

# LA DEPERDITION AU FIL DES TROIS VAGUES DE L'ENQUETE LONGITUDINALE ERFI-GGS

Arnaud Régnier-Loilier & Nelly Guisse<sup>1</sup>

*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.

Lors de la première vague d'enquête (2005), 10 079 personnes âgées de 18 à 79 ans vivant en ménage ordinaire ont été interrogées. Parmi elles, seules celles ayant donné leur accord pour que l'on reprenne contact trois ans plus tard étaient éligibles pour participer à la vague suivante, soit 89 % de l'échantillon. Mais, au final, seules 6 534 personnes ont pu être réinterrogées en 2008. Parmi les répondants à la deuxième vague, 97 % ont accepté de nouveau d'être recontactés en 2011 pour la 3<sup>e</sup> et dernière édition de l'enquête. Lors de la troisième vague, l'échantillon était donc constitué de ces répondants (6 300 environ) auxquels ont été ajoutées environ 1 200 personnes qui n'avaient pu être interrogées lors de la deuxième vague (impossibles à joindre, absence de longue durée mais aussi refus de participer). Au final, 5 773 personnes ont répondu en 2005 et en 2011, la plupart ayant également participé au questionnaire de 2008.

Comment les personnes ont-elles été suivies entre chaque vague d'enquête ? Quelles sont les sources de la déperdition entre chacune des vagues (perdue de vue, hors champ, refus, etc.) ? Quels sont les facteurs associés à l'attrition ? Les facteurs liés à l'attrition sont-ils identiques d'une vague à l'autre ou observe-t-on un effet de sélection de l'attrition au fil des vagues ?

---

<sup>1</sup> Nelly Guisse (Credoc ; à l'époque chargée d'études à l'Ined) a apuré la 3<sup>e</sup> vague de l'enquête Erfi et constitué le dictionnaire des codes. Elle a également réalisé la plupart des traitements statistiques présentés dans cet article. Je l'en remercie ainsi que Laurent Toulemon (Ined) pour son expertise et ses idées.

## Introduction

L'attrition représente l'érosion continue et sélective d'un échantillon initial au fil des vagues successives d'une enquête panel. Ce phénomène est « continu » dans la mesure où la probabilité qu'un individu sorte du champ d'observation augmente mécaniquement à mesure que les passages se répètent, les risques de mobilité géographique et de lassitude du répondant s'ajoutant aux facteurs de non réponse des enquêtes transversales (Laurie, Smith et Scott, 1999) ; il est « sélectif » car la probabilité de réinterrogation diffère en fonction des caractéristiques sociodémographiques des répondants et des conditions dans lesquelles les questionnaires précédents ont été administrés.

Outre le problème manifeste de la diminution de l'effectif qui peut compromettre la robustesse des tests statistiques, l'attrition peut conduire à une déformation de la structure de l'échantillon de départ telle qu'elle biaise les résultats et l'interprétation qui en est faite (Razafindratsima et Kishimba, 2004).

L'Ined et l'Insee ont réalisé entre 2005 et 2011 l'*Étude des relations familiales et intergénérationnelles*, version française de l'enquête internationale *Generations and Gender Survey*<sup>2</sup>. L'objectif de cette présentation se limite à la description de l'ampleur de l'attrition entre les 3 vagues de l'enquête et à l'identification de certains facteurs associés à l'attrition entre la première et la troisième vague.

## 1. Taux de réponse et attrition : tableau récapitulatif

L'échantillon initial de 2005 était composé de 18 000 fiches-adresses (FA). Parmi ces adresses, 15 % n'étaient pas utilisables ou n'ont pas été utilisées car le nombre total de répondants attendu (10 000) avait été atteint (soit 2 688 FA) ; 12 % ont donné lieu à un refus de participer (2 242). Les autres sont des hors-champs, des personnes impossibles à joindre (dont sans doute certaines correspondent à des refus déguisés), des absences de longue durée ou des personnes inaptes pour répondre (problème de santé, non francophones) ; enfin, certaines ont donné lieu à la description du ménage mais l'enquête n'a pas été réalisée ensuite (hors champ ? refus de la personne sélectionnée de répondre ?).

Parmi les 10 079 répondants à la première vague, 88 % (9 099) avaient accepté que l'on reprenne contact avec elles. Un courrier de remerciement-relance a été adressé à celles et ceux qui ne souhaitaient pas poursuivre. On leur demandait alors si « vraiment » ils ne voulaient pas que l'on reprenne contact avec eux dans trois ans : 150 ont finalement accepté.

Au total, 758 personnes ont été perdues de vue (y compris décès et hors champ, c'est-à-dire personnes parties vivre en institution ou à l'étranger) entre la vague 1 et la vague 2. L'échantillon pour la réalisation de la vague 2 était donc constitué de 8 341 personnes.

Parmi elles, 165 étaient « hors champ » lors du passage de l'enquêteur en 2008 (parties en institution, à l'étranger ou décédées) ; 794 ont refusé de répondre au questionnaire en 2008 et 547 n'ont pu être jointes (absence longue durée, déménagement, etc.) ; enfin, 42 répondants ont été identifiés comme n'étant pas le répondant de la première vague (*proxy* comme le conjoint par exemple, ou personne sans lien avec le ménage) et ont donc été supprimés de la base de données de la deuxième vague. Celle-ci compte au final 6 534 observations et, parmi elles, 97 % avaient accepté que l'on reprenne contact pour la 3<sup>e</sup> vague.

---

<sup>2</sup> Pour plus de détails sur l'enquête, voir Régnier-Loilier, 2012.

Entre les vagues 2 et 3, 48 personnes ont annoncé leur refus de poursuivre ou un proche nous a annoncé le décès du répondant. Pour la troisième vague, 6 296 répondants aux vagues 1 et 2 étaient donc éligibles. À ceux-ci a été ajouté un échantillon de 1 274 personnes n'ayant pas répondu à la deuxième vague (principalement les personnes impossibles à joindre lors de la deuxième vague et celles ayant refusé de répondre mais ne s'étant pas opposées à ce qu'un enquêteur les contacte pour la 2<sup>e</sup> vague). Plus précisément, pour ces derniers, un courrier leur avait été envoyé avant la troisième vague (printemps 2011) expliquant que nous n'avions pu les interroger en 2008 mais que nous souhaitions pouvoir les réinterroger en 2011. S'ils ne le souhaitaient pas, on leur offrait la possibilité de nous le faire savoir par mail, courrier ou téléphone. Au final, sur les 7 522 adresses, 5 781 ont donné lieu à un entretien lors de la 3<sup>e</sup> vague, avec un taux de succès évidemment beaucoup plus important parmi les personnes ayant répondu aux deux premières vagues (87 %) que parmi celles qui n'avaient pas été interrogées en 2008 (27 %).

L'attrition entre les deux premières vagues a donc été bien plus élevée (35 %) qu'entre les deux suivantes (17 %), résultat classique dans les enquêtes panel. Au final, sur l'ensemble de la période d'enquête (2005-2011), l'attrition s'élève à 43 %<sup>3</sup>, laquelle est dans environ un cas sur deux imputable à des refus (de poursuivre suite à la première ou la deuxième vague, ou de répondre lors du passage de l'enquêteur), dans un cas sur deux à des pertes de contact, des personnes impossibles à joindre, hors champ (départ en institutions, à l'étranger) ou décédées. Sur ce dernier point, nous ne connaissons pas le nombre exact de répondants décédés entre 2005 et 2011, mais on l'estime à 430<sup>4</sup>, soit 10 % de l'attrition totale.

