



Rapport de synthèse sur la révision des pondérations longitudinales SILC-2018

Considérations méthodologiques

Neuchâtel, 2022

Éditeur:	Office fédéral de la statistique (OFS)	Concept de mise en page:	Section DIAM
Renseignements:	info.silc@bfs.admin.ch, tél. 058 463 68 04	Numéro OFS :	be-f-20.03.02.05.03
Rédaction:	Ariane Wenger, OFS	Téléchargement:	www.statistique.ch
Contenu:	Ariane Wenger, OFS; Stéphane Fleury, OFS	Copyright:	OFS, Neuchâtel 2022 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée
Domaine:	20 Situation économique et sociale de la population		
Langue du texte original:	Français		
Traduction:	Services linguistiques de l'OFS		

Table des matières

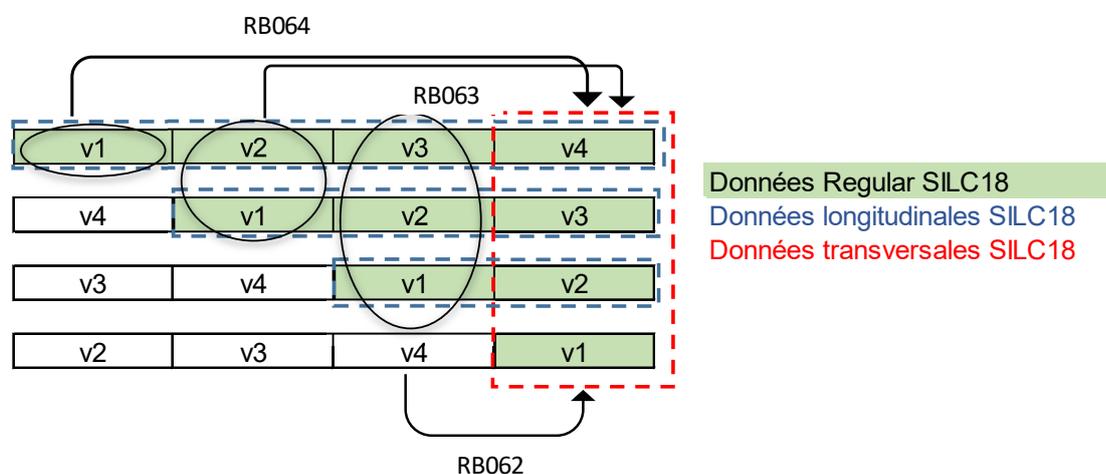
1	Introduction	3
2	Changements et améliorations apportés	4
3	Evaluation des améliorations apportées sur les poids longitudinaux	5
4	Conséquences sur les indicateurs publiés	9
5	Conclusion	10

1 Introduction

Depuis 2014, l'échantillon de l'enquête sur les revenus et les conditions de vie (SILC) est tiré dans un nouveau cadre de sondage qui est le cadre de sondage pour des enquêtes auprès des personnes et des ménages (SRPH) de l'OFS. Ce changement avait engendré une révision de la pondération transversale sur les données SILC14¹. La méthode de calcul des poids longitudinaux², qui permettent des analyses des transitions sur deux, trois ou quatre ans n'avait pas été touchée par cette révision. SILC est un panel rotatif sur quatre ans, ce qui signifie que chaque année un quart de l'échantillon est renouvelé et interrogé pendant quatre ans (cf. G1.1). Ce n'est qu'à partir de SILC17 que les données sont composées de quatre panels tirés dans le cadre de sondage SRPH, soit les panels 2014 à 2017. Bien que la révision des poids longitudinaux soit donc possible depuis SILC17, elle n'a été mise en œuvre qu'à partir de SILC18. Cependant, la révision sera appliquée de façon rétroactive sur les données longitudinales SILC17 dans le courant de l'année 2022.

