



Arnaud RÉGNIER-LOILIER\*

---

## Incohérence du nombre d'enfants déclarés entre les vagues de l'enquête française *Generations and Gender Survey*

*Comment savoir combien d'enfants ont eu les personnes qui répondent à une enquête ? L'article d'Arnaud RÉGNIER-LOILIER compare le nombre d'enfants déclarés par les personnes ayant répondu aux trois vagues de l'Étude des relations familiales et intergénérationnelles (Erfi) en 2005, 2008 et 2011, version française de l'enquête européenne Générations et genre. En l'absence d'une question unique sur le nombre d'enfants nés ou adoptés lors des deux premières vagues, la détermination de celui-ci se fait par une succession de questions sur le nombre d'enfants en vie ou décédés, vivant dans ou hors du ménage. Cette procédure a introduit des erreurs que l'auteur mesure ici grâce à une question supplémentaire lors de la troisième vague de l'enquête récapitulant le nombre total d'enfants : certains répondants ont tendance à ne pas déclarer les enfants vivant hors du ménage. L'article attire donc l'attention des utilisateurs et concepteurs d'enquêtes sur ces biais de recueil et l'utilité d'une question récapitulative sur l'ensemble des enfants.*

Entre 2005 et 2011, l'Institut national d'études démographiques (Ined) et l'Institut national de la statistique et des études économiques (Insee) ont réalisé l'Étude des relations familiales et intergénérationnelles (Erfi), volet français du programme d'enquêtes internationales *Generations and Gender Survey* (GGS). Comme pour toute enquête, il importe d'évaluer la qualité des données préalablement à leur exploitation. S'agissant d'une enquête à passages répétés (les mêmes personnes ont été sollicitées pour répondre à trois reprises), la question se pose d'autant plus : outre les possibles biais liés au taux de participation lors des différentes vagues, c'est également la cohérence des informations collectées à chacune d'elle qui doit être interrogée. C'est sur ce dernier point que se focalise la présente étude, en s'intéressant plus spécifiquement au nombre d'enfants déclarés par les répondants au fil des vagues d'enquête.

---

\* Institut national d'études démographiques, Paris.

Correspondance : Arnaud Régnier-Loilier, Institut national d'études démographiques, 133 boulevard Davout, 75980 Paris Cedex 20, tél : 33 (0)1 56 06 20 71, courriel : arnaud.regnier-loilier@ined.fr

### *La qualité des données en question*

La fiabilité des statistiques issues d'enquêtes sur échantillon dépend d'un ensemble d'éléments. Le taux de participation influe d'abord sur la représentativité des données. Certaines catégories de personnes sont en effet plus difficiles à joindre, d'autres refusent plus fréquemment de participer. La correction de la non-réponse et le calcul de variables de pondération sont alors indispensables afin de rendre l'échantillon représentatif. Mais la qualité des statistiques dépend aussi de la fiabilité de l'information recueillie. Celle-ci est étroitement liée à la formulation des questions (qui doit éviter toute équivoque) et aux possibilités qu'a la personne interrogée d'y répondre précisément. Les enquêtes rétrospectives tendent notamment à remonter de plus en plus loin dans le passé des personnes, augmentant le risque de défaillance de la mémoire ou de réinterprétation du passé (Auriat, 1996). Concernant plus spécifiquement les données rétrospectives des événements familiaux (naissances, mariages, etc.), sujet qui nous intéresse principalement ici, différentes recherches concluent à une plus grande fiabilité des dates collectées auprès des femmes (Poulain *et al.*, 1991) et soulignent une omission plus fréquente d'enfants de la part des hommes<sup>(1)</sup>. C'est en particulier le cas quand ces derniers ne vivent plus avec leurs enfants (Toulemon, 1997) et qu'ils sont donc moins impliqués dans leur éducation quotidienne (Joyner *et al.*, 2012). Le phénomène est par ailleurs amplifié si le père n'est pas marié (Rendall *et al.*, 1999).

Le mode de collecte ou encore les conditions de passation jouent également de manière significative sur les réponses. La présence du conjoint conduit par exemple à une sous-déclaration des relations multipartenariales ou de la consommation de drogues (Firdion, 1993). Par ailleurs, la présence de jeunes enfants durant l'entretien entraîne une moindre précision dans la datation d'événements passés : les enfants réclament parfois l'attention du répondant et réduisent sa concentration. À l'inverse, notamment lorsqu'un homme est interrogé, la présence de sa conjointe améliore la précision des dates (les conjoints se concertent alors, aidant au travail de mémoire) ainsi que la qualité des informations relatives à la répartition des tâches domestiques et parentales au sein du couple. La conjointe joue alors un rôle dans le contrôle de l'information donnée à l'enquêteur, limitant les réponses qui surestiment la participation de l'homme aux activités quotidiennes (Régnier-Loilier, 2007 ; Régnier-Loilier et Guisse, 2009).

Les enquêtes longitudinales, qui consistent à interroger les mêmes personnes à plusieurs reprises, soulèvent d'autres questions concernant la qualité des données. La déperdition de l'échantillon au fil des vagues peut d'abord venir compromettre la robustesse des tests statistiques. En outre, l'attrition

---

(1) On évoque ici le cas des pays développés. S'intéressant au Mali, Véronique Hertrich (1991) a montré que la société s'organisait autour des hommes et que ceux-ci disposaient d'une aussi bonne connaissance que les femmes pour certains événements (les omissions dans l'histoire génésique des hommes ne sont pas plus nombreuses que dans celle des femmes), voire meilleure comme pour le mariage.

étant rarement aléatoire mais sélective (pour l'enquête Erfi-GGS, voir Régnier-Loilier et Guisse, 2012), elle entraîne une déformation de la structure de l'échantillon de départ qui peut affecter les variables d'intérêt de l'enquête et biaiser les résultats et leur interprétation (Razafindratsima et Kishimba, 2004 ; Mazuy *et al.*, 2005). Par ailleurs, la réinterrogation de mêmes personnes lors des différentes vagues d'enquête pose la question de la cohérence des réponses collectées. Au-delà du risque d'erreurs de mémoire qui s'accroît avec le temps, le changement de situation d'une personne peut aussi modifier sa perception des choses, sa manière de se raconter (Mazuy et Lelièvre, 2005 ; Molinié et Laville, 2000). Par exemple, une rupture d'union s'accompagne le plus souvent d'une résidence des enfants chez la mère et parfois du relâchement voire de la rupture des liens entre le père et ses enfants. Ce type de situation peut alors le conduire à ne plus déclarer les enfants qu'il ne voit plus lors de la vague d'enquête suivante. Si les enfants partagent leur temps entre les domiciles des deux parents, le répondant peut, au contraire, considérer que ses enfants vivent à la fois chez lui et ailleurs, ce qui génère des doubles comptes.

### *Objectifs et délimitation du champ de l'étude*

L'objectif de cet article n'est pas d'évaluer la qualité de l'histoire génésique collectée dans l'enquête française GGS en la confrontant à des sources externes<sup>(2)</sup>. Il s'agit ici de confronter, pour un même répondant, le nombre d'enfants déclarés dans les différentes vagues afin de voir s'il est cohérent d'une vague à l'autre et de mieux comprendre les éventuels décalages observés. Le champ de l'étude est donc limité aux personnes ayant participé aux trois vagues de l'enquête, soit 5 436 observations.

Quelle est l'ampleur des incohérences entre ces différents indicateurs ? Ces décalages proviennent-ils d'une sous-déclaration des enfants en vague 2 et/ou en vague 3 ou de doubles comptes d'enfants lors de la première vague ? Quels enfants disparaissent d'une vague à l'autre : des enfants appartenant au ménage lors de la vague précédente, des enfants vivant ailleurs, des enfants décédés ? Les personnes ayant sous-déclaré un ou des enfants en deuxième vague sous-déclarent-elles aussi leurs enfants à la troisième vague ? Y a-t-il un profil spécifique des personnes pour lesquelles des incohérences ont été relevées ?

L'ampleur de la sous-déclaration étant importante et les données désormais diffusées auprès de la communauté des chercheurs, l'objectif de cette étude est d'attirer l'attention des utilisateurs de l'enquête française Erfi-GGS sur ce problème de qualité des données. Plus largement, elle vise à susciter de la part des autres pays participants au programme GGS une évaluation similaire et à attirer plus largement l'attention des concepteurs de futures enquêtes sur le risque de ne s'appuyer que sur des questions indirectes pour déterminer le

---

(2) Sur cette question, on pourra se référer à l'étude conduite sur 13 pays ayant réalisé la première vague de l'enquête GGS (Neels *et al.*, 2011).

nombre d'enfants d'une personne, qui plus est lorsqu'il s'agit d'une thématique centrale de l'étude.