Le taux d'attrition observé dans Erfi est au final assez semblable de celui observé dans d'autres enquêtes panel après 6 ans : 42 % pour le *panel européen* (1994-2001) et 45 % pour *les statistiques sur les revenus et les conditions de vie* (2004-...). Le taux d'attrition de ces dernières enquêtes, qui sont annuelles, étaient cependant bien moindre à 3 ans (environ 25 % contre 35 %), semblant indiquer un meilleur suivi à court terme lorsque les vagues sont rapprochées. Notons cependant qu'un certain nombre de facteurs influent sur l'attrition, comme l'autorisation de *proxys* ou le caractère obligatoire ou non de l'enquête. Sur ce dernier point, la déperdition de l'échantillon de l'enquête *SRCV* était très mesurée durant les 4 premières vagues alors que l'enquête était obligatoire mais s'est sensiblement accentuée ensuite lorsqu'elle a perdu son caractère d'obligation. Dans le même ordre d'idée, l'attrition observée dans l'enquête *Santé et Itinéraire professionnel* (SIP) était de 19 % (contre 35 % pour Erfi après 3 ans), différence pour une bonne part expliquée par le caractère obligatoire de l'enquête *SIP* mais aussi par l'annonce dès la première vague que l'enquête comptait deux vagues, ouvrant peu la possibilité aux répondants d'abandonner après la première vague (très peu de refus suite à la première interrogation contre près de 10 % dans Erfi) (Mermilliod, 2012) ou créant un biais de sélection à l'entrée avec peut-être davantage de refus de participer à l'enquête dès la première vague.

Pour plus de détails sur l'attrition dans Erfi, cf. **figure 1** ci-après.

---

<sup>3</sup> Si l'on considère les répondants aux vagues 1 et 3, sans tenir compte du fait d'avoir ou non répondu à la vague 2 ; mais 46 % des répondants de 2005 n'ont pas répondu aux trois vagues (soit à la vague 1, soit aux vagues 1 et 2, soit aux vagues 1 et 3).

<sup>4</sup> Estimation de l'auteur à partir des tables de mortalité par groupes d'âges quinquennaux appliquées à la structure de l'échantillon de 2005.

**Figure 1.** Détail du taux de réponse et de l'attrition entre les 3 vagues de l'enquête Érfi

		Solde		A retrancher		Repêchage	
		n	%	n	%	n	%
VAGUE 1		<b>Echantillon initial pour Vague 1</b>					
	Hors champ	18019		2430	13,5		
				258	1,4		
	Echec			1657	9,2		
				695	3,9		
				658	3,7		
				2242	12,4		
	Réussite	<b>n (enquête complète V1)</b>		<b>10079</b>	<b>55,9</b>		
ENTRE VAGUES 1 ET 2	Au terme de la vague 1			1130	11,2		
		9099				150	13,3
	Perte V1-V2			758	8,3		
VAGUE 2		<b>Echantillon pour Vague 2</b>					
	Hors champ	8341		165	2,0		
				45	0,5		
	Echec			547	6,6		
				190	2,3		
				8	0,1		
				794	9,5		
				16	0,2		
				42	0,5		
	Réussite	<b>n (enquête complète V2)</b>		<b>6534</b>	<b>78,3</b>		
ENTRE VAGUES 2 ET 3	Au terme de la vague 2			238	3,6		
		6296				1274	
	Perte V2-V3			48	0,8		
	Sous-Ech 1	1274		16,9			
	Sous-Ech 2	6248		83,1			
		<b>Echantillon pour Vague 3 (TOTAL)</b>		<b>7522</b>	<b>100,0</b>		
Sous-Ech 1 VAGUE 3	Hors champ			80	6,3		
				23	1,8		
	Echec			354	27,8		
				73	5,7		
				4	0,3		
				392	30,8		
	Réussite	<b>n (enquête complète répondants V1 et V3)</b>		<b>348</b>	<b>27,3</b>		
Sous-Ech 2 VAGUE 3	Hors champ			150	2,4		
				36	0,6		
	Echec			190	3,0		
				117	1,9		
				6	0,1		
				316	5,1		
	Réussite	<b>n (enquête complète répondant V1, V2 et V3)</b>		<b>5433</b>	<b>87,0</b>		
Total VAGUE 3	Réussite	<b>n (enquête complète V3)</b>		<b>5781</b>	<b>76,9</b>		

Source : Ined-Insee, Erfi-GGS1-3, 2005-2011

Note : classification issue des données brutes de l'enquête ; regroupements opérés par l'auteur.

Note 2 : parmi les 5 781 répondants de 2011, 8 ont été au final retirés car ils ne correspondaient pas aux individus « panel » (répondants assurément différents par rapport aux vagues précédentes)

## 2. Les facteurs liés à l'attrition : éléments descriptifs

### 2.1. Géographie de l'attrition

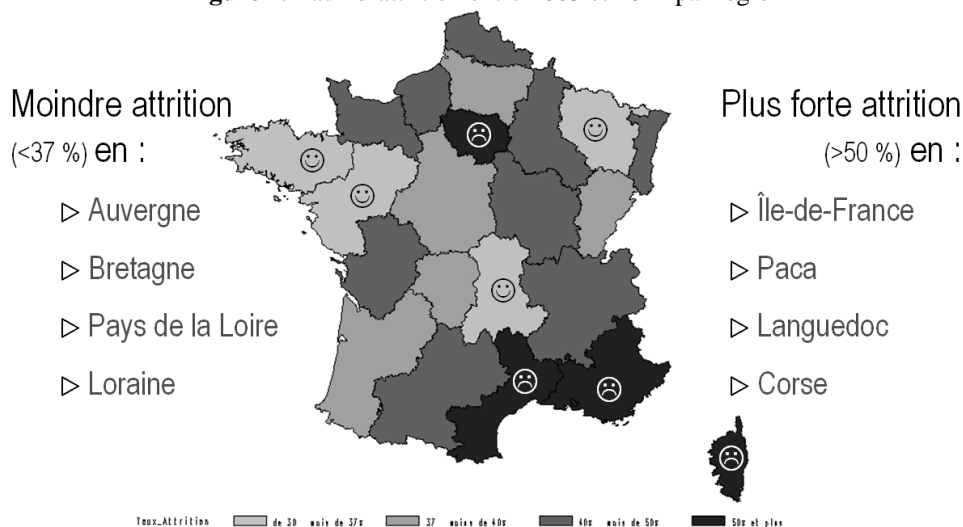
L'attrition a été plus ou moins marquée selon les différentes régions (**figure 2**), avec au final une déperdition plus forte de l'échantillon dans les régions méditerranéennes et en Île-de-France (attrition supérieure à 50 %). Mais c'est surtout entre la première et la deuxième vague que les disparités régionales se sont jouées (**figure 3**). En effet, entre 2008 et 2011, les différences interrégionales sont plus faibles et certaines régions qui avaient subies de fortes pertes entre 2005 et 2008 comblent même une partie de leur retard avec un taux de participation à la troisième vague légèrement supérieur à la moyenne : c'est le cas de la Corse et du Languedoc-Roussillon. En revanche, l'Île-de-France qui avait déjà connu une attrition plus élevée en 2008 creuse son écart en 2011 en enregistrant le plus faible taux de suivi.