Schéma du panel rotatif SILC et des poids longitudinaux (RB062/63/64)

G1.1



Le modèle de pondération longitudinale a été révisé en intégrant des nouvelles données issues des registres administratifs ainsi que d'autres améliorations méthodologiques. Ces changements impliquent une rupture de série entre les résultats longitudinaux SILC17 (ou antérieurs) et SILC18 (ou postérieurs), comme cela a été le cas entre les résultats transversaux SILC13 et SILC14.

Ce document a pour but de synthétiser les améliorations méthodologiques dans les pondérations longitudinales à partir de SILC18 et de documenter les principales modifications qui en découlent au niveau des résultats. Une documentation méthodologique détaillée sera disponible dans le rapport qualité SILC18³.

¹ Plus de détails sous <http://www.bfs.admin.ch/asset/fr/be-f-20.03-SILC-05>

² Plus de détails sous <http://www.bfs.admin.ch/asset/fr/be-f-20.03.02.05.01>

³ <https://www.bfs.admin.ch/asset/en/be-e-20-quality-report-2018> (publication printemps 2022)

2 Changements et améliorations apportés

Cette révision des pondérations longitudinales s'opère d'une part sur l'ajustement de la non-réponse totale⁴ (NRT) et d'autre part sur le calage final.

Avant SILC14, les personnes vivant dans un ménage ayant répondu au questionnaire Grille⁵ en vague 1⁶ étaient suivies pendant quatre ans, même si le ménage n'était pas complet (il est complet si le ménage a répondu aux questionnaires Grille et Ménage et qu'au moins une personne a répondu au questionnaire Individu). A partir de SILC14, cette règle a changé : les individus doivent être dans un ménage complet en vague 1 pour être suivis. Ce changement permet de se rapprocher des règles de suivi Eurostat et d'augmenter le nombre de variables auxiliaires pour la correction de la NRT - tant longitudinale que transversale - en vagues 2, 3 et 4. Il s'agit de variables provenant des données relevées en vague 1, disponibles uniquement pour les ménages complets, comme le risque de pauvreté, le coût du logement, l'intérêt pour la politique, etc.

Pour la pondération longitudinale, l'ajustement de la NRT était, avant SILC18, modélisé à l'aide d'une segmentation: arborescence asymétrique des facteurs explicatifs de la NRT, dans leur ordre croissant d'importance. Cette modélisation était effectuée au niveau individuel, ce qui engendrait des poids différents entre individus d'un même ménage. Ceci ne correspondait pas réellement au processus de non-réponse qui lui s'opère tout d'abord au niveau ménage. L'érosion longitudinale la plus fréquente s'observe au niveau du refus de répondre au questionnaire Grille (niveau ménage). Avec la révision, la correction de la NRT se fait au niveau des ménages mais sur la base d'un dataset Individus, cela grâce à un algorithme de calage intégré au niveau ménage⁷, ce qui n'était pas possible avec la segmentation. Cette méthode permet d'obtenir un poids égal pour toutes les personnes du même ménage. Cela permet d'être plus proche du processus réel de réponse.

L'utilisation du cadre de sondage SRPH permet, grâce à l'appariement à l'aide de numéros AVS, d'obtenir des informations démographiques de base et des informations sur le type de ménage issues du SRPH ainsi que des informations d'autres registres pour toute la population et ainsi également pour les individus en NRT (par exemple Caisses de Compensation avec certaines composantes de revenu). Ces informations, qui ont permis d'améliorer la pondération transversale, ont été utilisées de la même manière pour la révision de la pondération longitudinale.

Comme pour la pondération transversale, le calage final des poids longitudinaux a été révisé en intégrant des nouvelles données des registres. Toutes ces variables étant connues pour l'ensemble de l'échantillon ainsi que pour la population suisse, le calage final peut être fait de façon plus performante qu'auparavant.