## I. L'enquête Erfi-GGS et le dénombrement des enfants

### *Les données : l'enquête Generations and Gender Survey*

Dans le cadre du programme d'enquêtes longitudinales et comparatives *Generations and Gender* auquel une vingtaine de pays participe ([www.ggp-i.org](http://www.ggp-i.org)), la France a réalisé à l'automne 2005 la première vague de l'*Étude des relations familiales et intergénérationnelles*<sup>(3)</sup> (Erfi-GGS). Celle-ci porte de manière générale sur la famille. Âgée de 18 à 79 ans, la personne tirée au sort dans le logement pour répondre était notamment interrogée sur son enfance, sa situation conjugale passée et présente, ses enfants, sa situation professionnelle et le cas échéant celle de son conjoint, son bien-être, ses opinions (pour plus de détails sur le questionnaire international, voir Vikat *et al.*, 2007 ; pour plus de détails sur l'enquête française, voir Régnier-Loilier, 2012). Les mêmes personnes étaient recontactées trois et six ans plus tard (automne 2008 et automne 2011) pour répondre à un nouveau questionnaire. Aucun proxy n'était autorisé : si la personne ne souhaitait plus répondre, si elle était décédée, etc., l'entretien n'avait pas lieu avec un autre membre du ménage.

D'une vague à l'autre de l'enquête, le questionnaire est très proche, tant dans son architecture que dans son contenu ou dans la formulation des questions afin de garantir la comparabilité entre les vagues (Régnier-Loilier, 2006, 2010, 2012). De la même manière, le mode de collecte a été identique aux trois vagues en France : questionnaire administré sous CAPI<sup>(4)</sup> par des enquêteurs de l'Insee, en face-à-face au domicile des personnes sur l'ensemble du territoire métropolitain. Dans la mesure du possible, les mêmes enquêteurs étaient sollicités pour les différentes vagues, même s'il n'est malheureusement pas possible de repérer pour une personne donnée si elle a été interrogée ou non par le même enquêteur lors des différentes vagues.

À la première vague de l'enquête (2005) 10 079 personnes ont répondu, 6 534 à la deuxième (2008) et 5 781 à la troisième (2011)<sup>(5)</sup>. Parmi les répondants de 2011, 5 436 personnes avaient répondu à la fois en 2005 et en 2008 tandis que 345 personnes n'avaient pas pu être interrogées lors de la deuxième vague. Si une variable de pondération corrige la déformation de l'échantillon liée à

(3) Dénomination française de l'enquête *Generations and Gender Survey*. Ce changement de nom s'explique par l'imprécision du terme « gender », jugé trop scientifique et peu éloquent par les personnes interrogées lors des tests de l'enquête.

(4) Computer-assisted personal interviewing : collecte assistée par ordinateur, notamment gestion automatique des filtres.

(5) Voir Régnier-Loilier et Guisse (2012) pour une description précise de l'attrition et des facteurs associés.

l'attrition, elle n'est pas utilisée dans cet article, aucune représentativité n'étant ici recherchée.

### *La détermination du nombre total d'enfants dans l'enquête Erfi-GGS*

De manière générale, le questionnaire a été conçu de sorte à ne pas redemander en vagues 2 et 3 des informations déjà collectées lors de la première vague, tout en évitant de confronter directement le répondant à ses propres réponses aux vagues précédentes<sup>(6)</sup>. Toutefois, dans la mesure où les situations familiales sont susceptibles d'évoluer entre les différentes interrogations, la description des habitants du logement, des enfants vivant hors du ménage et des enfants décédés est reposée à l'identique à chaque vague d'enquête. Les questions doivent ainsi permettre de repérer les naissances et les décès d'enfants survenus entre les différentes vagues d'enquête, le départ d'enfants du foyer parental ou leur éventuel retour dans le ménage.

Le questionnaire ne contient donc pas de question directe sur le nombre total d'enfants d'une personne. Celui-ci est déterminé *a posteriori* par décompte des enfants décrits dans trois modules distincts :

- 1) le module sur les « personnes vivant habituellement dans le logement ». On connaît alors précisément le lien entre chacune des personnes décrites dans le logement et le répondant : conjoint/ami du répondant, enfant eu avec le conjoint actuel, enfant eu avec un précédent conjoint, bel-enfant, enfant adopté, enfant accueilli (placement), frère ou sœur du répondant, frère ou sœur du conjoint, parent du répondant, parent du conjoint, grand-parent, petit-enfant ou arrière-petit-enfant, autre membre de la famille du répondant, autre membre de la famille du conjoint, autre personne n'appartenant pas à la famille ;
- 2) le module « enfants non cohabitants », introduit par une question dont la formulation variait selon deux situations :

Si le répondant avait décrit un ou des enfants dans le logement : « 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 aujourd'hui décédés, nous en parlerons après : Oui/Non » et, si oui, « Combien ? ».

Si le répondant n'avait pas décrit d'enfants dans le logement : « Avez-vous adopté ou avez-vous eu vous-même des enfants ? Ne prenez pas en compte les beaux-enfants, les enfants accueillis ni les enfants aujourd'hui décédés, nous en parlerons après : Oui/Non » et, si oui, « Combien ? ».

(6) Par exemple, une confrontation sous la forme de questions telle que « La dernière fois, vous aviez indiqué vivre avec *n* enfants cohabitants / non cohabitants. Est-ce encore le cas ? » n'était pas prévue dans le cadre du questionnaire de référence de GGS (au niveau international) dans la mesure où celui-ci était conçu pour être également administrable en version « papier », donc sans récupération d'informations issues des vagues précédentes.

Chaque enfant était ensuite décrit précisément : sexe, date de naissance, lien filial de l'enfant avec le conjoint actuel, date de départ du foyer parental, fréquence des rencontres avec le répondant, note de satisfaction de la relation, situation parentale de l'enfant ;

- 3) le module « enfants décédés », introduit par la question « Il arrive que l'on perde un enfant. Avez-vous eu d'autres enfants qui sont aujourd'hui décédés ? Oui/Non » et, si oui, « Combien ? ».

Chaque enfant décédé était ensuite décrit de manière plus succincte que les enfants en vie : sexe, date de naissance, de décès, lien filial de l'enfant avec le conjoint actuel, situation parentale de l'enfant.

Si l'addition des enfants recensés dans ces différentes parties du questionnaire doit théoriquement permettre de déterminer le nombre total d'enfants qu'a eu une personne, l'appariement des données des deux premières vagues de l'enquête (2005 et 2008) a révélé que, pour une proportion non négligeable de répondants, le nombre total d'enfants différait en 2005 et en 2008 (Régnier-Loilier *et al.*, 2011), ce malgré un questionnement identique aux deux vagues. S'il est logique que le nombre total d'enfants soit plus élevé en 2008 qu'en 2005 (suite aux naissances), on observe de nombreux cas pour lesquels le répondant avait déclaré moins d'enfants à la deuxième vague qu'à la première (8 % de l'ensemble des répondants à la deuxième vague). L'ampleur du phénomène pouvait faire craindre des erreurs techniques (mauvais appariement entre les vagues, erreur de calcul du nombre d'enfants, erreur de filtrage dans le questionnaire) mais les vérifications apportées ont permis d'écarter cette hypothèse<sup>(7)</sup>.

En examinant de plus près ce problème d'incohérence, la différence semblait provenir de personnes qui avaient déclaré en première vague des enfants non cohabitants et qui répondaient ne pas avoir d'enfant lors du deuxième entretien, mais sans qu'une explication convaincante ne puisse être apportée. Stratégie d'évitement de la part d'enquêtés se rappelant suite à l'expérience de la première vague que chaque enfant donnait lieu à une interrogation fouillée (une dizaine de questions pour chaque enfant) ? Sous-déclaration involontaire de la part des enquêtés estimant que leur situation n'avait pas changé depuis la vague précédente ? Ou alors, sur-déclaration d'enfants lors de la première vague, par exemple déclarés comme membre du foyer mais également décrits comme vivant en dehors du ménage (hypothèse envisageable en première vague dans

---

(7) Plusieurs rapprochements ont été effectués entre les différentes vagues afin de s'assurer qu'il n'y ait pas d'erreur d'appariement : nombre de frères et sœurs, caractéristiques des parents mais aussi prénom des habitants du logement (à partir des données brutes issues de la collecte). En outre, le nombre total d'enfants étant déduit du décompte de ceux vivant dans le ménage, hors ménage et décédés, la construction du compteur a été vérifiée à chaque vague. Nous sommes également retournés aux programmes informatiques afin de nous assurer que les questions visant à dénombrer les enfants avaient bien été posées systématiquement. Enfin, un éventuel « effet enquêteur » (certains enquêteurs auraient pu indiquer qu'une personne n'avait pas d'enfant sans lui poser la question pour gagner du temps, ceux-ci étant rémunérés au questionnaire et non à l'heure) semble pouvoir être écarté, la sous-déclaration touchant la plupart des enquêteurs et jamais l'ensemble de ses fiches-adresses.

la mesure où la personne tirée au sort dans le ménage pour répondre à l'enquête n'était pas nécessairement celle ayant complété le tableau de composition du ménage)<sup>(8)</sup>.

