945	12 615	12 752	1692	1700	145	12 897
Total	-	31 603	31 814	3 612	3 617	145	31 959

TABLE 10 Tailles, par strate, de l'enquête thématique ELRC 2019.

Strate (h)	Taille échantillon						
	net	brut		réserve			total
	prévu (r _h ⁰)	prévu (n _h ⁰)	tiré (\tilde{n}_h^0)	prévu (l _h ⁰)	tiré (\tilde{l}_h^0)	activé (\bar{l}_h^0)	activé ($\tilde{n}_h^0 + \bar{l}_h^0$)
AG	759	1 934	1 948	220	218	0	1 948
AI + AR	80	156	150	20	19	0	150
BE	1 163	2 774	2 713	318	299	0	2 713
BL	327	825	794	96	99	0	794
BS	220	615	618	70	74	0	618
FR	347	736	751	82	74	0	751
GE	1 607	4 379	4 299	502	486	0	4 299
GL	45	149	163	18	13	0	163
GR	225	541	567	60	66	0	567
JU	82	224	207	28	26	0	207
LU	455	1 135	1 139	134	133	0	1 139
NE	197	446	432	54	58	0	432
NW + OW	92	195	188	20	16	0	188
SG	568	1 429	1 467	160	158	0	1 467
SH	93	235	242	28	34	0	242
SO	309	858	863	100	110	0	863
SZ	178	435	436	48	54	0	436
TG	310	630	680	70	73	0	680
TI	906	1 785	1 802	208	222	0	1 802
UR	40	103	103	14	11	0	103
VD	1 966	4 304	4 342	484	470	96	4 438
VS	767	1 873	1 925	214	224	49	1 974
ZG	141	341	361	38	32	0	361
Ville de Zurich	911	2 497	2 527	284	288	0	2 527
ZH sans ville de Zurich	1 232	3 004	3 097	342	360	0	3 097
Total	13 020	31 603	31 814	3 612	3 617	145	31 959

3.11 Plan et tirage MRMT 2021

3.11.1 Définition de la population cible

L'enquête MRMT aurait normalement dû être réalisée en 2020. Du fait des conditions particulières liées à la pandémie de Covid-19, cette enquête a été repoussée en 2021. La population cible de l'enquête MRMT 2021 est restreinte aux personnes âgées de 6 ans ou plus, les âges étant calculés au 31 décembre 2020.

La taille de la population cible prise en considération pour le plan de sondage et qui correspond au SRPH avec la date de référence du 30 juin 2020, créé le 11 août 2020, est alors $N = 7\,897\,647$.

3.11.2 Tailles nettes, stratification et allocation de l'échantillon OFS

Voir le document *Microrecensement mobilité et transports 2021* [Muralti \[2023\]](#), section 2. *Plan d'échantillonnage*, page 2 et suivantes.

3.11.3 Densification de l'échantillon OFS

Voir le document *Microrecensement mobilité et transports 2021* [Muralti \[2023\]](#).

Au final la taille totale est ainsi $r^0 = \sum_h r_h + \sum_h r_h^+ = 56\,652$.

3.11.4 Tailles brutes et estimation des taux de réponse

Les tailles brutes n_h^0 par strate sont estimées à partir :

- des taux t^{TEL} , par strate, du SRPH avec la date de référence du 30 juin 2020, créé le 11 août 2020, augmenté de 3% pour tenir compte des numéros supplémentaires obtenus par une recherche avec le logiciel SpeederMan ;
- des taux de réponse tr^{TEL} , par strate, de l'enquête MRMT 2015 ;
- des taux de réponse tr^{ALTEL} , par grande région, de l'enquête MRMT 2015.

Le choix de ne pas utiliser les taux tr^{TEL} et tr^{ALTEL} de la dernière enquête thématique réalisée, mais ceux de la dernière enquête MRMT, est motivé par la population cible et la stratification particulières de cette enquête.

La taille brute totale est ainsi $n^0 = \sum_h n_h^0 = 111\,757$.

3.11.5 Réserves et fractionnement en plusieurs vagues

Deux réserves de 10% sont prévues par strate, d'une taille brute totale $l^0 = \sum_h l_h^k = 11\,177$.

Les tailles brutes sont fractionnées en quatre vagues $i, i = 1, 2, 3, 4$, avec environ un quart de ces tailles pour chaque vague (environ, car, comme cette enquête tient compte de l'attribution d'un jour de référence, le fractionnement en vagues tient compte du nombre de jours de référence de la période d'enquête de chaque vague).

Au vu des taux de réponse des trois premières vagues, le plan brut est adapté pour la quatrième et dernière vague, en tenant compte des taux de réponse observés lors des vagues précédentes

et une réserve supplémentaire de 10% y est ajoutée.

Les tailles brutes n_h^0 et les tailles brutes des réserves l_h^0 , après adaptation, par vague, sont données dans la table 11.

3.11.6 Tirage coordonné et tailles d'échantillon effectives

Le tirage de la première vague est effectué dans le SRPH avec la date de référence du 30 septembre 2020, créé le 10 novembre 2020.

Le tirage de la deuxième vague est effectué dans le SRPH avec la date de référence du 31 décembre 2020, créé le 9 février 2021.

Le tirage de la troisième vague est effectué dans le SRPH avec la date de référence du 31 mars 2021, créé le 11 mai 2021.

Le tirage de la quatrième vague est effectué dans le SRPH avec la date de référence du 30 juin 2021, créé le 15 août 2021.

Les tailles des populations cibles de tirage N_i , par vague, sont données dans la table 11.

Les tailles brutes obtenues \tilde{n}_i^0 , les taille brutes des réserves obtenues \tilde{l}_i^0 et les tailles des réserves activées \bar{l}_i^0 , par vague, sont données dans la table récapitulative 11.

Il est à noter que, comme cette enquête tient compte de l'attribution d'un jour de référence, lors de la décision d'activer une réserve, certains jours attribués se sont déjà écoulés et qu'il s'en suit que l'activation effective sur le terrain d'enquête ne concerne pas la totalité des réserves données dans la table récapitulative 11.

3.11.7 Table récapitulative

La table récapitulative 11 donne, par vague, les différentes tailles de l'enquête.

TABLE 11 Tailles, par vague, de l'enquête thématique MRMT 2021.

Vague (i)	Taille population cible (N)	Taille échantillon						total activé ($\tilde{n}_i^0 + \bar{l}_i^0$)
		net prévu (r^0)	brut		réserve		activé	
			prévu (n_i^0)	tiré (\tilde{n}_i^0)	prévu (l_i^0)	tiré (\tilde{l}_i^0)	activé (\bar{l}_i^0)	
1	7 891 339		27 862	27 655	5 576	5 706	0	27 655
2	7 887 376	-	28 164	27 698	5 642	5 624	72	27 770
3	7 882 559		27 863	28 062	5 576	5 517	1 322	29 384
4	7 876 312		33 263	33 167	9 996	9 830	6 153	39 320
Total	-	56 652	117 152	116 582	26 790	26 677	7 547	124 129

4 Pondération

Cette section décrit les principes de pondération utilisés pour les enquêtes thématiques. Certaines de ces enquêtes ont des échantillons constitués de sous-échantillons sélectionnés dans plusieurs cadres de sondage trimestriels. Le cas d'une enquête auprès de personnes est considéré, pour laquelle une pondération pour les ménages doit également éventuellement être produite et dans laquelle deux phases d'enquête se succèdent, par exemple une première phase téléphonique suivie d'une deuxième phase auto-administrée pour les répondants à la première phase. Une seule personne cible peut être sélectionnée dans chaque ménage (selon leur constitution dans le cadre de sondage).

4.1 Pondération d'une enquête en une seule vague

Le calcul d'une pondération pour une enquête en une seule vague suit un schéma standard :

1. reprise des probabilités de tirage π_k des personnes sélectionnées et calcul d'un poids initial pour l'échantillon brut : $\delta_k = 1/\pi_k$ [voir [Horvitz & Thompson, 1952](#)],
2. estimation de probabilités de réponse p_k pour chaque personne répondante, en utilisant des informations disponibles sur tout l'échantillon brut et calcul des poids de l'estimateur par dilatation, $d_k = 1/\pi_k p_k$. La non-réponse est considérée comme une deuxième phase de sélection aléatoire dont les vraies probabilités de sélection sont inconnues,
3. calage des poids obtenus sur des marges connues [voir [Deville & Särndal, 1992](#)] : les poids calés $w_k = g_k d_k$ sont tels que les estimateurs des totaux de certaines variables disponibles pour les unités répondantes sont égales à des totaux imposés et connus pour toute la population. Les facteurs g_k sont ceux qui sont les plus proches de 1 pour une distance donnée et qui permettent de satisfaire les contraintes de calage.

Pour calculer les poids des ménages le même principe est utilisé. Puisqu'une personne au plus est sélectionnée par ménage, la probabilité de tirage d'un ménage m , π_m est obtenue en sommant les probabilités de tirage des personnes qui le composent,

$$\pi_m = \sum_{k \in m} \pi_k.$$

La probabilité d'observer un ménage dans l'échantillon des répondants est donnée par

$$\pi_m p_m = \sum_{k \in m} \pi_k p_k,$$

et dépend donc des probabilités de réponse p_k pour toutes les personnes sélectionnables ($\pi_k > 0$) du ménage. La probabilité de réponse prédite du ménage s'en déduit :

$$p_m = \frac{\sum_{k \in m} \pi_k p_k}{\sum_{k \in m} \pi_k}.$$

Les poids $d_m = 1/\pi_m p_m$ peuvent ensuite être calés avec des facteurs g_m en utilisant des caractéristiques des ménages et de leurs membres pour lesquelles les effectifs dans la population sont connus. Le programme CALMAR [voir [Sautory, 1993](#), [Le Guennec & Sautory, 2002](#)] développé à l'INSEE permet de réaliser simultanément le calage pour les personnes cibles, pour les ménages et leurs membres. En pratique, le rapport r_k entre le poids de la personne cible k et celui de son ménage est fixé, par exemple par l'inverse de la probabilité conditionnelle d'observer la personne cible sachant que son ménage est observé,

$$r_k = \sum_{j \in m} \pi_j p_j / \pi_k p_k.$$

Le programme de calage produit alors des facteurs de calage g_m proches de 1 et tels que les poids ménages $w_m = g_m d_m$ satisfont les contraintes de calage pour les ménages et leurs membres et les poids des personnes cibles $w_k = g_m d_m r_k = g_m d_k$ satisfont les contraintes de calage pour les personnes cibles.

4.2 Estimation des probabilités de réponse

La non-réponse est classiquement modélisée par un processus aléatoire par lequel chaque unité décide de répondre ou non indépendamment des autres. La vraie probabilité de réponse est inconnue et des estimateurs sans biais (sous le plan de tirage et de non-réponse) n'existent que si toutes les probabilités de réponse (et de tirage) sont strictement positives. Dans la pratique ces probabilités de réponse sont estimées (ou prédites) en utilisant les informations disponibles à la fois pour les répondants et pour les non-répondants [voir par exemple [Haziza & Lesage, 2016](#)].

Les méthodes de prédiction usuelles peuvent être considérées comme des techniques de régression logistiques et se différencient par la sélection des prédicteurs. Une partie de ces méthodes impliquent de partitionner l'échantillon en groupes homogènes de réponse en utilisant les informations pertinentes, par un choix raisonné ou à l'aide de techniques de segmentation. Il est aussi possible de simplement ajuster un modèle logistique en utilisant certaines des variables continues et/ou catégorielles comme prédicteurs. Une autre solution consiste à se servir de cette dernière méthode comme d'une étape préliminaire pour créer des groupes homogènes de réponse. Il n'existe pas de bonne pratique universellement supérieure à toute autre.

En présence de non-réponse, les estimateurs par dilatation ne sont pas exactement sans biais car les vraies valeurs des probabilités de réponse p_k sont inconnues. Mais une bonne estimation des p_k permet de limiter le risque de biais. Toutefois, si des probabilités estimées pour certains répondants sont très faibles, soit que leur estimation est mauvaise, soit que leur estimation est bonne mais que la vraie probabilité p_k est faible, le poids d_k obtenu est élevé et cela entraîne une variance élevée de l'estimateur par dilatation, particulièrement sensible pour les domaines d'intérêt dans lesquels ces répondants se trouvent.

Le choix d'un modèle pour prédire les probabilités de réponse relève donc du compromis à faire entre biais et variance. Il est impossible par nature d'estimer le biais de manière fiable : si une variable disponible permettait de détecter un biais, elle devrait vraisemblablement être utilisée pour la prédiction des probabilités de réponse. Un choix pragmatique conduit à retenir un modèle relativement complet mais pour lequel la variance d'estimation reste supportable.

[Haziza \[2018\]](#) développe l'idée de choisir un modèle de non-réponse en fonction des variables d'intérêt de l'enquête. En effet, si la non-réponse est totalement étrangère aux variables d'intérêt,

$$\sum_{k \in U} y_k p_k = \sum_{k \in U} y_k \sum_{k \in U} p_k / N,$$

où N est la taille de la population U , corriger cette non-réponse a pour seul effet d'augmenter la variance d'estimation. Il serait alors préférable d'appliquer un coefficient unique de dilatation (diviser les poids δ_k par le taux de réponse comme estimateur de $\sum_{k \in U} p_k / N$). Or, toutes les utilisations futures des données ne sont pas connues lors du calcul de la pondération. Il paraît dès lors difficile d'écarter un bon prédicteur de la non-réponse au risque de subir un biais excessif pour certaines estimations.

4.3 Deuxième phase d'enquête

Dans le cas où les répondants à une première phase d'enquête sont invités à répondre à une deuxième phase, il faut prévoir une deuxième étape de non-réponse et prédire des probabilités

conditionnelles de réponse ρ_k à la deuxième phase. Le poids de l'estimateur par dilatation devient $d_k = 1/\pi_k p_k \rho_k$ pour l'échantillon des répondants à la seconde phase. Le jeu de données disponible pour estimer cette probabilité est réduit car il contient seulement les répondants à la première phase. Les informations collectées à la première phase pourraient être utilisées dans un modèle de prédiction des ρ_k . Il est également possible de modéliser et prédire directement $p_k \rho_k$ sur l'échantillon complet avec le risque que les poids de la deuxième phase ne soient pas cohérents avec ceux de la première. En effet, des unités qui répondent aux deux phases pourraient avoir un poids à la deuxième phase inférieur à celui de la première.

