

## Chapitre 2



# Attrition et déformation de l'échantillon au fil des vagues de l'enquête Érfi

*Arnaud Régnier-Loilier  
et Nelly Guisse\**

### Introduction

La question de la non-réponse est un enjeu majeur dans les enquêtes par questionnaire, le taux de participation présageant de la représentativité des données. Dans le cas des enquêtes longitudinales, l'enjeu est d'autant plus important que le risque de déperdition s'accroît à mesure que le nombre de vagues augmente. L'attrition consiste en cette érosion continue et sélective de l'échantillon initial au fil des vagues. 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, le risque 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 engendrer une déformation de la structure de l'échantillon de départ et aller jusqu'à biaiser les résultats et l'interprétation qui en est faite (Razafindratsima et Kishimba, 2004).

Cette question de l'attrition se pose particulièrement dans l'enquête *Generations and Gender Survey* en cours dans une vingtaine de pays. Celle-ci prévoit en effet d'interroger à trois reprises les mêmes personnes, à trois ans d'intervalle entre les vagues, afin de permettre à terme des comparaisons internationales. Nous proposons ici un bilan de la déperdition observée dans l'enquête française (Érfi), sur laquelle repose l'essentiel du présent ouvrage<sup>(1)</sup>.

---

\* Les auteurs tiennent à remercier Laurent Toulemon (Ined) pour ses conseils et suggestions d'analyses, ainsi que les deux relecteurs anonymes de ce chapitre.

(1) Pour plus d'informations concernant l'attrition dans l'enquête hongroise mobilisée dans le **chapitre 13**, se rapprocher de l'auteur.

Après avoir proposé une revue de la littérature sur le sujet et présenté les données, cette recherche mesure l'attrition observée entre les différentes vagues de l'enquête française, en prenant soin de distinguer différents types de sortie d'échantillon : refus, perte de contact, sortie du champ. On caractérise dans un deuxième temps l'attrition, en considérant différentes familles de facteurs explicatifs : variables sociodémographiques, lieu de vie des personnes, état de santé et sociabilité des personnes, déroulement des entretiens précédents. Enfin, l'enquête comptant trois vagues, nous cherchons à voir si la déperdition observée entre les deux premières vagues de l'enquête et entre les deux dernières tient aux mêmes facteurs. Dit autrement, nous souhaitons voir si la déformation de l'échantillon s'accroît au fil des vagues.

Si l'étude de l'attrition et des facteurs qui y sont associés est cruciale en aval de la collecte puisqu'elle sert de base au calcul des variables de pondération longitudinales (en particulier lors de l'étape de correction de la non-réponse<sup>(2)</sup>), il est également intéressant de s'y attarder dans un contexte où les enquêtes longitudinales tendent à se développer. L'objectif de ce chapitre n'est donc pas de mesurer les biais occasionnés par l'attrition sur les variables d'intérêt de l'enquête (trop nombreuses, comme l'illustre la variété des études proposées dans cet ouvrage)<sup>(3)</sup>, mais d'apporter des éléments visant à mieux comprendre l'érosion de l'échantillon observée au fil des vagues de l'enquête Érfi, dans la perspective de futures enquêtes à passages répétés. Ces éléments pourront être mobilisés afin de mieux définir le tirage initial d'un échantillon. Les préconisations internationales de l'enquête *Generations and Gender Survey* présentaient par exemple la manière de déterminer la taille initiale de l'échantillon à interroger en vague 1 pour parvenir à l'objectif d'au moins 3 000 femmes en âge reproductif (18-44 ans en vague 1) réinterrogées en vague 3. Cette estimation s'appuyait sur un taux d'attrition attendu de 10 % entre chaque vague (Unece, 2005, p. 11). Ce chapitre offre des éléments chiffrés plus précis. Par ailleurs, le repérage de certaines situations liées au refus de participer aux vagues successives pourrait permettre d'affûter l'argumentaire pour convaincre plus efficacement ces personnes de participer. De même, mieux connaître le profil des personnes perdues de vue entre les vagues pourrait permettre la mise en place d'un suivi plus individualisé des personnes (avec par exemple l'envoi plus régulier de courriers pour les personnes les plus mobiles). Plus globalement, le repérage des profils de personnes pour lesquelles l'attrition est la plus forte peut permettre de sur-échantillonner de manière raisonnée certaines catégories de population, dans la perspective de disposer *in fine* d'une bonne représentativité des répondants au terme des différentes vagues d'enquête.