L'enjeu des pondérations consiste à déterminer les caractéristiques des personnes / ménages qui ne répondent pas et à augmenter le poids des personnes / ménages répondants qui leur ressemblent le plus. De manière similaire, les poids des personnes / ménages qui correspondent à des profils répondant plus volontiers est diminué. La dernière étape, appelée calage final, permet d'ajuster les poids des répondants pour que les totaux pondérés des caractéristiques socioéconomiques des répondants correspondent aux totaux connus de ces caractéristiques dans la population de référence. Ainsi, le nombre important d'informations disponibles dans le SRPH ou d'autres registres contenant aussi les numéros AVS a permis d'améliorer la qualité de la pondération en diminuant le risque de biais engendré par la NRT.

⁴ Dans le longitudinal, on ne considère que la non-réponse totale au niveau ménage (NRTm), indiquée dans cette note uniquement par NRT.

⁵ Le questionnaire de l'enquête SILC se déroule en 3 parties. Le questionnaire Grille auquel répond une personne du ménage d'au moins 18 ans concerne essentiellement la composition du ménage ainsi que des données sociodémographiques. La deuxième partie est le questionnaire Ménage auquel peuvent s'ajouter des questionnaires Proxy Adultes (lorsqu'une personne n'est pas éligible au questionnaire individuel) et des questionnaires Proxy Enfants (enfants de 0-12 ans). La troisième partie est le questionnaire Individuel pour les individus de 16 ans et plus.

⁶ Individus dans un ménage interrogé la première des quatre années.

⁷ Un calage intégré au niveau ménage est un calage auquel une contrainte au niveau ménage est ajoutée. Celle-ci impose que les poids de la correction de la NRT soient les mêmes pour tous les membres d'un même ménage.

3 Evaluation des améliorations apportées sur les poids longitudinaux

3.1 Distribution

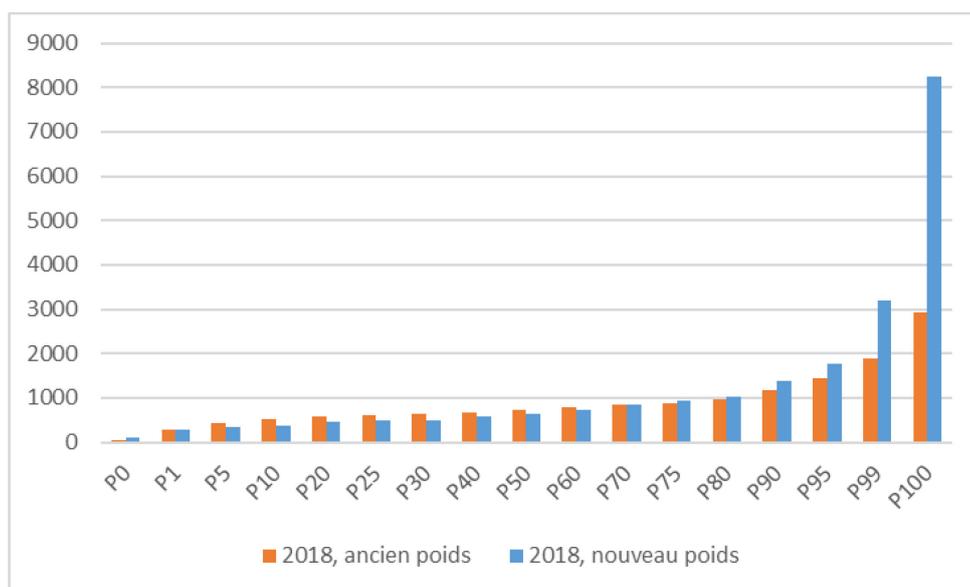
Distribution des poids longitudinaux selon l'ancienne et la nouvelle méthode

T3.1

	RB062		RB063		RB064	
	ancienne pondération	nouvelle pondération	ancienne pondération	nouvelle pondération	ancienne pondération	nouvelle pondération
Nobs	10273	10272	6563	6563	3246	3246
mean	794	806	1229	1249	2458	2498
stddev	305	574	437	908	992	1587
max	2922	8242	3719	16220	8775	16273
P95	1429	1762	2110	2787	4522	5428
Q3	892	934	1390	1448	2757	3039
median	728	648	1106	1006	2258	2125
Q1	607	476	946	718	1839	1426
P5	441	346	750	505	1412	960
min	57	95	148	127	515	513
Augmentation relative des écarts types avec la nouvelle méthode	88%		108%		60%	

Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018, sans prise en compte du loyer fictif
 Note : La différence du nombre d'observations pour le poids RB062 est due à un ménage qui était considéré comme complet lors de l'ancienne pondération et qui ne l'est plus dans la nouvelle pondération.