4.4 Discussion sur le calage

Selon [Deville & Särndal \[1992\]](#), tous les choix de distance de calage produisent des estimateurs asymptotiquement équivalents à l'estimateur par la régression [voir [Cassel et al., 1976](#)]. En pratique, les variables de calage peuvent être choisies comme de bons prédicteurs des variables d'intérêt de l'enquête dans un modèle de régression linéaire. Toutefois, un nombre élevé de contraintes de calage peut aussi conduire à obtenir des facteurs de calage g_k et g_m élevés et éloignés de 1. Ceci est un double inconvénient : la proximité entre g_k et 1 est importante pour limiter le biais de l'estimateur calé ; un g_k élevé peut conduire à un poids w_k élevé, ce qui entraîne une variance importante pour l'estimateur calé pour des variables d'intérêt qui seraient mal corrélées aux variables de calage.

Comme pour la modélisation de la non-réponse, les variables de calage choisies devraient être liées aux thèmes de l'enquête mais ne pas conduire à augmenter exagérément la dispersion des poids d'extrapolation ou à en créer de trop grands.

Le calage simultané des ménages et des personnes cibles est intéressant car il permet d'assurer une certaine cohérence. Par exemple, le poids d'une personne seule dans son ménage est égal au poids de ce ménage et le poids d'un ménage est toujours inférieur ou égal au poids de la personne cible. En cas de calage simultané des poids ménages et des poids des personnes cibles après et/ou avant combinaison, pour chaque $k \in m \in s_i$, on a de plus $g_m = g_k$ et/ou $g_m^i = g_k^i$. Mais un calage simultané implique aussi un plus grand nombre de contraintes pour un seul jeu de facteurs d'ajustement g_m et donc des g_m et g_k plus dispersés et éloignés de 1 que si les deux calages étaient réalisés séparément.

Le choix des options pour le calage résulte donc d'un arbitrage entre d'une part les gains attendus pour les variables proches des variables de calage et les relations souhaitées entre les différentes pondérations et d'autre part les risques acceptés pour l'estimation de variables qui ne sont pas bien corrélées aux variables de calage.

4.5 Pondération d'une enquête sélectionnée en plusieurs vagues

Dans le cas d'une enquête sélectionnée dans plusieurs cadres SRPH, il n'y a pas une population cible ou une population de référence naturelle : des personnes entrent et sortent de la population, des ménages changent de composition. Ainsi, le choix des caractéristiques à estimer est déterminant pour une stratégie de pondération.

La réunion des populations des cadres correspond à l'ensemble dans lequel l'échantillon complet a été sélectionné et il est possible de calculer les probabilités de sélection des personnes et des ménages dans la réunion des vagues d'enquête. Cependant, cette population a peu d'intérêt : elle est plus grande que la population à chaque date, un répondant qui n'est présent dans la population qu'à une seule vague a une probabilité de sélection plus faible et donc un poids plus important qu'un répondant qui est présent à chaque vague. Pour les ménages, le problème est encore plus important : des personnes dans des ménages qui changent de composition apparaissent plusieurs fois, ce qui complique encore le choix des caractéristiques à estimer.

4.5.1 Coefficients de combinaison des vagues

À l'instar de la solution retenue pour le pooling des échantillons du relevé structurel [voir [Potterat et al., 2019](#)], une possibilité est de calculer une *moyenne pondérée* des différentes enquêtes. Par exemple, pour une enquête en quatre vagues, un choix naturel serait de définir l'estimateur

$$\widehat{t}_y = \frac{1}{4} (\widehat{t}_y^1 + \widehat{t}_y^2 + \widehat{t}_y^3 + \widehat{t}_y^4),$$

où $\widehat{t}_y^1, \dots, \widehat{t}_y^4$ sont des estimateurs du total d'une variable y lors des quatre vagues d'enquête respectives. Si ces estimateurs sont sans biais, alors l'espérance mathématique de \widehat{t}_y est simplement égale à la moyenne arithmétique des totaux de la variable y lors des quatre occasions d'enquête.

Soit les estimateurs $\widehat{t}_y^i, i = 1, \dots, 4$ définis par

$$\widehat{t}_y^i = \sum_{k \in s_i} w_k^i y_k^i,$$

où y_k^i est la valeur observée dans l'échantillon (qui peut changer avec i , c.à.d. entre les vagues), w_k^i est un poids calculé pour la personne k lorsque seul l'échantillon s_i de la vague i dans laquelle la personne $k \in U_i$ était sélectionnée est pondéré et U_i est la population dans laquelle l'échantillon s_i a été sélectionné. Alors le poids w_k d'une unité k dans l'estimateur \widehat{t}_y peut s'écrire en tout généralité comme :

$$w_k = a_k^i w_k^i,$$

où a_k^i est un coefficient qui permet de combiner les vagues d'enquête.

S'il y a v vagues et si l'échantillon complet s est la réunion de s_1, \dots, s_v , alors

$$w_k I_{k \in s} = \sum_{i=1}^v a_k^i w_k^i I_{k \in s_i},$$

où $I_{k \in s}, I_{k \in s_i}$ sont des indicatrices d'appartenance dans les échantillons respectifs. Lorsque les a_k^i et w_k^i ne dépendent pas de l'échantillon, l'espérance sous le plan d'un estimateur de total pour y_k est donnée par

$$E \left(\sum_{k \in s} w_k I_{k \in s} y_k \right) = \sum_{i=1}^v \sum_{k \in U_i} a_k^i w_k^i \pi_k^i p_k^i y_k^i, \quad (1)$$

où $\pi_k^i p_k^i$ est la probabilité d'observer la personne k dans l'échantillon s_i . L'indice i est omis abusivement dans la partie gauche de l'équation (1).

Le choix d'un coefficient de combinaison des vagues constant égal à $1/v$ pour les personnes et les ménages permet d'estimer les moyennes simples des totaux sur les différentes vagues. Cela présente un double avantage : c'est compréhensible et facilite l'interprétation des estimateurs (tout du moins pour des totaux) et l'estimateur possède les propriétés de lissage des variations saisonnières d'une moyenne mobile.

Cependant, ce choix n'est pas judicieux si les vagues d'enquête sont de tailles fortement différentes. Dans ce cas, donner un poids aussi important à la plus petite vague qu'aux autres entraîne une variance d'estimation plus élevée que dans le cas où les coefficients sont déterminés par les variances des estimateurs des différentes vagues. Le choix des facteurs de combinaison de vagues devrait donc être raisonné entre une cible d'estimation intelligible et des préoccupations de précision, liées aux taux de sondages et aux taux de réponse dans les différentes vagues. Des facteurs différents selon les «strates» de tirage et les caractéristiques des unités peuvent être envisagés ainsi que des facteurs différents pour les personnes cibles et les ménages.

4.5.2 Estimation des probabilités de réponse

Un choix de modèle ou un ajustement séparé d'un modèle de réponse pour les différentes vagues peut être considéré. L'inconvénient par rapport à une utilisation de l'échantillon complet sans distinction des vagues est que le nombre de prédicteurs est plus élevé. Les avantages sont une possible meilleure correction si la non-réponse dépend de la vague et un calcul de poids w_k^i pour les différentes vagues qui suit le schéma habituel.

4.5.3 Calage séparé pour chaque vague et/ou calage des poids finaux

Il est possible de calculer des poids de vague w_k^i calés sur des marges qui peuvent naturellement dépendre de la vague. Par exemple, chaque vague peut être calée sur des totaux du cadre de sondage correspondant. Dans le cas de coefficients de combinaison uniformes (égaux à $1/v$), le poids final sera alors calé sur la combinaison correspondante des totaux de calage. Il est aussi possible de simplement effectuer un calage des poids après la combinaison des poids des vagues, ou encore de faire un calage avant et un autre après la combinaison des poids des vagues. Le choix de la procédure à utiliser et le choix des valeurs pour les marges de calage devrait résulter du choix des paramètres cibles à estimer. À noter que les valeurs des variables de calage des observations dépendent de la vague de sélection. Ce sont des variables aléatoires, ce qui ne correspond pas exactement au cadre classique du calage sur marges.

4.5.4 Étapes de pondération pour une enquête en plusieurs vagues

Les étapes de pondération effectuées sont les suivantes :

1. reprise des probabilités de sélection π_k^i dans chaque vague et calcul d'un poids initial de vague δ_k^i pour les personnes et un poids initial $\delta_m^i = 1 / \sum_{k \in m} \pi_k^i$ pour les ménages, reprise des variables utilisées pour modéliser la non-réponse et pour les calages.
2. estimation des probabilités de réponse p_k^i (modèle ajusté séparément dans chaque vague ou sur un échantillon constitué des échantillons des différentes vagues) des personnes cibles et membres des ménages. Calcul de poids de l'estimateur par dilatation $d_k^i = \delta_k^i / p_k^i$ pour les personnes, calcul de probabilités de réponse p_m^i pour les ménages et de poids $d_m^i = \delta_m^i / p_m^i$.
3. le cas échéant, prédiction de probabilités de réponse conditionnelles ρ_k^i à une deuxième phase d'enquête et calcul des poids correspondants.
4. (facultatif) Calage séparé des poids de vague (avec choix d'un calage simultané ménages-personnes cibles ou de calages séparés), calcul des facteurs $g_k^i = w_k^i / d_k^i$, $g_m^i = w_m^i / d_m^i$ pour les personnes et les ménages.
5. choix de coefficients de combinaison a_k^i pour les personnes cibles et a_m^i pour les ménages, combinaison des vagues et calcul des poids combinés w_k et w_m pour les personnes et les ménages, respectivement.
6. (facultatif) calage des poids finaux sur des marges choisies, éventuellement en reprenant des variables et marges séparées pour les différentes vagues (avec choix d'un calage simultané ménages-personnes cibles ou de calages séparés), calcul des facteurs $g_k = w_k / d_k$, $g_m = w_m / d_m$.

Le poids final pour une personne cible répondante k à la première phase est :

$$w_k I_{k \in s} = g_k \left(\sum_{i=1}^v \frac{a_k^i g_k^i}{\pi_k^i p_k^i} I_{k \in s_i} \right), \quad (2)$$

et le poids final pour un ménage répondant m à la première phase est :

$$w_m I_{m \in s} = g_m \left(\sum_{i=1}^v \frac{a_m^i g_m^i}{\sum_{k \in m} \pi_k^i p_k^i \rho_k^i} I_{m \in s_i} \right). \quad (3)$$

Pour l'échantillon de deuxième phase, les poids finaux sont respectivement pour une personne cible k :

$$w_k I_{k \in s} = g_k \left(\sum_{i=1}^v \frac{a_k^i g_k^i}{\pi_k^i p_k^i \rho_k^i} I_{k \in s_i} \right), \quad (4)$$

et pour un ménage m :

$$w_m I_{m \in s} = g_m \left(\sum_{i=1}^v \frac{a_m^i g_m^i}{\sum_{k \in m} \pi_k^i p_k^i \rho_k^i} I_{m \in s_i} \right). \quad (5)$$

Le schéma général des étapes qui constituent une pondération est représenté sur la figure 1. La combinaison des vagues peut s'effectuer dans une des trois positions :

1. avant la correction de la non-réponse. Dans ce cas les échantillons des différentes vagues sont combinés en un seul échantillon sur lequel est effectuée la pondération. Même si la pondération s'effectue sur un seul échantillon, les corrélations entre les tirages des différentes vagues rentrent en ligne de compte pour le calcul de la précision,
2. après la correction pour la non-réponse et avant le calage (sans un calage pour chaque vague),
3. après la correction pour la non-réponse et après un calage par vague (avec la possibilité d'effectuer un deuxième calage après la combinaison des vagues). Chacun des calages peut être soit simultané (personne-ménage) soit séparé. Si le dernier calage effectué est simultané alors la cohérence des poids dans les ménages avec une seule personne est garantie.

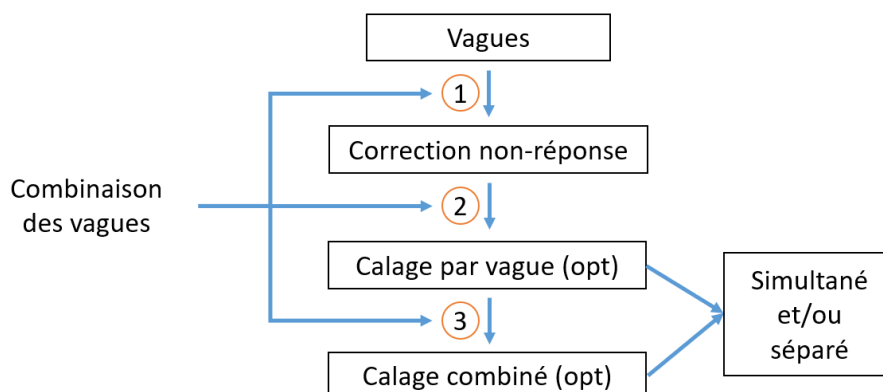


FIGURE 1 Présentation du schéma général de pondération.

5 Pondération des enquêtes thématiques

5.1 Pondération MZB 2021

L'enquête MZB 2021 a été effectuée pour la première fois en multimode CAWI/CATI en une seule vague. Par le passé, seul le mode CATI était utilisé.

Soit une personne k appartenant à une strate h . Notons aussi n_h la taille de l'échantillon activé dans h . Le poids initial de k est noté par δ_k et est donné par :

$$\delta_k = \frac{N_h}{n_h}, k \in h,$$

où N_h est la taille de la population dans la strate h .

Variables utilisées

La modélisation de la non-réponse est effectuée à l'aide d'un algorithme de segmentation implémenté dans la macro CHAIDBIN [voir [Kass, 1980](#)]. Les variables utilisées dans la modélisation de la non-réponse et du calage sont listées dans la table 12.

TABLE 12 Variables utilisées dans le modèle de non-réponse ou dans le calage, MZB 2021

Variable	Non-réponse	Calage
ALTEL	X	
classe d'âge	X	X
état civil	X	X
langue de la commune	X	
nationalité	X	X
sexe	X	X
canton	X	X
taille de ménage	X	
typologie de ménage	X	

Les distributions des poids pour les différentes étapes de la pondération ainsi que les modalités des variables utilisées se trouvent respectivement dans les tables 23 et 28 en annexe.

5.2 Pondération ESS 2022

L'enquête est constituée d'une enquête téléphonique (CATI) et d'une enquête auto-administrée en ligne ou par écrit (CAWI/PAPI). Les personnes ayant répondu à l'enquête CATI sont invitées à répondre à l'enquête CAWI/PAPI. Les personnes proxy (personne de substitution proche de la personne cible) ayant répondu pour la personne cible à l'enquête CATI ne participent pas à l'enquête CAWI/PAPI. Les répondants à l'enquête CAWI/PAPI sont donc un sous-ensemble des répondants à l'enquête CATI sans prendre en compte les personnes proxy. Toutefois, elles sont incluses dans l'échantillon brut de cette phase et sont utilisées comme non-répondants dans la modélisation de la non-réponse.