À l'inverse, les répondants des Pays de la Loire, les mieux suivis entre les deux premières vagues, confirment leur plus grande « fidélité » par rapport aux autres régions en 2011 ; au final, il s'agit de la région où l'attrition a été la plus faible (31 %).

Ces disparités régionales sont à rapprocher de la densité urbaine, avec une attrition plus faible dans les petites villes et dans les zones rurales (33 % dans les communes de moins de 5 000 habitants ; 39 % dans les communes rurales). Les habitants des grandes villes, et en particulier à Paris et dans les villes méditerranéennes, vivent majoritairement en appartement (en 2005, 73 % des répondants Érfi vivant à Paris sont dans ce cas, plus de 50 % des habitants de villes comptant au moins 100 000 habitants, contre 5 % des habitants de communes rurales), logements souvent plus difficiles d'accès pour les enquêteurs que les quartiers pavillonnaires (code d'accès à l'entrée des immeubles, personnes plus difficiles à joindre). Ces populations très urbaines sont par ailleurs plus mobiles géographiquement ; elles ont plus souvent déclaré avoir l'intention de déménager dans les trois années à venir (20 % d'intentions fermes de déménager dans les trois ans en 2005 parmi les répondants des villes de plus de 100 000 habitants contre moins de 8 % dans les communes rurales).

Toutefois, pour l'Île-de-France, l'attrition plus importante tient à des refus plus fréquents (plus forte défiance, moindre disponibilité ?) contrairement aux régions méditerranéennes où elle s'explique davantage par d'autres motifs<sup>5</sup> (les résidences secondaires y sont plus nombreuses pouvant expliquer davantage d'impossibles à joindre et d'absences de longue durée ; la population est également plus âgée, ceci pouvant expliquer davantage de départs en institution et de décès).

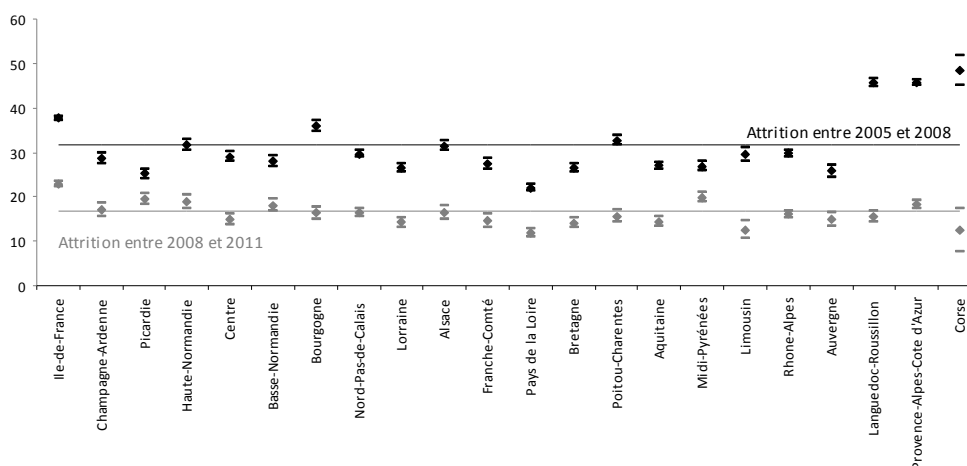
**Figure 2.** Taux d'attrition entre 2005 et 2011 par région



Source : Ined-Insee, Erfi-GGS1-3, 2005-2011

<sup>5</sup> Résultats non présentés ici.

**Figure 3.** Taux d'attrition par région entre 2005 et 2008 et entre 2008 et 2011



Source : Ined-Insee, Erfi-GGS1-3, 2005-2011

Note : intervalles de confiance à 95 %

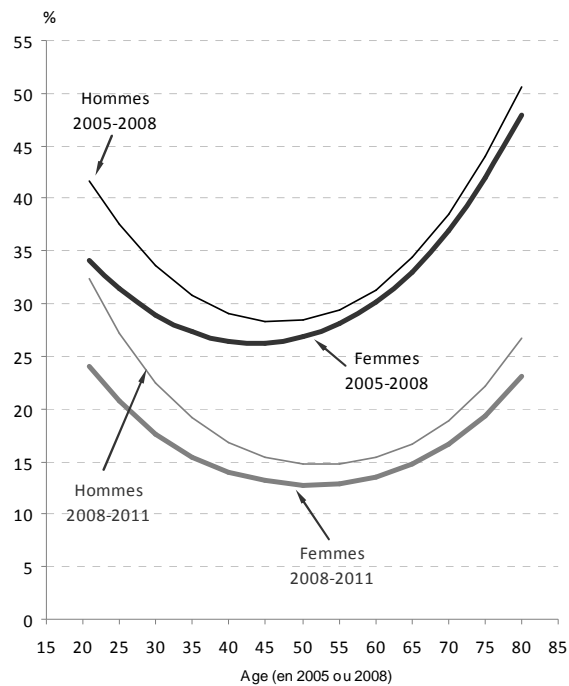
## 2.2. Une attrition plus élevée aux âges extrêmes et d'importantes différences selon le sexe aux jeunes âges

À la fois entre les deux premières vagues et les deux dernières, l'attrition est bien plus élevée aux âges extrêmes, avant 30 ans et au-delà de 60 ans (**figure 4**). Pour les plus jeunes, l'attrition peut s'interpréter par une probabilité plus grande de mobilité : décohabitation de chez les parents, déménagement pour poursuivre des études, se mettre en couple ou suite à l'arrivée d'un enfant, changement de situation vis-à-vis de l'emploi. Pour les plus âgés, si la perte de contact peut également être attribuable à des déménagements suite au passage à la retraite par exemple, elle s'explique aussi par une plus forte propension à avoir refusé de poursuivre l'étude au terme du questionnaire de la première vague (Régnier-Loilier, 2009). À cela s'ajoutent également des sorties de champ plus fréquentes (départs en institution, en maison de retraite par exemple) et à une probabilité accrue de décès.

Plus dans le détail, la comparaison des taux d'attrition entre les vagues 1 et 2 et entre les vagues 2 et 3 laisse apparaître une inversion de tendance entre les plus jeunes et les plus âgés. Ces derniers ont bien moins fréquemment participé à la deuxième vague que les plus jeunes mais ils ont un peu plus souvent répondu qu'eux à la troisième vague. Cela tient en partie à un effet de sélection des plus âgés suite à la première vague : les refus, on l'a rappelé, avaient été nettement plus nombreux de leur part mais, une fois cette sélection opérée, ils ne sortent pas plus fréquemment de l'échantillon que les plus jeunes.

Enfin, quelle que soit la période observée (de la vague 1 à 2 ou de la vague 2 à 3), l'attrition est plus élevée du côté des hommes, mais l'écart est principalement marqué avant 50 ans : les jeunes hommes participent nettement moins à la vague suivante que les jeunes femmes.

