

# PROBLEME DE COHERENCE D'UNE VAGUE A L'AUTRE DE L'ENQUETE ÉRFI-GGS : L'EXEMPLE DU NOMBRE D'ENFANTS

Arnaud Régnier-Loilier

*Institut national d'études démographiques, 133 boulevard Davout, 75980 Paris Cedex 20*

[arnaud.regnier-loilier@ined.fr](mailto:arnaud.regnier-loilier@ined.fr)

**Résumé.** L'enquête Érfi (*Étude des relations familiales et intergénérationnelles*) est la déclinaison française de l'enquête internationale GGS (*Generations and Gender Survey*) actuellement en cours dans une vingtaine de pays. Elle visait à interroger, à trois reprises, les mêmes personnes, chaque vague d'enquête devant être espacée de 3 années.

La qualité des données longitudinales dépend d'un côté de l'attrition mais aussi, d'un autre, de la cohérence des réponses entre les vagues. La collecte sous Capi permet de filtrer certaines questions d'une vague à une autre selon les informations recueillies lors des vagues précédentes afin de limiter le temps de passation, les incohérences, etc. Toutefois, quelques questions doivent être reposées à chaque vague, comme la description des enfants dont la situation est susceptible d'évoluer (le répondant peut avoir eu un nouvel enfant, un enfant peut décohabiter, décéder, etc.). On décrit ainsi à chaque vague les enfants du répondant vivant dans le ménage, hors ménage et décédés. De ces différents modules du questionnaire est déduit par comptage le nombre total d'enfants (aucune question directe sur le nombre total d'enfants du répondant n'étant posée).

S'il est logique que ce nombre augmente au fil du temps (suite à une naissance), il est plus surprenant qu'il diminue. Certes, il existe classiquement une sous-déclaration des enfants décédés ou de ceux avec qui les liens sont rompus, pouvant expliquer que certains enfants « disparaissent » d'une vague à l'autre. Toutefois, l'écart observé dans l'enquête française est de trop forte ampleur pour ne relever que de cette seule explication.

L'objectif de cet article est de montrer l'ampleur de cette sous-déclaration, de repérer quels enfants « disparaissent » et de voir si cela correspond à des profils de répondants spécifiques, à un effet enquêteur, un effet panel, etc.

Ce problème interroge plus généralement la méthode du décompte des enfants décrits dans différentes parties d'un même questionnaire pour déterminer le nombre total d'enfants d'une personne lors d'une enquête à passages répétés et invite à procéder à des vérifications analogues dans d'autres enquêtes de même type, notamment dans les autres enquêtes GGS.

## Introduction

Lors de l'appariement des données des deux premières vagues de l'enquête *Étude des relations familiales et intergénérationnelles*<sup>1</sup> (2005 et 2008), il s'est avéré que pour une proportion non négligeable de répondants, le nombre total d'enfants eus ou adoptés différait en 2005 et 2008 (Régnier-Loilier, Saboni et Valdes, 2011). Que cette différence soit positive (plus d'enfants en 2008 qu'en 2005) est logique pour les plus jeunes, ceux-ci ayant pu avoir un ou des enfants, mais que la différence soit négative au-delà de 50 ans questionnait davantage.

Ce décalage a d'abord donné lieu à un ensemble de vérifications sur le plan « technique ». L'ampleur de la différence pouvait laisser craindre un mauvais appariement des répondants entre les vagues. Toutefois, les vérifications apportées ont permis d'écarter cette hypothèse. Autre piste, le nombre total d'enfants n'étant pas issu d'une question unique (par exemple, « Combien d'enfants avez-vous eus ou adoptés ? ») mais du décompte des enfants décrits dans le ménage, des enfants non cohabitants à la date de l'enquête et des enfants décédés, la différence pouvait être due à une erreur de construction des compteurs de nombre total d'enfants. Là aussi, suite aux vérifications apportées, l'hypothèse d'une méthode de calcul différente aux deux vagues a pu être écartée. Autre explication également rejetée, l'erreur de programmation informatique (par le jeu des filtres, les questions visant à dénombrer les enfants hors ménage ou décédés auraient pu ne pas être posées systématiquement, mais tel ne fut pas le cas).

En creusant davantage ce problème de cohérence, la différence semblait provenir d'une sous-déclaration des enfants non cohabitants, malgré un questionnement identique à celui de la première vague, mais sans que l'on puisse apporter d'explication tranchée au phénomène. Stratégie d'évitement de la part d'enquêtés qui auraient déclaré ne pas avoir d'enfants hors ménage, se rappelant suite à l'expérience de la première vague que chaque enfant donnerait ensuite lieu à une description fouillée (une dizaine de questions pour chaque enfant) ? Sous-déclaration involontaire de la part des enquêtés estimant que rien n'avait changé à ce niveau depuis la vague précédente ? Sous-déclaration volontaire de la part d'enquêteurs afin de limiter le temps de passation du questionnaire<sup>2</sup> ? Ou alors, sur-déclaration d'enfants lors de la première vague, par exemple comptés à la fois comme faisant partie du foyer et également décrits comme vivant en dehors du ménage (double compte possible notamment en cas de résidence alternée ou de cohabitation ponctuelle) ?

Lors de la mise en place de la troisième vague, une attention particulière a été portée à ce problème de qualité des données. En premier lieu, un effort particulier a été fait lors des formations afin de sensibiliser les enquêteurs à ce problème qui touche à une partie centrale de l'enquête, et nous attirions leur attention dans le questionnaire avec une « consigne enquêteur » spécifique : « TRES IMPORTANT : recenser tous les enfants non cohabitants (en vie) du répondant, même si ceux-ci étaient déjà non cohabitants lors de la vague précédente ». Par ailleurs, afin de connaître le « véritable » nombre d'enfants eus ou adoptés du répondant, il a été décidé d'ajouter, en plus des questions déjà présentes dans les deux premières vagues, une question récapitulative plus directe : « Pour récapituler, combien d'enfants avez-vous eu vous-même en tout, qu'il s'agisse d'enfants eus ou adoptés, avec votre conjoint actuel ou avec un précédent conjoint ? Merci de prendre également en compte les enfants qui ne vivent plus avec vous ou qui sont aujourd'hui décédés ». Au final, au terme de la troisième vague, nous disposons donc de 4 indicateurs de nombre d'enfants : 3 issus du décompte des enfants du

---

<sup>1</sup> Pour plus de détails sur l'enquête, voir Régnier-Loilier, 2012.