Les personnes proxy reçoivent un ensemble de questions réduit par rapport au questionnaire reçu par les autres participants à la phase CATI. Pour cette première phase, deux pondérations ont alors été effectuées, la première sans considérer les 686 proxy dans l'échantillon et la deuxième en considérant ces données supplémentaires. Les poids obtenus de la pondération avec proxy sont utilisés dans la pondération de la deuxième phase CAWI/PAPI. Le poids initial ou poids de tirage δ_k^i d'une unité sélectionnée k est défini comme suit :

$$\delta_k^i = \begin{cases} \frac{1}{\pi_k^i} & \text{si } k \notin (EESE, ESO), \\ \frac{1}{\pi_k^i + \pi_k'^i} & \text{si } k \in EESE, \\ \frac{1}{\pi_k^i + \pi_k''^i} & \text{si } k \in ESO. \end{cases}$$

où π_k^i est la probabilité que l'unité k fasse partie de l'échantillon s_i , $\pi_k'^i$ la probabilité qu'elle fasse partie de l'échantillon OFSP I (EESE) pour la vague i et $\pi_k''^i$ est la probabilité que l'unité fasse partie de l'échantillon OFSP II (ESO) pour la vague i . Ce poids initial⁵ est, dans une première étape corrigé par un facteur qui permet de compenser la non-réponse, puis ajusté pour tenir compte du découpage en vagues et enfin réajusté à l'aide d'un calage pour intégrer des informations connues pour toute la population.

La combinaison des vagues est ainsi effectuée après la correction de la non-réponse et avant l'étape de calage sur marges. Les quatre vagues de l'enquête ne sont pas de tailles similaires. On calcule dans ce cas les facteurs de combinaisons proportionnellement à la taille de l'échantillon activé par strate de tirage et par vague. Soit une unité k appartenant à une vague i et à une strate h . Notons aussi $n_{h,i}$ la taille de l'échantillon activé dans h dans la vague i et n_h la taille de l'échantillon total activé dans la strate h . Le poids de k avant la combinaison des vagues est noté par d_k^i . On définit le poids de k après la combinaison des vagues comme suit :

$$d_k = \frac{n_{h,i}}{n_h} d_k^i. \quad (6)$$

Pour la phase CAWI/PAPI, le poids initial d'une unité k dans la vague i est donné par d_k^i , soit l'inverse du poids de l'unité k obtenu lors de la première phase avec proxy suite à la correction de la non-réponse et avant la combinaison des vagues. Similairement à la première phase, ces poids sont corrigés, pour chaque vague, par un facteur de non-réponse. Les poids obtenus sont ensuite corrigés par un facteur de combinaison des vagues d'après (6).

Variabes utilisées

Les variables utilisées dans la modélisation de la non-réponse et du calage sont listées dans la table 13. Pour la phase CATI, la modélisation de la non-réponse est effectuée séparément pour les groupes TEL (personnes avec un numéro de téléphone fixe) et ALTEL (personnes sans numéro de téléphone fixe) à l'aide des mêmes variables.

5. En prenant en compte la taille de l'échantillon brut par rapport à celle de la population cible, ainsi que l'allocation proportionnelle dans les strates, les probabilités d'inclusion π_k^i , $\pi_k'^i$ et $\pi_k''^i$ ont un ordre de grandeur de 10^{-2} , garantissant ainsi des poids δ_k^i largement supérieurs à 1.

TABLE 13 Variables utilisées dans le modèle de non-réponse ou dans le calage, ESS 2022

Variable	Non-réponse CATI	Non-réponse CAWI/PAPI	Calage personnes cibles
CH	X	X	
classe d'âge	X	X	
classe Marges ⁶			X
déciles des revenus CdC	X	X	
état civil	X	X	X
langue de la commune	X	X	
nationalité			X
niveau de formation ⁷		X	
sexe	X	X	X
strate	X	X	X
taille de ménage	X	X	X
typologie de ménage	X	X	X
vague	X		

Les distributions des poids pour les différentes étapes de la pondération ainsi que les modalités des variables utilisées se trouvent respectivement dans les tables 24 et 29 en annexe.

5.3 Pondérations EFG 2018 et 2023

L'enquête de 2018 est constituée d'une enquête téléphonique (CATI) et d'une enquête en ligne ou par écrit (CAWI/PAPI). Les personnes ayant répondu à l'enquête CATI sont invitées à répondre à l'enquête CAWI/PAPI. Pour la phase CATI, deux pondérations sont effectuées, l'une pour les personnes et l'autre pour les ménages, comme cela est expliqué dans la section 4.

La combinaison des vagues est effectuée après la correction de la non-réponse et avant l'étape de calage simultané sur marges. Comme les trois vagues de l'enquête sont de tailles similaires, les poids après la correction de la non-réponse sont multipliés par $\frac{1}{3}$. En d'autres termes, pour une unité k appartenant à une vague i et à une strate h avec un poids avant la combinaison des vagues noté par d_k^i , on définit le poids de k après la combinaison des vagues comme suit :

$$d_k = \frac{1}{3}d_k^i. \quad (7)$$

Les poids ménages s'obtiennent de manière analogue.

Pour la phase CAWI/PAPI, les poids initiaux des unités sont donnés par d_k^i . Similairement à la première phase, ces poids sont corrigés, pour chaque vague, par un facteur de non-réponse. Les poids obtenus sont ensuite corrigés par un facteur de combinaison des vagues d'après (7).

Variables utilisées

Les variables utilisées dans la modélisation de la non-réponse et du calage sont listées dans la table 14. Pour la phase CATI, la modélisation de la non-réponse est effectuée séparément pour les groupes TEL (personnes avec un numéro de téléphone fixe) et ALTEL (personnes sans numéro de téléphone fixe) pour chacune des 3 vagues.

6. Croisement entre classes d'âges, sexe, état civil et CH.

7. La variable niveau de formation provient des réponses au questionnaire de la phase CATI.

TABLE 14 Variables utilisées dans le modèle de non-réponse ou dans le calage, EFG 2018/2023

Variable	Non-réponse CATI	Non-réponse CAWI/PAPI	Calage personnes cibles	Calage ménage
classe d'âge	X	X		
état civil	X		X	
CH ⁸	X	X	X	
langue de la commune	X			
taille de ménage	X			
typologie de ménage	X	X		X
sexe	X	X		
strate_canton	X	X		
strate_men13_canton ⁹				X
strate_cadre ¹⁰			X	
age_sexe ¹¹			X	
Revenus CdC ¹²	X	X	X	X
Ind_knd ¹²	X	X	X	
Ind_ahh ¹²	X	X	X	
Ind_enk ¹²	X	X	X	

Le plan de sondage utilisé pour sélectionner l'échantillon de l'enquête EFG 2023 est similaire à celui utilisé pour l'EFG 2018 en ce qui concerne la répartition de cet échantillon par zones géographiques (grandes régions, cantons et villes de densification), population concernée (résidents permanents âgés de 15 à 79 ans au 31 décembre précédant l'enquête) et types de ménage (comme en 2018, les ménages qui comprennent un enfant de 13 ans ou moins sont sur-échantillonnés). Une différence notable entre les deux enquêtes concernant l'échantillonnage est le passage de trois tirages dans trois cadres successifs en 2018 à un seul tirage en 2023. En 2018, l'échantillon total était réparti de manière à peu près égale entre un premier échantillon tiré en décembre 2017, un deuxième en mars 2018 et un troisième en juin 2018. La période de collecte s'étendait ainsi de début 2018 à l'automne. En 2023, pour des raisons organisationnelles, le tirage a été effectué en une seule fois en février 2023, pour une période de collecte essentiellement au printemps 2023. L'échantillon final a été mis à disposition de la section METH le 21 décembre 2023.

Le déroulement de la collecte a quant à lui été complètement revu puisqu'en 2018 le questionnement était réalisé par entretien téléphonique (CATI), complété par un questionnaire électronique ou papier (CAWI-PAPI) pour une partie des questions. En 2023, la collecte était réalisée en première intention via un questionnaire électronique. Sur demande du répondant, un entretien téléphonique pouvait être substitué au questionnaire électronique. Seuls 1 468 des 18 912 répondants (7.8%) ont utilisé cette possibilité. En 2023, 18 912 personnes ont répondu à l'enquête parmi les 33 203 personnes sélectionnées. Le taux de réponse était de 57%, en nette amélioration par rapport au taux obtenu en 2018 : 46.7% (16 815 répondants sur 36 029 sélectionnées) pour la partie CATI et 42.1% (15 171 répondants pour l'ensemble CATI et CAWI/PAPI). Une étape supplémentaire de préparation des données, décrite ci-dessous, a conduit à l'élimi-

8. Indicatrice si la personne est Suisse.

9. Croisement entre strate_canton et une indicatrice sur la présence ou non d'un enfant de moins de 13 ans dans le ménage.

10. Croisement entre strates canton et classe d'âge d'une personne cible.

11. Croisement entre sexe et classe d'âge d'une personne cible.

12. Variables utilisées seulement dans la pondération de EFG 2023.

nation de 595 questionnaires de l'EFG 2023, pour un échantillon final de 18 317 observations et un taux de réponse de 55.2%.

Deux adaptations notables ont été apportées à la pondération de l'EFG 2023 par rapport à EFG 2018. La première concerne l'utilisation des données sur les revenus assurés des personnes travaillant en Suisse, mises à disposition de l'OFS par les caisses de compensation AVS. Ces données sont utilisées depuis 2019 lors de la pondération de l'échantillon du relevé structurel de la population suisse [voir [Qualité, 2023](#)]. En effet, le taux de réponse aux enquêtes des personnes dont le revenu est dans le haut de la distribution est généralement plus élevé que celui des personnes dont le revenu est dans le bas de la distribution. C'est le cas pour le relevé structurel et également pour l'enquête EFG 2023, comme le montre le tableau 15.

TABLE 15 Taille de l'échantillon brut activé, de l'échantillon net et taux de réponse en fonction du décile du revenu AVS de la personne cible dans l'enquête EFG 2023

Décile	Éch. brut	Éch. net	Tx. rép. (en %)
Hors AVS	9 237	4 828	52.3
1er (0 - 11 879 CHF/an)	2 021	968	47.9
2e (11 879 - 25 000 CHF/an)	2 242	1 157	51.6
3e (25 000 - 38 728 CHF/an)	2 302	1 232	53.5
4e (38 728 - 51 683 CHF/an)	2 460	1 381	56.1
5e (51 683 - 62 400 CHF/an)	2 332	1 255	53.8
6e (62 400 - 72 800 CHF/an)	2 427	1 312	54.1
7e (72 800 - 84 500 CHF/an)	2 391	1 382	57.8
8e (84 500 - 101 149 CHF/an)	2 479	1 472	59.4
9e (101 149 - 133 250 CHF/an)	2 544	1 604	63.1
10e (133 250 et + CHF/an)	2 768	1 726	62.4
Total	33 203	18 317	55.2

Les données AVS utilisées pour la pondération de l'enquête EFG 2023 étaient les plus récentes disponibles lors de ce calcul et portaient sur les revenus de l'année 2021. Deux variables catégorielles ont été construites à partir des revenus AVS : une variable catégorielle définie au niveau des personnes, avec les modalités du tableau 15 et une autre variable définie au niveau du ménage de manière similaire mais en considérant les déciles du *revenu équivalent du ménage*. Le revenu équivalent du ménage est obtenu en divisant la somme des revenus AVS du ménage par la *taille équivalent du ménage* (1 pour la personne la plus âgée plus 0.5 pour chaque personne de 14 ans ou plus, plus 0.3 pour chaque enfant de moins de 14 ans). Le modèle de non-réponse utilise les deux variables. Les poids personnes sont calés sur les marges de la variable au niveau personne tandis que les poids ménages sont calés sur les marges de la variable au niveau ménage.

La deuxième adaptation concerne l'utilisation des données sur les relations familiale disponibles dans la Statistique de la population et des ménages (STATPOP) et issues des données des registres de l'état civil. Lors des travaux préparatoires à la production des résultats, la section DEM, responsable de l'enquête, a constaté une sensible diminution, par rapport aux enquêtes précédentes, de la proportion des personnes qui déclarent avoir des enfants, petits-enfants ou un partenaire qui vivent dans un autre ménage.

Si les données de STATPOP ne fournissent pas d'information sur des éventuels partenaires (hors mariages ou partenariats enregistrés), elles contiennent par contre des identifiants (numéros AVS) des parents et des enfants pour une grande part de la population. Ces données, ne sont pas exhaustives, en particulier en ce qui concerne les personnes issues de la migration dont les parents ou les enfants peuvent vivre à l'étranger et ne disposent ainsi pas de numéro

AVS. Elles permettent toutefois de créer des variables indicatrices d'existence, selon STATPOP, d'enfants ou petits-enfants vivant dans le ménage ou en dehors de celui-ci. La comparaison de ces variables avec les réponses à l'enquête EFG permet de mettre en évidence une nette augmentation du nombre de répondants qui déclarent ne pas avoir d'enfant alors que les données de STATPOP en indiquent l'existence. Par exemple, parmi les répondants de nationalité Suisse ou bien étrangère et arrivés en Suisse avant d'avoir 25 ans, seulement 107 (0.9% de 12 239 répondants) déclaraient ne pas avoir d'enfants dans l'EFG 2018 alors qu'ils en avaient selon STATPOP, tandis qu'ils étaient 564 (3.9% de 14 446 répondants) dans cette situation dans l'EFG 2023. Une explication avancée pour cette évolution est le changement de mode de collecte entre les deux enquêtes.

Il a été décidé d'exclure de l'échantillon des répondants de l'EFG 2023 les personnes qui ont déclaré ne pas avoir d'enfants ou de petits-enfants alors que les données STATPOP indiquent le contraire. L'échantillon final pondéré compte 18 317 personnes. Trois variables indicatrices ont été ajoutées au modèle de non-réponse et au calage des poids : «avoir un enfant selon les données de STATPOP», «avoir un enfant vivant dans un autre ménage selon les données de STATPOP», «avoir un petit-enfant selon les données de STATPOP». Les distributions des poids pour les différentes étapes des pondérations de EFG 2018 et de EFG 2023 se trouvent en annexe dans les tables 25 et 26 respectivement. Les modalités des variables utilisées pour EFG 2018 et EFG 2023 se trouvent dans la table 30 en annexe.

5.4 Pondération ELRC 2019

L'enquête est constituée d'une enquête téléphonique (CATI) et d'une enquête en ligne ou par écrit (CAWI/PAPI). Les personnes ayant répondu à l'enquête CATI sont invitées à répondre à l'enquête CAWI/PAPI. Les personnes proxy ayant répondu à l'enquête CATI ne participent pas à l'enquête CAWI/PAPI. Les répondants à l'enquête CAWI/PAPI sont donc un sous-ensemble des répondants à l'enquête CATI.