**Figure 4.** Taux d'attrition par sexe et âge entre 2005 et 2008 et entre 2008 et 2011



Source : Ined-Insee, Erfi-GGSI-3, 2005-2011

Note : L'âge est celui de 2005 pour l'observation de l'attrition entre 2005 et 2008 ; il correspond à celui observé en 2008 pour l'observation de l'attrition entre 2008 et 2011. Lissage des séries par modélisation incluant l'âge en continu (rapporté à la moyenne d'âge de l'échantillon) et son carré. Ces courbes « résument bien » les valeurs observées pour chaque âge (non présentées ici), en gommant les variations aléatoires.

### 3. Les facteurs d'attrition : « toutes choses égales »

Bien des effets se conjuguent pour « expliquer » l'attrition. Il y a par exemple un lien fort entre la taille de l'unité urbaine, le type de logement et l'intention de déménager. Afin de mesurer l'effet « net » des différents facteurs sur la probabilité d'attrition, un ensemble de modélisations a été réalisé, en jouant sur les variables prises en compte dans la modélisation. Seules trois sont ici présentées (**figure 5**). Le choix des variables retenues fait suite à des hypothèses construites sur la base de la littérature existante sur le sujet (nous ne détaillons pas ici ce point). Il s'agit des caractéristiques des personnes en 2005 (première vague). Pour simplifier la présentation, nous considérons quatre grands groupes de facteurs et nous nous limitons à une description sommaire des résultats.

#### 3.1. L'habitat

Toutes choses égales par ailleurs, notamment en tenant compte du type de logement, de la taille de l'unité urbaine et de l'intention de déménager dans les trois ans, on retrouve un effet spécifique pour l'Île-de-France et les régions méditerranéennes avec une moindre participation à la vague 3. L'explication est cependant différente, l'attrition tenant en Île-de-France à davantage de refus, contrairement à la région méditerranéenne<sup>6</sup>. Les locataires,

<sup>6</sup> Résultat confirmé « toutes choses égales par ailleurs » suite à une modélisation de la probabilité d'avoir « refusé de répondre » *versus* « ne pas avoir répondu pour une autre raison » (résultats non présentés ici ;

personnes vivant en appartement ou encore ayant exprimé leur intention de déménager dans les trois ans ont également été plus souvent perdus entre 2005 et 2011, en raison principalement d'une perte de contact plus fréquente (n'habite pas à l'adresse indiquée, impossible à joindre). À l'inverse, les propriétaires accédants et les résidants de communes de moins de 5 000 habitants ont davantage pris part à la troisième vague : la mobilité des populations y est plus faible (point que l'on « retrouve » par une plus faible attrition chez les agriculteurs, population très peu mobile) et par un accès plus simple au logement.

### **3.2. Les caractéristiques sociodémographiques**

Les hommes, déjà sous-représentés dès la première vague de l'enquête<sup>7</sup>, ont connu une attrition plus forte, confirmant les résultats descriptifs (cf. plus haut) ; il en va de même des personnes plus jeunes et plus âgées. Il est à noter que leur moindre participation ne tient pas à des refus plus fréquents de participer mais davantage à d'autres motifs (impossible à joindre, partis sans laisser d'adresse, partis en institution, etc.). Les peu diplômés, les personnes au chômage et les étrangers ont également moins pris part à la troisième vague de l'enquête. Si une précédente étude a relevé un refus plus fréquent des étrangers d'être réinterrogés (Watson et Wooden, 2009), attribuable à une moindre facilité à s'exprimer en français lorsqu'il ne s'agit pas de leur langue maternelle mais aussi au fait qu'ils peuvent se sentir moins concernés par une enquête nationale, cela ne semble pas être le cas ici. Leur moindre participation à la troisième vague de l'enquête tient davantage à d'autres facteurs qu'un refus explicite.

### **3.3. Santé perçue, sociabilité**

Les personnes s'estimant en mauvaise santé, vivant seules<sup>8</sup> ou étant en situation de monoparentalité avec un enfant en bas âge (moins de trois ans) en 2005 ont moins souvent répondu au questionnaire de 2011 ; cela ne tient pas à des refus plus fréquents de leur part mais à d'autres motifs. On peut faire l'hypothèse de départs en institution ou de décès plus fréquents pour les personnes en mauvaise santé ; pour les personnes seules, on peut penser à une prise de contact plus difficile pour l'enquêteur (plus le logement compte d'habitants, plus l'enquêteur a de chances de tomber sur quelqu'un lorsqu'il s'y rend).

La « sociabilité » joue également un rôle significatif : les personnes ayant déclaré ne pas avoir été confidentes de leur entourage au cours des douze derniers mois<sup>9</sup> ont ainsi moins souvent pris part à la troisième vague de l'enquête, avec des refus plus fréquents de poursuivre l'étude.

À l'inverse, et suite au constat d'enquêteurs lors des bilans de tests ou des collectes, nous avons émis l'hypothèse d'une plus forte participation des personnes éprouvant un sentiment de solitude, la visite de l'enquêteur étant parfois perçue comme une occasion d'échanger avec

---

disponibles auprès des auteurs). Dans la suite de cette partie, on se réfèrera à cette même modélisation dès lors qu'elle permet « d'éclairer » nos résultats.

<sup>7</sup> La moindre participation des hommes est « classique » dans les enquêtes (moindre disponibilité et/ou intérêt pour répondre, plus difficiles à joindre, etc.).

<sup>8</sup> L'effet de cette variable est gommé dans les modèles 2 et 3 par la prise en compte de la durée du questionnaire (les deux variables sont en partie liées, l'entretien ayant été plus court pour les personnes seules en raison du filtrage du questionnaire : pas de description du conjoint, pas de description de l'organisation quotidienne du ménage et de la prise de décisions au sein du couple, etc.).

<sup>9</sup> L'intitulé de la question était « Au cours des douze derniers mois, quelqu'un vous a-t-il parlé de sa vie privée ou de son moral ? ».



l'extérieur. Toutefois, l'inclusion d'indicateurs de cet ordre<sup>10</sup> dans le modèle ne laisse pas apparaître d'effets significatifs sur la probabilité d'avoir participé à la troisième édition de l'enquête (résultats non présentés ici).

### 3.4. Posture par rapport à l'enquête

Les études méthodologiques mettent l'accent sur l'importance de la posture du répondant par rapport à l'étude, son intérêt pour le sujet de l'enquête et la manière dont s'est déroulé l'entretien. Si nous ne disposons que de peu d'informations à ce niveau, quelques indicateurs permettent cependant d'éprouver cette hypothèse pour l'enquête Erfi : le fait d'avoir refusé de répondre à certaines questions (revenu par exemple) ou que la réponse donnée aux questions dites « sensibles » (religion et Pacs) soit enregistrée<sup>11</sup> ; le fait de ne pas avoir souhaité être destinataire des résultats de l'enquête ; ou encore la durée du questionnaire de 2005.

Une posture de défiance par rapport à la collecte, avec par exemple le refus que l'enquêteur conserve les réponses données aux questions sensibles, conduit à un refus plus fréquent de participer aux vagues suivantes de l'enquête. Il en est de même pour les personnes ayant refusé de répondre à certaines questions (revenu mensuel par exemple). La plus forte attrition pour ces personnes tient pour beaucoup à des refus bien plus fréquents de poursuivre l'étude, sans doute parce qu'elles ont estimé que le questionnaire était trop indiscret.