<sup>2</sup> Les enquêteurs sont payés au questionnaire et non à l'heure, sur la base d'une durée de passation estimée lors de l'enquête pilote.

ménage, hors ménage et décédés à chaque vague, et un indicateur issu de la question récapitulative incluse en vague 3.

Nous confrontons ici ces 4 indicateurs afin de voir si le même phénomène que celui relevé entre les vagues 1 et 2 s'est rejoué lors de la vague 3 et de mieux comprendre les décalages observés. Cette présentation a pour objectif d'attirer l'attention des utilisateurs sur la qualité des données concernant les indicateurs de nombre d'enfants. Une étude plus fouillée de cette question est en cours de préparation.

## **1. Importante sous-déclaration des enfants chez les répondants de 45 ans et plus aux vagues 2 et 3**

De manière agrégée, nous pouvons d'abord confronter le nombre moyen d'enfants par groupe d'âges et sexe, issu des quatre indicateurs dont nous disposons (**figure 1**) : le nombre d'enfants issu du décompte des enfants du ménage, hors ménage et décédés en 2005, 2008 et 2011 (« décompte V1 », « décompte V2 », « décompte V3 »), et le nombre total d'enfants décrit au travers de la question récapitulative posée en 2011 (« récapitulatif V3 »). Le champ est réduit aux personnes ayant répondu aux trois vagues de l'enquête afin de contourner d'éventuels biais qui seraient liés à l'attrition<sup>3</sup>. Les données ne sont pas pondérées, l'objectif n'étant pas une mesure de la fécondité par groupe d'âges mais simplement la comparaison d'indicateurs pour un même groupe de personnes.

Entre 18 et 44 ans, tant chez les femmes que chez les hommes, la comparaison des courbes « décompte V1 » (—), « décompte V2 » (—o—) et « décompte V3 » (--Δ--) matérialise l'arrivée d'enfants (naissances) : à un âge donné, le nombre moyen d'enfants est plus élevé en V3 qu'en V2, et plus élevé en V2 qu'en V1. Par contre, à partir de 45 ans, les courbes « décompte V2 » (—o—) et « décompte V3 » (--Δ--) se superposent très bien mais passent sous la courbe « décompte V1 » (—) : entre 45 et 79 ans, les personnes ont donc en moyenne déclaré moins d'enfants en 2008 et 2011 qu'en 2005 (- 0,25 enfant). La parfaite superposition des courbes 2008 et 2011 pourrait laisser penser à une sur-déclaration d'enfants lors de la première vague. Toutefois, l'indicateur « récapitulatif V3 » (—) offre une information précieuse et plaide plutôt en faveur d'une sous-déclaration d'enfants aux vagues 2 et 3. En effet, cette courbe se superpose presque parfaitement avant 40 ans à la courbe « décompte V3 » (--Δ--), ce qui est logique puisque les deux indicateurs sont calés sur la même année (2011), et presque parfaitement à la courbe « décompte V1 » (—) à partir de 45 ans (léger décalage chez les hommes entre 40 et 54 ans cependant). Cette confrontation semble indiquer que la méthode du décompte des enfants cohabitants, non cohabitants et décédés est plutôt fiable lors de la première vague passés 45-50 ans<sup>4</sup>, et lors de la troisième vague chez les groupes d'âges encore féconds. La méthode du décompte semble en revanche induire une importante sous-déclaration aux vagues 2 et 3 (de 0,25 enfants en moyenne, chez les personnes âgées de 45 à 79 ans en 2005). On pourrait faire l'hypothèse d'une sous-déclaration d'enfants qui seraient décédés depuis 2005 mais la différence est d'ampleur trop importante et touche presque à l'identique tous les groupes d'âges à partir de 50 ans (or la

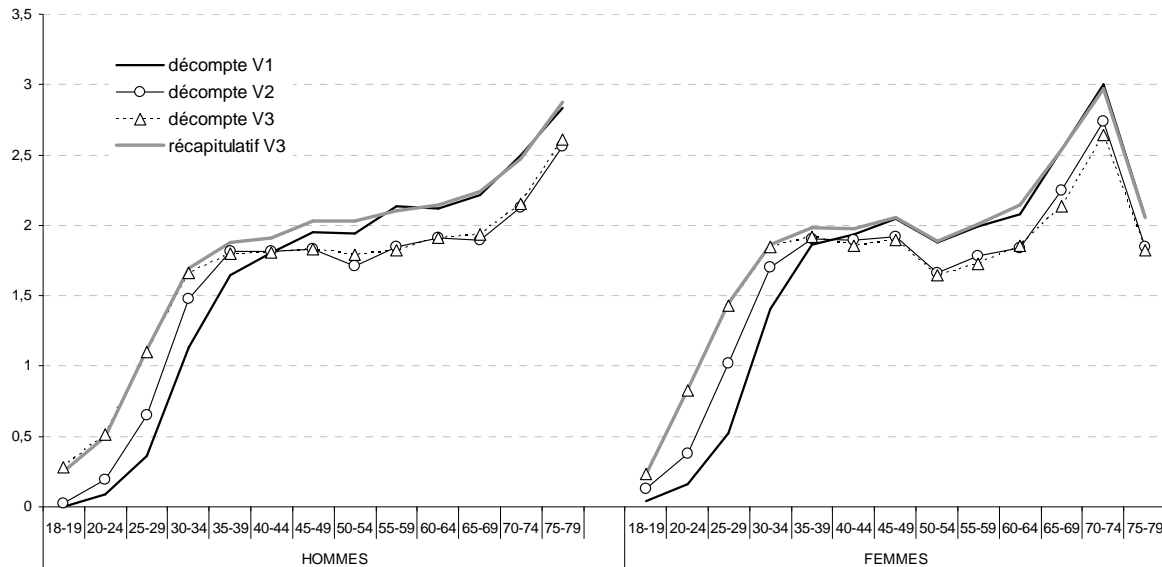
---

<sup>3</sup> Précisions que dans l'étude de l'attrition entre les vagues 2 et 3 (cf. partie II), nous avons inclus comme variable explicative le fait d'avoir ou non déclaré moins d'enfants en V2 qu'en V1 avec l'idée qu'une sous-déclaration pouvait s'être accompagnée d'une attrition plus forte entre les vagues 2 et 3 (lassitude de l'enquête), mais aucun effet significatif n'a été relevé.