54 ■

(2) Le calcul des poids longitudinaux s'est fait en deux temps : après examen des facteurs liés à l'attrition (présentés dans ce chapitre), nous avons procédé à la correction de la non-réponse puis au calage sur les marges issues de la première vague (voir Charrance et Razafindratsima, 2012).

(3) Ce chantier reste entier et justifierait à lui seul une étude complète. Bien que chaque chercheur ait été invité à s'assurer que l'attrition n'affectait pas ses résultats, nous avons choisi dans cet ouvrage de nous focaliser sur la publication des résultats.

## ■ Brève revue de la littérature sur les déterminants de la participation à une enquête

Lorsqu'une personne appartient à l'échantillon d'une enquête, deux étapes conditionnent sa participation : la prise de contact (réussite ou échec à entrer en contact avec elle) et, si le contact est établi, la coopération de l'enquêté(e) (acceptation ou refus de répondre). Si les études portant sur l'attrition se sont longtemps restreintes à comprendre le phénomène dans son ensemble, la plupart des chercheurs s'accordent aujourd'hui sur le fait que les déterminants de l'attrition diffèrent selon que la non-réponse provienne d'une perte de contact ou d'un refus explicite à participer de nouveau à l'enquête (Groves et Couper, 1998 ; Watson et Wooden, 2009). La richesse des travaux menés sur la déperdition dans les enquêtes longitudinales permet d'appréhender le phénomène dans toute sa complexité. Deux catégories de facteurs sont habituellement mobilisées pour comprendre l'attrition. Il y a, d'une part, les caractéristiques du répondant, de son ménage ou de sa zone de résidence et, d'autre part, les spécificités de l'enquête : la nature du questionnaire, mais aussi les caractéristiques de l'enquêteur et la qualité de la relation qui s'établit entre lui et l'enquêté, notamment autour de la question fondamentale de la confiance. Pour chacune de ces catégories, il est essentiel de distinguer ce qui joue sur la perte de contact de ce qui influe sur le refus de coopération, en gardant à l'esprit qu'un même facteur peut jouer sur les deux tableaux et que l'effet peut alors se faire dans des sens opposés.

■ 55

### 1. Les déterminants *sociodémographiques de la prise de contact*

On observe traditionnellement que la prise de contact se fait plus facilement auprès des femmes, des personnes âgées et des parents de jeunes enfants : ces individus passent en moyenne plus de temps à leur domicile (Groves et Couper, 1998). Le contact est plus souvent établi dans les ménages comptant plusieurs adultes qu'auprès de personnes vivant seules en raison d'une probabilité accrue qu'au moins l'un des habitants du logement soit présent au moment du passage de l'enquêteur. En revanche, les personnes résidant dans des zones à forte densité urbaine sont plus difficiles à appréhender. Ces zones concentrent une plus forte proportion de locataires, vivant en appartement dans des ensembles parfois très sécurisés et difficiles d'accès (Uhrig, 2008 ; Blom, de Leeuw et Hox, 2011). Les habitants des grandes villes ont également une plus faible probabilité d'être trouvés chez eux car ils y passent en moyenne moins de temps en raison de durées de transport plus longues que dans les petites villes ou à la campagne, mais aussi d'une plus grande offre d'activités sociales et culturelles (Groves et Couper, 1998).

Dans le contexte d'enquêtes longitudinales, la probabilité de perdre le contact avec un répondant augmente à mesure que le risque de mobilité géographique s'élève. Une étude conduite à partir du Panel européen des ménages montre que le principal facteur explicatif de l'attrition tient à la difficulté à suivre des personnes qui déménagent (Breuil-Genier et Valdelièvre,

2001). Ce risque est accru chez les répondants qui déclarent envisager un déménagement, qui n'aiment pas leur quartier ou qui ont déjà connu un déménagement dans les années précédant l'enquête (Uhrig, 2008). Enfin, Philippe Collomb (1979) puis, à sa suite, Benoît Riandey (1988) ont montré le poids prépondérant joué par la détermination des enquêteurs à retrouver la trace des répondants perdus de vue. Dans le cas d'un changement d'adresse, la qualité du suivi dépend alors essentiellement de l'investissement des enquêteurs dans ce qui peut devenir une véritable enquête.