Le refus de recevoir les premiers résultats de l'enquête marque un moindre intérêt pour l'objet de l'étude et conduit, sans surprise, à davantage de refus de poursuivre et, *in fine*, à une attrition nettement plus marquée.

Par ailleurs, si l'on s'attendait à une attrition plus forte pour les personnes ayant eu un premier entretien sensiblement plus long que la moyenne, suite à un possible effet de lassitude, c'est l'inverse que l'on observe : les personnes dont le questionnaire de la première vague a duré entre 1h15 et 2h30 ont une propension plus forte à avoir participé à la troisième vague. Ce résultat, qui semble aller à l'encontre des idées préconçues que l'on se fait (on se limite généralement à une heure de questionnement craignant un effet de déconcentration mais aussi de lassitude chez le répondant), s'interprète en réalité assez bien dans le cas Erfi. Par le jeu des filtres, la variance de la durée du questionnaire est importante, et un questionnaire était d'autant plus long que la situation personnelle du répondant était en adéquation avec la thématique de l'enquête. Ces personnes (pour lesquelles l'entretien est long) ont ainsi pu se sentir davantage intéressées et concernées par l'enquête, expliquant leur meilleure participation aux vagues successives. À l'inverse, le fait d'intégrer la durée du questionnaire (Modèle 2, **figure 5**) annule en partie l'effet de la variable « type de ménage » observé dans le Modèle 1 : le fait de vivre seul(e) sans enfant dans le ménage diminue la probabilité d'avoir répondu aux vagues suivantes si l'on ne tient pas compte de la durée du questionnaire mais cet effet disparaît lorsque la durée du questionnaire est prise en compte.

---

<sup>10</sup> Plusieurs questions permettaient d'approcher ce sentiment de solitude. Par exemple : « Je vais maintenant vous lire plusieurs phrases sur votre vie actuelle. Pouvez-vous m'indiquer pour chacune dans quelle mesure elles vous correspondent... Vous ne vous sentez pas suffisamment entouré(e) », ou « Pouvez-vous me dire à quelle fréquence vous avez ressenti les choses suivantes au cours de la semaine dernière... Vous vous êtes senti(e) seul(e) ».

<sup>11</sup> Le questionnaire de 2005 comportait deux questions sensibles au sens de la loi informatique et liberté, qui impliquaient que l'on avertisse le répondant de la présence de ces deux questions et qu'on lui demande en fin d'entretien s'il acceptait que l'on enregistre ses réponses ; le cas échéant, il devait alors signer un document (autorisation expresse), nominatif, confirmant qu'il acceptait l'enregistrement de ses réponses à ces deux questions. Pour plus d'explications sur ce point, voir Régnier-Loilier, 2006.

Outre la durée du questionnaire, c'est aussi la disponibilité au quotidien des personnes qui peut avoir un effet sur le suivi. En particulier, nous posons l'hypothèse que le fait d'exercer une activité professionnelle à temps plein, avec un temps de travail élevé, pouvait conduire à une moindre propension à avoir participé aux vagues suivantes. À cette fin, le Modèle 3 (**figure 5**) intègre une variable de situation d'activité tenant compte du statut d'activité (salarié ou non) et du temps de travail de la personne lors de la première vague (variable qui se substitue à la catégorie socioprofessionnelle). Ni l'exercice d'une activité à temps partiel, ni une durée hebdomadaire de travail supérieur à 40 heures n'ont d'impact.

#### **4. D'une vague à l'autre : accentuation de la déformation de l'échantillon ou effet de sélection ?**

Indépendamment des facteurs d'attrition observés entre les vagues 1 et 3, on peut se poser la question de la déformation de l'échantillon au fil des vagues. Entre les vagues 1 et 2, l'attrition n'avait pas été aléatoire (Régnier-Loilier, 2009) et entre les vagues 1 et 3, elle ne l'est pas non plus (cf. section précédente). Se pose alors la question de savoir s'il y a eu accentuation de la déformation de l'attrition entre les vagues 2 et 3 (l'influence des facteurs d'attrition observés entre les vagues 1 et 2 se serait donc rejouée entre les vagues 2 et 3) ou si l'attrition avait été sélective entre les deux premières vagues mais qu'elle serait devenue plus aléatoire entre les vagues 2 et 3.

Dit autrement, on peut faire l'hypothèse que certains facteurs ont pu jouer lors de la première vague mais ne plus jouer ensuite, par effet de sélection. Par exemple, les personnes peu intéressées par la thématique de l'enquête ou ayant trouvé les questions trop indiscretes ont pu refuser de poursuivre dès le terme de la première vague ; l'échantillon constitué pour la deuxième vague était donc de ce point de vue sélectionné, les personnes encore dans l'échantillon étant les plus volontaires et les plus intéressées pour poursuivre. À l'inverse, on peut faire l'hypothèse que d'autres variables, liées par exemple à la probabilité de mobilité ou de sortie de champ (décès, départ en institution), n'ont pas lieu de jouer différemment entre les vagues et se sont rejouées à l'identique entre les vagues 1 et 2 et entre les vagues 2 et 3.

À partir de quelques indicateurs, il semble que les deux effets se combinent. L'ampleur de l'attrition observée entre les vagues 1 et 2 est bien supérieure à celle observée entre les vagues 2 et 3, plaidant en faveur d'un effet de sélection ; par contre, la dispersion de la variable de pondération de la vague 3 est supérieure à celle de la vague 2<sup>12</sup>, indiquant plutôt une accentuation de la déformation de la structure de l'échantillon.

Afin d'affiner l'analyse, nous avons dans un premier temps mis en regard deux modélisations calquées sur le même principe que le modèle 2 de la **figure 5**, mesurant la probabilité de suivi entre 2005 et 2008 et entre 2008 et 2011 (**figure 6**)<sup>13</sup>. La comparaison des R<sup>2</sup>, classiquement faibles en sciences sociales, laisse cependant apparaître une moins bonne qualité du modèle « Suivi V2 V3 ». Bien que les modèles ne soient pas directement comparables (effectifs différents notamment), cela semble plaider en faveur d'un effet de sélection, du moins sur la base des variables retenues dans les modélisations : elles expliquent

---

<sup>12</sup> La comparaison a été effectuée après normalisation des poids de 2008 et de 2011, afin de les rendre comparables, indépendamment de l'ampleur de l'attrition très différente entre 2005 et 2008 et entre 2008 et 2011.

<sup>13</sup> Sauf mention contraire dans le tableau, les caractéristiques sont celles observées en 2005 pour le premier modèle (suivi V1 – V2) et en 2008 pour le second modèle (V2 – V3).

« moins » le suivi entre V2 et V3 qu'entre V1 et V2. Toutefois, la comparaison des modalités significatives laisse apparaître peu de différences, la plupart des facteurs liés à l'attrition entre 2005 et 2008 se jouant entre 2008 et 2011.

Quelle que soit la période inter-vague, le suivi a été moindre en Île-de-France, pour les personnes ayant l'intention de déménager, aux âges extrêmes (les jeunes sont plus mobiles, les plus âgés conservent une plus forte propension à sortir du champ), refusant de répondre à certaines questions (revenu du ménage notamment), ne souhaitant pas recevoir les premiers résultats<sup>14</sup> ou encore s'estimant en mauvaise santé ; le suivi est en revanche meilleur chez les agriculteurs, les habitants de communes comptant moins de 5 000 habitants et pour les personnes plus diplômées.