<sup>4</sup> Et probablement aux plus jeunes âges mais la confrontation du nombre d'enfants en 2005 et 2011 n'a de sens que pour les personnes ayant achevé leur vie féconde.

probabilité d'avoir perdu un enfant entre les vagues n'est pas la même si l'on a 50 ans ou si l'on en a 79).

**Figure 1.** Nombre moyen d'enfants par âge (en 2005) d'après 4 indicateurs



Source : Ined-Insee, Érfi-GGS123, 2005-2011

Champ : répondants aux trois vagues d'enquête (n = 5 436)

Note : données non pondérées<sup>5</sup> ; l'âge est celui du répondant en 2005

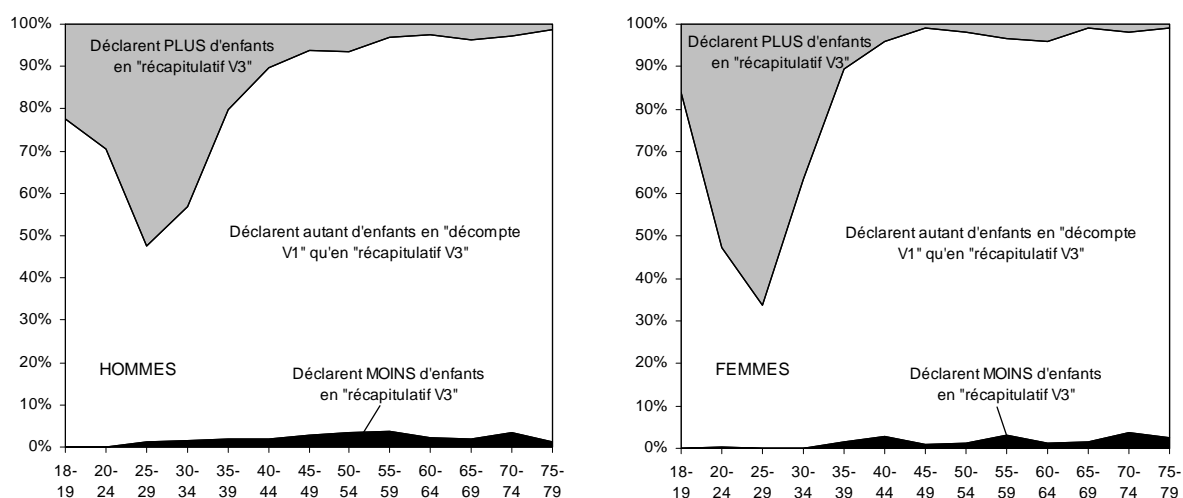
Lecture (exemple pour les hommes de 55-59 ans) : d'après le décompte des enfants (du ménage, hors ménage et décédés), les hommes de 55-59 ans ont déclaré 2,14 enfants en 2005, 1,82 enfants en 2008, 1,82 enfants en 2011 et 2,10 enfants en 2011 d'après leur réponse à la question récapitulative du nombre total d'enfants.

Cette confrontation reste cependant limitée dans la mesure où on ne se place pas au niveau individuel mais à un niveau agrégé (comparaison de moyennes). Afin de valider la fiabilité de la méthode du décompte lors de la vague 1, on peut comparer pour un même individu le nombre total d'enfants eus ou adoptés, d'après la méthode du décompte, avec l'indicateur récapitulatif en V3 (**figure 2**). Ainsi, trois configurations existent :

- pour un individu  $i$ , le nombre d'enfants est identique ( $\square$ ) s'il est déduit du « décompte en V1 » ou du « récapitulatif V3 » ;
- le nombre d'enfants obtenu par le « décompte en V1 » est inférieur ( $\blacksquare$ ) au nombre d'enfants déclaré au « récapitulatif V3 » (plus d'enfants en 2011 qu'en 2005 : naissances, sous-déclaration en V1, sur-déclaration ou double-compte en V3) ;
- le nombre d'enfants obtenu par le « décompte en V1 » est supérieur ( $\blacksquare$ ) au nombre d'enfants déclaré au « récapitulatif V3 » (moins d'enfants en 2011 qu'en 2005 : sur-déclaration en V1 ou sous-déclaration en V3).

<sup>5</sup> L'idée étant simplement de confronter, pour un même groupe de personnes, leurs réponses à différents moments (2005, 2008 et 2011) et non de proposer une mesure représentative du nombre moyen d'enfants par âge.

**Figure 2.** Proportion de personnes déclarant autant, plus ou moins d'enfants d'après 2 indicateurs : décompte en V1 et question récapitulative en V3



Source : Ined-Insee, Érfi-GGS123, 2005-2011

Champ : répondants aux trois vagues d'enquête (n = 5 436)

Note : données non pondérées

Lecture (exemple pour les hommes de 60-64 ans) : 2 % des hommes de 60-64 ans (en 2005) déclarent moins d'enfants vague 3 à la question récapitulative du nombre total d'enfants que d'après le décompte en vague 1, 95 % autant et 3 % davantage.

Logiquement, la proportion de femmes pour qui le nombre d'enfants est supérieur en V3 (■) est importante jusque 40 ans environ (maximal chez les 25-29 ans en 2005) illustrant la naissance d'enfants à ces âges, puis devient infime (2 % entre 45 et 79 ans). On retrouve le même phénomène chez les hommes, bien que plus étalé sur les différents groupes d'âges, traduisant notamment une fertilité qui se prolonge au-delà de 45-49 ans, contrairement aux femmes. La proportion de femmes et d'hommes pour qui le nombre d'enfants est inférieur en V3 par rapport à V1 (■) est infime (1,4 % pour l'ensemble des femmes de 18 à 79 ans ; 1,9 % pour les hommes). Au final, pour les 45 ans et plus, la cohérence est donc très forte entre l'indicateur de décompte en V1 et l'indicateur récapitulatif en V3 (□ : plus de 95 %), indiquant que le nombre total d'enfants déduit du décompte des enfants du ménage, hors ménage et décédés en 2005 est fiable.