Les personnes proxy reçoivent un ensemble de questions réduit par rapport au questionnaire reçu par les autres participants à la phase CATI. Pour cette première phase, deux pondérations ont alors été effectuées, la première avec un échantillon réduit (en enlevant les 246 proxy de l'échantillon) et la deuxième avec l'échantillon complet contenant ces données supplémentaires. Pour l'estimation d'un indicateur lié à une question posée à tout l'échantillon, les poids de la pondération avec proxy peuvent être utilisés. D'un autre côté, pour l'estimation d'un indicateur lié à une question posée seulement aux non-proxy, ou bien un croisement avec une telle question, les poids de la pondération sans-proxy doivent être utilisés.

Pour chacune des deux pondérations (avec ou sans les proxy), deux pondérations sont effectuées, l'une pour les personnes et l'autre pour les ménages, comme cela est expliqué dans la section 4.

La combinaison des vagues est effectuée après la correction de la non-réponse et avant l'étape de calage sur marges. Les trois vagues de l'enquête ne sont pas de tailles similaires. On calcule dans ce cas les facteurs de combinaisons proportionnellement à la taille de l'échantillon activé par strate de tirage et par vague. Soit une unité k appartenant à une vague i et à une strate h . Notons aussi $n_{h,i}$ la taille de l'échantillon activé dans h dans la vague i et n_h la taille de l'échantillon total activé dans la strate h . Le poids de k avant la combinaison des vagues est noté par d_k^i . On définit le poids de k après la combinaison des vagues comme suit :

$$d_k = \frac{n_{h,i}}{n_h} d_k^i. \quad (8)$$

Les poids ménages s'obtiennent de manière analogue.

Pour la phase CAWI/PAPI, les poids initiaux des unités sont donnés par d_k^i . Similairement à la

première phase, ces poids sont corrigés, pour chaque vague, par un facteur de non-réponse. Les poids obtenus sont ensuite corrigés par un facteur de combinaison des vagues d'après (8).

Variabes utilisées

Les variables utilisées dans la modélisation de la non-réponse et du calage sont listées dans la table 16. Pour la phase CATI, la modélisation de la non-réponse est effectuée séparément pour les groupes TEL (personnes avec un numéro de téléphone fixe) et ALTEL (personnes sans numéro de téléphone fixe) pour chacune des 3 vagues.

TABLE 16 Variables utilisées dans le modèle de non-réponse ou dans le calage, ELRC 2019

Variable	Non-réponse CATI	Non-réponse CAWI/PAPI	Calage personnes cibles	Calage ménage
CH	X	X		
classe d'âge	X	X	X	
classe Marges			X	
état civil	X	X	X	
langue de la commune			X	X
nationalité	X		X	
sexe	X	X	X	
strate	X	X	X	X
taille de ménage	X	X		X
typologie de ménage	X			X

Les distributions des poids pour les différentes étapes de la pondération ainsi que les modalités des variables utilisées se trouvent respectivement dans les tables 27 et 31 en annexe.

5.5 Pondération MRMT 2021

La pondération de l'enquête MRMT 2021 est décrite dans le document [Muralti \[2023\]](#). La pondération MRMT 2021 ne suit pas exactement le même schéma de pondération présenté dans la section 4. Les échantillons des quatre vagues sont combinés dans un seul échantillon sur lequel on va appliquer un modèle de non-réponse et un calage. La probabilité de sondage d'un individu k dans la vague i est $\pi_k^i = n_h^i / N_h^i$, où N_h^i est le nombre de personnes dans le cadre SRPH selon la strate h à la vague i et n_h^i est la taille de l'échantillon brut activé selon la strate h à la vague i . Les poids de sondage des personnes $\delta_k^i = 1/\pi_k^i = N_h^i/n_h^i$ dans toutes les vagues $i = 1, \dots, v$, sont combinés en un seul jeu de poids selon l'équation suivante :

$$\delta_k = \frac{\sum_{i=1}^v \delta_k^i \tilde{n}_h^i \frac{D_i}{365}}{\sum_{i=1}^v \tilde{n}_h^i} = \frac{\sum_{i=1}^v \tilde{n}_h^i \frac{N_h^i}{n_h^i} \frac{D_i}{365}}{\sum_{i=1}^v \tilde{n}_h^i}, \quad (9)$$

où \tilde{n}_h^i est la taille de l'échantillon net visée selon la strate h à la vague i et D_i est le nombre de jours de référence de la vague, équivalant à 91, 92, 91 et 91 jours respectivement.

La correction de la non-réponse au niveau des personnes cibles est réalisée sur l'échantillon combiné. Pour chaque personne, on obtient donc un poids corrigé pour la non-réponse $d_k = \delta_k/p_k$, où p_k est la probabilité de réponse estimée. Ensuite, un calage au niveau des personnes de ces poids est effectué pour obtenir les poids finaux au niveau des personnes $w_k = g_k d_k$, où g_k

est le facteur de calage pour la personne k . Les variables de calage sont des variables catégorielles qui reflètent des caractéristiques socio-démographiques, géographiques et temporelles liées à l'enquête.

Les probabilités de réponse pour les ménages sont les mêmes que les probabilités de réponse pour les personnes cibles, c.à.d. $p_m = p_k$. Ainsi, la pondération au niveau des ménages est obtenue en partageant les poids corrigés pour la non-réponse $d_k = \delta_k/p_k$ de chaque personne cible par le nombre de personnes sélectionnables du ménage, t_m :

$$d_m = \frac{d_k}{t_m}. \quad (10)$$

En d'autres termes, le poids personne est partagé entre les personnes du ménage.

Un calage au niveau des ménages est ensuite effectué pour obtenir les poids finaux au niveau des ménages $w_m = g_m d_m$, où g_m est le facteur de calage pour le ménage m . Les calages au niveau des personnes et des ménages sont effectués séparément. Plus de détails peuvent être trouvés dans [Muralti \[2023\]](#).

6 Estimation de variance

L'estimation de variance est une étape essentielle pour évaluer la précision des résultats des enquêtes. Le but de cette section est de développer un estimateur de variance pour des statistiques portant sur les personnes et les ménages dans le cadre des enquêtes thématiques. Y sont discutés les défis spécifiques liés à l'estimation de la variance pour les enquêtes menées en plusieurs vagues en prenant en considération la structure de l'échantillon et des plans d'échantillonnage complexes, les étapes de la pondération, c.à.d. la correction pour la non-réponse, la combinaison des vagues, le calage, la phase de l'enquête et l'estimateur souhaité (total ou moyenne).

6.1 Précision d'estimation d'un total pour une enquête en plusieurs vagues

Un cas particulier important pour justifier des calculs de variance approchés est celui où la population ne change pas entre les vagues d'enquêtes et toutes les vagues ont le même plan de sondage ainsi que les mêmes probabilités de réponse, où les échantillons s_i sont disjoints, les a_k^i valent $1/v$ et les poids ne sont pas calés. Dans ce cas, l'estimateur du total pour les personnes $\hat{Y} = \sum_{k \in s} w_k I_{k \in s} y_k$ est exactement l'estimateur obtenu en sélectionnant en une seule fois l'échantillon avec des probabilités de sélection égales à v fois celles d'une des vagues. Ainsi, sous ces hypothèses, une variance approchée est obtenue en traitant l'échantillon comme s'il avait été tiré en une seule fois.

Dans le cas général, lorsque les a_k^i et w_k^i ne dépendent pas de l'échantillon, la variance d'estimation d'un total $\hat{Y} = \sum_{k \in s} w_k I_{k \in s} y_k$ est égale à :

$$\text{var}(\hat{Y}) = \sum_{i=1}^v \sum_{j=1}^v \sum_{k \in U_i} \sum_{\ell \in U_j} a_k^i w_k^i y_k^i a_\ell^j w_\ell^j y_\ell^j \text{cov}(I_{k \in s_i}, I_{\ell \in s_j}). \quad (11)$$

Lorsque les poids w_k^i, w_ℓ^j sont obtenus par calage à partir des poids d_k^i, d_ℓ^j , une variance approchée peut-être obtenue en remplaçant dans l'équation (11) les termes $w_k^i y_k^i, w_\ell^j y_\ell^j$ par $w_k^i z_k^i, w_\ell^j z_\ell^j$ où z_k^i, z_ℓ^j sont des variables appelées linéarisées, obtenues par régression linéaire des y_k^i, y_ℓ^j sur les variables de calage [voir à ce sujet [Deville & Särndal, 1992](#), [Deville, 1999](#)]. Dans un premier temps, l'effet du calage sur la précision des estimations sera négligé pour obtenir un estimateur approché de variance.

Les termes $\text{cov}(I_{k \in s_i}, I_{\ell \in s_j})$ dépendent des plans de sondage joints qui ont été utilisés pour sélectionner les différents s_i . Pour les enquêtes thématiques de l'OFS, ils prennent en principe ¹³ les valeurs suivantes :

- $\pi_k^i p_k^i (1 - \pi_k^i p_k^i)$ si $i = j$ et $k = \ell$, où $\pi_k^i p_k^i$ est la probabilité d'observer k dans s_i ,
- $-\pi_k^i p_k^i \pi_\ell^i p_\ell^i$ si $i = j, k \neq \ell$ mais k et ℓ sont dans le même ménage,
- 0 si $i = j, k \neq \ell$ et k et ℓ sont dans des ménages différents,
- $-\pi_k^i p_k^i \pi_\ell^j p_\ell^j$ si $i \neq j, k = \ell$ et le ménage de k a conservé le même identifiant dans U_i et U_j ,
- 0 si $i \neq j, k = \ell$ et le ménage de k a changé d'identifiant entre U_i et U_j ,
- $-\pi_k^i p_k^i \pi_\ell^j p_\ell^j$ si $i \neq j, k \neq \ell$ et le ménage de k dans U_i a le même identifiant que le ménage de ℓ dans U_j ,
- 0 si $i \neq j, k \neq \ell$ et le ménage de k dans U_i a un identifiant différent de celui du ménage de ℓ dans U_j .

13. Les taux de sondage π_k^i aux différentes vagues et aux autres enquêtes sélectionnées entre celles-ci devraient rester assez faibles pour éviter de sélectionner un même ménage à plusieurs vagues.

En regroupant les termes, la variance devient :

$$\begin{aligned} \text{var}(\widehat{Y}) &= \sum_{i=1}^v \sum_{k \in U_i} (a_k^i w_k^i y_k^i)^2 \pi_k^i p_k^i (1 - \pi_k^i p_k^i) - \sum_{i=1}^v \sum_{m \in U_i} \sum_{k \neq \ell \in m} a_k^i w_k^i y_k^i \pi_k^i p_k^i a_\ell^i w_\ell^i y_\ell^i \pi_\ell^i p_\ell^i \\ &\quad - \sum_{i=1}^v \sum_{j \neq i} \sum_{m \in U_i \cap U_j} \left(\sum_{k \in m \in U_i} a_k^i w_k^i y_k^i \pi_k^i p_k^i \right) \cdot \left(\sum_{\ell \in m \in U_j} a_\ell^j w_\ell^j y_\ell^j \pi_\ell^j p_\ell^j \right). \end{aligned} \quad (12)$$

La liste des individus $\{k \in m \in U_i\}$ n'est pas nécessairement égale à $\{\ell \in m \in U_j\}$. Le troisième terme dans l'équation (12) ne peut pas être estimé sans biais sous le plan car un même ménage m ne peut pas être sélectionné dans deux vagues différentes. Le deuxième terme ne peut pas non plus être estimé car une seule personne par ménage est sélectionnée dans chaque vague. Si les taux de sondage sont faibles, les termes $\pi_k^i p_k^i \pi_\ell^j p_\ell^j$ sont négligeables devant $\pi_k^i p_k^i (1 - \pi_k^i p_k^i)$. Ainsi, le premier terme de (12) est usuellement le terme dominant. Une approximation de la variance est alors donnée par :

$$\text{var}_a(\widehat{Y}) \approx \sum_{i=1}^v \sum_{k \in U_i} (a_k^i w_k^i y_k^i)^2 \pi_k^i p_k^i (1 - \pi_k^i p_k^i). \quad (13)$$

Si les coefficients a_k^i ne dépendent pas de k mais seulement de la vague i , on reconnaît ici

$$\text{var}_a(\widehat{Y}) = \sum_{i=1}^v (a_i^i)^2 \text{var}_a^i(\widehat{Y}_i),$$

où les termes $\text{var}_a^i(\widehat{Y}_i)$ sont des variances approchées des totaux estimés sur les différentes vagues d'enquête.

L'approximation (13) peut être estimée approximativement sans biais par :

$$\widehat{\text{var}}_a(\widehat{Y}) = \sum_{i=1}^v \sum_{k \in s_i} (a_k^i w_k^i y_k^i)^2 (1 - \pi_k^i p_k^i). \quad (14)$$

Dans le cas où le tirage se fait en une seule vague, la formule (14) reste valide et devient :

$$\widehat{\text{var}}(\widehat{Y}) = \sum_{k \in s} (w_k y_k)^2 (1 - \pi_k p_k). \quad (15)$$

Pour estimer la variance d'un total sur l'échantillon de deuxième phase, il suffit de remplacer les p_k^i par $p_k^i \rho_k^i$.

Dans le cas d'un total pour les ménages, la variance de l'estimateur peut encore s'écrire :

$$\text{var}(\widehat{Y}) = \sum_{i=1}^v \sum_{j=1}^v \sum_{m \in U_i} \sum_{m' \in U_j} a_m^i w_m^i y_m^i a_{m'}^j w_{m'}^j y_{m'}^j \text{cov}(I_{m \in s_i}, I_{m' \in s_j}). \quad (16)$$

Pour les enquêtes de l'OFS, la variance des estimateurs portant sur les ménages est plus simple que celle des estimateurs sur les personnes : les sélections de ménages distincts sont indépendantes, donc $\text{cov}(I_{m \in s_i}, I_{m' \in s_j}) = 0$ si $m \neq m'$, quels que soient i et j et un ménage ne peut habituellement être sélectionné qu'à une seule vague : $\text{cov}(I_{m \in s_i}, I_{m \in s_j}) = -\pi_m^i p_m^i \pi_m^j p_m^j$ si $i \neq j$. Il en résulte que :

$$\text{var}(\widehat{Y}) = \sum_{i=1}^v \sum_{m \in U_i} (a_m^i w_m^i y_m^i)^2 \pi_m^i p_m^i (1 - \pi_m^i p_m^i) - \sum_{i=1}^v \sum_{j \neq i} \sum_{m \in U_i \cap U_j} a_m^i w_m^i y_m^i \pi_m^i p_m^i a_m^j w_m^j y_m^j \pi_m^j p_m^j. \quad (17)$$

Là encore, les données manquent pour estimer le deuxième terme de (17), mais si les taux de sondage sont faibles, de manière analogue à l'estimation de la variance sur les personnes, la variance d'une estimation d'un total peut être approchée par :

$$\text{var}_a(\hat{Y}) \approx \sum_{i=1}^v \sum_{m \in U_i} (a_m^i w_m^i y_m^i)^2 \pi_m^i p_m^i (1 - \pi_m^i p_m^i).$$

Cette variance approchée peut être estimée par :

$$\widehat{\text{var}}_a(\hat{Y}) = \sum_{i=1}^v \sum_{m \in s_i} (a_m^i w_m^i y_m^i)^2 (1 - \pi_m^i p_m^i). \quad (18)$$

Comme pour le cas d'un total au niveau des personnes, cette formule reste valable pour le cas d'une seule vague. Pour estimer la variance d'un total sur l'échantillon de deuxième phase, il suffit de remplacer les p_m^i par $p_m^i \rho_m^i$.