D'autres variables pour lesquelles on posait l'hypothèse d'un effet de sélection (donc d'un moindre effet entre 2008 et 2011) se jouent cependant : c'est le cas du sexe et de la nationalité. Sur ce dernier point, le résultat n'est guère surprenant relativement à ce qui a été mis en avant plus haut : dans Erfi, l'attrition pour ces personnes ne s'explique pas plus par davantage de refus que la moyenne, contrairement aux résultats d'autres études laissant percevoir davantage de refus et un moindre intérêt pour une enquête nationale.

Certains facteurs ne jouent plus entre les vagues 2 et 3 : c'est le cas des plus bas diplômes, de la région méditerranéenne et des locataires (sur ce dernier point, on peut penser à un effet de sélection : les locataires les plus mobiles ont déménagé entre les deux premières vagues et ont été plus souvent perdus ; ceux qui restent sont les moins mobiles limitant le risque de perdre le contact<sup>15</sup>). Enfin, la durée du questionnaire ne joue plus sur le suivi entre 2008 et 2011, mais il est à noter qu'il s'agit de la durée observée en 2005 dans les deux modèles (celle de 2008 n'étant pas disponible).

Au final, donc, la plupart des facteurs liés à une attrition plus forte entre les deux premières vagues se jouent entre les deux vagues suivantes, accentuant la déformation de la structure de l'échantillon.

Une seconde méthode a été testée afin de voir si la différence d'effet d'un même facteur était significativement différente entre V1 et V2 et entre V2 et V3. L'effet de deux caractéristiques peut jouer dans le même sens ou être différent mais on ne sait pas si la différence est significative. Pour ce faire, les échantillons de 2005 et 2008 ont été « empilés » dans un même fichier, avec l'inclusion d'une variable permettant de repérer s'il s'agit de l'échantillon de la première ou de la deuxième vague (indicatrice de vague). On évalue ensuite la probabilité d'avoir participé à la vague suivante<sup>16</sup> sur la base des modèles présentés en **figure 6**, en incluant en plus cette indicatrice de vague ainsi que le facteur d'interaction entre celle-ci et chacune des variables du modèle. Une interaction statistiquement significative indique que l'effet de la variable est différent entre les vagues 1 et 2 et entre les vagues 2 et 3 ; un effet non significatif indique que l'effet est globalement de même ampleur (soit la variable ne jouait pas entre V1 et V2 et son effet est du même ordre entre V2 et V3, soit elle était significative entre V1 et V2 et joue dans le même sens entre V2 et V3). Cette modélisation (**figure 7**) confirme un effet significativement différent pour les non diplômés et en région

---

<sup>14</sup> Un net effet de sélection apparaît au sujet du souhait de recevoir les premiers résultats : au terme du questionnaire de la première vague (2005), 7,7 % n'ont pas souhaité recevoir les premiers résultats contre 3,1 % au terme de la deuxième vague (2008) ; par contre, la proportion de personnes refusant de renseigner la fourchette de revenu du ménage est identique, quelle que soit la vague, et très faible (1,5 %) comparativement à d'autres enquêtes (TeO, Ined-Insee, par exemple).

<sup>15</sup> Dans les faits, les locataires occupent leur logement depuis 8 ans en moyenne en 2005 (médiane de 4 ans) contre 9 ans en moyenne en 2008 (médiane de 5 ans).

<sup>16</sup> Pour l'échantillon de la vague 1, d'avoir répondu à la vague 2 ; pour l'échantillon de la vague 2 d'avoir répondu à la vague 3.

méditerranéenne (attrition plus forte en vague 2 qui disparaît en vague 3), pour l'âge, mais aussi pour la durée du questionnaire et le fait de souhaiter recevoir les premiers résultats (effet moindre en V3 allant dans le sens d'un effet de sélection au fil des vagues pour cette variable).

La question de la déformation de l'échantillon dans le cadre d'une enquête longitudinale questionne en creux les biais induits sur les variables d'intérêt. Ainsi, en dehors d'une modification significative de la structure de l'échantillon de l'enquête *Intentions de fécondité* (Ined, 1998-2003) liée à une attrition élevée (70 %), la déperdition a également entraîné une modification du lien entre la variable d'intérêt (intention d'avoir un enfant) et ses facteurs explicatifs (Mazuy *et al.*, 2005).

Les variables d'intérêt dans l'enquête Érfi sont nombreuses en raison de la variété des thématiques abordées dans le questionnaire et il n'est pas ici question de répliquer l'étude précédemment citée mais simplement de sensibiliser à l'existence de possibles biais de ce type. Si les quelques explorations faites de manière descriptive semblent indiquer un faible impact de la déformation de la structure de l'échantillon sur la répartition des réponses à un certain nombre de variables d'intérêt après application des poids, il appartient aux chercheurs utilisant les données longitudinales de s'assurer dans une première phase de leurs travaux de la robustesse des données.

## Bibliographie

- Laurie H., Smith R., Scott L., 1999, « Strategies for Reducing Non-response in a Longitudinal Panel Survey », *Journal of Official Statistics*, Vol. 15: 269-282.
- Mazuy M., Razafindratsima N., La Rochebrochard, É. de, 2005, *Déperdition dans l'enquête « Intentions de fécondité »*, Documents de travail, Ined, 129, 36 pages.
- Mermilliod C., 2012, *L'impact de la mise en place d'un suivi d'adresse entre les deux vagues de l'enquête santé et itinéraire professionnel (SIP)*, Document de travail, Série Sources et méthodes, n°37, Drees, 26 pages.
- Razafindratsima N., Kishimba N., 2004, « La déperdition dans la cohorte Cocon entre 2000 et 2002 », *Population*, 59 (3-4) : 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)*, Documents de Travail, 133, 231 pages.
- Régnier-Loilier A., 2009, « La déperdition dans les enquêtes longitudinales : l'exemple de l'enquête Érfi », *Courrier des statistiques*, 128, 55-62.
- 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.
- Watson N., Wooden M., 2009, « Identifying Factors Affecting Longitudinal Survey Response », in Lynn P. (dir.), *Methodology of Longitudinal Surveys*, Chichester, UK : John Wiley & Sons, chapitre 10.