La variabilité des poids longitudinaux révisés est entre 1,6 et 2 fois plus importante que celle des anciens poids. L'ajout de nouvelles variables auxiliaires dans la modélisation du processus de réponse a permis de l'améliorer par une différenciation plus détaillée des caractéristiques sous-jacentes qui engendre naturellement une plus grande variabilité des poids finaux (T3.1). On peut ainsi espérer baisser le biais de la NRT avec un risque d'augmenter la variance des estimateurs.



Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018, sans prise en compte du loyer fictif

Les distributions sont semblables pour les trois poids longitudinaux RB062, RB063 et RB064, ancienne et nouvelle méthode. Jusqu'au percentile P80, les deux méthodes produisent des poids dont la distribution est proche (G3.1). Ce n'est qu'à partir du P90, i.e. pour les grands poids, que la différence de distribution est plus marquée. Cela soutient l'hypothèse que grâce aux nouvelles variables auxiliaires, la nouvelle pondération tient mieux compte des plus faibles probabilités de réponse (que ce soit au travers d'une meilleure correction de la non-réponse ou d'un meilleur calage).

3.2 Effet structurel

L'effet structurel le plus marqué est une augmentation de la taille de la population résidente permanente au domicile principal, estimée par SILC à 125'518 personnes pour l'année 2017 lorsqu'on considère la transition entre 2017 et 2018 (RB062), à 129'641 pour celle entre 2016 et 2018 (RB063) et de 128'332 pour la transition entre 2015 et 2018 (RB064). Dans l'ancienne méthode de pondération, le calage final se faisait sur la population résidente permanente au domicile principal selon la Statistique de la population et des ménages (STATPOP) au 31 décembre de l'année précédant le début de la transition alors qu'avec la nouvelle méthode, le calage final se fait sur le SRPH qui comprend en plus les résidents non permanents vivant dans un ménage avec au moins un résident permanent.

3.3 Erosion et ventilations sociodémographiques

Afin d'évaluer la qualité de la pondération longitudinale, nous proposons de ré-estimer un indicateur de vague 1 en utilisant les répondants de la vague 4 avec le poids longitudinal correspondant, cela en figeant la variable d'intérêt de la vague 1. Dans le cas d'une pondération corrigeant parfaitement l'érosion corrélée à nos variables d'intérêt, le résultat devrait donner une valeur similaire qu'avec les poids transversaux de vague 1 (avant l'érosion longitudinale).

Par exemple, le taux de risque de pauvreté⁸ de 2015 calculé avec le poids transversal RB050 est de 15,6%. Avec une pondération longitudinale parfaite, ce taux devrait être similaire en le calculant sur les individus dans les ménages complets de vague 4 en 2018 avec la caractéristique « A risque de pauvreté » de 2015 fixée et le poids longitudinal RB064. Si ce n'est pas le cas, cela indique que le poids longitudinal corrige imparfaitement le biais dû à l'érosion.