6.2 Linéarisation pour l'estimation d'un total/une moyenne global(e) ou sur un domaine

Pour estimer la variance d'une estimation autre que le total global, la variable d'intérêt y_k (observée sur une ou plusieurs vagues) est remplacée par une variable z_k appelée variable linéarisée.

- Pour un total dans un domaine d , la variable z_k est égale à y_k à l'intérieur du domaine et 0 ailleurs.
- Pour une moyenne globale d'une variable d'intérêt,

$$z_k = (y_k - \hat{Y}) / \sum_k w_k, \text{ avec } \hat{Y} = \sum_k y_k w_k / \sum_k w_k.$$

- Pour une moyenne dans un domaine d , la variable linéarisée z_k est égale à

$$z_k = (y_k - \hat{Y}_d) / \sum_{k \in d} w_k,$$

à l'intérieur du domaine et 0 ailleurs [voir Tillé, 2019].

La table 17 résume tous ces cas.

TABLE 17 Linéarisation pour l'estimation d'un total/moyenne global(e) ou sur domaine.

	global(e)	domaine
total	$z_k = y_k$	$z_k = y_k$ si $k \in d$ et 0 sinon
moyenne	$z_k = (y_k - \hat{Y}) / \sum_k w_k$	$z_k = (y_k - \hat{Y}_d) / \sum_{k \in d} w_k$ si $k \in d$ et 0 sinon

6.3 Prise en considération du calage

Il est important de prendre en compte les différents calages effectués lors de la pondération dans l'estimation de la variance pour refléter correctement leur impact sur la précision des estimateurs. Les calages séparés ou simultanés, qu'ils soient effectués avant ou après la combinaison des vagues, ont souvent pour effet de réduire la variance des estimations. En ajustant les estimations pour tenir compte des différences entre les groupes et les structures de ménages complexes, en utilisant par exemple des caractéristiques individuelles telles que l'âge, le sexe,

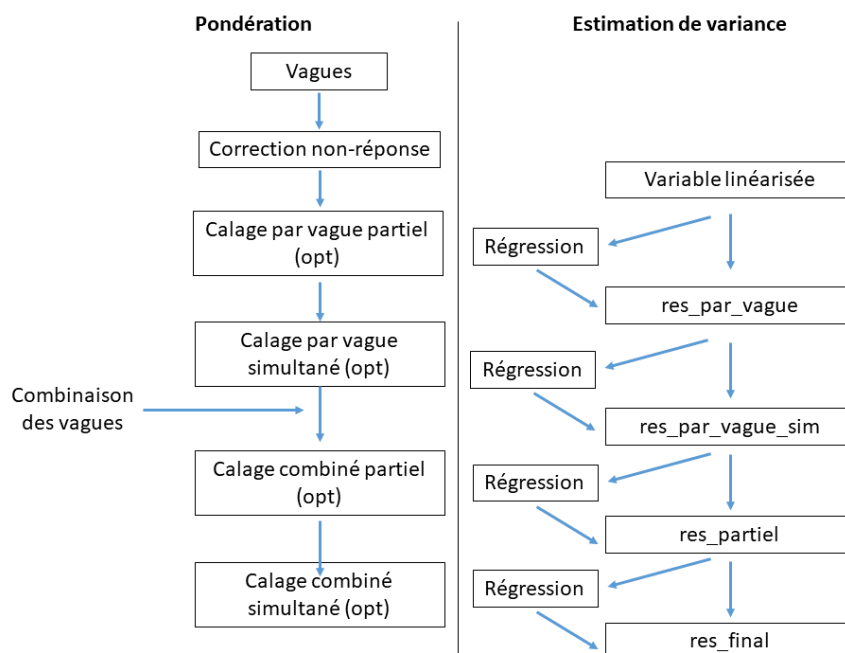


FIGURE 2 Lien entre l'estimation de variance et les différentes options de calages de la pondération.

le statut de résidence, ou le niveau de formation, ou des caractéristiques au niveau des ménages, telles que le revenu du ménage, la taille ou la composition du ménage, le calage permet de tirer profit des informations auxiliaires disponibles et d'améliorer la précision des résultats. Cependant, le calage peut également introduire des biais si les variables de calage sont nombreuses et mal choisies ou si les modèles ne capturent pas correctement la relation entre les variables de calage et les caractéristiques d'intérêt. Une sélection prudente des variables et une évaluation rigoureuse des modèles sont nécessaires pour garantir la qualité des estimations. Comme présenté dans la section 4, différents types de calage sont possibles : séparé (au niveau des personnes ou au niveau des ménages) ou simultané. Chacun de ces calages peut être effectué avant et après la combinaison des vagues. Au total quatre calages peuvent être effectués en respectant un ordre précis :

- (cal1) séparé (personnes ou ménages) avant la combinaison des vagues (un calage par vague) ;
- (cal2) simultané (personnes et ménages) avant la combinaison des vagues (un calage par vague) ;
- (cal3) séparé (personnes ou ménages) après la combinaison des vagues (un seul calage) ;
- (cal4) simultané (personnes et ménages) après la combinaison des vagues (un seul calage).

Tous ces calages sont optionnels, cependant le dernier, c.à.d. le calage simultané après combinaison des vagues, a l'avantage de permettre de garantir une cohérence entre les poids personnes et les poids des ménages constitués d'un seul individu.

Le processus d'estimation de variance (voir figure 2) prend en compte les calages en utilisant la variable d'intérêt ou sa linéarisée z_k (ou z_m pour les ménages), selon le cas (voir table 17) comme première entrée d'un processus en quatre étapes (reg1,..., reg4) correspondant aux quatre calages (cal1,..., cal4) décrits précédemment comme suit :

$$y_k \xrightarrow{\text{linéarisation}} z_k \xrightarrow{\text{reg1}} \epsilon_k^i \xrightarrow{\text{reg2}} e_k^i \xrightarrow{\text{reg3}} \epsilon_k \xrightarrow{\text{reg4}} e_k$$

Si un de ces calages optionnels n'est pas effectué, aucune régression n'est appliquée et la sortie de l'étape est identique à l'entrée. Cependant, si le calage est effectué, une régression est

appliquée sur la sortie de l'étape précédente en prenant comme covariables les variables utilisées dans le calage correspondant. Dans le cas d'un calage avant la combinaison des vagues (séparé ou simultané), une régression est effectuée pour chaque vague séparément. Les résidus de régression sont ensuite récupérés à la sortie. Dans chaque régression, des poids sont utilisés et ils dépendent de la méthode de calage (voir table 32 dans l'annexe). Dans le cas d'un calage simultané, la régression correspondante utilise des poids combinés calculés par la macro CALMAR [Le Guennec & Sautory, 2002].

La sortie finale du processus est alors la sortie de la quatrième étape, appelée résidu final (e_k pour les personnes ou e_m pour les ménages) et c'est cette variable qui est utilisée dans la formule d'estimation de la variance (équation 15 pour les personnes et équation 18 pour les ménages) avec le poids final de la pondération (w_k^i pour les personnes ou w_m^i pour les ménages), c'est-à-dire que si les w_m^i sont obtenus par calage, les termes $w_k^i y_k$ et $w_m^i y_m$ sont remplacés par $w_k^i e_k$ et $w_m^i e_m$, respectivement.

En résumé, l'estimation de variance des estimations obtenues en utilisant le schéma de pondération décrit dans la section 4 s'effectue comme suit :

- construire la variable linéarisée selon l'estimateur souhaité (voir table 17),
- effectuer la/les régression(s) selon le/les calage(s) utilisés dans la pondération et récupérer le résidu final e_k ou e_m ,
- utiliser les équations (15) et (18) en remplaçant y_k par e_k et y_m par e_m , respectivement.

Ce processus d'estimation de la variance pour les enquêtes thématiques a récemment été implémenté dans une macro SAS et mis à disposition des différentes sections.

6.4 Précision pour les estimateurs de l'enquête MRMT 2021

Comme l'enquête MRMT 2021 ne suit pas exactement le schéma de pondération décrit dans la section 4, un estimateur de variance différent a été développé pour cette enquête. Même si dans la pondération de cette enquête les sous-échantillons des différentes vagues sont combinés dans un seul échantillon sur lequel la correction de non-réponse et le calage sont effectués (voir section 5.5), l'estimation de variance doit considérer les poids initiaux π_k^i des sous-échantillons à cause des corrélations entre les sous-échantillons des vagues.

Les coefficients de combinaisons des poids des différentes vagues peuvent être déduits de l'équation (9) :

$$a_k^i = \frac{\sum_{h=1}^v \tilde{n}_h^i \frac{D_i}{365}}{\sum_{h=1}^v \tilde{n}_h^i} \text{ pour } k \in h. \quad (19)$$

La somme sur toutes les vagues de ces coefficients pour un individu k n'est pas forcément égale à 1. Néanmoins, les rapports $\frac{D_i}{365}$ pour $i = 1, \dots, 4$ sont tous très proches de 1/4, ce qui implique que la somme $\sum_{i=1}^v a_k^i \approx 1$.

Puisque toute la pondération de cette enquête s'effectue sur un échantillon combiné et utilise un seul calage (séparé après combinaison des vagues), l'estimateur de variance du total global \hat{Y} d'une variable d'intérêt, au niveau des personnes, donné par l'équation (14) peut s'écrire comme :

$$\widehat{\text{var}}(\hat{Y}) = \sum_{i=1}^v \sum_{k \in S_i} \left(\frac{a_k^i g_k}{\pi_k^i p_k} e_k \right)^2 (1 - \pi_k^i p_k), \quad (20)$$

où les π_k^i sont les probabilités de tirage dans chaque vague et $e_k = \epsilon_k$ sont les résidus obtenus par une régression linéaire de la variable d'intérêt par les variables de calage.

Cette dernière équation peut s'exprimer comme :

$$\widehat{\text{var}}(\hat{Y}) = \sum_{i=1}^v \sum_{k \in S_i} (w_k e_k)^2 (1 - \pi_k^i p_k). \quad (21)$$

La variance d'un total sur un domaine ou d'une moyenne (globale ou sur un domaine) d'une variable d'intérêt, est donnée par l'équation (21), en utilisant les résidus e_k obtenus par une régression linéaire de la variable z_k par les variables de calage, où la variable linéarisée z_k est définie selon le cas par la table 17.

Puisque la méthode de calage utilisée dans la pondération est la méthode linéaire tronquée, dans chacune des régressions les poids utilisés sont égaux à d_k si $L < d_k < U$ et 0 sinon, où L et U sont les bornes inférieures et supérieures choisies pour le facteur de calage g_k (voir annexe E).

L'estimation de la variance au niveau des ménages pour un total (global ou sur un domaine), ou pour une moyenne (globale ou sur un domaine), s'obtient de la même façon que pour les personnes, mais en remplaçant l'indice k des personnes par l'indice m des ménages et en utilisant les variables de calage au niveau des ménages à la place des variables de calage au niveau des personnes.

7 Conclusions

Ce rapport décrit le calcul du plan de sondage et les tirages des cinq enquêtes thématiques de l'OFS sur la période 2018 à 2023, les méthodes utilisées pour calculer les poids d'extrapolation et donne des éléments sur l'estimation de la précision de ces enquêtes. Les procédures utilisées pour les différentes enquêtes sont largement similaires. En particulier, les mêmes programmes sont utilisés pour sélectionner les échantillons de toutes les enquêtes et les programmes de pondération sont les mêmes pour les enquêtes ESS, EFG et ELRC.

Les estimateurs de variance proposés dans ce rapport sont bien adaptés aux enquêtes thématiques et sont implémentés dans une macro SAS que nous mettons à disposition des différentes sections. Ils peuvent être utilisés pour de futures enquêtes ou pour comparer les résultats avec les estimateurs de variance simplifiés utilisés auparavant.

Le système des enquêtes thématiques, s'il repose sur un ensemble de pratiques et de méthodes relativement homogène, n'est pas figé dans le temps. Il connaît des évolutions liées à des contraintes organisationnelles par exemple en ce qui concerne le tirage de l'échantillon dans un ou plusieurs cadres des sondage trimestriels. Il y a aussi des évolutions techniques telles que le passage à un questionnaire électronique. Il y a enfin des évolutions méthodologiques, telles que l'utilisation de nouvelles données pour améliorer la précision ou, à l'instar de l'enquête EFG, l'adaptation des plans de sondage pour certains objectifs spécifiques d'estimation.

Dans un futur proche, il est prévu d'utiliser les mêmes programmes de pondération pour les enquêtes MZB et MRMT que ceux des enquêtes ESS, EFG et ELRC (qui devient en 2024 deux enquêtes : ELR et EC). L'utilisation des données de l'AVS sur les revenus devrait aussi être étendue aux enquêtes MZB, ELRC et MRMT. Enfin, comme mentionné plus haut, des nouvelles variables auxiliaires, non disponibles actuellement, pourraient être intégrées dans les différentes procédures sans que la logique devrait être adaptée. Ainsi, ce rapport de méthodes décrit un procédé générique dont pourraient encore profiter d'autres enquêtes auprès des personnes et des ménages en l'adaptant à leurs conditions spécifiques.