**Figure 5.** Modélisation de la probabilité d'avoir répondu à la troisième vague de l'enquête Érfi (paramètres estimés de la régression logistique)

	MODELE 1	MODELE 2	MODELE 3
<b>Constante</b>	0,96 ***	0,95 ***	0,95 ***
<b>Sexe</b>			
<i>Homme</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
<i>Femme</i>	0,14 ***	0,15 ***	0,15 ***
<b>Age</b>			
Age	-0,14 *	-0,17 **	-0,15 *
Age <sup>2</sup>	-0,63 ***	-0,63 ***	-0,64 ***
<b>Niveau de diplôme</b>			
Aucun diplôme	-0,23 ***	-0,23 ***	-0,23 ***
<i>Diplôme inférieur au Bac</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Diplôme de niveau Bac et plus	0,29 ***	0,29 ***	0,31 ***
<b>Catégorie socioprofessionnelle</b>			
Agriculteur	0,40 *	0,37 *	
Artisan, commerçant et chef d'entreprise	-0,17	-0,16	
Cadre et profession intellectuelle supérieure	0,09	0,09	
Profession intermédiaire	-0,01	-0,01	
<i>Employé</i>	<i>ref</i>	<i>ref</i>	
Ouvrier	-0,05	-0,06	
Chômeur	-0,36 ***	-0,36 ***	
Retraité	0,08	0,09	
Au foyer	-0,08	-0,08	
Etudiant	-0,10	-0,07	
Autre inactif	-0,44 ***	-0,44 ***	
<b>Situation d'activité</b>			
Indépendants, chef d'entreprise			-0,13
Salarié (temps plein, 40h et plus par semaine)			0,06
<i>Salarié (temps plein, moins de 40h par semaine)</i>			<i>ref</i>
Salarié (temps partiel)			0,01
au foyer			-0,08
Chômeur			-0,35 ***
Retraité			0,07
Etudiant			-0,05
Autre inactif			-0,44 ***
<b>Taille de la localité de résidence</b>			
Ville de 100 000 habitants et plus	-0,01	-0,01	0,00
<i>Ville de 5 000 à 99 999 habitants</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Ville de moins de 5 000 habitants	0,33 ***	0,34 ***	0,34 ***
Commune rurale	0,00	0,00	0,02
<b>Région</b>			
Ile-de-France	-0,51 ***	-0,49 ***	-0,48 ***
Bassin parisien	-0,06	-0,05	-0,05
<i>Nord</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Est	0,05	0,04	0,04
Ouest	0,07	0,08	0,08
Sud Ouest	-0,02	-0,03	-0,02
Centre Est	0,06	0,05	0,05
Méditerranée	-0,58 ***	-0,59 ***	-0,58 ***
<b>QUALITE MODELE</b>			
Likelihood Ratio	<.0001	<.0001	<.0001
R <sup>2</sup>	0.1635	0.1670	0.1666
Percent Concordant	69.3	69.6	69.6

	MODELE 1	MODELE 2	MODELE 3
<b>Nationalité</b>			
<i>Française</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Etrangère	-0,58 ***	-0,60 ***	-0,60 ***
<b>Réponse donnée concernant le revenu du ménage</b>			
<i>Réponse</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Ne sait pas	-0,31 **	-0,29 **	-0,28 **
Refus de répondre	-0,57 ***	-0,56 ***	-0,57 ***
<b>Statut d'occupation du logement</b>			
<i>Propriétaire, usufruitier</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Accédant à la propriété	0,14 **	0,12 *	0,11
Locataire, logé gratuitement	-0,12 *	-0,12 *	-0,13 *
<b>Intention de déménager dans les 3 ans</b>			
Non, certainement pas	<i>ref</i>	<i>ref</i>	<i>ref</i>
Ne sait pas, probablement, probablement pas	-0,05	-0,07	-0,07
<i>Oui, certainement</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Oui, certainement	-0,36 ***	-0,37 ***	-0,37 ***
<b>Etat de santé perçue</b>			
<i>Très bon, bon</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Moyen, mauvais	-0,09	-0,09 *	-0,09 *
<i>Très mauvais</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Très mauvais	-0,47 **	-0,51 **	-0,51 **
<b>Type de logement</b>			
<i>Maison</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Appartement, autre	-0,18 ***	-0,17 ***	-0,17 ***
<b>Souhaite recevoir les résultats de l'enquête ?</b>			
<i>Oui</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Non	-1,30 ***	-1,27 ***	-1,27 ***
<b>Signature de l'autorisation express ?</b>			
<i>Signée</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Refus de signer	-1,07 ***	-1,05 ***	-1,05 ***
Non concerné	-0,72 ***	-0,69 ***	-0,69 ***
<b>A écouté quelqu'un parler de sa vie</b>			
<i>Oui</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Non	-0,19 ***	-0,17 ***	-0,16 ***
<b>Type de ménage</b>			
<i>Couple sans enfant</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Couple avec enfant(s) de plus de 3 ans	0,05	0,04	0,04
Couple avec enfant(s) de moins de 3 ans	-0,07	-0,09	-0,09
Seul avec enfant(s) de plus de 3 ans	-0,02	0,00	0,00
Seul avec enfant(s) de moins de 3 ans	-0,54 **	-0,53 **	-0,52 **
Ménage de plusieurs personnes	-0,17	-0,13	-0,12
Seul	-0,12 **	-0,08	-0,08
<b>Durée du questionnaire de 2005</b>			
20 à 44 min		-0,24 ***	-0,24 ***
<i>45 min à moins de 1h</i>		<i>ref</i>	<i>ref</i>
1h à 1h15		-0,01	-0,01
1h15 à moins de 2h30		0,16 **	0,16 **
Non communiqué		-0,14	-0,14

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

Champ : tous les répondants de la première vague (en 2005).

Lecture : un coefficient de signe positif (resp. négatif), statistiquement significatif, indique que l'on est en présence d'un facteur qui accroît (resp. réduit) la probabilité d'avoir répondu à la troisième vague, en 2011.

\*\*\* : significatif au seuil de 1% ; \*\* : significatif au seuil de 5% ; \* : significatif au seuil de 10% ; *ref* : catégorie de référence.

**Figure 6.** Modélisation de la probabilité d'avoir répondu à la deuxième vague (« suivi V1-V2) / à la troisième vague (suivi V2-V3) de l'enquête Érfi

	Suivi V1-V2	Suivi V2-V3
<b>Constante</b>	1,32 ***	2,04 ***
<b>Sexe</b>		
Homme	ref	ref
Femme	0,07	0,21 ***
<b>Age</b>		
Age	-0,06	0,14
Age <sup>2</sup>	-0,47 ***	-0,86 ***
<b>Niveau de diplôme</b>		
Aucun diplôme	-0,27 ***	-0,05
Diplôme inférieur au Bac	ref	ref
Diplôme de niveau Bac et plus	0,23 ***	0,35 ***
<b>Catégorie socioprofessionnelle</b>		
Agriculteur	-0,02	0,37
Artisan, commerçant et chef d'entreprise	-0,20	0,05
Cadre et profession intellectuelle supérieure	0,13	0,10
Profession intermédiaire	0,00	0,19
Employé	ref	ref
Ouvrier	-0,06	-0,08
Chômeur	-0,21 **	-0,24
Retraité	0,00	0,29 *
Au foyer	-0,28 ***	0,07
Etudiant	-0,04	-0,36
Autre inactif	-0,23 *	-0,75 ***
<b>Taille de la localité de résidence</b>		
Ville de 100 000 habitants et plus	-0,08	0,09
Ville de 5 000 à 99 999 habitants	ref	ref
Ville de moins de 5 000 habitants	0,33 ***	0,42 ***
Commune rurale	0,07	0,06
<b>Région</b>		
Ile-de-France	-0,50 ***	-0,63 ***
Bassin parisien	-0,10	-0,16
Nord	ref	ref
Est	-0,08	0,03
Ouest	0,01	0,09
Sud Ouest	-0,11	-0,06
Centre Est	-0,08	-0,07
Méditerranée	-0,72 ***	-0,20