Aucune investigation plus approfondie n'a pu être menée avant la troisième vague de l'enquête en 2011. Toutefois, lors de la mise en place de celle-ci, une attention particulière a été portée sur ce point. En premier lieu, un effort a été fait lors de la formation des enquêteurs pour les sensibiliser à ce problème et une consigne spécifique a été ajoutée dans le questionnaire afin d'attirer leur attention au moment du module décrivant les enfants vivant en dehors du ménage : « Très 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 savoir si les différences observées tenaient à des omissions à une vague d'enquête ou, au contraire, à des doubles comptes, une question récapitulative plus directe a été ajoutée en vague 3 pour savoir combien d'enfants l'enquêté(e) avait eu au cours de sa vie : « 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 terme de la troisième vague, nous disposons donc, pour un même répondant, de quatre indicateurs de nombre total d'enfants : trois sont issus du décompte des enfants du ménage, hors ménage et décédés à chaque vague (en 2005, 2008 et 2011) et un quatrième indicateur provient de la question récapitulative posée à la troisième et dernière vague. Ce sont les quatre indicateurs mis en regard dans cet article.

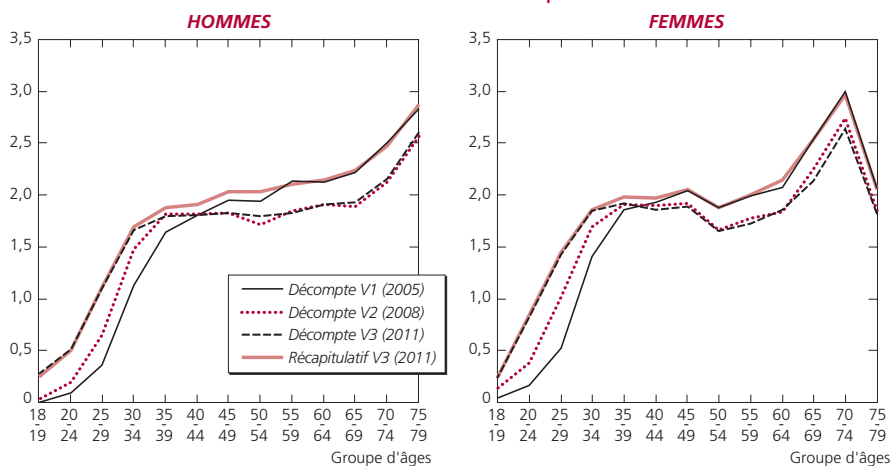
## **II. Importante sous-déclaration des enfants chez les répondants de 45 ans et plus aux vagues 2 et 3**

De manière agrégée, confrontons d'abord le nombre moyen d'enfants par groupe d'âges et sexe, issu des quatre indicateurs de décompte des enfants (figure 1) : enfants du ménage, hors ménage et décédés en 2005 (décompte V1), 2008 (décompte V2) et 2011 (décompte V3), et le nombre total d'enfants obtenu par la question récapitulative posée en 2011 (récapitulatif V3), indicateur qui peut être considéré comme le plus fiable. Rappelons que les données ne sont pas pondérées, l'objectif n'étant pas une mesure de la fécondité par groupe d'âges mais la comparaison d'indicateurs pour un même groupe de personnes.

---

(8) Le tirage au sort du répondant intervenait en effet après la description du ménage par l'un de ses habitants (Régnier-Loilier, 2006). Ainsi, un enfant cohabitait une partie du temps seulement 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 doubles comptes ont pu être repérés mais ils sont très peu nombreux.

Figure 1. Nombre moyen d'enfants par âge (en 2005) selon les 4 indicateurs de décompte des enfants



**Lecture** : d'après le décompte des enfants (du ménage, hors ménage et décédés), les hommes de 55-59 ans ont déclarés 2,14 enfants en 2005 (décompte V1), 1,82 en 2008 (décompte V2), 1,82 en 2011 (décompte V3) et 2,10 en 2011 d'après la question récapitulative (récapitulatif V3).

**Note** : données non pondérées ; âge du répondant en 2005.

**Champ** : répondants aux trois vagues d'enquête (N = 5 436).

**Source** : Ined-Insee, Erfi-GG5123, 2005-2011.

Entre 18 et 44 ans (âge du répondant lors de la première vague), la différence entre les courbes décompte V1, décompte V2 et décompte V3 correspond à des naissances : à chaque âge, 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 et décompte V3 se superposent et passent sous la courbe décompte V1. Entre 45 et 79 ans, les personnes ont donc déclaré en moyenne 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 surdéclaration d'enfants lors de la première vague. Toutefois, le niveau de l'indicateur récapitulatif V3 plaide plutôt en faveur d'une sous-déclaration d'enfants lors des vagues 2 et 3. En effet, la courbe de l'indicateur récapitulatif se superpose presque parfaitement avant 40 ans à la courbe décompte V3, 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 indique 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>(9)</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 enfant en moyenne, chez les personnes âgées de 45 à 79 ans en 2005.

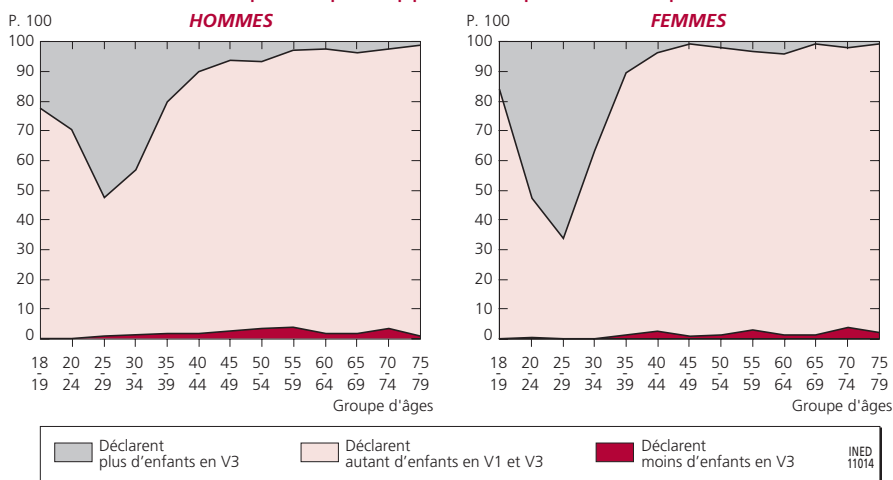
(9) Et probablement aux plus jeunes âges, mais la confrontation du nombre d'enfants en 2005 et 2011 n'a de sens que pour les personnes ayant achevé leur vie féconde.



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 et d'après l'indicateur récapitulatif en V3 (figure 2). Ainsi, trois configurations existent pour un individu  $i$  :

- le nombre d'enfants est identique s'il est déduit du décompte V1 ou du récapitulatif V3 ;
- le nombre d'enfants obtenu par le décompte V1 est inférieur au nombre d'enfants déclaré au récapitulatif V3 (plus d'enfants en 2011 qu'en 2005 : naissances, sous-déclaration en V1, surdéclaration ou double-compte en V3) ;
- le nombre d'enfants obtenu par le décompte V1 est supérieur au nombre d'enfants déclaré au récapitulatif V3 (moins d'enfants en 2011 qu'en 2005 : surdéclaration en V1 ou sous-déclaration en V3).

**Figure 2. Proportion de personnes déclarant autant, plus, moins d'enfants dans le décompte V1 par rapport à la question récapitulative V3**



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

**Note :** données non pondérées ; âge du répondant en 2005.

**Champ :** répondants aux trois vagues d'enquête (N = 5 436).

**Source :** Ined-Insee, Erfi-GGS123, 2005-2011.

Logiquement, la proportion de femmes dont le nombre d'enfants est supérieur en V3 est importante jusqu'à 40 ans environ (maximal chez les 25-29 ans en 2005) compte tenu de la naissance d'enfants à ces âges, puis elle 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, mais aussi une

moindre stabilité des réponses au fil des vagues (nous y revenons plus loin). La proportion de femmes et d'hommes dont 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). 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 à plus de 95 %), indiquant que le nombre total d'enfants déduit du décompte des enfants du ménage, hors ménage et décédés en 2005 est globalement fiable.