Annexe

A Notations

TABLE 18 Notations

Symbole	Description
h	strate, $h = 1, 2, \dots, H$
i	vague, $i = 1, \dots, V$
g	canton densifiant son échantillon OFS ou grande région restante
N, N_h, N_g	taille de population cible
r	taille nette de l'échantillon OFS
r^+	taille nette des densifications
r^0, r_h^0, r_g^0	taille nette de l'échantillon OFS augmenté des densifications
t^{TEL}	taux des personnes du cadre de sondage avec au moins un numéro de téléphone
tr^{TEL}	taux de réponse des personnes avec au moins un numéro de téléphone
tr^{ALTEL}	taux de réponse des personnes sans numéro de téléphone
tr^{EENE}	taux de réponse des personnes de nationalité D'Europe de l'est et du sud-est
tr^{ESO}	taux de réponse des personnes de nationalité D'Europe sud-ouest
u_h^0	taux de réponse global obtenu avec t^{TEL} , tr^{TEL} et tr^{ALTEL}
$n^0, n_h^0, n_g^0, n^1, n_h^1, n^2, n_h^2$	taille brute d'échantillon prévue
$l^0, l_h^0, l_g^0, l^1, l_h^1, l^2, l_h^2$	taille de réserve prévue
$\tilde{n}^0, \tilde{n}_h^0, \tilde{n}_g^0, \tilde{n}^1, \tilde{n}_h^1, \tilde{n}^2, \tilde{n}_h^2$	taille brute d'échantillon obtenue
$\tilde{l}^0, \tilde{l}_h^0, \tilde{l}_g^0, \tilde{l}^1, \tilde{l}_h^1, \tilde{l}^2, \tilde{l}_h^2$	taille de réserve obtenue
$\bar{l}^0, \bar{l}_h^0, \bar{l}_g^0, \bar{l}^1, \bar{l}_h^1, \bar{l}^2, \bar{l}_h^2$	taille de réserve activée
k	indice de la personne
m	indice du ménage
i	indice de la vague
v	nombre de vagues
s, s_i	échantillon sélectionné s , échantillon tiré lors de la vague i
U_i	population dans laquelle l'échantillon s_i a été sélectionné
π_k	probabilité de tirage de la personne k
π_k^i	probabilité de tirage de la personne k dans la vague i
a_k^i	coefficient de combinaison des vagues pour la personne k
δ_k^i	poids initial de la personne k
p_k^i	probabilité de réponse de la personne k
ρ_k^i	probabilités de réponse conditionnelle lors de la deuxième phase pour la personne k
d_k^i	poids corrigé pour la non-réponse de la personne k dans la vague i
d_k	poids de la personne k après combinaison des vagues
g_k^i	facteur de calage par vague pour la personne k

g_k	facteur de calage après combinaison des vagues pour la personne k
w_k^i	poids calés obtenus par calage par vague
w_k	poids calés obtenus par calage après combinaison des vagues
y_k, y_k^i	variable d'intérêt au niveau des personnes
\hat{t}_y^i	estimateur du total de y_k^i
z_k	variable linéarisée au niveau des personnes
ϵ_k^i	résidus de régression des y_k sur les variables du calage partiel (personnes) par vague
ϵ_k	résidus de régression des y_k sur les variables du calage partiel après combinaison des vagues
e_k^i	résidus de régression des y_k sur les variables du calage simultané (personnes, ménages) par vague
e_k	résidus de régression des y_k sur les variables du calage simultané après combinaison des vagues (résidus finaux)
π_m	probabilité de tirage pour le ménage m
a_m^i	coefficient de combinaison des vagues pour le ménage m
δ_m^i	poids initial pour le ménage m
p_m^i	probabilité de réponse pour le ménage m
ρ_m^i	probabilités de réponse conditionnelle lors de la deuxième phase pour le ménage m
d_m^i	poids corrigé pour la non-réponse pour le ménage m
d_m	poids du ménage m après combinaison des vagues
g_m^i	facteur de calage par vague pour le ménage m
g_m	facteur de calage après combinaison des vagues pour le ménage m
w_m^i	poids calés obtenus par calage par vague
w_m	poids calés obtenus par calage après combinaison des vagues
y_m, y_m^i	variable d'intérêt au niveau des ménages
z_m	variable linéarisée au niveau des ménages
ϵ_m^i	résidus de régression des y_m sur les variables du calage partiel par vague
ϵ_m	résidus de régression des y_m sur les variables du calage partiel après combinaison des vagues
e_m^i	résidus de régression des y_m sur les variables du calage simultané par vague
e_m	résidus de régression des y_m sur les variables du calage simultané après combinaison des vagues (résidus finaux)

B Tables récapitulatives des deux échantillons supplémentaires de l'OFSP pour l'enquête ESS 2022

TABLE 19 Tailles, par vague, de l'enquête ESS 2022, pour l'échantillon de l'OFSP des étrangers de nationalité d'Europe de l'est ou du sud-est

Vague (<i>i</i>)	Taille population cible (N_i)	Taille échantillon					
		brut		réserve			total activé
		prévu (n_i^0)	tiré (\tilde{n}_i^0)	prévu (l_i^0)	tiré (\tilde{l}_i^0)	activé (\bar{l}_i^0)	($\tilde{n}_i^0 + \bar{l}_i^0$)
1	452 229	634	666	65	46	46	712
2	454 609	634	608	130	134	134	742
3	457 861	634	654	325	368	368	1 022
4	460 769	634	640	195	212	66	706
Total	-	2 536	2 568	715	760	614	3 182

TABLE 20 Tailles, par strate, de l'enquête ESS 2022, pour l'échantillon de l'OFSP des étrangers de nationalité d'Europe de l'est ou du sud-est

Strate (h)	Taille échantillon						
	net	brut		réserve			total
	prévu (r_h^0)	prévu (n_h^0)	tiré (\tilde{n}_h^0)	prévu (l_h^0)	tiré (\tilde{l}_h^0)	activé (\bar{l}_h^0)	activé ($\tilde{n}_h^0 + \bar{l}_h^0$)
AG	94	368	366	99	105	85	451
AI	1	4	5	0	0	0	5
AR	3	12	16	0	0	0	16
BE	61	196	199	55	55	42	241
BL	28	112	120	33	29	25	145
BS	27	104	107	33	44	36	143
FR	19	60	54	22	25	24	78
GE	27	92	102	22	24	22	124
GL	5	16	17	11	14	11	28
GR	8	28	26	11	11	7	33
JU	2	8	8	0	0	0	8
LU	38	104	110	33	43	30	140
NE	6	20	18	11	12	11	29
NW	2	4	4	0	0	0	4
OW	2	4	2	0	0	0	2
SG	71	236	248	66	52	42	290
SH	10	36	33	11	13	12	45
SO	36	116	103	33	32	28	131
SZ	17	48	36	11	11	8	44
TG	33	112	99	33	33	26	125
TI	19	48	46	11	14	11	57
UR	2	4	4	0	0	0	4
VD	52	172	171	44	53	44	215
VS	14	48	56	11	15	10	66
ZG	14	36	37	11	11	7	44
Ville de Zurich	39	136	131	44	50	39	170
ZH sans ville de Zurich	119	412	450	110	114	94	544
Total	750	2 536	2 568	715	760	614	3 182

TABLE 21 Tailles, par vague, de l'enquête ESS 2022, pour l'échantillon de l'OFSP des étrangers de nationalité d'Europe du sud-ouest

Vague (<i>i</i>)	Taille population cible (N_i)	Taille échantillon					
		brut		réserve			total
		prévu (n_i^0)	tiré (\tilde{n}_i^0)	prévue (l_i^0)	tirée (\tilde{l}_i^0)	activée (\bar{l}_i^0)	activé ($\tilde{n}_i^0 + \bar{l}_i^0$)
1	55 8691	141	145	16	20	20	165
2	55 8565	141	137	32	31	31	168
3	55 9362	141	146	80	73	73	219
4	55 8830	141	155	48	36	10	165
Total	-	564	583	176	160	134	717

TABLE 22 Tailles, par strate, de l'enquête ESS 2022, pour l'échantillon de l'OFSP des étrangers de nationalité d'Europe du sud-ouest

Strate (<i>h</i>)	Taille échantillon						
	net prévu (r_h^0)	brut prévu tiré (n_h^0) (\tilde{n}_h^0)		réserve prévu tiré activé (l_h^0) (\tilde{l}_h^0) (\bar{l}_h^0)			total activé ($\tilde{n}_h^0 + \bar{l}_h^0$)
AG	16	44	48	11	8	5	53
AI	0	0	0	0	0	0	0
AR	1	0	0	0	0	0	0
BE	16	36	31	11	6	3	34
BL	6	16	17	11	10	8	25
BS	6	16	21	11	7	7	28
FR	12	28	31	11	8	7	38
GE	25	52	51	11	10	8	59
GL	1	4	5	0	0	0	5
GR	6	16	19	0	0	0	19
JU	2	4	5	0	0	0	5
LU	7	20	26	11	9	8	34
NE	8	16	13	11	10	10	23
NW	1	0	0	0	0	0	0
OW	1	0	0	0	0	0	0
SG	9	20	28	11	13	10	38
SH	1	4	6	0	0	0	6
SO	5	12	14	0	0	0	14
SZ	2	8	11	0	0	0	11
TG	6	16	9	0	0	0	9
TI	27	44	43	11	12	9	52
UR	0	0	0	0	0	0	0
VD	37	80	93	22	10	9	102
VS	14	28	23	11	18	15	38
ZG	3	8	8	0	0	0	8
Ville de Zurich	11	28	26	11	15	12	38
ZH sans ville de Zurich	27	64	55	22	24	23	78
Total	250	564	583	176	160	134	717

C Distribution des poids des pondérations des enquêtes thématiques

TABLE 23 Distribution des poids personnes pour les différentes étapes de la pondération MZB 2021

Etape	n	moyenne	total	P0	P1	P5	P10	P25	P50	P75	P90	P95	P99	P100
Poids initiaux, phase 1	32 004	201.0	6 434 338.9	42.2	42.2	42.2	42.2	42.2	156.7	334.9	348.7	357.7	364.4	401.6
Poids après correction de la non-réponse, phase 1	18 200	355.6	6 472 431.2	60.0	61.2	61.2	68.7	105.1	290.5	561.7	670.3	703.8	835.1	1 138.4
Poids après calage, phase 1	18 200	351.8	6 402 283.0	55.2	59.3	62.5	66.5	105.9	289.7	559.2	658.3	711.3	801.0	927.1

TABLE 24 Distribution des poids personnes (sans proxy) pour les différentes étapes de la pondération ESS 2022

Etape	vague	n	moyenne	total	P0	P1	P5	P10	P25	P50	P75	P90	P95	P99	P100
Poids initiaux, phase 1	1	13 926	512.8	7 140 657	103.6	121.8	154.4	169.7	256.0	435.0	666.2	1 194.0	1 218.9	1 253.2	1 395.2
	2	15 307	471.2	7 213 368	96.1	111.5	141.7	167.0	236.5	403.5	609.4	1 094.6	1 146.1	1 146.1	1 276.1
	3	17 036	425.6	7 250 594	89.5	102.7	124.5	143.6	216.7	362.7	562.8	1 009.5	1 048.2	1 057.2	1 173.4
	4	14 547	488.3	7 103 708	89.5	102.5	118.3	169.5	255.5	425.0	666.7	1 192.9	1 204.1	1 247.3	1 385.7
Poids après correction de la non-réponse (vagues combinées), phase 1		21 244	344.3	7 314 049	39.9	51.7	73.4	102.7	164.8	269.4	444	652.4	851.8	1 395.8	3 316.1
		21 244	338.1	7 182 252.0	38	49.3	73.6	101.1	165	270.5	441.8	650.3	823.9	1 301	3 052.5
Poids après correction de la non-réponse, phase 2	1	4 467	1 641.1	7 330 549.2	192.2	243.6	337.5	481.4	784.0	1 268.0	2 134.7	3 138.2	4 112.0	6 857.9	18 555.8
	2	4 937	1 490.3	7 357 741.8	177.3	220.2	307.0	430.2	716.1	1 163.3	1 951.8	2 807.5	3 726.6	6 195.1	15 209.5
	3	5 381	1 350.6	7 267 752.5	157.7	201.8	283.7	392.2	647.0	1 064.0	1 782.9	2 544.5	3 278.5	5 281.5	18 225.7
	4	4 352	1 692.4	7 365 300.4	164.9	204.2	309.2	456.0	733.2	1 267.5	2 215.4	3 296.1	4 423.0	7 613.4	26 385.7
Poids après correction de la non-réponse (vagues combinées), phase 2		19 137	383.3	7 335 606.7	41.7	53.7	77.6	109.3	179.7	295.9	499.8	732	961	1 599.6	6 133.5
		19 137	375.3	7 182 252.0	40.2	51.7	79	108.2	179.2	293.8	489.7	730.3	939.3	1 515.7	4 020

TABLE 25 Distribution des poids personnes et ménages pour les différentes étapes de la pondération EFG 2018

Poids	Etape	vague	n	moyenne	total	P0	P1	P5	P10	P25	P50	P75	P90	P95	P99	P100
Poids personne	Poids initiaux, phase 1	1	11 719	549.5	6 439 026.0	40.4	40.4	100.8	102.9	275.2	456.3	1 069.8	1 173.0	1 265.7	1 642.3	1 744.0
		2	12 170	538.2	6 549 983.3	41.0	41.0	100.3	104.3	274.1	447.5	959.6	1 169.4	1 259.4	1 645.7	1 743.0
		3	12 166	557.3	6 779 890.8	41.5	41.5	99.8	106.2	274.3	451.6	1 060.4	1 163.1	1 256.1	1 646.4	1 748.8
	Poids après correction de la non-réponse (vagues combinées), phase 1	1	16 815	388.6	6 534 877.5	19.6	23.2	52.5	65.1	146.9	267.3	581.0	809.4	1 057.8	1 668.2	4 187.8
		2	16 815	392.3	6 596 828	19.3	24.4	53.2	64.3	144.2	267.4	585.0	816.1	1 098.6	1 761.9	4 527.7
		3	5 092	426.3	2 170 603.1	21.4	26.1	55.3	67.2	158.3	291.2	626.6	906.2	1 179.5	1 886.4	5 060.1
	Poids après correction de la non-réponse, phase 2	1	5 044	427.4	2 155 718.3	20.9	23.1	54.5	72.0	152.8	285.9	632.8	901.3	1 211.5	2 011.9	4 872.9
		2	5 035	436.3	2 206 644.1	22.3	27.2	59.1	81.0	172.8	309.7	643.8	925.7	1 132.9	1 769.3	3 965.8
		3	15 171	430.6	6 532 965.4	20.9	25.4	56.5	72.0	160.3	297.9	635.4	907.1	1 177.6	1 910.6	5 060.1
	Poids après correction de la non-réponse (vagues combinées), phase 2	1	15 171	434.8	6 596 828	20.3	25.5	57.6	71.7	157.4	294.8	634.5	904.8	1 225.4	2 013.9	6 328.9
		2	11 719	286.1	3 353 060.1	8.1	20.2	42.4	51.5	93.4	213.2	346.6	586.5	1 069.8	1 256.7	1 744.0
		3	12 170	283.6	3 451 710.6	8.7	20.5	37.3	52.1	94.9	210.9	339.0	584.7	1 065.8	1 259.4	1 743.0
Poids ménage	Poids initiaux, phase 1	1	16 815	204.3	3 436 009.6	6.5	10.9	23.9	30.5	62.9	131.6	254.3	408.8	661.6	1 401.3	3 650.2
		2	16 815	206.3	3 468 536	3.2	11.1	22.9	29.6	60.7	129.5	255.2	421.5	681.1	1 449.8	4 058.4
		3	5 092	222.5	1 133 178	7.5	11.8	25.5	32.3	70.5	142.4	279.5	468.0	714.4	1 492.9	3 314.5
	Poids après correction de la non-réponse, phase 2	1	5 044	223.8	1 138 877	8.1	11.2	24.8	33.6	65.8	141.1	273.6	451.2	744.3	1 630.7	4 073.0
		2	5 035	229.5	1 155 590	8.6	13.0	27.5	36.6	75.1	153.5	297.1	460.6	691.9	1 432.5	2 590.5
		3	15 171	225.9	3 427 645.0	7.5	11.8	26.2	33.6	70.6	145.9	284.2	459.8	722.1	1 523.2	4 073.0
	Poids après correction de la non-réponse (vagues combinées), phase 2	1	15 171	228.6	3 468 536	3.3	11.7	24.4	32.5	66.9	142.0	282.2	470.3	744.1	1 611.6	5 490.0
		2	11 719	286.1	3 353 060.1	8.1	20.2	42.4	51.5	93.4	213.2	346.6	586.5	1 069.8	1 256.7	1 744.0
		3	12 170	283.6	3 451 710.6	8.7	20.5	37.3	52.1	94.9	210.9	339.0	584.7	1 065.8	1 259.4	1 743.0