Partant de ce constat, on peut désormais s'affranchir de l'indicateur « récapitulatif en V3 » et confronter les 3 indicateurs issus du décompte en V1, V2 et V3, toujours au niveau individuel. Si en moyenne (figure 1) la sous-déclaration semble de même ampleur en vague 2 et en vague 3, l'objectif est ici de voir si ce sont les mêmes personnes qui ont sous-déclaré en vague 2 et en vague 3 ou si certaines sous-déclarent à une vague mais pas à l'autre. À partir de ces trois indicateurs, un grand nombre de configurations sont possibles<sup>6</sup>. Pour simplifier, nous confrontons ici 5 situations (figure 3) :

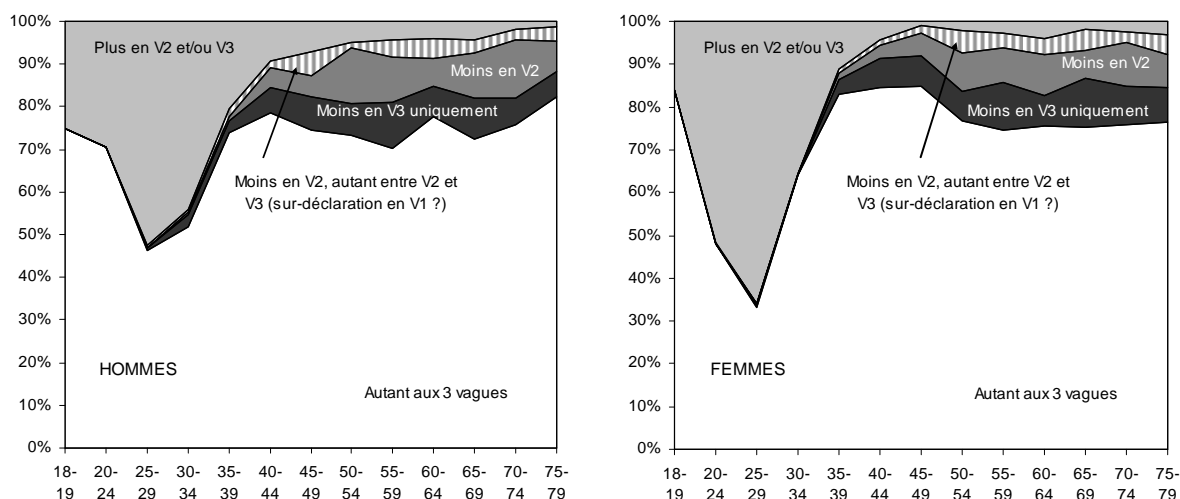
- pour un individu i, le nombre d'enfants est identique (□) aux trois vagues (d'après le « décompte en V1 », le « décompte en V2 » et le « décompte en V3 ») ;
- « plus en V2 et/ou V3 » (■) : le nombre d'enfants obtenu par le « décompte en V1 » est inférieur au nombre d'enfants déclaré au « décompte en V2 » et/ou au « décompte en V3 » ; (plus d'enfants en 2008 et/ou 2011 qu'en 2005 : naissances) ;

<sup>6</sup> V1 = V2 = V3 ; V1 > V2 = V3 ; V1 < V2 = V3 ; V1 = V2 > V3 ; V1 = V2 < V3 ; V1 > V2 > V3 ; V1 < V2 < V3 ; V1 > V2 < V3 et V1 = V3 ; V1 > V2 < V3 et V1 < V3 ; V1 > V2 < V3 et V1 > V3 ; V1 < V2 > V3 et V1 = V3 ; V1 < V2 > V3 et V1 < V3 ; V1 < V2 > V3 et V1 > V3.

- « Moins en V3 uniquement » (■) : le nombre d'enfants obtenu par le « décompte en V3 » est inférieur à celui obtenu en « décompte en V1 » et « décompte en V2 », ces deux derniers étant identiques ;
- « Moins en V2, autant entre V2 et V3 » (▨) : le nombre d'enfants obtenu par le « décompte en V2 » est inférieur à celui obtenu en « décompte en V1 », mais le « décompte en V2 » est identique au « décompte en V3 ». La sous-déclaration serait donc identique aux vagues 2 et 3, ou il y aurait eu une sur-déclaration en V1 (double compte par exemple) ;
- « Moins en V2 » (■) : le nombre d'enfants obtenu par le « décompte en V2 » est inférieur à celui obtenu en « décompte V1 » et le « décompte en V2 » n'est pas identique au « décompte en V3 ».

Avant 40 ans, la confrontation des trois indicateurs offre des résultats cohérents : le nombre d'enfants est le même ou il a augmenté au fil des vagues (naissances). Par contre, au-delà de 40 ans, tant chez les femmes que chez les hommes, 1 personne sur 6 (environ 16 %) sous-déclare à la vague 2 et/ou à la vague 3. Et, fait notable, la sous-déclaration est « volatile » dans le sens où ce n'est que rarement qu'une même personne sous-déclare en vague 2 et en vague 3, par rapport à la vague 1 (« Moins en V2, autant entre V2 et V3 » : ▨). La plupart des personnes qui sous-déclarent ne le font qu'à une seule vague (2 ou 3). Par ailleurs, dans la mesure où les incohérences touchent surtout les 50 ans et plus, la présomption est forte qu'il s'agisse d'enfants non cohabitants ou décédés.

**Figure 3.** Proportion de personnes déclarant autant, plus ou moins d'enfants aux différentes vagues d'après les 3 indicateurs de décompte



Source : Ined-Insee, Érfi-GGS123, 2005-2011

Champ : répondants aux trois vagues d'enquête (n = 5 436)

Note : données non pondérées

Lecture (exemple pour les hommes de 60-64 ans) : d'après le décompte des enfants aux trois vagues, 78 % des hommes de 60-64 ans (en 2005) déclarent autant d'enfants aux trois vagues, 7 % en déclarent moins en vague 3 (mais autant aux vagues 1 et 2), 7 % en déclarent moins en vague 2 (mais autant en vague 1 et 3), 5 % en déclarent moins en vague 2 mais autant en vagues 2 et 3, et 4 % en déclarent plus en vague 2 et/ou vague 3.