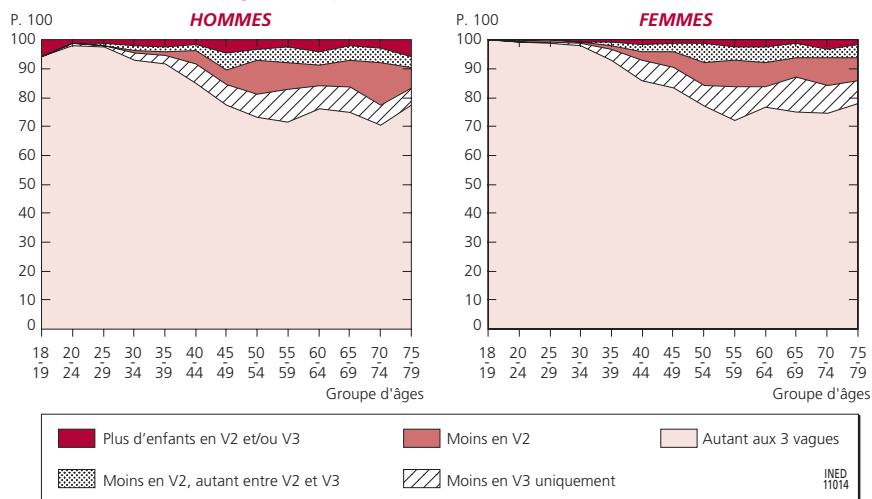
Partant de ce constat, on peut désormais s'affranchir de l'indicateur récapitulatif en V3 et confronter les 3 indicateurs issus du décompte en V1, V2 et V3 au niveau individuel. Toutefois, afin de se focaliser désormais sur la question de la sous-déclaration des enfants déclarés en vague 1 sans que les comparaisons entre les vagues soient perturbées par l'effet des éventuelles naissances aux âges féconds, on ne s'intéresse désormais qu'au décompte des enfants nés jusqu'en septembre 2005, c'est-à-dire avant le début de la collecte de la première vague. 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>(10)</sup>. Pour simplifier, cinq situations sont ici confrontées pour un individu *i* (figure 3) :

- le nombre d'enfants nés avant octobre 2005 est identique aux trois vagues (décompte V1 = décompte V2 = décompte V3) ;
- le nombre d'enfants est inférieur en V3 uniquement : décompte V3 < décompte V1 et décompte V3 < décompte V2, avec décompte V2 = décompte V1 ;
- le nombre d'enfants est inférieur en V2 uniquement : décompte V2 < décompte V1 et décompte V3 = décompte V1 ;
- le nombre d'enfants est inférieur en V2, et égal en V2 et V3 : décompte V2 < décompte V1, et décompte V2 = décompte V3. La sous-déclaration serait donc identique aux vagues 2 et 3, ou bien il y aurait eu une surdéclaration en V1 (possible double compte, par exemple la déclaration d'un même enfant à la fois dans le tableau des habitants du logement et résidant ailleurs, par exemple des étudiants ou des enfants en résidence alternée) ;
- le nombre d'enfants est supérieur en V2 et/ou en V3 : décompte V1 < décompte V2 et/ou décompte V1 < décompte V3 (plus d'enfants en 2008 et/ou 2011 qu'en 2005 : naissance, adoption, déclaration d'un enfant oublié en vague 1).

Avant 40 ans, la confrontation des trois indicateurs offre des résultats relativement cohérents. Par contre, au-delà de 40 ans, tant chez les femmes

(10)  $V1 = V2 = V3$ ;  $V1 > V2 = V3$ ;  $V1 < V2 = V3$ ;  $V1 = V2 > V3$ ;  $V1 = V2 < V3$ ;  $V1 > V2 > V3$ ;  $V1 < V2 < V3$ ;  $V1 > V2 < V3$  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  $V1 > V3$ .

**Figure 3. Proportion de personnes déclarant autant, plus, moins d'enfants nés avant 2005, aux différentes vagues d'après les 3 indicateurs de décompte V1, V2 et V3**



**Lecture :** 76 % des hommes de 60-64 ans (en 2005) déclarent autant d'enfants (nés avant 2005) aux trois vagues, 8 % en déclarent moins en vague 3 (mais autant aux vagues 1 et 2), 7 % en déclarent moins en vague 2 (mais autant aux vagues 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.

**Note :** données non pondérées ; âge du répondant en 2005.

**Champ :** répondants aux trois vagues d'enquête (N = 5 436).

**Source :** Ined-Insee, Erfi-GGS123, 2005-2011.

que chez les hommes, plus d'une personne sur cinq (environ 22 %) sous-déclare des enfants à la vague 2 et/ou à la vague 3. Et, fait notable, la sous-déclaration est relativement volatile : la majorité des personnes qui sous-déclarent ne le font qu'à une seule vague (en vague 2 ou en vague 3).

Bien que l'attention des enquêteurs ait été particulièrement attirée sur ce phénomène lors de la troisième vague (formations des enquêteurs, instructions de collecte affichées à l'écran dans Capi), la sous-déclaration d'enfants en vague 3 a été d'une ampleur similaire à celle observée en vague 2 (respectivement 9 % et 8 %, tableau 1). Par ailleurs, même si le fait d'avoir déclaré moins d'enfants en vague 2 par rapport à la vague 1 conduit plus souvent à en déclarer aussi moins en vague 3 par rapport à la vague 1 (39 % des sous-déclarants en V2 sous-déclarent en V3 contre seulement 7 % des personnes n'ayant pas sous-déclaré en V2)<sup>(11)</sup>, le plus souvent la sous-déclaration n'a pas touché les mêmes personnes : parmi les « sous-déclarants » à l'une des vagues, 42 % n'ont sous-déclaré qu'en vague 3<sup>(12)</sup>, 35 % qu'en vague 2 et 22 % ont sous-déclaré à la fois en vague 2 et en vague 3 (par rapport à la vague 1).

(11) La comparaison entre ces deux proportions pourrait être discutée dans la mesure où le champ n'est pas ici limité aux seules personnes susceptibles d'avoir « sous-déclaré » un ou des enfants, c'est-à-dire celles ayant au moins un enfant lors de la première vague. Toutefois, un écart d'ampleur comparable demeure en limitant l'observation aux parents d'au moins un enfant en vague 1 (39 % contre 9 %).

(12) 327 personnes sur 770 sous-déclarants à l'une des vagues (327 + 171 + 272). La même méthode de calcul a été appliquée pour les autres proportions citées par la suite.

**Tableau 1. Proportion (%) et effectifs de personnes ayant sous-déclaré des enfants en vague 2 et/ou 3 par rapport à la vague 1**

	Pas de sous-déclaration en V3		Sous-déclaration en V3		Ensemble	
	%	N	%	N	%	N
Pas de sous-déclaration en V2	93,4	4 666	6,6	327	91,8	4 993
Sous-déclaration en V2	61,4	272	38,6	171	8,2	443
Ensemble	90,8	4 938	9,2	498	100,0	5 436

*Lecture* : 38,6 % des personnes ayant déclaré moins d'enfants (nés avant 2005) en vague 2 par rapport à la vague 1 ont également déclaré moins d'enfants en vague 3 par rapport à la vague 1.  
*Note* : données non pondérées.  
*Champ* : répondants aux trois vagues d'enquête (N = 5 436).  
*Source* : Ined-Insee, Erfi-GGS123, 2005-2011.

### III. Les omissions : principalement des enfants hors ménage

Les indicateurs utilisés décomptent trois catégories d'enfants : ceux vivant dans le ménage, ceux vivant en dehors du ménage et les enfants décédés. La sous-déclaration observée porte-t-elle sur l'ensemble des enfants déclarés en première vague ou sur une catégorie particulière d'enfants ? On sait par exemple que les enfants décédés tendent à être sous-déclarés dans les enquêtes ou que les hommes sous-déclarent parfois les enfants avec lesquels ils ne vivent plus, notamment suite à une séparation (Toulemon, 2005). La sous-déclaration pourrait également être liée au départ du foyer parental d'enfants qui étaient cohabitants lors de la première vague. Mais la sous-déclaration étant très nette chez les 55 ans et plus (en 2005), donc âgés d'au moins 58 ans en vague 2 et 61 ans en vague 3, âge auquel de moins en moins de personnes vivent encore avec leurs enfants<sup>(13)</sup>, cela laisse supposer que la sous-déclaration porte principalement sur des enfants non cohabitants.

Si l'on compare les personnes ayant déclaré moins d'enfants en vague 2 et/ou 3 qu'en vague 1 à celles ayant déclaré le même nombre d'enfants aux trois vagues<sup>(14)</sup>, les sous-déclarants se caractérisent par une proportion beaucoup plus importante de parents avec au moins un enfant non cohabitant : 84 % des sous-déclarants en vague 2 et 76 % en vague 3 étaient dans ce cas en 2005 contre 51 % des personnes ayant indiqué le même nombre d'enfants aux trois vagues. Ils sont également plus souvent parent d'un enfant décédé en 2005 (12 % contre 4 % des non sous-déclarants). À l'inverse, les sous-déclarants sont proportionnellement moins nombreux à cohabiter avec un ou des enfants en 2005. La sous-déclaration semble donc liée aux enfants non cohabitants et aux enfants décédés en première vague.