TABLE 26 Distribution des poids personnes et ménages pour les différentes étapes de la pondération EFG 2023

Poids	Etape	n	moyenne	total	P0	P1	P5	P10	P25	P50	P75	P90	P95	P99	P100
Poids personne	Poids initiaux, phase 1	33 297	201.9	6 722 183.6	17.3	17.3	41.1	42.1	83.5	118.8	250.2	508.5	557.1	659.6	716.5
	Poids après correction de la non-réponse, phase 1	18 317	364.5	6 677 265.3	22.7	26.5	64.6	71.4	125.8	241.4	443.8	904.3	1 056.7	1 386.5	2 962.8
	Poids après calage, phase 1	18 317	372.7	6 826 476	10.9	24.8	60.4	70.3	122.6	238.1	452.7	940.6	1 114.5	1 537.1	3 372.4
Poids ménage	Poids initiaux, phase 1	33 297	109.6	3 650 703.8	3.3	8.7	14.8	20.6	29.4	61.2	125.1	254.3	468.8	562.0	716.5
	Poids après correction de la non-réponse, phase 1	18 317	198.2	3 629 778.1	6.8	12.8	27.0	33.5	52.8	121.8	224.7	451.4	645.6	1 153.0	2 962.8
	Poids après calage, phase 1	18 317	200.7	3 675 856	2.5	11.9	24.7	32.0	52.3	121.2	229.6	470.2	672.3	1 156.4	3 264.2

TABLE 27 Distribution des poids personnes et ménages (sans proxy) pour les différentes étapes de la pondération ELRC 2019

Poids	Etape	vague	n	moyenne	total ¹⁴	P0	P1	P5	P10	P25	P50	P75	P90	P95	P99	P100	
						1	2	3	1	2	3	1	2	3	1	2	3
Poids initiaux, phase 1		1	9 626	746.8	7 189 151.1	309.2	309.2	309.2	309.2	468.8	935.6	997.4	1 071.6	1 071.6	1 129.7	1 151.2	
		2	9 436	740.6	6 988 205.2	308.3	308.3	308.3	308.3	468.3	935.6	999.8	1 068.6	1 068.6	1 126.5	1 147.9	
		3	12 897	546.1	7 043 478.1	198.6	198.6	198.6	198.6	351.9	566.4	737.5	775.6	775.6	898.8	1 175.9	
Poids après correction de la non-réponse (vagues combinées), phase 1		13 171	539.3	7 102 762.9	127.0	149.3	183.1	233.3	295.9	501.8	652.0	909.8	1 165.8	1 165.8	1 717.5	4 115.4	
		13 171	534.3	7 037 630.0	43.9	144.5	179.7	219.9	289.7	482.5	652.9	917.2	1 158.7	1 158.7	1 730.6	4 263.6	
		13 171	534.3	7 037 630.0	614.9	693.9	862.7	962.2	1 236.7	2 097.2	2 646.3	3 810.1	4 775.4	4 775.4	7 085.4	13 740.8	
Poids après correction de la non-réponse, phase 2		2	3 356	2 037.9	6 839 124.0	586.5	641.5	800.5	884.8	1 140.0	1 936.8	2 413.5	3 452.5	3 452.5	4 234.0	5 740.2	
		3	4 601	1 541.98	7 094 652.2	356.4	386.2	475.4	643.4	853.6	1 354.0	1 811.4	2 643.1	2 643.1	3 450.9	5 459.8	
		11 255	632.4	7 117 734.1	161.4	179.3	230.3	269.7	350.1	573.7	753.7	1 079.4	1 368.8	1 368.8	2 071.9	5 008.4	
Poids après calage (vagues combinées), phase 2		11 255	625.3	7 037 630.0	63.3	169.2	218.8	252.5	336.4	551.9	751.2	1 085.5	1 380.0	1 380.0	2 197.0	5 717.6	
		11 255	392.9	3 781 773.4	28.1	61.8	103.1	142.9	222.1	330.5	498.7	820.3	996.4	996.4	1 071.6	1 151.2	
		2	9 436	391.1	3 690 651.3	28.0	63.4	102.8	142.5	221.9	332.2	499.9	570.1	570.1	996.6	1 147.9	
Poids après correction de la non-réponse (vagues combinées), phase 1		3	12 897	292.2	3 769 055.7	22.1	39.7	66.2	97.9	167.1	235.2	372.4	584.8	584.8	737.5	898.8	
		13 171	284.5	3 747 664.9	18.2	41.5	64.7	80.4	130.1	224.9	329.8	578	757.9	757.9	1 265.2	3 735.4	
		13 171	283.8	3 737 861	9.6	40.3	62	78.5	125.8	218.3	331.3	581	771.8	771.8	1 352	4 263.6	
Poids après correction de la non-réponse, phase 2		1	3 298	1 175.0	3 875 321.7	132.5	211.4	290.8	365.2	537.9	931.7	1 323.1	2 446.1	2 446.1	3 177.2	5 458.3	
		2	3 356	1 057.0	3 547 259.6	95.1	191.1	256.0	331.4	503.8	907.7	1 195.9	2 135.5	2 135.5	2 678.9	4 341.6	
		3	4 601	824.2	3 792 164.5	88.5	119.6	187.3	226.5	366.0	633.3	934.2	1 629.0	1 629.0	2 267.1	4 244.2	
Poids après correction de la non-réponse (vagues combinées), phase 2		11 255	333.9	3 758 337.5	26.4	54.5	79.3	99	153.3	265.4	380.5	668.5	897.2	897.2	1 571.8	4 419.5	
		11 255	332.1	3 737 861	9.4	49.9	74.8	94.6	148.1	253.7	381.9	670.8	914.5	914.5	1 603.4	5 717.6	

14. La taille de la population cible a légèrement diminué entre la vague 1 et la vague 3 ce qui explique les variations au niveau des totaux des poids initiaux.

D Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage

TABLE 28 Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage, MZB 2021

Variable	Modalités
classe d'âge	15-19 ans 20-24 ans 30-34 ans 40-44 ans 50-54 ans 60-64 ans 65-69 ans 70-74 ans
État civil	Célibataire Marié Veuf Divorcé
Langue de la commune	Allemand/Romanche Français Italien
Nationalité ¹⁵	Suisse Autre
Canton	AG AI/AR BE BL BS FR GE GL GR JU LU NE NW/OW SG SH SO SZ TG TI UR VD VS ZG Ville de Zurich Reste ZH
Taille du ménage	1 personne 2 personnes 3 personnes 4 personnes 5 personnes et plus
Typologie de ménage	Homme seul >=25 ans Femme seule >=25 ans Deux personnes de plus de 25 ans de sexe différent Deux personnes de plus de 25 ans de même sexe Trois personnes ou plus de plus de 25 ans Une personne ou plus < 25 ans Homme seul >=25 ans avec personne(s) < 25 ans Femme seule >=25 ans avec personne(s) < 25 ans Deux personnes >= 25 ans avec personne(s) < 25 ans Trois personnes ou plus >=25 ans avec personne(s) < 25 ans

15. Pour la modélisation de la non-réponse, une modalité supplémentaire englobant Allemagne, France et Autriche a été utilisée.

TABLE 29 Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage, ESS 2022

Variable	Modalités
CH	Suisse Étranger
classe d'âge	15-19 ans 20-24 ans 25-34 ans 35-44 ans 45-54 ans 55-64 ans 65-74 ans 75-84 ans 85 ans et plus
État civil	Marié ou en partenariat Non-marié et pas en partenariat
Classe Marges	Croisement entre classes d'âges, sexe, état civil et CH
Déciles des revenus CdC	10 modalités pour les déciles des revenus soumis à une cotisation AVS Une catégorie à part pour les valeurs manquantes
Langue de la commune	Allemand/Romanche Français Italien
Nationalité	Europe de l'est ou du sud-est Europe du sud-ouest Suisse/reste
Niveau de formation ¹⁶	École obligatoire Degré secondaire II Degré tertiaire
Grandes régions	Région lémanique : GE, VD, VS Espace Mittelland : BE, FR, JU, NE, SO Suisse du Nord-Ouest : AG, BL, BS Zurich : ZH Suisse orientale : AI, AR, GL, GR, SG, SH, TG Suisse centrale : LU, NW, OW, SZ, UR, ZG Tessin : TI
Strates	Les 7 grandes régions avec AG, AR, BL, FR, GE, GR, LU, NE, NW, SG, SZ, TG, TI, VD, VS, ZG et la ville de Zurich considérés comme des strates à part.
Taille du ménage	1 personne 2 personnes 3 personnes 4 personnes 5 personnes et plus
Typologie de ménage	Homme seul majeur/ Femme seule majeure/ Une ou plusieurs personnes mineures Trois personnes majeures ou plus/ Deux personnes majeures de même sexe Homme/Femme seul(e) majeur(e), avec personne(s) mineure(s) Trois personnes majeures ou plus, avec ou sans personne(s) mineure(s)/Deux personnes majeures de même sexe avec personne(s) mineure(s) Deux personnes majeures de sexe opposé avec personne(s) mineure(s) Deux personnes majeures de sexe opposé/ Deux mineures ou plus Autre

16. Relevé à la première phase de l'enquête.

TABLE 30 Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage, EFG 2018/2023

Variable	Modalités
CH	Suisse Étranger
classe d'âge	15-19 ans 20-24 ans 25-34 ans 35-44 ans 45-54 ans 55-64 ans 65-74 ans 75-80 ans 80 ans et plus
État civil	Marié ou en partenariat Non-marié et pas en partenariat
Langue de la commune	Allemand/Romanche Français Italien
Ménage ₁₄	Personnes âgées de 15 à 25 ans uniquement Personnes de 25 ans ou plus, dont le ménage comprend au moins un enfant < 13 ans Personnes de 25 ans ou plus, dont le ménage ne comprend pas d'enfant < 13 ans
Taille du ménage	1 personne 2 personnes 3 personnes 4 personnes 5 personnes et plus
Typologie de ménage	Deux personnes de plus de 25 ans de même sexe/Trois personnes ou plus > 25 ans Femme seule >=25 ans avec personne(s) < 25 ans/ Deux personnes de sexe différent >= 25 ans avec personne(s) < 25 ans Homme seul >=25 ans ou 2 personnes de même sexe >25 ans ou trois personnes ou plus >=25 ans, avec personne(s) < 25 ans Autre
strate_canton	AI/AR/GL/GR/SH AG BE BL BS FR GE JU LU NE NW/OW/SZ/UR/ZG SG SO TG TI VD VS Ville de Zürich Reste ZH
strate_men13_canton	Croisement entre strate_canton et une indicatrice sur la présence ou non d'un enfant de moins de 13 ans dans le ménage
strate_cadre	Croisement entre strates canton et Ménage ₁₄
age_sexe	Croisement entre sexe et classe d'âge d'une personne cible
Revenus CdC ¹⁷	10 modalités pour les déciles des revenus soumis à une cotisation AVS Une catégorie à part pour les valeurs manquantes hors AVS Similairement pour les ménages, avec les déciles du revenu équivalent du ménage
Ind_knd ¹⁷	Indicatrice si la personne a un enfant identifié dans STATPOP
Ind_ahh ¹⁷	Indicatrice si l'un des enfants de la personne est lui aussi dans STATPOP, mais pas dans le même ménage que la personne
Ind_enk ¹⁷	Indicatrice si la personne a un petit-enfant identifié dans STATPOP

17. Variables utilisées seulement dans la pondération de EFG 2023.

TABLE 31 Modalités des variables auxiliaires utilisées pour la non-réponse ou le calage, ELRC 2019

Variable	Modalités
CH	Suisse Étranger
classe d'âge	15-19 ans 20-24 ans 25-34 ans 35-44 ans 45-54 ans 55-64 ans 65-74 ans 75 ans et plus
État civil	Marié ou en partenariat Non-marié et pas en partenariat
Classe Marges	Croisement entre classes d'âges, sexe, état civil et CH
Langue de la commune	Allemand/Romanche Français Italien
Nationalité	Suisse Balkans + Turquie Belgique, Allemagne, France, Italie, Liechtenstein, Luxembourg, Autriche, Pays-Bas Espagne, Portugal, Croatie Royaume-Uni, Pologne, Hongrie, Russie, Erythrée, Chine, Etats-Unis, Brésil, Sri Lanka Autres
Grandes régions	Région lémanique : GE, VD, VS Espace Mittelland : BE, FR, JU, NE, SO Suisse du Nord-Ouest : AG, BL, BS Zurich : ZH Suisse orientale : AI, AR, GL, GR, SG, SH, TG Suisse centrale : LU, NW, OW, SZ, UR, ZG Tessin : TI
Strates	Les 7 grandes régions avec GE, VS, VD et la ville de Zurich considérés comme des strates à part.
Taille du ménage	1 personne 2 personnes 3 personnes 4 personnes 5 personnes et plus
Typologie de ménage	Homme seul majeur/ Femme seule majeure/ Une ou plusieurs personnes mineures Deux personnes majeures de sexe différent, avec ou sans personne(s) mineure(s) Deux personnes majeures du même sexe Trois personnes majeures ou plus, avec ou sans personne(s) mineure(s) Homme/Femme seul(e) majeur(e)/ Deux personnes majeures de même sexe, avec personne(s) mineure(s)

E Variance d'un estimateur calé

E.1 Linéarisation d'un estimateur calé

Une méthode d'estimation de variance par linéarisation est présentée dans Vallée & Tillé [2019]. Pour une statistique $\hat{\theta}$ calculée sur un échantillon s , l'estimateur de variance s'écrit

$$\widehat{\text{var}}(\hat{\theta}) = \sum_{k, \ell \in s} \frac{\pi_{k\ell} - \pi_k \pi_\ell}{\pi_{k\ell}} \hat{z}_k \hat{z}_\ell, \quad (22)$$

où π_k et π_ℓ sont les probabilités d'inclusion du premier ordre d'unités k et ℓ , $\pi_{k\ell}$ est leur probabilité d'inclusion jointe et \hat{z}_k, \hat{z}_ℓ sont les valeurs d'une variable obtenue en dérivant θ par rapport respectivement à a_k et a_ℓ , les indicatrices de sélection des unités k et ℓ .