QUALITE MODELE		
Likelihood Ratio	<.0001	<.0001
R <sup>2</sup>	0.1507	0.0992
Percent Concordant	68.8	67.4

	Suivi V1-V2	Suivi V2-V3
<b>Nationalité</b>		
Française	ref	ref
Etrangère	-0,42 ***	-0,49 ***
<b>Réponse donnée concernant le revenu du ménage</b>		
Réponse	ref	ref
Ne sait pas	-0,37 ***	-0,30 *
Refus de répondre	-0,63 ***	-0,81 ***
<b>Statut d'occupation du logement</b>		
Propriétaire, usufruitier	ref	ref
Accédant à la propriété	0,13 *	0,07
Locataire, logé gratuitement	-0,25 ***	-0,14
<b>Intention de déménager dans les 3 ans</b>		
Non, certainement pas	ref	ref
Ne sait pas, probablement, probablement pas	-0,04	-0,13
Oui, certainement	-0,40 ***	-0,31 ***
<b>Etat de santé perçue</b>		
Très bon, bon	ref	ref
Moyen, mauvais	-0,05	-0,21 **
Très mauvais	-0,63 ***	-1,26 ***
<b>Souhaite recevoir les résultats de l'enquête ?</b>		
Oui	ref	ref
Non	-1,62 ***	-1,30 ***
<b>A écouté quelqu'un parler de sa vie</b>		
Oui	ref	ref
Non	-0,19 ***	-0,14 *
<b>Type de ménage</b>		
Couple sans enfant	ref	ref
Couple avec enfant(s) de plus de 3 ans	0,00	0,04
Couple avec enfant(s) de moins de 3 ans	0,09	0,14
Seul avec enfant(s) de plus de 3 ans	0,01	-0,20
Seul avec enfant(s) de moins de 3 ans	-0,10	-0,25
Ménage de plusieurs personnes	-0,05	0,16
Seul	-0,08	-0,05
<b>Durée du questionnaire de 2005</b>		
20 à 44 min	-0,21 ***	-0,21 **
45 min à moins de 1h	ref	ref
1h à 1h15	0,09	0,02
1h15 à moins de 2h30	0,19 ***	-0,15
Non communiqué	-0,04	-0,51 ***

Source : Érfi-GGS, Ined-Insee, 2005-2008-2011

Champ : tous les répondants de la première vague (en 2005) pour le modèle 'suivi V1-V2' ; tous les répondants de la deuxième vague (2008) pour le modèle 'suivi V2-V3'.

Lecture : un coefficient de signe positif (resp. négatif), statistiquement significatif, indique que l'on est en présence d'un facteur qui accroît (resp. réduit) la probabilité d'avoir répondu à la vague suivante.

\*\*\* : significatif au seuil de 1% ; \*\* : significatif au seuil de 5% ; \* : significatif au seuil de 10% ; ref : catégorie de référence.

**Figure 7.** Modélisation de la probabilité d'avoir répondu à la vague suivante et facteurs d'interaction avec la vague

	VAGUE 1	INTERACTION
<b>Constante</b>	1,4421	
<b>Vague d'observation</b>		
Première vague (2005)		
Deuxième vague (2008)	0,601 **	
<b>Sexe</b>		
<i>Homme</i>		
Femme	0,10 *	0,11
<b>Age</b>		
Age	-0,09	0,23
Age <sup>2</sup>	-0,44 ***	-0,42 **
<b>Niveau de diplôme</b>		
Aucun diplôme	-0,30 ***	0,25 *
<i>Diplôme inférieur au Bac</i>		
Diplôme de niveau Bac et plus	0,20 ***	0,15
<b>Catégorie socioprofessionnelle</b>		
Agriculteur	0,16	0,21
Artisan, commerçant et chef d'entreprise	-0,16	0,22
Cadre et profession intellectuelle supérieure	0,08	0,02
Profession intermédiaire	0,03	0,16
<i>Employé</i>		
Ouvrier	-0,04	-0,04
Chômeur	-0,20	-0,05
Retraité	0,00	0,28
Au foyer	-0,19 *	0,26
Etudiant	-0,04	-0,32
Autre inactif	-0,21	-0,54 **
<b>Taille de la localité de résidence</b>		
Ville de 100 000 habitants et plus	-0,11 *	0,19 *
<i>Ville de 5 000 à 99 999 habitants</i>		
Ville de moins de 5 000 habitants	0,33	0,09
Commune rurale	0,05	0,01
<b>Région</b>		
Ile-de-France	-0,46 ***	-0,17
Bassin parisien	-0,07	-0,09
<i>Nord</i>		
Est	-0,04	0,07
Ouest	0,01	0,08
Sud Ouest	-0,05	-0,02
Centre Est	0,00	-0,07
Méditerranée	-0,75 ***	0,56 ***

	VAGUE 1	INTERACTION
<b>Nationalité</b>		
<i>Française</i>		
Etrangère	-0,47 ***	-0,03
<b>Réponse donnée concernant le revenu du ménage</b>		
<i>Réponse</i>		
Ne sait pas	-0,43 ***	0,12
Refus de répondre	-0,67 ***	-0,14
<b>Statut d'occupation du logement</b>		
<i>Propriétaire, usufruitier</i>		
Accédant à la propriété	0,20 ***	-0,13
<i>Locataire, logé gratuitement</i>		
Locataire, logé gratuitement	-0,17 ***	0,03
<b>Intention de déménager dans les 3 ans</b>		
Non, certainement pas		
Ne sait pas, probablement, probablement pas	-0,05	-0,08
Oui, certainement	-0,41 ***	0,10
<b>Etat de santé perçue</b>		
<i>Très bon, bon</i>		
Moyen, mauvais	-0,05	-0,16
Très mauvais	-0,71 ***	-0,54
<b>Souhaite recevoir les résultats de l'enquête ?</b>		
<i>Oui</i>		
Non	-1,70 ***	0,40 **
<b>A écouté quelqu'un parler de sa vie</b>		
Oui		
Non	-0,17 ***	0,03
<b>Type de ménage</b>		
<i>Couple sans enfant</i>		
Couple avec enfant(s) de plus de 3 ans	0,03	0,01
Couple avec enfant(s) de moins de 3 ans	0,01	0,12
Seul avec enfant(s) de plus de 3 ans	-0,01	-0,18
Seul avec enfant(s) de moins de 3 ans	-0,35	0,10
Ménage de plusieurs personnes	0,01	0,15
Seul	-0,11 *	0,05
<b>Durée du questionnaire de 2005</b>		
20 à 44 min	-0,25 ***	0,04
<i>45 min à moins de 1h</i>		
1h à 1h15	0,04	-0,03
1h15 à moins de 2h30	0,21 ***	-0,35 **
Non communiqué	-0,11	-0,40 **

QUALITE MODELE	
Likelihood Ratio	<.0001
R <sup>2</sup>	0.1748
Percent Concordant	71.3

Source : Érfi-GGS, Ined-Insee, 2005-2008-2011

Champ : tous les répondants de la première vague (en 2005) et de la deuxième vague (2008).

\*\*\* : significatif au seuil de 1% ; \*\* : significatif au seuil de 5% ; \* : significatif au seuil de 10% ; ref : catégorie de référence.