## *2. Les déterminants sociodémographiques de la coopération*

Les recherches sur l'attrition montrent en outre qu'une fois le contact établi entre le répondant et l'enquêteur, la propension à accepter de répondre varie en fonction des caractéristiques sociodémographiques de la personne sollicitée.

Le niveau social et le niveau d'éducation peuvent s'avérer déterminants dans le suivi des enquêtés (Uhrig, 2008), les plus diplômés se montrant plus souvent coopératifs. L'investissement cognitif est parfois plus lourd pour les personnes moins diplômées qui peuvent être moins à l'aise pour répondre à un questionnaire. Les habitants des grandes villes, en plus d'être plus difficiles à contacter, se montrent aussi moins enclins à répondre en raison d'un sentiment diffus d'insécurité et d'un contexte de moindre cohésion sociale dans les zones de grande densité urbaine. Cette peur de l'inconnu permet aussi d'expliquer une moindre coopération des personnes âgées (Holbrook, Green et Krosnick 2003; Uhrig, 2008). Les étrangers refusent aussi plus souvent d'être réinterrogés (Watson et Wooden, 2009) : ils ont une moindre facilité à s'exprimer en français lorsqu'il ne s'agit pas de leur langue maternelle et peuvent aussi se sentir moins concernés par une enquête nationale. Par ailleurs, la participation à des activités sociales, l'implication politique ou associative sont autant de facteurs liés positivement au suivi des répondants (Uhrig, 2008; Stoop, 2005; Lipps, 2007). Enfin, plusieurs chercheurs se sont intéressés au lien entre le fait de connaître des changements dans sa vie et le taux de réponse aux enquêtes. Il semblerait que l'instabilité conjugale, professionnelle ou géographique soit liée positivement au refus de participer. Outre l'effet mécanique d'un déménagement qui peut rendre plus difficile la prise de contact, ces personnes sont moins susceptibles de répondre à une enquête parfois intrusive et coûteuse en temps (Voorpostel et Lipps, 2011). L'effet est d'autant plus marqué que le changement se fait dans un sens défavorable : une rupture conjugale, une période de chômage ou une baisse de salaire par exemple (Fitzgerald, Gottschalk et Moffit, 1998).

## *3. Les caractéristiques du questionnaire et de l'enquêteur sur la réinterrogation*

Les conditions de déroulement de l'enquête influent également sur la probabilité d'accepter d'être réinterrogé. C'est ainsi que la thématique abordée

dans le questionnaire, la qualité des passations précédentes et les caractéristiques de l'enquêteur peuvent ainsi être déterminantes sur le risque qu'un individu sorte de l'échantillon.

D'abord, les personnes plus éloignées du sujet de l'enquête peuvent se sentir peu concernées par l'étude et donc moins disposées à y consacrer du temps. Dans l'enquête Intentions de fécondité (Ined, 1998-2003), l'étude de la déperdition a ainsi mis en évidence que les répondants les plus éloignés des questions liées à un projet d'enfant, à savoir ceux sans enfant et n'en souhaitant pas, avaient une plus grande probabilité de refuser d'être recontactés (Mazuy *et al.*, 2005).

La part des données manquantes aux vagues précédentes peut être un indicateur de ce moindre intérêt porté à l'enquête, si ce n'est d'une certaine attitude de défiance de la part de l'enquêté. Uhrig (2008), Lipps (2007), mais aussi Watson et Wooden (2009) ont ainsi montré que les répondants qui refusent de renseigner les questions sur leur salaire ont une plus grande probabilité de sortir de l'échantillon à la vague suivante. Plus globalement, Loosvedt, Pickery et Billiet (2002) montrent que la non-réponse aux questions dites « sensibles » est liée à une moindre coopération à la vague suivante.

De même, la durée du questionnaire peut présager de la probabilité de réponse aux vagues suivantes. Un questionnaire plus court ou plus long que la moyenne peut augmenter le risque de refus à la vague suivante. Dans le premier cas, il peut traduire un manque d'intérêt ou une moindre proximité du répondant avec la thématique de l'enquête ; dans le second cas, il s'accompagne d'une augmentation du coût d'opportunité en nécessitant un investissement plus lourd de la part du répondant (Watson et Wooden, 2009).