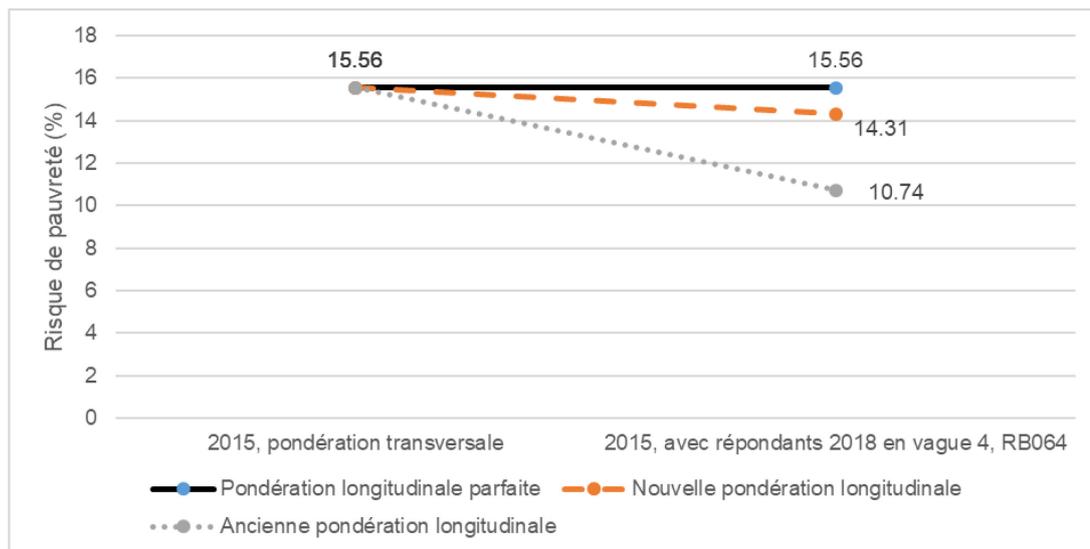
⁸ Le taux de risque de pauvreté est calculé ici sans le loyer fictif.

Avec la nouvelle pondération longitudinale, le biais imparfaitement corrigé par les poids a fortement diminué. Le taux de risque de pauvreté de 2015 (15,6%) était recalculé à 10,7% avec l'ancienne pondération longitudinale sur la vague 4 et à 14,3% avec la nouvelle pondération.

Le graphique G3.2 montre que la nouvelle pondération n'est pas parfaite, mais qu'elle corrige nettement mieux la plus faible probabilité de réponse en vague 4 des personnes à risque de pauvreté en vague 1 par rapport à celles qui ne l'étaient pas.

Taux de risque de pauvreté de 2015 en vague 1, estimé avec la pondération transversale, avec l'ancienne et la nouvelle pondération longitudinale

G3.2



Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018 et données transversales 2015, sans prise en compte du loyer fictif

De la même façon que pour le taux de risque de pauvreté, l'érosion est analysée sur certaines variables sociodémographiques relevées avec SILC15. Le pourcentage est calculé sur la population répondante de SILC15 en utilisant le poids transversal RB050 de 2015. Ce pourcentage est également calculé sur ces mêmes variables sociodémographiques de SILC15 mais uniquement pour les répondants de SILC18 en utilisant l'ancien et le nouveau poids RB064.