(13) En 2005, 28 % des 55-59 ans vivent encore avec au moins un enfant dans le ménage, proportion qui tombe à 16 % chez les 60-64 ans.

(14) Le champ est limité aux personnes ayant déclaré au moins un enfant lors de la première vague, seule population soumise au risque de sous-déclaration lors des vagues suivantes.

Afin d'affiner l'observation et de repérer quels enfants sont sous-déclarés, le nombre total d'enfants nés avant la première vague d'enquête est décomposé en trois compteurs dénombant les enfants du ménage, les enfants hors ménage et les enfants décédés. L'observation est limitée aux personnes ayant sous-déclaré un ou plusieurs enfants en vague 2 d'une part (N = 443 personnes), et en vague 3 d'autre part (N = 498 personnes)<sup>(15)</sup>. Nous comparons le nombre d'enfants non cohabitants déclarés en vague 2 par rapport à la vague 1 puis les déclarations à la vague 3 par rapport à la vague 1 (tableau 2A). Nous procédons de même pour les enfants du ménage (tableau 2B) et pour les enfants décédés (tableau 2C).

La sous-déclaration apparaît d'abord majoritairement liée au fait d'avoir déclaré en 2005 un ou plusieurs enfants non cohabitants. En effet, 9 sous-déclarants sur 10 n'ont indiqué aucun enfant décédé en 2005. Pour eux, la sous-déclaration ne peut donc porter que sur les enfants cohabitants ou non cohabitants en 2005. De même 6 sous-déclarants sur 10 n'ont indiqué aucun enfant cohabitant (première colonne des tableaux 2B et 2C). À l'inverse, 8 sous-déclarants sur 10 avaient en 2005 au moins un enfant non cohabitant.

Les diagonales des tableaux 2 indiquent que le nombre d'enfants déclaré (nés avant la vague 1) est le même aux deux vagues mises en regard (2005 et 2008, ou 2005 et 2011) ; les valeurs situées au-dessus de cette diagonale indiquent que le nombre d'enfants déclaré en 2008 (ou 2011) est inférieur à celui déclaré en 2005 ; les valeurs situées sous cette diagonale indiquent 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 assez rarement déclaré moins d'enfants décédés (tableau 2C) 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 ». Ainsi, 35 enfants décédés déclarés en 2005 ne sont plus déclarés en 2008 et 34 ne le sont plus en 2011. Si ce nombre d'enfants décédés non déclarés en vague 2 ou en vague 3 est important relativement à l'ensemble des enfants décédés décrits par les sous-déclarants en vague 1 (plus de la moitié), il est cependant loin d'expliquer la sous-déclaration dans son ensemble : cette trentaine de cas représente moins de 10 % du nombre total de sous-déclarants.

En réalité, 80 % des sous-déclarants ont indiqué moins d'enfants non cohabitants en 2008 et en 2011 qu'en 2005 (tableau 2A) : les effectifs se situent 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. Ceci signifie que la sous-déclaration ne concerne pas un enfant non cohabitant parmi la descendance mais l'ensemble des enfants non cohabitants. Par exemple, sur les 443 personnes ayant déclaré moins d'enfants en 2008 qu'en 2005, la moitié (219) avaient déclaré au moins 2 enfants non cohabitants en 2005 et ont indiqué ne pas en avoir du tout en 2008. Les sous-déclarants sont donc dans près de 8 cas sur 10 des personnes

(15) Parmi eux, 134 ont sous-déclaré à la fois en vague 2 et en vague 3.

**Tableau 2. Comparaison du nombre d'enfants (A) non cohabitants, (B) cohabitants, (C) décédés, déclarés en 2005 et 2008 (vagues 1-2), et en 2005 et 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	71	115	132	60	27	405
	1	0	8	6	2	1	17
	2	0	1	2	3	1	7
	3	0	1	0	4	2	7
	4 et +	0	0	0	0	7	7
Total		71	125	140	69	38	443
A. Vagues 1 - 3		Nombre d'enfants hors ménage en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants hors ménage en 2011	0	119	128	132	53	28	460
	1	3	5	6	0	1	15
	2	0	1	4	2	2	9
	3	0	0	1	5	2	8
	4 et +	0	0	2	0	4	6
Total		122	134	145	60	37	498
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	278	35	4	4	1	322
	1	8	36	23	6	0	73
	2	0	1	16	16	1	34
	3	0	0	2	5	2	9
	4 et +	0	0	1	0	4	5
Total		286	72	46	31	8	443
B. Vagues 1 - 3		Nombre d'enfants du ménage en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants du ménage en 2011	0	284	57	24	7	2	374
	1	5	22	38	12	0	77
	2	0	0	6	27	1	34
	3	0	0	0	6	2	8
	4 et +	0	0	0	1	4	5
Total		289	79	68	53	9	498
C. Vagues 1 - 2		Nombre d'enfants décédés en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants décédés en 2008	0	388	30	1	0	0	419
	1	2	16	3	0	0	21
	2	0	0	2	1	0	3
	3	0	0	0	0	0	0
	4 et +	0	0	0	0	0	0
Total		390	46	6	1	0	443
C. Vagues 1 - 3		Nombre d'enfants décédés en 2005					
		0	1	2	3	4 et +	Total
Nombre d'enfants décédés en 2011	0	435	32	0	0	0	467
	1	3	23	0	1	0	27
	2	0	1	2	1	0	4
	3	0	0	0	0	0	0
	4 et +	0	0	0	0	0	0
Total		438	56	2	2	0	498

**Lecture** : parmi les 443 sous-déclarants à la vague 2, 132 personnes ont déclaré avoir 2 enfants hors ménage en 2005 et 0 enfant hors ménage en 2008 (tableau 2A).

**Note** : données non pondérées.

**Champ** : répondants ayant participé aux trois vagues d'enquête, parent d'au moins un enfant en 2005 (cohabitant, non cohabitant ou décédé) et ayant déclaré moins d'enfants (nés avant 2005) en 2008 qu'en 2005 (N = 443) / moins d'enfants (nés avant 2005) en 2011 qu'en 2005 (N = 498).

**Source** : Ined-Insee, Erfi-GGS123, 2005-2011.

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 ». Le même phénomène s'observe pour les sous-déclarants en vague 3.

Par contre, si l'on compare le nombre d'enfants cohabitants décrits aux différentes vagues, la mécanique apparaît assez différente (tableau 2B). Lorsque le nombre d'enfants cohabitants observé en vague 2 et/ou 3 est inférieur à celui observé en vague 1, les effectifs ne sont pas concentrés sur la première ligne du tableau. Par exemple, sur les 46 sous-déclarants de 2008 (vague 2) qui cohabitaient avec 2 enfants en 2005, 23 cohabitent encore avec un enfant en 2008. Dans ces cas de figure, la sous-déclaration pourrait être liée au départ du foyer parental d'un des enfants et donc à une omission dans la description des enfants vivant hors du ménage en 2008.

#### IV. Les sous-déclarants n'ont pas de profil spécifique

Afin de mieux comprendre les omissions observées, nous avons établi un portrait « type » des répondants ayant sous-déclaré le nombre de leurs enfants à la deuxième et/ou à la troisième vague d'enquête, à partir de trois régressions logistiques (tableau 3) : le modèle A modélise la probabilité d'avoir déclaré moins d'enfants en vague 2 qu'en vague 1, le modèle b modélise la probabilité d'avoir décrit moins d'enfants en vague 3 qu'en vague 1, le modèle c modélise la probabilité d'avoir décrit moins d'enfants en vague 2 et/ou 3 qu'en vague 1.

Différents facteurs sont pris en compte, la plupart relatifs à la situation observée lors de la première vague<sup>(16)</sup> :

- caractéristiques individuelles des répondants en 2005 : sexe, âge et niveau de diplôme (ce dernier est considéré comme un proxy d'une plus ou moins grande aisance à comprendre les attendus d'une question) ;
- rupture conjugale depuis la première vague : le fait de s'être séparé peut conduire à une rupture des liens avec les enfants ;
- nombre d'enfants cohabitants, non cohabitants et décédés décrits en 2005 afin d'apprécier l'effet des configurations familiales plus ou moins propices à la sous-déclaration ;
- présence d'enfant(s) non cohabitant(s) avec qui le lien est rompu en 2005 (aucune rencontre par an) : la rupture du lien et le temps qui passe ont pu conduire le répondant à ne plus déclarer ses enfants. En outre, les

(16) Sauf mention contraire dans le tableau. 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.