Dans un autre registre, l'obtention d'adresses-relais à la fin d'un questionnaire peut être décisive. Non seulement cela permet de retrouver plus facilement les nouvelles coordonnées du répondant en cas de déménagement mais, au-delà de cet effet mécanique, l'acceptation de fournir ces renseignements peut aussi s'interpréter comme un indicateur d'une disposition d'esprit plus favorable à l'enquête et, par là, d'une plus grande propension à accepter le suivi (Riandey, 1988 ; Laurie, Smith et Scott, 1999).

Enfin, certains auteurs ont montré que l'enquêteur lui-même avait un impact sur le taux de réponse. Au-delà de sa capacité à retrouver l'enquêté, les stratégies mises en œuvre dans l'argumentaire déployé pour le convaincre de répondre sont décisives. Cet effet se retrouve dans l'âge et le niveau d'expérience de l'enquêteur, tous deux liés positivement au taux de participation des répondants (Blom, de Leeuw et Hox, 2011), mais aussi dans ses traits de personnalité, les enquêteurs les plus ouverts et les plus extravertis obtenant de meilleurs résultats (Jäckle *et al.*, 2013).

## II ■ Les données et la méthode

### 1. Quelques rappels sur l'enquête

La France a réalisé à l'automne 2005 la première vague de l'Étude des relations familiales et intergénérationnelles, laquelle a été reconduite trois et six ans plus tard (automne 2008 et automne 2011) auprès des mêmes répondants (voir chapitre 1 pour une présentation détaillée de l'enquête), avec néanmoins une érosion de l'échantillon au fil des vagues. Dans la mesure du possible, les mêmes enquêteurs étaient sollicités pour les différentes vagues, le rôle positif de cette continuité dans des enquêtes longitudinales en face-à-face ayant été démontré (voir par exemple Behr *et al.*, 2005 ; Nicoletti et Peracchi, 2005 ; Burricand et Lorgnet, 2014). Malheureusement, il est impossible ici de déterminer la part d'enquêteurs ayant participé aux différentes vagues, aucune information n'ayant été collectée en ce sens dans l'enquête.

Suite à la première vague d'enquête, seules les personnes ayant donné leur accord écrit (formulaire signé de leur part, sur lequel figurait leur adresse et d'éventuelles informations complémentaires : téléphone, coordonnées de personnes-relais) pour qu'un enquêteur reprenne contact avec elles trois ans plus tard étaient éligibles pour les vagues suivantes. L'actualisation de leurs coordonnées entre les vagues était assurée « artisanalement » par le service des Enquêtes et sondages de l'Ined (aucun rapprochement avec des fichiers administratifs n'était possible). La probabilité d'avoir été perdu de vue entre l'une des vagues d'enquête a été bien moindre pour les répondants pour lesquels nous disposions d'au moins une personne-relais (sans toutefois qu'il n'y ait de différences significatives entre le fait de ne disposer des coordonnées que d'une seule personne-relais ou de deux<sup>(4)</sup>). En outre, le nombre de numéros de téléphone laissés par la personne (jusque trois : domicile, professionnel, portable) s'avère positivement corrélé à la probabilité de garder le contact au fil des vagues<sup>(5)</sup>.

Au final, 10 079 personnes ont répondu à la première vague de l'enquête (2005), 6 534 à la deuxième (2008) et 5 781 à la troisième (2011).

### 2. Les objectifs de l'étude et la méthode choisie

Cette recherche poursuit différents objectifs. Après avoir rappelé les stratégies mises en œuvre afin de limiter la déperdition de l'échantillon entre les vagues, le premier objectif vise à décrire précisément l'ampleur de l'attrition entre chacune d'elles. Nous veillons en particulier à distinguer les différentes sources d'attrition : les refus de prendre part à la vague suivante, la perte de contact, les personnes sortant du champ de l'enquête<sup>(6)</sup> mais aussi la part de l'attrition imputable à la mortalité.

(4) La « fiche de suivi » prévoyait le recueil de plusieurs numéros de téléphone du répondant ainsi que les coordonnées de deux personnes relais.

(5) Résultats non présentés ici mais disponibles auprès des auteurs.

(6) Les personnes partant vivre à l'étranger ou en institution (maison de retraite par exemple) n'étaient pas suivies.