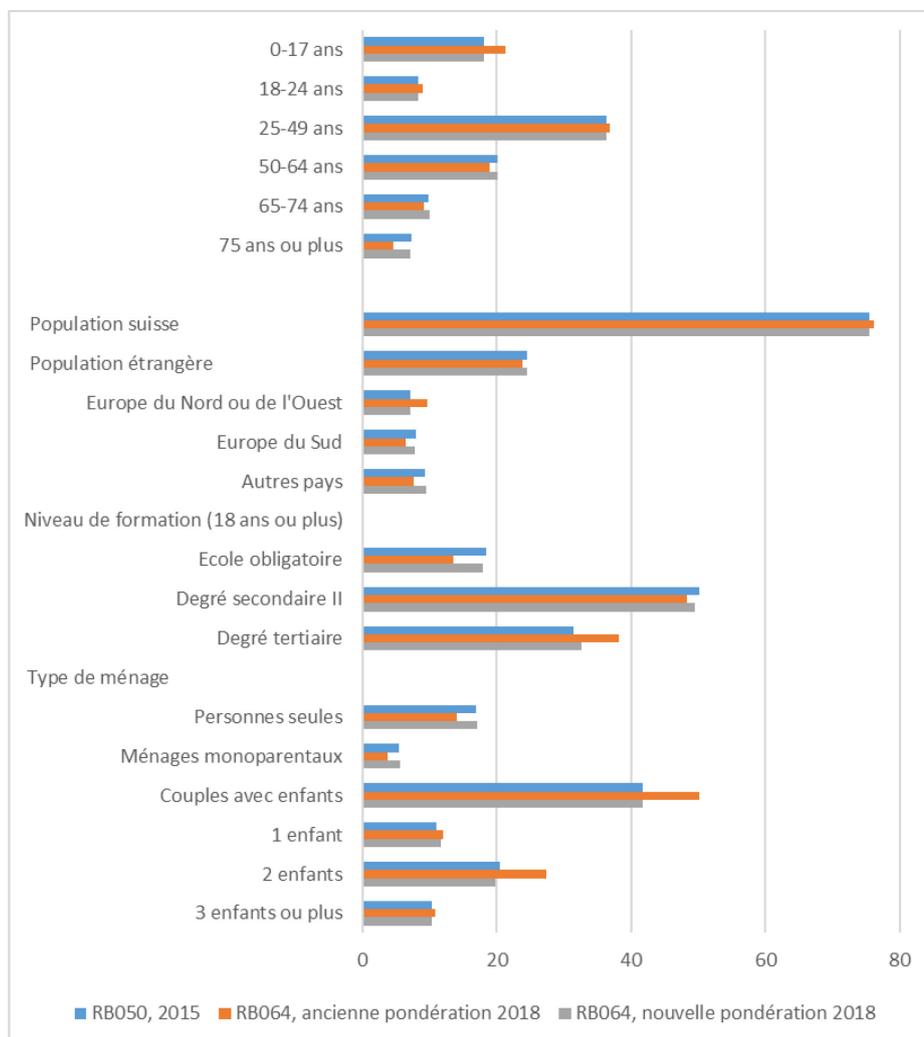
L'ancienne pondération longitudinale sous-corrigeait l'érosion des 50-64 ans et des 75 ans et plus (y c. les décès survenus entre SILC15 et SILC18) surtout au profit des 0-17 ans, ce qui n'est plus le cas avec la nouvelle pondération (G3.3).

En ce qui concerne la formation des personnes, la nouvelle pondération corrige mieux la sous-représentation des bas niveaux de formation. Le niveau de formation n'est utilisé que dans l'algorithme de la correction de la NRT. Le niveau de formation n'est pas disponible pour le calage final, en revanche, certaines variables tirées des registres sont fortement corrélées avec ce dernier.

Globalement, pour toutes les variables sociodémographiques utilisées dans les tables standards, une amélioration est constatée avec la nouvelle pondération.

Pourcentage pondéré d'individus dans la population, avec le poids transversal RB050 de 2015 et les poids longitudinaux RB064 ancienne et nouvelle méthode.
Selon ventilations sociodémographiques de SILC15

G3.3



Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018 et données transversales 2015, sans prise en compte du loyer fictif