Tableau 3. Probabilité d'avoir déclaré moins d'enfants en 2008 et/ou 2011 par rapport à 2005 (modèle *logit*)

		Modèle A	Modèle B	Modèle C	Effectif
		Sous-déclaration en vague 2 vs vague 1	Sous-déclaration en vague 3 vs vague 1	Sous-déclaration en vague 2 ou en vague 3 vs vague 1	
Constante		-1,48 ***	-2,24 ***	-0,93 ***	-
Sexe du répondant	Homme	0,00 Réf.	0,00 Réf.	0,00 Réf.	1 670
	Femme	-0,03 -	-0,03 -	-0,05 -	2 371
Groupe d'âges en 2005	18-29 ans	-0,87 *	-1,12 **	-1,31 ***	151
	30-34 ans	-1,48 ***	-0,99 ***	-1,55 ***	366
	35-39 ans	-1,01 ***	-0,44 *	-0,96 ***	530
	40-44 ans	-0,51 **	0,08 -	-0,31 *	475
	45-49 ans	-0,31 -	0,14 -	-0,21 -	442
	50-54 ans	0,00 Réf.	0,00 Réf.	0,00 Réf.	462
	55-59 ans	-0,17 -	0,38 **	0,15 -	494
	60-64 ans	-0,32 -	0,10 -	-0,19 -	364
	65-69 ans	-0,41 *	0,26 -	-0,09 -	335
Niveau d'instruction en 2005	70-74 ans	-0,13 -	-0,11 -	-0,03 -	239
	75-79 ans	-0,40 -	-0,21 -	-0,35 -	183
	Sans diplôme	-0,16 -	0,22 -	-0,02 -	478
	CEP	-0,35 -	0,13 -	-0,23 -	521
	BEPC, brevet élémentaire	0,18 -	0,09 -	0,05 -	316
	CAP, BEP	-0,01 -	0,05 -	-0,08 -	1 141
	Bac technologique ou professionnel	-0,44 -	0,29 -	-0,06 -	283
	Bac général	0,00 Réf.	0,00 Réf.	0,00 Réf.	252
Évolution de la situation conjugale entre 2005 et 2011	Bac + 2	-0,22 -	-0,27 -	-0,20 -	387
	Diplôme supérieur	0,01 -	-0,18 -	-0,15 -	663
	Pas de changement	0,00 Réf.	0,00 Réf.	0,00 Réf.	3 601
Nombre d'enfants cohabitants en 2005	Séparation	-0,34 -	0,64 ***	0,18 -	265
	Mise en couple	-0,09 -	0,24 -	0,03 -	175
Nombre d'enfants non cohabitants en 2005	0	0,00 Réf.	0,00 Réf.	0,00 Réf.	1 855
	1 et +	0,24 -	0,20 -	0,25 *	2 186
Nombre d'enfants décédés en 2005	0	-1,20 ***	-0,46 **	-0,86 ***	1 768
	1	0,07 -	0,07 -	0,02 -	709
Nombre d'enfants avec lien rompu en 2005	2	0,00 Réf.	0,00 Réf.	0,00 Réf.	874
	3 et +	-0,07 -	-0,31 **	-0,25 **	690
Interaction Sexe (femme) * Enfants avec lien rompu en 2005 (1 ou +)	0	0,00 Réf.	0,00 Réf.	0,00 Réf.	3 825
	1 et +	0,88 ***	0,75 ***	0,87 -	216
Souhaite recevoir les résultats (2005)	0	0,00 -	0,00 Réf.	0,00 Réf.	3 859
	1 et +	0,96 ***	0,66 ***	0,99 ***	182
Réponse au revenu en 2005	Oui	0,00 Réf.	0,00 Réf.	0,00 Réf.	3 930
	Non	-0,20 -	-0,01 -	-0,19 -	111
Acceptation enregistrement questions sensibles	A répondu	0,45 -	-0,67 -	0,31 -	3 995
	A refuser de répondre	0,00 Réf.	0,00 Réf.	0,00 Réf.	46
Durée du questionnaire en 2005	A accepté	0,00 Réf.	0,00 Réf.	0,00 Réf.	3 927
	A refusé	0,77 ***	-0,25 -	0,31 -	114
	20-44 minutes	0,02 -	0,17 -	0,08 -	423
	45-59 minutes	0,00 Réf.	0,00 Réf.	0,00 Réf.	1 301
	60-74 minutes	-0,16 -	-0,07 -	-0,16 -	1 126
Évolution des conditions de passage	75 minutes ou plus	0,01 -	-0,22 -	-0,06 -	949
	Non connu	0,04 -	0,18 -	0,11 -	242
Avoir sous-déclaré en 2008 (vague 2)	Quelqu'un présent en V1, plus après	0,30 **	-0,04 -	0,20 **	832
	Autres situations	0,00 Réf.	0,00 Réf.	0,00 Réf.	3 209
R <sup>2</sup>	Non		0,00 Réf.		
	Oui		1,62 ***		
Effectif et %	R <sup>2</sup>	0,06	0,09	0,08	
	Sous-déclaration	443 11,0	498 12,3	770 19,1	4 041
		Même nombre d'enfants	3 598 89,0	3 271 80,9	

**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.

**Note** : données non pondérées.

**Seuils de significativité** : \* 10 % ; \*\* 5 % ; \*\*\* 1 % ; - facteur non significatif.

**Champ** : répondants âgés de 18 à 79 ans en 2005, ayant participé aux trois vagues d'enquête et ayant déclaré au moins un enfant en 2005.

**Source** : Ined-Insee, Erfi-GGS123, 2005-2011.



relations enfants-parents après une séparation étant genrées (Régnier-Loilier, 2013), une interaction avec le sexe du répondant a été introduite dans le modèle ;

- posture du répondant 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 dites « sensibles » au sens de la Cnil (pacs et religion), refus de recevoir les premiers résultats (marquant un moindre intérêt pour l'objet de l'étude), durée du questionnaire (le souvenir d'un questionnaire trop long a pu conduire le répondant à vouloir en limiter la durée lors des vagues suivantes) ;
- comparaison des conditions de passation (présence ou non d'une tierce personne lors de l'entretien) par rapport à la vague 1 : la présence du conjoint peut limiter les risques de mauvaise compréhension des questions et le conjoint peut, le cas échéant, rectifier une réponse, mentionner un oubli, etc.

Le champ est limité aux répondants 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 générale, le pouvoir explicatif des modèles A, B et C est extrêmement faible. Les facteurs pris en compte pèsent donc peu sur le phénomène de sous-déclaration. Ce sont principalement les caractéristiques liées aux « types » d'enfants décrits par le répondant en 2005 qui jouent. Dans les trois modèles, la probabilité de sous-déclaration est significativement plus faible si le répondant n'avait aucun enfant non cohabitant en 2005, ce qui confirme nos premières 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) n'a pas d'influence. La sous-déclaration apparaît également plus probable si le répondant a mentionné avoir un ou plusieurs enfants décédés au moment de la première vague de l'enquête. Ceux-ci tendent donc à ne plus être évoqués lors des vagues suivantes. La présence d'enfants cohabitants en 2005 n'a que peu d'effet sur la probabilité de sous-déclarer des enfants lors des vagues suivantes (seul un effet apparaît dans le modèle C)<sup>(17)</sup>. Enfin, le fait d'avoir décrit un ou plusieurs enfants en 2005 avec qui le lien était rompu (absence de rencontres dans l'année) augmente la probabilité d'avoir déclaré moins d'enfants lors de la ou des vagues suivantes. Le facteur d'interaction avec le sexe du répondant est lui aussi significatif (à l'exception du modèle B) et montre un effet de sous-déclaration bien moins marqué du côté féminin. Ce résultat renforce l'hypothèse d'une omission d'enfants par les hommes suite à une rupture conjugale observée dans *l'Enquête famille de 1999* (Toulemon, 2005).

(17) Une interaction entre le nombre d'enfants cohabitants et le nombre d'enfants non cohabitants a été testée. Aucun effet significatif n'apparaît.

L'évolution des conditions de passation par rapport à la première vague a également un effet significatif dans les modèles A et C : si une tierce personne assistait à l'entretien en 2005 mais que l'entretien se déroule en tête-à-tête à l'une des vagues suivantes, la probabilité d'omission est alors plus grande. La présence d'un tiers, le plus souvent le conjoint, pourrait ainsi jouer un rôle de régulateur des réponses fournies par le répondant, en évitant le risque d'omission ou de mauvaise compréhension des questions.

Nous indiquions précédemment que les sous-déclarants étaient partiellement les mêmes aux deux vagues. Afin de repérer un éventuel effet cumulatif de sous-déclaration d'une vague à l'autre, nous avons inclus dans le modèle B une indicatrice de sous-déclaration lors de la vague 2. L'effet est significatif et il s'agit de la variable dont le paramètre estimé est le plus élevé du modèle (1,62), mettant à jour une probabilité bien supérieure des sous-déclarants à la vague 2 d'omettre à nouveau certains enfants en vague 3, toutes autres choses étant égales par ailleurs.

Au-delà de l'effet de ces deux variables et de la situation parentale du répondant, presque aucune des autres caractéristiques incluses dans les modélisations n'influence la probabilité de sous-déclaration : le diplôme ne joue pas, pas plus que le sexe, ou l'âge passé 50 ans (la sous-déclaration est moindre chez les plus jeunes, cf. les observations de la figure 3) ; on ne relève pas non plus d'incidence 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. Ces derniers résultats ne plaident pas en faveur de l'hypothèse d'une possible « stratégie d'évitement » de la part des enquêtés.