Notre deuxième objectif consiste à repérer différents facteurs liés à l'attrition. En nous appuyant sur la littérature existante sur le sujet et en fonction des données disponibles dans l'enquête, nous avons formulé différentes hypothèses de recherche que nous avons ensuite éprouvées, d'abord de manière descriptive, puis en construisant des modélisations afin de repérer l'effet des différents facteurs, *ceteris paribus* (modèles *logit*). Nous avons considéré différentes familles de facteurs. En premier lieu, les caractéristiques sociodémographiques des personnes : sexe, âge, nationalité, niveau d'études, situation d'activité, situation familiale. Deuxièmement, leur situation résidentielle : taille de l'unité urbaine, type d'habitat (maison ou appartement), statut d'occupation du logement (propriétaire, locataire, etc.) et intention de déménager, avec l'hypothèse d'une mobilité géographique plus ou moins forte et une facilité d'accès au logement par l'enquêteur différente selon les situations. Troisièmement, des facteurs liés à la santé du répondant et à sa sociabilité, que l'on peut approcher dans l'enquête par la fréquence des échanges de confidences avec l'entourage. Enfin, les spécificités du déroulement de l'entretien lors de la première vague représentent une quatrième famille de facteurs : durée du questionnaire, mais aussi attitude de défiance par rapport au questionnaire (mesurée par le refus de répondre à certaines questions) et intérêt porté à l'étude (mesuré par le souhait de recevoir ou non les premiers résultats de l'enquête). Dans cette partie, notre analyse porte principalement sur l'attrition observée entre la première et la dernière vague afin de proposer un bilan général de l'enquête.

Afin de répondre au troisième objectif de notre recherche, nous décomposons ensuite l'attrition en comparant les facteurs explicatifs de l'attrition entre les deux premières vagues de l'enquête et entre les deux dernières. Cette décomposition nous permet de voir si les mêmes facteurs se rejouent d'une vague à l'autre, renforçant alors la déformation de la structure de l'échantillon au fil des vagues, ou si au contraire un « effet de sélection » peut être mis à jour (l'attrition serait sélective dans la première période inter-vague, moins voire plus du tout dans la seconde). Toutefois, si la comparaison des seuils de significativité entre ces deux modèles est éclairante, elle est discutable puisque chacun des modèles ne porte pas sur les mêmes effectifs. En outre, si l'effet d'un facteur est le même dans les deux modélisations, il se peut toutefois qu'il ne soit pas de même ampleur. Ainsi, un modèle unique a ensuite été construit après avoir empilé dans un même fichier les échantillons des vagues 1 et 2 (10 079 + 6 534). On estime, pour ces 16 613 observations, leur probabilité d'avoir participé à la vague suivante en fonction des mêmes variables explicatives présentes dans les modèles disjoints mais, d'une part, en contrôlant par l'échantillon d'origine (indicateur « vague 1 » ou « vague 2 ») et, d'autre part, en incluant un facteur d'interaction entre cette indicatrice et chacune des dimensions explicatives. C'est le résultat de ces interactions qui nous intéresse alors en premier lieu : une interaction statistiquement significative indique alors 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 au contraire que l'effet est du même ordre entre

les différentes vagues. Nous posons ici l'hypothèse d'une moindre influence entre les vagues 2 et 3 des facteurs liés au refus de répondre (les personnes plus méfiantes par rapport à l'enquête ou qui se sentent peu concernées par la thématique seront sorties de l'échantillon dès la fin de la première vague : par exemple, celles ayant refusé de répondre à certaines questions ou n'ayant pas souhaité recevoir les résultats), mais d'un effet cumulatif au fil des vagues pour les facteurs davantage liés à la perte de contact (c'est-à-dire à la mobilité géographique, au temps de présence dans le logement, au risque de sortir du champ de l'enquête : départ en institution ou décès).

### III ■ Les résultats

#### 1. Taux de réponse et attrition : une vue d'ensemble

Le tableau 1 présente la taille de l'échantillon initial et les différentes sources d'attrition (refus, perte de contact, etc.) au fil des trois vagues de l'enquête. Parmi les 18 019 fiches-adresses (FA) échantillonnées en 2005, 15 % n'étaient pas utilisables ou n'ont pas été mobilisées car le nombre total de répondants attendus (10 000) avait été atteint (soit 2 688 FA)<sup>(7)</sup> ; 12 % ont donné lieu à un refus de participer (2 242). Les autres sont des personnes hors champ, impossibles à joindre (dont sans doute certaines correspondent à des refus déguisés), des absences de longue durée ou des personnes incapables à 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 (refus de la personne sélectionnée parmi les membres du ménage de répondre).