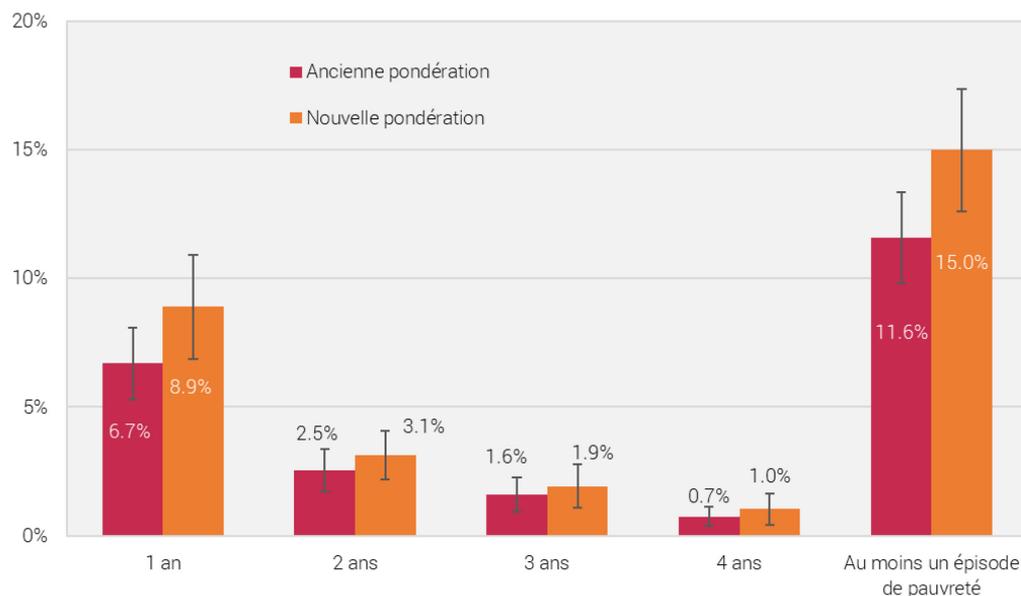
4 Conséquences sur les indicateurs publiés

4.1 Dynamique de la pauvreté

Personnes touchées par la pauvreté pendant une période de quatre ans

Selon le nombre d'années, en % de la population

G4.1

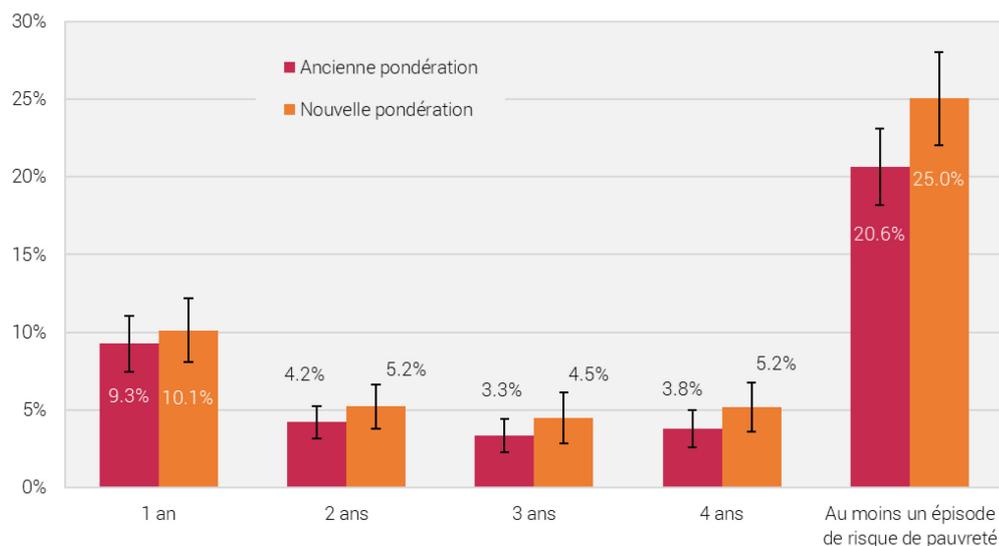


Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018, sans prise en compte du loyer fictif
Exemple de lecture: avec l'ancienne pondération, 2.5% de la population était pauvre selon le concept absolu au cours de deux des quatre années considérées et 0.7% l'était au cours de quatre années.