On notera que malgré l'attention attirée auprès des enquêteurs lors de la troisième vague sur ce phénomène (durant les formations ainsi que dans les instructions de collecte affichées à l'écran dans Capi), la sous-déclaration en vague 3 a été d'une ampleur similaire à celle observée en vague 2 (8,7 % contre 7,8 % : **figure 4**). Par ailleurs, même si le fait d'avoir déclaré moins d'enfants en vague 2 conduit plus souvent à en déclarer moins en vague 3 par rapport à la vague 1 (37 % contre 6 %), majoritairement, la sous-déclaration n'a pas touché les mêmes personnes : parmi les « sous-déclarants » à l'une des vagues, 43 % n'ont sous-déclaré qu'en vague 3<sup>7</sup>, 36 % qu'en vague 2 et 21 % ont sous-déclaré à la fois en vague 2 et en vague 3 (par rapport à la vague 1).

**Figure 4.** Proportion de personnes ayant sous-déclaré à la vague 3 selon le fait d'avoir sous-déclaré en vague 2

% ligne ( <i>effectifs</i> )	Pas de sous-déclaration en V3	Sous déclaration en V3	Ensemble
Pas de sous-déclaration en V2	93,7 (4697)	6,3 (315)	92,2 (5012)
Sous déclaration en V2	62,7 (266)	37,3 (158)	7,8 (424)
Ensemble	91,3 (4963)	8,7 (473)	100 (5436)

Source : Ined-Insee, *Érfti-GGS123*, 2005-2011

Champ: répondants aux trois vagues d'enquête (n = 5 436)

Note : données non pondérées

Lecture : 37,3 % des personnes ayant déclaré moins d'enfants en vague 2 ont déclaré moins d'enfants en vague 3

## 2. Les enfants sous-déclarés : principalement des enfants hors ménage

Les indicateurs utilisés décomptent trois catégories d'enfants : ceux vivant dans le ménage, en dehors du ménage et les enfants décédés. On peut s'interroger sur le fait de savoir si la sous-déclaration observée porte sur l'ensemble des enfants ou sur une catégorie particulière. On sait par exemple que les enfants décédés tendent à être sous-déclarés dans les enquêtes ou que les hommes sous-déclarent parfois leurs enfants avec qui ils ne vivent plus, notamment suite à une séparation quand le lien a été rompu (Toulemon, 2005). En outre, la sous-déclaration est particulièrement nette à partir de 50 ans, âge auquel la plupart des personnes ne vivent plus avec leurs enfants. Nous posons donc pour hypothèse que les enfants sous-déclarés sont des enfants non cohabitants (en vie ou décédés).

Pour éprouver cette hypothèse, nous décomposons le nombre total d'enfants en trois compteurs dénombrant les enfants du ménage, hors ménage et décédés. L'observation est limitée aux personnes de 50 ans et plus en 2005 (âge à partir duquel la sous-déclaration est nette) ayant sous-déclaré un ou des enfants en vague 2 d'une part (n = 321 personnes), et en vague 3 d'autre part (n= 320 personnes). Nous mettons ensuite en regard le nombre d'enfants non cohabitants déclaré en vague 2 *versus* vague 1 et en vague 3 *versus* vague 1 (**figure 5a**). Nous faisons de même pour les enfants du ménage (**figure 5b**) et pour les enfants décédés (**figure 5c**).

<sup>7</sup> 315 personnes sur 739 sous-déclarants à l'une des vagues (315+158+266). Même méthode de calcul pour les autres proportions citées à la suite.

Dans les différents tableaux, les diagonales indiquent que le nombre d'enfants déclaré est le même aux deux vagues mises en regard (2005 et 2008 ou 2005 et 2011) ; les valeurs situées au-dessus de cette diagonale indique que le nombre d'enfants déclaré en 2008 ou 2011 est inférieur à celui déclaré en 2005 ; les valeurs situées sous cette diagonale indique que le nombre d'enfants déclaré en 2008 ou 2011 est supérieur à celui déclaré en 2005.

Qu'il s'agisse de 2008 ou de 2011, les sous-déclarants ont très rarement déclaré moins d'enfants cohabitants (**figure 5b**) ou décédés (**figure 5c**) qu'en 2005 : la plupart des effectifs se situent en effet sur la diagonale et même à l'intersection de la ligne « 0 enfant » et de la colonne « 0 enfant », à l'exception de quelques personnes qui vivaient en 2005 avec un enfant et qui ne vivent plus avec en 2008 (24) ou 2011 (23). On note quelques cas d'enfants décédés en 2005 et qui ne sont plus déclarés en 2008 (24 sur 321) ou 2011 (14 sur 320) mais cette sous-déclaration est loin d'expliquer la proportion de personnes ayant déclaré moins d'enfants qu'en 2005. En réalité, environ 90 % des sous-déclarants de 2008 et de 2011 ont déclaré moins d'enfants non cohabitants qu'en 2005 (**figure 5a**) : les effectifs se concentrent en effet massivement au-dessus de la diagonale. Plus encore, on observe une concentration très forte des effectifs sur la première ligne du tableau, quel que soit le nombre d'enfants déclaré en 2005, indiquant que la sous-déclaration ne concerne pas un enfant non cohabitant parmi la descendance mais l'ensemble des enfants de cette catégorie. Par exemple, sur les 321 personnes ayant déclaré moins d'enfants en 2008, plus de la moitié (113 + 54) avaient déclaré avoir 2 ou 3 enfants non cohabitants en 2005 et disent ne pas en avoir du tout en 2008. Les sous-déclarants sont donc au final dans près de 9 cas sur 10 des personnes ayant répondu négativement à la question « (le cas échéant : *Nous avons déjà parlé des enfants vivant au sein de votre ménage*). Avez-vous adopté ou avez-vous eu vous-même d'autres enfants ? Ne prenez pas en compte les beaux-enfants, les enfants accueillis ni les enfants décédés. Nous en parlons après »).

On pourrait être tenté de dire que la formulation de la question est en cause. Il n'est pas fait référence directement aux enfants « vivant ailleurs » mais aux « autres » enfants ; par ailleurs, l'accent est d'abord mis sur le fait d'avoir « adopté » des enfants. Toutefois, comme nous l'avons précisé précédemment, la formulation n'a pas changé par rapport à la première vague et l'indicateur récapitulatif en V3 colle presque parfaitement avec le nombre d'enfants issu du décompte des enfants du ménage, hors ménage et décédés de la vague 1 (pour les 50 ans et plus). Ou alors, la question est moins bien comprise dans le cadre d'une deuxième ou troisième interrogation, les répondants pouvant penser qu'on leur demande s'ils n'ont pas « d'autres » enfants que ceux évoqués précédemment, entendu comme lors de la vague précédente. Dit autrement, certains auraient peut-être interprété la question comme faisant écho à ce qui s'était passé au cours des trois dernières années.