## Conclusion et discussion

La qualité des données relatives aux enfants collectées dans les vagues successives de l'enquête Erfi-GGS doit être interrogée. En premier lieu, cette analyse confirme bien une sous-déclaration des enfants décédés dans les enquêtes. Par ailleurs, un important décalage dans le dénombrement des enfants non cohabitants apparaît entre les trois vagues d'enquête, et ce quel que soit le sexe de la personne interrogée. 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 pouvait suspecter des doubles comptes d'enfants en vague 1, cette piste doit être rejetée. L'introduction d'une question récapitulative du nombre total d'enfants du répondant lors de la dernière vague permet en effet de valider l'information recueillie lors de la première vague et plaide ainsi en faveur d'un phénomène d'omission lors des vagues 2 et 3.

Une récente étude conduite à partir du *continuous British household survey* montrant que, la proportion de personnes sans enfant tend à augmenter au fil des vagues, propose plusieurs pistes explicatives au phénomène : le changement

dans les instructions aux enquêteurs, la réorganisation de l'enquête, la modification du mode de collecte, mais aussi l'allongement substantiel du questionnaire ont sans doute contribué à un moins bon report des enfants (Ni Bhrolcháin *et al.*, 2011). Toutefois, dans l'enquête française Erfi-GGS, aucun de ces facteurs ne semble pouvoir expliquer les omissions observées : l'architecture du questionnaire est identique aux trois vagues, la formulation des questions similaire, le mode de collecte est demeuré inchangé. Seul le fait que les conditions de passation aient pu évoluer, avec la présence d'un tiers lors de la première vague puis un entretien en tête-à-tête lors de la deuxième vague, augmente la probabilité d'omission.

L'hypothèse d'un « effet d'apprentissage » ou « effet de lassitude » peut être proposée : au fil des vagues, les participants se familiariseraient avec le filtrage du questionnaire en anticipant qu'une réponse positive à la question « Avez-vous eu des enfants ? » déclenche ensuite une série de questions pour chaque enfant. Les omissions seraient alors à mettre sur le compte de stratégies d'évitement de la part des enquêtés. Mais cette hypothèse paraît fragile pour plusieurs raisons. En premier lieu, on ne note aucun effet de la durée du questionnaire de la première vague sur la probabilité de sous-déclaration aux vagues suivantes. En outre, la description des enfants intervenant plutôt en début de questionnaire, un effet de lassitude du répondant paraît peu probable à ce moment-là de l'entretien. Par ailleurs, la majorité des omissions relevées lors de la deuxième vague ne se retrouvent pas lors de la troisième vague. Enfin, lors de l'étude des caractéristiques liées à l'attrition entre les vagues 2 et 3 (Régnier-Loilier et Guisse, 2012), nous avons inclus comme variable explicative le fait d'avoir ou non déclaré moins d'enfants en 2008 qu'en 2005 avec l'hypothèse qu'une sous-déclaration pouvait traduire un effet de lassitude de l'enquêté(e) et être accompagnée d'une attrition plus forte, mais aucun effet significatif n'a été relevé.

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) montrent un phénomène plutôt aléatoire, difficile à comprendre. Par élimination, l'explication qui semble à ce jour la plus convaincante est l'imprécision de la question utilisée pour recenser les enfants non cohabitants dans un contexte de réinterrogation. Dans les questionnaires des vagues 2 et 3, on demande aux enquêtés s'ils ont « adopté ou eu des enfants » (ou « d'autres enfants » pour ceux ayant décrit au moins un enfant dans le ménage) mais il n'est pas fait mention explicitement au fait d'avoir « des enfants ne vivant pas dans le ménage » ou « vivant ailleurs ». 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 reparler ou, plus simplement, qu'ils n'avaient pas eu d'enfant depuis la dernière vague. Cette piste est d'autant plus plausible que

les omissions concernent principalement des personnes qui ne cohabitent plus au moment de l'enquête avec leurs enfants, et pour lesquelles la question posée était alors « Avez-vous adopté ou avez-vous eu vous-même des enfants ? » ; tandis qu'elle était précédée de « Nous avons déjà parlé des enfants vivant au sein de votre ménage », pour celles et ceux qui avaient décrit au moins un enfant dans le foyer. Plus encore, cette question suivait alors sans transition une question faisant référence aux 12 derniers mois écoulés : « Au cours des 12 derniers mois, avez-vous gardé des enfants ne faisant pas partie de votre ménage ? ». Certains enquêtés ont alors pu comprendre la question comme « au cours des 12 derniers mois... avez-vous adopté ou avez-vous eu vous-même des enfants ? ». Toutefois, cette explication est fragile dans la mesure où l'enchaînement des questions est resté inchangé au fil des vagues. Comment expliquer alors que ce phénomène d'omission ne se retrouve pas lors de la première vague d'enquête ?

Si cette étude ne permet pas de donner une explication tranchée au phénomène d'omission d'enfants dans l'enquête Erfi-GGS, elle présente 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 exploitations et analyses. À ce jour, aucune imputation n'a en effet été réalisée dans les données diffusées, ni au niveau national, ni dans les bases internationales harmonisées<sup>(18)</sup>. Ainsi, il importe de récupérer les enfants « manquants » de la première vague d'enquête pour toutes les études s'appuyant sur les données de la deuxième ou de la troisième vague. De même, la description des petits-enfants (partie introduite en vague 3) est également biaisée dans la mesure où cette question n'était posée qu'aux répondants ayant au moins un enfant de 16 ans ou plus. Le second intérêt est d'attirer l'attention des concepteurs de futures enquêtes longitudinales sur ce risque de sous-déclaration dès lors que l'on n'introduit pas une question directe pour dénombrer les enfants mais que l'on procède par cumul d'informations collectées à différents moments du questionnaire. La méthode du décompte des enfants cohabitants, non cohabitants et décédés pour déterminer le nombre total d'enfants paraît certes fiable dans l'enquête française, mais uniquement pour la première vague. La collecte de l'enquête française ayant eu lieu sous CAPI, il aurait peut-être été judicieux de mettre en place aux vagues 2 et 3 une procédure de « validation / correction » des informations collectées lors de la vague précédente (par exemple, rappeler la composition du ménage en demandant si la situation avait évolué)<sup>(19)</sup>. Il restera à vérifier si ce problème se retrouve dans d'autres enquêtes, notam-

(18) L'objectif de cette investigation était précisément de comprendre la nature des écarts de réponses entre vagues, en vue d'éventuelles imputations ou corrections. Aucune directive n'est cependant prévue à ce sujet au niveau des consignes d'harmonisation et de diffusion internationale des données.

(19) Ce type de procédure n'est possible que lors d'une collecte assistée par ordinateur. Elle peut aussi avoir des effets pervers. Certaines personnes pourraient douter que l'enquête soit vraiment confidentielle alors que l'enquêteur rappelle des informations collectées trois ans auparavant, et par conséquent ne pas souhaiter participer à la vague suivante. D'autres pourraient remettre en cause ce qui a été collecté lors de la vague précédente.

ment dans les enquêtes GGS qui ont eu lieu dans les autres pays quand les données seront disponibles. Une étude méthodologique réalisée à la suite de la première vague de l'enquête GGS allemande, critique à l'égard de la validité de l'histoire génésique recueillie, concluait également à la trop grande complexité du processus de collecte des enfants non cohabitants et regrettait l'absence d'une question récapitulative simple (Kreyenfeld *et al.*, 2013), point de vue que nous partageons ici.

**Remerciements** : L'auteur remercie vivement les referees et le comité de rédaction pour leurs relectures et suggestions constructives. L'enquête Erfi-GGS a été financée par l'Ined, l'Insee, l'ANR, le COR, la Cnaf, la Cnav, la Dares, la Drees et iPOPs.