Dans le cas de l'estimateur calé du total de la variable $(y_k)_{k \in s}$, Vallée & Tillé [2019] constatent que l'on peut choisir

$$\hat{z}_k = w_k e_k,$$

où e_k est le résidu de la régression des y_k sur les variables de calage \mathbf{x}_k , pondérée par les poids $a_k d_k F'_k(\mathbf{x}_k^t \boldsymbol{\lambda})$, $d_k = 1/\pi_k$, F_k est la fonction de calage de l'unité k , F'_k sa dérivée et $\boldsymbol{\lambda}$ est le vecteur solution du calage (tel que $w_k/d_k = g_k = F_k(\mathbf{x}_k^t \boldsymbol{\lambda})$ pour tout k dans s).

Remarque On fera attention à l'appellation "variable linéarisée" ou "linéarisée". Dans Vallée & Tillé [2019], ce terme désigne \hat{z}_k , mais ailleurs il est parfois utilisé pour désigner \hat{z}_k/d_k , \hat{z}_k/w_k , e_k , ou encore $g_k e_k$: une variable dont l'estimateur de Horvitz-Thompson du total a approximativement la même variance que l'estimateur dont on cherche à estimer la variance.

Dans le cas d'un plan de Poisson, l'estimateur de variance (22) est simplement :

$$\sum_{k \in s} (1 - \pi_k) \hat{z}_k^2 = \sum_{k \in s} d_k (d_k - 1) g_k^2 e_k^2.$$

Remarque Dans certains cas, par exemple lors d'investigations ou de simulations, il peut être utile de calculer un estimateur de variance sans avoir à calculer les poids de calage. Si l'on utilise la méthode linéaire le calcul des e_k ne nécessite pas de connaître les g_k . On peut alors simplement retenir l'approximation $\sum_{k \in s} d_k (d_k - 1) e_k^2$.

E.2 Application aux fonctions de calage usuelles

La macro CALMAR permet de choisir parmi cinq fonctions : linéaire, raking-ratio, logit, linéaire tronquée et sinus hyperbolique. Ce sont surtout les quatre premières fonctions qui sont utilisées à l'OFS. Elles sont définies comme suit :

1. *Linéaire*

$$F_k(t) = 1 + q_k t,$$

$$q_k > 0,$$

2. *Raking-ratio*

$$F_k(t) = \exp(q_k t),$$

3. *Logit*

$$F_k(t) = \frac{L(U - 1) + U(1 - L) \exp(Aq_k t)}{(U - 1) + (1 - L) \exp(Aq_k t)},$$

$$\text{où } A = (U - L) / [(U - 1)(1 - L)] \text{ et } L < 1 < U,$$

4. Linéaire tronquée

$$F_k(t) = 1 + q_k t \text{ si } L < 1 + q_k t < U,$$

$$L \text{ si } 1 + q_k t \leq L \text{ et } U \text{ si } 1 + q_k t \geq U.$$

Les poids de régression utilisés pour le calcul de la linéarisée figurent dans le tableau 32 pour ces quatre fonctions de calage.

TABLE 32 Dérivée de la fonction de calage et poids utilisé dans le calcul des résidus pour les principales fonctions de calage

Fonction	$F'_k(t)$	poids : $d_k F'_k(\mathbf{x}_k^t \boldsymbol{\lambda})$
Linéaire	q_k	$q_k d_k$
Raking-ratio	$q_k \exp(q_k t)$	$q_k w_k$
Logit	$q_k \frac{(U-L)^2 \exp(Aq_k t)}{(U-1)+(1-L) \exp(Aq_k t)}$	$q_k d_k \frac{(U-g_k)(g_k-L)}{(U-1)(1-L)}$
Linéaire tronquée	q_k ou 0	$q_k d_k$ si $L < g_k < U$ et 0 sinon

On remarque que les unités pour lesquelles $F'_k(\mathbf{x}_k^t \boldsymbol{\lambda}) = 0$ ne rentrent pas dans le calcul des coefficients de régression. Cela concerne, dans le cas de la fonction linéaire tronquée, les unités telles que $g_k = L$ ou $g_k = U$. Avec la méthode Logit, le terme $(U - g_k)(g_k - L)$ peut devenir petit mais ne s'annule pas.

Ces unités ne rentrent pas dans l'ajustement de la régression, mais il faut tout de même calculer les erreurs de prédiction e_k correspondantes et les utiliser dans l'estimation de variance.

Bibliographie

- CASSEL, C. M., SÄRNDAL, C.-E. & WRETMAN, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite population. *Biometrika* **63**, 615–620.
- DEVILLE, J.-C. (1999). Estimation de variance pour des statistiques et des estimateurs complexes : techniques de résidus et de linéarisation. *Techniques d'enquête* **25**, 219–230.
- DEVILLE, J.-C. & SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.
- HAZIZA, D. (2018). Cibler les variables d'intérêt importantes à l'étape du traitement de la non-réponse dans les enquêtes. Journées de méthodologie de l'INSEE.
- HAZIZA, D. & LESAGE, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics* **32**, 129–145.
- HORVITZ, D. G. & THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.
- KASS, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **29**, 119–127.
- LE GUENNEC, J. & SAUTORY, O. (2002). CALMAR2 : une nouvelle version de la macro CALMAR de redressement d'échantillon par calage. In *Actes des Journées de Méthodologie, Insee, Paris*. Paris.
- MURALTI, J.-L. (2023). Rapport méthodologique : plan d'échantillonnage, taux de réponse et pondération. Office Fédéral de la Statistique, Neuchâtel.
- POTTERAT, J., QUALITÉ, L. & ASSOULIN, D. (2019). Strukturhebung der eidgenössischen Volkszählung Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. Bundesamt für Statistik, Neuchâtel.
- QUALITÉ, L. (2023). Relevé structurel de la population suisse 2018 : Révision de la pondération et sélections répétées. Office fédéral de la statistique, Neuchâtel.
- SAUTORY, O. (1993). La macro calmar, redressement d'échantillon par calage sur marges. Séries des documents de travail de la Direction des Statistiques Démographiques et Sociales, F9310, Insee, Paris.
- TILLÉ, Y. (2019). *Théorie des sondages : échantillonnage et estimation en populations finies*. Paris : Dunod, 2nd ed.
- VALLÉE, A.-A. & TILLÉ, Y. (2019). Linearisation for variance estimation by means of sampling indicators: Application to non-response. *International Statistical Review* **87**, 347–367.

Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodenberichte der Sektion Statistische Methoden des BFS
Methodology reports published by the FSO's Statistical Methods Section

Meskaldji, D.-E., Panchard, C., Qualité, L., Saliba, J. (2024). Enquêtes thématiques 2018-2023. Plans de sondage, tirages, pondération et estimation de variance. Numéro OFS: 338-0090

Chevalier, C. (2024). Statistique de la valeur ajoutée 2021. Révision du plan de sondage. Numéro OFS: 338-0089

Nedyalkova, D., Assoulin, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalents. FSO number: 338-0088

Qualité, L. (2023). Relevé structurel de la population suisse 2018. Révision de la pondération et sélections répétées. Numéro OFS: 338-0087

Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0086

Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS: 338-0085

Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS: 338-0084

Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS: 338-0083

Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS: 338-0082

Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS: 338-0081

Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer: 338-0080

Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015: Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0079

Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS: 338-0078

Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics. FSO number: 338-0077

Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. BFS-Nummer: 338-0076

Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. BFS-Nummer: 338-0073

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009: Statistische Datenaufbereitung und Hochrechnung. BFS-Nummer: 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro OFS: 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009: Stichprobenrahmen und Stichprobenplan. BFS-Nummer: 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007: Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010: Stichprobenplan und Hochrechnung. BFS-Nummer: 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO number: 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro OFS: 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. BFS-Nummer: 338-0064

Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0063

Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0062

Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro OFS: 338-0061

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO number: 338-0060

Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro OFS: 338-0059

Graf, E. (2009). Weightings of the Swiss Household Panel: SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. FSO number: 338-0058

Graf, E. (2009). Pondérations du Panel Suisse de Ménages: PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro OFS: 338-0057-05

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

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

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

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

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

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

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer: 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 OFS: 338-0049

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

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

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer: 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 OFS: 338-0045

- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro OFS: 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer: 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. BFS-Nummer: 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS: 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 OFS: 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS: 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. FSO number: 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer: 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer: 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer: 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer: 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer: 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer: 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS: 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS: 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer: 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer: 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number: 338-0027
- Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. BFS-Nummer: 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS: 338-0025
- Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS: 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 OFS: 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS: 338-0022

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

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer: 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 OFS: 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. BFS-Nummer: 338-0018

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

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

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

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

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro OFS: 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 OFS: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer: 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 OFS: 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 OFS: 338-0009

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

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

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

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

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

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

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

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

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

Programme des publications de l'OFS

En tant que service statistique central de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public. Il utilise plusieurs moyens et canaux pour diffuser ses informations statistiques par thème.

Les domaines statistiques

- 00 Bases statistiques et généralités
- 01 Population
- 02 Espace et environnement
- 03 Travail et rémunération
- 04 Économie nationale
- 05 Prix
- 06 Industrie et services
- 07 Agriculture et sylviculture
- 08 Énergie
- 09 Construction et logement
- 10 Tourisme
- 11 Mobilité et transports
- 12 Monnaie, banques, assurances
- 13 Sécurité sociale
- 14 Santé
- 15 Éducation et science
- 16 Culture, médias, société de l'information, sport
- 17 Politique
- 18 Administration et finances publiques
- 19 Criminalité et droit pénal
- 20 Situation économique et sociale de la population
- 21 Développement durable, disparités régionales et internationales

Les principales publications générales

L'Annuaire statistique de la Suisse



L'Annuaire statistique de la Suisse de l'OFS constitue depuis 1891 l'ouvrage de référence de la statistique suisse. Il englobe les principaux résultats statistiques concernant la population, la société, l'État, l'économie et l'environnement de la Suisse.

Le Mémento statistique de la Suisse



Le mémento statistique résume de manière concise et attrayante les principaux chiffres de l'année. Cette publication gratuite de 52 pages au format A6/5 est disponible en cinq langues (français, allemand, italien, romanche et anglais).

Le site Internet de l'OFS: www.statistique.ch

Le portail «Statistique suisse» est un outil moderne et attrayant vous permettant d'accéder aux informations statistiques actuelles. Nous attirons ci-après votre attention sur les offres les plus prisées.

La banque de données des publications pour des informations détaillées

Presque tous les documents publiés par l'OFS sont disponibles gratuitement sous forme électronique sur le portail Statistique suisse (www.statistique.ch). Pour obtenir des publications imprimées, vous pouvez passer commande par téléphone (+41 58 463 60 60) ou par e-mail (order@bfs.admin.ch).
www.statistique.ch → Trouver des statistiques → Catalogues et banques de données → Publications

Vous souhaitez être parmi les premiers informés?



Abonnez-vous à un Newsmail et vous recevrez par e-mail des informations sur les résultats les plus récents et les activités actuelles concernant le thème de votre choix.
www.news-stat.admin.ch

STAT-TAB: la banque de données statistiques interactive



La banque de données statistiques interactive vous permet d'accéder simplement aux résultats statistiques dont vous avez besoin et de les télécharger dans différents formats.
www.stattab.bfs.admin.ch

Statatlas Suisse: la banque de données régionale avec ses cartes interactives



L'atlas statistique de la Suisse, qui compte plus de 4500 cartes, est un outil moderne donnant une vue d'ensemble des thématiques régionales traitées en Suisse dans les différents domaines de la statistique publique.
www.statatlas-suisse.admin.ch

Pour plus d'informations

Centre d'information statistique

+41 58 463 60 11, info@bfs.admin.ch

Ce rapport présente les méthodes statistiques appliquées par l'Office fédéral de la statistique (OFS) pour les cinq enquêtes thématiques menées entre 2018 et 2023 auprès des personnes et des ménages dans le cadre du système de recensement de la population suisse: l'Enquête sur les Familles et les Générations (EFG 2018 et 2023), l'Enquête sur la Langue, la Religion et la Culture (ELRC 2019), le Microrecensement Formation de Base et Formation Continue (MZB 2021), le Microrecensement Mobilité et Transports (MRMT 2021) et l'Enquête Suisse sur la Santé (ESS 2022).

Le rapport détaille les plans de sondage, la sélection et la préparation des échantillons, ainsi que le calcul des poids d'extrapolation. Pour ces derniers, il décrit les différentes étapes, comme la modélisation de la non-réponse et le calage, en prenant en compte les spécificités propres à chaque enquête. Une attention particulière est accordée à l'estimation de la variance, notamment pour les enquêtes en plusieurs vagues.

Les enquêtes thématiques, bien que largement homogènes, évoluent régulièrement grâce à l'intégration de nouvelles techniques, comme les questionnaires électroniques et à des améliorations méthodologiques, telles que l'utilisation de nouvelles variables auxiliaires (par exemple, les données de l'AVS) pour améliorer la précision des estimations. Ce rapport fournit ainsi une base des méthodes statistiques visant à standardiser dans la mesure du possible ces dernières pour les enquêtes thématiques et dont pourraient bénéficier d'autres enquêtes de l'OFS.

En ligne

www.statistique.ch

Imprimés

www.statistique.ch
Office fédéral de la statistique
CH-2010 Neuchâtel
order@bfs.admin.ch
tél. +41 58 463 60 60

Numéro OFS

338-0090

ISBN

978-3-303-00762-4

Les informations publiées ici contribuent à mesurer la réalisation des objectifs de développement durable (ODD).



Système d'indicateurs MONET 2030

www.statistique.ch → Trouver des statistiques → Développement durable → Système d'indicateurs MONET 2030

**La statistique
compte pour vous.**

www.la-statistique-compte.ch