**Figure 5.** Nombre d'enfants **a.** non cohabitants **b.** cohabitants **c.** décédés déclarés en 2005 et en 2008 (VAGUES 1-2), en 2005 et en 2011 (VAGUES 1-3)

a. VAGUES 1 - 2		Nombre d'enfants hors ménage en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants hors ménage en 2008	0	17	<b>81</b>	<b>113</b>	<b>54</b>	<b>23</b>	288
	1	0	5	6	0	0	11
	2	0	1	2	1	2	6
	3	0	1	0	4	7	12
	4 et +	0	0	0	0	4	4
Total		17	88	121	59	36	321
a. VAGUES 1 - 3		Nombre d'enfants hors ménage en 2005 / 2008					
		0	1	2	3	4 et +	Total
Nombre d'enfants hors ménage en 2011	0	19	<b>89</b>	<b>113</b>	<b>49</b>	<b>27</b>	297
	1	0	1	3	2	2	8
	2	0	0	2	3	1	6
	3	0	0	1	4	2	7
	4 et +	0	0	0	0	2	2
Total		19	90	119	58	34	320

b. VAGUES 1 - 2		Nombre d'enfants du ménage en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants du ménage en 2008	0	246	<b>25</b>	<b>1</b>	<b>1</b>	<b>1</b>	274
	1	6	22	8	1	0	37
	2	0	0	5	2	0	7
	3	0	0	0	1	0	1
	4 et +	0	0	1	0	1	2
Total		252	47	15	5	2	321
b. VAGUES 1 - 3		Nombre d'enfants du ménage en 2005 / 2008					
		0	1	2	3	4 et +	Total
Nombre d'enfants du ménage en 2011	0	246	<b>40</b>	<b>7</b>	<b>1</b>	<b>1</b>	295
	1	3	10	7	2	0	22
	2	0	0	0	2	0	2
	3	0	0	0	0	0	0
	4 et +	0	0	0	0	1	1
Total		249	50	14	5	2	320

c. VAGUES 1 - 2		Nombre d'enfants décédés en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants décédés en 2008	0	281	<b>19</b>	<b>1</b>	<b>0</b>	<b>0</b>	301
	1	2	12	3	0	0	17
	2	0	0	2	1	0	3
	3	0	0	0	0	0	0
	4 et +	0	0	0	0	0	0
Total		283	31	6	1	0	321
c. VAGUES 1 - 3		Nombre d'enfants décédés en 2005 / 2008					
		0	1	2	3	4 et +	Total
Nombre d'enfants décédés en 2011	0	282	<b>12</b>	<b>0</b>	<b>0</b>	<b>0</b>	294
	1	2	19	0	1	0	22
	2	0	1	2	1	0	4
	3	0	0	0	0	0	0
	4 et +	0	0	0	0	0	0
Total		284	32	2	2	0	320

Source : Ined-Insee, Érfi-GGS123, 2005-2011

Champ: répondants âgés de 50 à 79 ans en 2005, ayant participé aux trois vagues d'enquête et ayant déclaré moins d'enfants en 2008 qu'en 2005 (n = 321) et moins d'enfants en 2011 qu'en 2005 et 2008 (n = 320)

Note : données non pondérées

Lecture (exemple tableau « a. VAGUES 1-2 », ligne « 0 » colonne « 2 ») : parmi les 321 sous-déclarants à la vague 2 de 50 ans et plus, 113 personnes ont déclaré avoir 2 enfants hors ménage en 2005 et 0 enfant hors ménage en 2008.

### 3. Pas de profil spécifique des sous-déclarants

Une manière de faire pour essayer de mieux comprendre cette sous-déclaration consiste à établir un portrait « type » des sous-déclarants. Afin de déterminer si certaines caractéristiques sont associées à la sous-déclaration d'enfants lors de la vague 2 et/ou de la vague 3, plusieurs régressions logistiques ont été mises en place (**figure 6**). Elles visaient à modéliser la probabilité d'avoir sous-déclaré en vague 2 par rapport à la vague 1 ; la probabilité d'avoir sous-déclaré en vague 3 par rapport à la vague 1 ou à la vague 2 ; la probabilité d'avoir sous-déclaré en vague 2 ou en vague 3 par rapport à la vague 1. Seule cette dernière est présentée, les résultats étant les mêmes que si l'on considère séparément les différentes vagues. Elle inclut différents facteurs, pour la plupart relatifs à la situation observée lors de la première vague<sup>8</sup> :

- caractéristiques individuelles des répondants en 2005 (sexe, âge, niveau de diplôme, santé perçue du répondant),
- indicateur de rupture conjugale entre les vagues 1 et 3 (le fait de s'être séparé pouvant conduire à une rupture des liens avec ses enfants),
- nombre d'enfants cohabitants, non cohabitants, décédés en 2005 (afin de faire émerger des configurations plus ou moins liées à la sous-déclaration),
- indicateurs de posture par rapport à l'enquête en 2005 : refus de répondre à certaines questions, en l'occurrence au revenu du ménage et à l'enregistrement des réponses aux questions « sensibles » (pacs et religion), refus de recevoir les premiers résultats (marquant un moindre intérêt pour l'objet de l'étude) ou encore la durée du questionnaire (avec l'hypothèse que le souvenir d'un questionnaire trop long a pu conduire le répondant à vouloir en limiter la durée lors des vagues suivantes),
- présence du conjoint durant l'entretien de 2008 ou 2011 (avec l'idée que la présence du conjoint peut limiter les risques de mauvaise compréhension de la question ou limiter les risques de sous-déclaration des enfants),
- indicateur de présence d'enfants non cohabitants avec qui le lien est rompu en 2005 (aucune rencontre par an), avec l'idée que la rupture du lien et le temps qui passe ont pu conduire le répondant à ne plus déclarer ces enfants ; les relations enfants-parents après une séparation étant très « genrés » (Régnier-Loilier, 2006), une interaction avec le sexe du répondant a été introduite dans le modèle.