## RÉFÉRENCES

- AURIAT N., 1996, *Les défaillances de la mémoire humaine. Aspects cognitifs des enquêtes rétrospectives*, Ined/PUF, Cahier n° 136, 204 p.
- FIRDION J.-M., 1993, « L'effet du rang d'appel et de la présence du conjoint dans une enquête par téléphone », *Population*, 48(5), p. 1281-1314.
- HERTRICH V., 1997, « Les réponses des hommes valent-elles celles des femmes? Une double collecte sur les questions génésiques et matrimoniales dans une population au Mali », *Population*, 52(1), p. 45-62.
- JOYNER K., PETERS H. E., HYNES K., SIKORA A., RUBENSTEIN TABER J., RENDALL M. S., 2012, « The quality of male fertility data in major U.S. Surveys », *Demography*, 49(1), p. 101-124.
- KREYENFELD M., HORNING A., KUBISCH K., 2013, « The German Generations and Gender Survey: Some critical reflections on the validity of fertility histories », *Comparative Population Studies*, 38(1), p. 3-28.
- MAZUY M., LELIÈVRE É., 2005, « Déclarer ses enfants, déclarer ses conjoints : rationalité des locuteurs et mode de questionnement », in Lefèvre C., Filhon A. (dir.), *Histoires de familles, histoires familiales. Les résultats de l'enquête Famille de 1999*, Ined, Cahier n° 156, p. 573-588.
- MAZUY M., RAZAFINDRATSIMA N., LA ROCHEBROCHARD É. DE, 2005, *Déperdition dans l'enquête « Intentions de fécondité »*, Ined, Document de travail n° 129, 36 p.
- NEELS K., DE WINTER T., VERMANT G., 2011, « The quality of demographic data in GGS Wave 1 », communication à la *First User Conference of the GGP "Demographic and social challenges in an aging Europe"*, 23-24 mai, Budapest, Hongrie.
- NÍ BHROLCHÁIN M., BEAUJOUAN É., MURPHY M., 2011, « Sources of error in reported childlessness in a continuous British household survey », *Population Studies*, 65(3), p. 305-318.
- POULAIN M., RIANDEY B., FIRDION J.-M., 1991, « Enquête biographique et registre belge de population : une confrontation des données », *Population*, 46(1), p. 65-88.
- RAZAFINDRATSIMA N., KISHIMBA N., 2004, « La déperdition dans la cohorte Cocon entre 2000 et 2002 », *Population*, 59(3-4), p. 419-448.
- 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)*, Ined, Document de travail n° 133, 232 p.
- RÉGNIER-LOILIER A., 2007, « Conditions de passation et biais occasionnés par la présence d'un tiers sur les réponses obtenues à l'enquête Erfi », *Économie et statistique*, n° 407, p. 27-50.
- RÉGNIER-LOILIER A., 2010, *Présentation, questionnaire et documentation de l'Étude des relations familiales et intergénérationnelles (Erfi-GGS2)*, Ined, Document de travail n° 165, 212 p.
- RÉGNIER-LOILIER A., 2012, *Présentation, questionnaire et documentation de la troisième vague de l'Étude des relations familiales et intergénérationnelles (Erfi-GGS 2011)*, Ined, Document de travail n° 187, 328 p.

- RÉGNIER-LOILIER A., 2013, « Quand la séparation des parents s'accompagne d'une rupture du lien entre le père et l'enfant », *Population et sociétés*, n° 500, 4 p.
- RÉGNIER-LOILIER A., GUISSÉ N., 2009, « Mise en scène de la vie quotidienne. Dit-on la même chose en présence de son conjoint ? », in Régnier-Loilier A. (dir.), *Portraits de familles. L'Étude des relations familiales et intergénérationnelles*, Paris, Ined, Grandes enquêtes n° 2, p. 195-218.
- RÉGNIER-LOILIER A., GUISSÉ N., 2012, « Attrition entre les trois vagues d'enquête », in Régnier-Loilier A. (dir.), *Présentation, questionnaire et documentation de la troisième vague de l'Étude des relations familiales et intergénérationnelles (Erfi-GGS 2011)*, Document de travail n° 187, p. 29-43.
- RÉGNIER-LOILIER A., SABONI L., VALDES B., 2011, *Presentation and modifications to the GGS questionnaire in France (wave 2)*, Ined, Document de travail n° 173, 146 p.
- RENDALL M. S., CLARKE L., PETERS H. E., RANJIT N., VERROPOULOU G., 1999, « Incomplete reporting of men's fertility in the United States and Britain: A research note », *Demography*, 36(1), p. 135-144.
- TOULEMON L., 1997, « The fertility of step-families: The impact of childbearing before the current union », Communication à la PAA (*Population Association of America*), 29-31 mars, New-York.
- TOULEMON L., 2005, « Enfants et beaux-enfants des hommes et des femmes », in Lefèvre C. et Filhon A. (dir.), *Histoires de familles, histoires familiales. Les résultats de l'enquête Famille de 1999*, Cahier n° 156, p. 59-77.
- VIKAT A., SPÉDER Z., BEETS G., BILLARI F., BÜHLER C. *et al.*, 2007, « Generations and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course », *Demographic Research*, 17(14), p. 389-440.

**Arnaud RÉGNIER-LOILIER • INCOHÉRENCE DU NOMBRE D'ENFANTS DÉCLARÉS ENTRE LES VAGUES DE L'ENQUÊTE FRANÇAISE GENERATIONS AND GENDER SURVEY**

L'enquête Erfi (*Étude des relations familiales et intergénérationnelles*) est la déclinaison française de l'enquête internationale et longitudinale GGS (*Generations and Gender Survey*). Les mêmes personnes ont été interrogées à trois reprises, en 2005, 2008 et 2011. Si l'enquête est conçue de façon à éviter une redondance du questionnement d'une vague à l'autre, la situation familiale de la personne est susceptible d'évoluer au cours du temps. On décrit ainsi à chaque vague les enfants du répondant vivant dans son logement, ceux qui résident ailleurs et les enfants décédés. De ces différents modules du questionnaire est déduit par addition le nombre total d'enfants (aucune question directe sur le nombre total d'enfants du répondant n'est prévue). Un important phénomène de sous-déclaration des enfants est mis à jour lors des vagues 2 et 3 de l'enquête française. L'objectif de cet article est de mesurer l'ampleur des omissions afin d'avertir les utilisateurs potentiels des données, de repérer quels enfants « disparaissent » et de voir si cela correspond à des profils de répondants spécifiques. Cette étude interroge plus généralement la méthode du décompte des enfants décrit dans différentes parties d'un même questionnaire pour déterminer le nombre total d'enfants d'une personne et invite à procéder à des vérifications analogues dans les autres enquêtes GGS.

**Arnaud RÉGNIER-LOILIER • INCONSISTENCIES IN REPORTED NUMBER OF CHILDREN IN SUCCESSIVE WAVES OF THE FRENCH GENERATIONS AND GENDER SURVEY**

The ERFI survey (*Étude des relations familiales et intergénérationnelles*) is the French version of the international Generations and Gender Survey. The same respondents were interviewed three times, in 2005, 2008 and 2011. Although the survey was designed to avoid redundant questions from one wave to the next, a respondent's family situation is likely to change over time. In each wave respondents were therefore asked about their children living in the dwelling, those living elsewhere and those deceased. The results of these separate questionnaire modules were then summed to give the total number of children (no direct question about the total was included). Substantial under-reporting of children in Waves 2 and 3 of the French survey was observed. This article aims to measure the extent of the omissions in order to alert potential users of the data, to identify which children "disappear" and establish whether this corresponds to particular respondent profiles. More broadly, the study calls into question the method of summing the children recorded in various parts of the same questionnaire in order to determine the respondent's total number of children, and suggests that similar verifications should be made in other GGS surveys.

**Arnaud RÉGNIER-LOILIER • INCOHERENCIA DEL NÚMERO DE NIÑOS DECLARADOS EN LAS TRES OLEADAS DE LA ENCUESTA FRANCESA GENERATIONS AND GENDER SURVEY**

La encuesta Erfi (*Étude des relations familiales et intergénérationnelles*) es la declinación francesa de la encuesta internacional y longitudinal GGS (*Generation and Gender Survey*). Las mismas personas han sido interrogadas tres veces consecutivas, en 2005, 2008 y 2011. Bien que la encuesta esté concebida para evitar una redundancia en las preguntas de una oleada a otra, la situación familiar de la persona puede variar en el curso del tiempo. Se pregunta así en cada oleada sobre los hijos del encuestado que viven en su alojamiento, los que viven en otro domicilio y los que han muerto. De estos diferentes módulos del cuestionario se deduce por adición el número total de hijos (en ningún momento se hace una pregunta directa sobre dicho número). En las dos últimas oleadas de la encuesta francesa se ha producido un número importante de omisiones en la declaración de los hijos. El objetivo de este artículo es medir la importancia de estas omisiones a fin de advertir a los usuarios potenciales de los datos, de identificar qué niños "desaparecen" y ver si ello corresponde a perfiles específicos de encuestados. De manera más general, este estudio cuestiona el método de cómputo de los hijos descrito en las diferentes partes de un mismo cuestionario para determinar el número total de hijos de persona e invita a proceder a verificaciones similares en las otras encuestas GGS.

---

**Mots-clés :** enquête *Generations and Gender Survey* (GGS), enquête *Étude des relations familiales et intergénérationnelles* (Erfi), méthodologie d'enquête, enquête longitudinale, qualité des données, nombre d'enfants déclarés, collecte.

**Keywords:** *Generations and Gender Survey* (GGS), *Étude des relations familiales et intergénérationnelles* (ERFI), survey methodology, longitudinal survey, data quality, reported number of children, data collection.