Personnes à risque de pauvreté à 60% de la médiane pendant une période de quatre ans

Selon le nombre d'années, en % de la population

G4.2

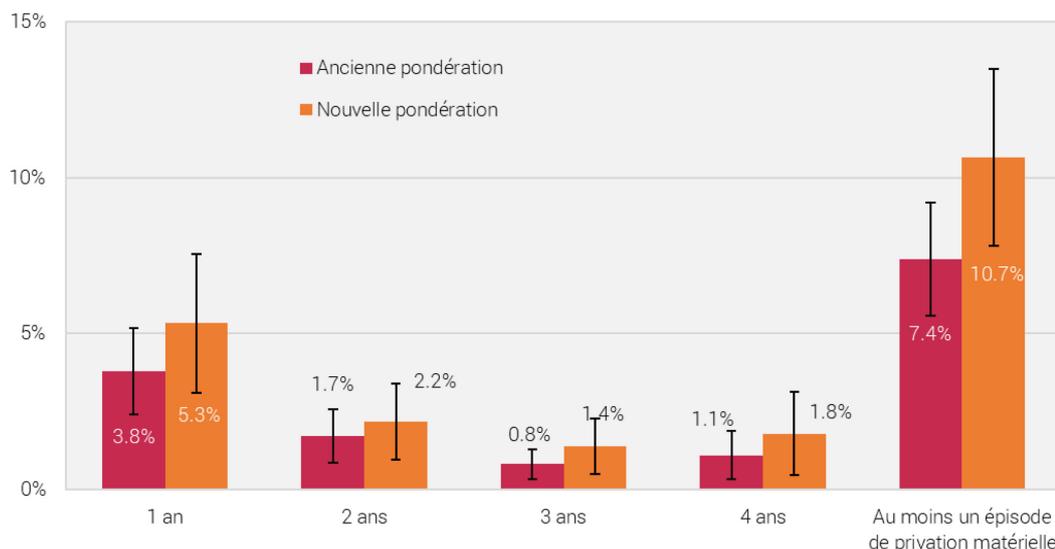


Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018, sans prise en compte du loyer fictif

Personnes privées matériellement pendant une période de quatre ans

Selon le nombre d'années, en % de la population

G4.3



Source: OFS – Enquête sur les revenus et les conditions de vie (SILC), données longitudinales 2015-2018, sans prise en compte du loyer fictif

Pour les trois indicateurs⁹, il y a une augmentation non négligeable des taux avec la nouvelle pondération, mais qui n'est jamais significative.

4.2 Pauvreté persistante

Le taux de risque de pauvreté persistant¹⁰ avec SILC18 a augmenté de 2,2 points de pourcentage, passant de 6,6% avec l'ancienne pondération à 8,8% avec la nouvelle. L'augmentation la plus marquée (3,5 points de pourcentage) concerne les personnes de 18-24 ans, dont le taux passe de 3,8% avec l'ancienne pondération à 7,3% avec la nouvelle, suivie par les moins de 18 ans (de 15,0% à 18,0%). Une augmentation non négligeable est également constatée chez les personnes âgées de plus de 65 ans, dont le taux passe de 7,4% à 10,9%.

5 Conclusion

Le changement de règle de suivi des ménages a permis d'utiliser pour l'ajustement de la non-réponse totale ménage des variables provenant des données relevées en vague 1, disponibles uniquement pour les ménages complets. La nouvelle pondération modélise ainsi mieux les probabilités de réponse. Le cadre de sondage SRPH, en offrant une couverture nettement plus complète de la population de référence et l'accès à un plus grand nombre de données sur la population, a permis d'améliorer le calage final. Il en résulte une révision fondamentale des méthodes de pondération longitudinale et par conséquent une amélioration importante de la qualité des estimations longitudinales produites depuis SILC 2018. La dispersion des poids révisés est plus grande ce qui entraîne une tendance à la hausse des intervalles de confiance.

L'analyse des impacts de la révision sur les principaux indicateurs montre qu'il est indispensable d'indiquer une rupture de série entre les résultats longitudinaux avant et après la révision.

⁹ Les définitions de ces trois indicateurs sont disponibles sous [Pauvreté et privations matérielles | Office fédéral de la statistique \(admin.ch\)](#) et les résultats sous [Dynamique de la pauvreté | Office fédéral de la statistique \(admin.ch\)](#)

¹⁰ Part des personnes ayant un revenu inférieur au seuil de risque de pauvreté (60 % de la médiane) pendant l'année en cours et pendant au moins deux des trois années précédentes.