Le champ est limité aux répondants âgés de 50 ans et plus en 2005 (raisons exposées précédemment<sup>9</sup>), ayant participé aux trois volets de l'enquête et ayant déclaré avoir au moins un enfant en 2005 (la sous-déclaration n'étant pas possible dans le cas contraire).

De manière globale, le pouvoir explicatif du modèle est extrêmement faible. Les facteurs introduits jouent donc peu sur la probabilité d'avoir déclaré moins d'enfants aux vagues 2 et 3. Plus dans le détail, seules les caractéristiques liées aux « types » d'enfants décrits par le répondant en 2005 jouent. Ainsi, la probabilité de sous-déclaration est significativement plus faible si le répondant n'avait aucun enfant non cohabitant en 2005, confirmant nos premières

---

<sup>8</sup> Sauf mention contraire dans la figure. À noter que dans la mesure où la sous-déclaration peut avoir eu lieu en vague 2 ou en vague 3, le même modèle a été répliqué en incluant les caractéristiques observées à la vague précédant la sous-déclaration (caractéristiques de 2005 si la sous-déclaration a eu lieu en 2008, caractéristiques de 2008 si la sous-déclaration a eu lieu en 2011). Les résultats étant identiques, seul le modèle relatif aux caractéristiques observées en 2005 est ici présenté.

<sup>9</sup> La même modélisation a été réalisée sur les 18-79 ans (en 2005) : les résultats obtenus sont identiques, avec cependant une probabilité moindre de sous-déclaration entre 18 et 49 ans.

observations : la sous-déclaration du nombre total d'enfants est étroitement liée à la présence d'enfants non cohabitants lors de la première vague. Par contre, le nombre d'enfants non cohabitants en tant que tel (avoir 1, 2 ou 3 enfants et plus) ne joue pas. La sous-déclaration apparaît également plus probable si le répondant avait mentionné avoir un ou des enfants décédés au moment de la première vague de l'enquête. Ceux-ci tendent donc à ne plus être évoqués lors des vagues suivantes. Par rapport à nos précédentes observations, un autre phénomène se dessine assez nettement : la présence d'enfants cohabitants en 2005 augmente la probabilité de sous-déclarer des enfants lors des vagues suivantes. On peut faire l'hypothèse selon laquelle le départ du foyer parental de ces enfants depuis 2005 a conduit à ce qu'ils ne soient plus déclarés dans les vagues suivantes, pas même en tant qu'enfants hors ménage. Enfin, le fait d'avoir décrit un ou des enfants en 2005 avec qui le lien était rompu (absence de rencontres dans l'année) augmente la probabilité d'avoir décrit moins d'enfants lors de la ou des vagues suivantes. Le facteur d'interaction avec le sexe du répondant est lui aussi significatif et montre un effet bien moins marqué du côté féminin. Ce résultat renforce l'hypothèse émise par Laurent Toulemon (2005) à partir de l'enquête *Étude de l'histoire familiale* de 1999 d'une sous-déclaration des enfants par les hommes suite à une rupture conjugale.

Au-delà de la situation parentale du répondant, aucune des caractéristiques incluses dans les modélisations ne jouent : le fait de s'être séparé ne joue pas ; pas d'influence du diplôme, de l'état de santé, de l'âge ; pas d'incidence non plus de la posture par rapport à l'enquête ni de la manière dont l'entretien s'est déroulé, notamment pas d'effet de la durée du questionnaire, contredisant l'hypothèse d'une possible « stratégie d'évitement » de la part des enquêtés.

**Figure 6.** Modélisation de la probabilité d'avoir déclaré moins d'enfants en 2008 ou 2005 par rapport à 2005 (paramètres estimés de la régression logistique)

		Sous-déclaration en V2 ou V3	
Constante		-0,91	**
SEXE DU REpondANT	Homme	0,00	Ref.
	Femme	0,04	-
NB D'ENFANTS COHABITANTS EN 2005	0	0,00	Ref.
	1 et +	0,35	**
NB D'ENFANTS NON COHABITANTS EN 2005	0	-0,85	***
	1	0,00	Ref.
	2	0,00	-
	3 et +	-0,21	-
NB D'ENFANTS DECEDES EN 2005	0	0,00	Ref.
	1 et +	0,55	***
NB D'ENFANTS AVEC LIEN ROMPU EN 2005	0	0,00	Ref.
	1 et +	0,80	***
Interaction SEXE (femme) * ENFANTS LIEN ROMPU (1 ou +)		-0,75	*
AGE DU REpondANT EN 2005	50-54 ans	0,00	Ref.
	55-59 ans	0,14	-
	60-64 ans	-0,12	-
	65-69 ans	0,05	-
	70-74 ans	0,01	-
	75-79 ans	-0,13	-
SOUHAITE RECEVOIR LES RESULTATS (2005)	Oui	0,00	Ref.
	Non	-0,14	-
EVOLUTION SITUATION CONJUGALE 2005-2011	Séparation	-0,10	-
	Mise en couple	-0,16	-
	Autres situations	0,00	Ref.
REponse AU REVENU EN 2005	A répondu	-0,26	-
	A refuser de répondre	0,00	Ref.
ACCEPTATION ENREGISTREMENT	A accepté	0,00	Ref.
	A refusé	-0,18	-
NIVEAU DE DIPLÔME EN 2005	sans diplôme	0,00	-
	CEP	-0,02	-
	BEPC, brevet élémentaire	0,21	-
	CAP, BEP	-0,07	-
	Bac technologique ou pro.	0,00	Ref.
	Bac général	0,26	-
	Bac + 2	0,23	-
	Diplôme supérieur	0,13	-
SANTÉ PERçUE EN 2005	Très bon	0,22	-
	Bon	-0,03	-
	Moyen	0,00	Ref.
	Mauvais	-0,13	-
	Très mauvais	0,23	-
CONDITIONS DE PASSATION	Présence du conjoint	0,15	-
	Absence du conjoint	0,00	Ref.
DUREE QUESTIONNAIRE EN 2005	20-44 min	-0,06	-
	45-59 min	0,00	Ref.
	60-74 min	-0,14	-
	75 min ou plus	-0,05	-
	Non connu	0,12	-
AVOIR SOUS-DECLARE EN 2008 (vague 2)	Oui		
	Non		
R <sup>2</sup>		0,02	
EFFECTIF (et %)		517	25,5
		1511	74,5

Source : Ined-Insee, Érfi-GGS123, 2005-2011

Champ : répondants âgés de 50 à 79 ans en 2005, ayant participé aux trois vagues d'enquête et ayant déclaré au moins un enfant en 2005

Note : données non pondérées

Légende : Ref. = situation de référence ; - = facteur non significatif ; \* = sign. à 10 % ; \*\* = à 5 % ; \*\*\* = à 1 %

Lecture : un coefficient positif (resp. négatif), statistiquement significatif, indique que l'on est en présence d'un facteur qui accroît (resp. décroît) la probabilité d'avoir déclaré moins d'enfants qu'en 2005

## Conclusion

Concernant les enfants décrits dans l'enquête Érfi, la qualité des données collectées dans les vagues successives doit être questionnée. Un important phénomène de sous-déclaration, notamment des enfants non cohabitants décrits lors de la première vague, apparaît en effet aux vagues 2 et 3. En moyenne, les répondants âgés de 50 ans et plus en 2005 déclarent 0,25 enfant de moins en vague 2 et en vague 3, par rapport à la première vague. Si l'on aurait pu suspecter non pas une sous-déclaration aux vagues 2 et 3 mais une sur-déclaration d'enfants en vague 1 (possibles double-comptes, notamment en vague 1 où le répondant à l'enquête n'était pas nécessairement la personne ayant complété le tableau de composition du ménage<sup>10</sup>), cette hypothèse doit être rejetée. L'introduction lors de la dernière vague d'une question récapitulative du nombre total d'enfants du répondant permet en effet de valider l'information recueillie lors de la première vague. Ainsi, la méthode du décompte des enfants cohabitants, non cohabitants et décédés pour déterminer le nombre total d'enfants paraît fiable ; mais uniquement pour la première vague.

Une explication « technique » des décalages observés semble raisonnablement pouvoir être rejetée. La méthode de détermination du nombre total d'enfants est identique aux différentes vagues. Les vérifications apportées à ce jour n'ont pas non plus permis de mettre à jour d'erreurs de filtrage dans le questionnaire des vagues 2 et 3. Enfin, l'éventualité d'un mauvais appariement des répondants entre les vagues doit également être écartée.

S'il est assez difficile d'étudier un éventuel « effet enquêteur » (les enquêteurs auraient par exemple pu indiquer qu'une personne n'avait pas d'enfants non cohabitants sans lui poser la question afin de gagner en temps de passation) dans la mesure où le nombre d'enquêteurs était important (environ 400) et que chacun avait peu de fiches-adresses à traiter, les quelques investigations menées à ce niveau semble plutôt conduire à rejeter cette piste explicative : la sous-déclaration touche la plupart des enquêteurs et en aucun cas elle ne concerne la majorité des fiches-adresses d'un même enquêteur.

L'absence d'effet des caractéristiques individuelles sur la sous-déclaration (en dehors du nombre d'enfants cohabitants, non cohabitants et décédés en vague 1) et la volatilité de celle-ci selon les vagues (les sous-déclarants de la vague 2 et de la vague 3 ne sont que partiellement les mêmes personnes) semblent plaider en faveur d'un phénomène plutôt aléatoire, qui reste donc difficilement explicable.

Bien que l'architecture du questionnaire soit identique aux trois vagues et que la formulation des questions soit similaire, la sous-déclaration pourrait peut-être tenir à l'imprécision de la question utilisée pour recenser les enfants non cohabitants dans le contexte des deuxième et troisième vagues. Il n'est en effet pas fait mention explicitement au fait d'avoir « des enfants ne vivant pas dans le ménage » ou « vivant ailleurs » mais simplement à « d'autres enfants ». Certains répondants ont pu estimer qu'ils avaient déjà décrit leurs enfants dans la ou les vagues précédentes et qu'il n'y avait plus lieu d'en parler. Cette piste explicative reste cependant très fragile en raison de la durée écoulée entre les questionnaires (3 et 6 ans).

Les résultats de cette étude sont donc peu concluants mais ils ont au moins deux intérêts. Le premier est de sensibiliser les utilisateurs des données à ce problème afin qu'ils en tiennent compte dans la réalisation de leurs travaux. Le second est d'attirer l'attention des concepteurs

---

<sup>10</sup> Tirage au sort du répondant après description du ménage (pour plus de détails : Régnier-Loilier, 2006). Ainsi, un enfant partiellement cohabitant aurait pu être décrit dans le ménage par la personne remplissant le tableau des habitants du logement, puis considéré par le répondant à l'enquête comme ne faisant pas partie du ménage. Dans les faits, quelques cas de double-comptes ont pu être repérés mais ils sont très peu nombreux.

de futures enquêtes panels sur ce risque de sous-déclaration dès lors que l'on ne part pas d'une question directe pour dénombrer les enfants mais du cumul d'informations collectées à différents moments du questionnaire. Il reste aujourd'hui à voir si le problème rencontré dans Érfi se retrouve dans d'autres enquêtes, notamment dans les enquêtes GGS qui ont eu lieu dans les autres pays. Les données de la deuxième vague viennent seulement d'être rendues accessibles, et pour quatre pays seulement dont l'Allemagne, pays pour lequel l'attrition entre les deux premières vagues frôle les 70 %.

## Bibliographie

- Régnier-Loilier A., 2006, *Présentation, questionnaire et documentation de l'étude des relations familiales et intergénérationnelles (Erfi). Version française de l'enquête Generations and Gender Survey (GGS)*, Documents de Travail, 133, 231 pages.
- Régnier-Loilier A., Saboni L., Valdes B., 2011, *Présentation and modifications to the Generations and Gender Survey Questionnaire in France (Wave 2)*, Documents de Travail, 173, 146 pages.
- Régnier-Loilier A., Guisse N., 2012, *Dictionnaire des codes. Troisième vague de l'enquête Érfi*.
- Régnier-Loilier A., 2012, *Présentation, questionnaire et documentation de la troisième vague de l'étude des relations familiales et intergénérationnelles (Erfi-GGS 2011)*, Document de travail, 187, Ined, 328 pages.
- Toulemon L., 2005, « Enfants et beaux-enfants des hommes et des femmes », in Lefèvre C. et Filhon A. (dir.), *Histoires de familles, histoires familiales*, Chapitre 3, les Cahiers de l'Ined, 156 : 59-77.