Parmi les 10 079 répondants à la première vague, 88 % (8 949) avaient accepté que l'on reprenne contact avec eux trois ans plus tard pour la deuxième vague. Un courrier de remerciements-relance a été adressé à celles et ceux qui ne souhaitaient pas poursuivre. On leur demandait alors si « vraiment » ils refusaient que l'on reprenne contact avec eux dans trois ans : 150 ont finalement accepté. Outre ce premier courrier de « relance », une procédure plus générale de suivi et d'actualisation des adresses avait été appliquée sur l'ensemble de la période d'enquête (encadré 1).

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 (2008) 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.), 42 personnes ayant répondu au

(7) Les deux échantillons de réserve ont été débloqués en fin de collecte car un effet de saturation de l'échantillon apparaissait, laissant craindre que l'objectif des 10 000 répondants ne pourrait être atteint. Toutefois, en raison de contraintes budgétaires, la collecte a été stoppée une fois l'objectif atteint.

## Encadré 1. Procédure de suivi et d'actualisation des adresses entre les vagues

Entre 2005 et 2011, chaque répondant était destinataire de deux courriers par an (liste des envois ci-dessous), avec pour objectif de garder le contact avec le répondant, de l'intéresser à l'objet de l'étude et d'actualiser ses coordonnées. Ainsi, différents types de courriers ont été envoyés : remerciements pour avoir participé à la vague précédente, premiers résultats, cartes de vœux, lettre d'annonce de la vague suivante. En voici le détail :

- mars 2006 : lettre de remerciements<sup>(1)</sup> avec trois modèles de courrier : 1) une relance pour les personnes ayant « refusé » de poursuivre l'enquête ; 2) des remerciements avec demande de coordonnées de personnes-relais à contacter en cas de déménagement, si nous n'en disposions pas pour le répondant ; 3) des remerciements « simples » pour les personnes ayant accepté d'être recontactées et ayant donné les coordonnées d'au moins une personne-relais ;
- décembre 2006 : envoi des premiers résultats de l'enquête sur la fréquence des rencontres entre parents et enfants, sous la forme d'un « 4 pages »<sup>(2)</sup> ;
- juin 2007 : envoi de nouveaux résultats sur la répartition des tâches domestiques au sein du couple, sous la forme d'un « 8 pages »<sup>(3)</sup> ;
- janvier 2008 : envoi d'une carte de vœux à l'effigie de l'enquête et accompagnée d'un petit sachet de graines de fleurs à semer chez soi ;
- juin 2008 : envoi d'un courrier d'annonce de la deuxième vague de l'enquête suivi, fin 2008, d'une lettre-avis annonçant le passage prochain d'un enquêteur ;
- mars 2009 : lettre de remerciements selon deux déclinaisons : 1) si nous n'avions pour le répondant aucune coordonnée de personnes-relais (50 % des cas environ), le courrier rappelait l'importance d'en disposer, son rôle, et était accompagné d'une fiche à compléter ; 2) si l'on disposait des coordonnées d'au moins une personne-relais, la lettre envoyée était un remerciement « simple » ;
- décembre 2009 : courrier accompagnant les premiers résultats utilisant les données longitudinales et montrant l'évolution de la répartition des tâches domestiques suite à la naissance d'un enfant<sup>(4)</sup> ;
- juin 2010 : courrier accompagné d'une brochure de présentation de l'ouvrage *Portraits de familles* (ouvrage collectif publié à partir des données de la première vague sur les différentes thématiques abordées dans l'enquête) et d'un marque-page sous forme d'un « QCM » ;
- janvier 2011 : envoi d'une carte de vœux à l'effigie de l'enquête, accompagnée d'un sachet de graines de tomates-cerises ;
- juin 2011 : courrier d'annonce de l'enquête, avec deux versions : 1) l'une adressée aux participants aux deux premières vagues ; 2) l'autre aux non-répondants à la deuxième vague, dans laquelle était précisé le moyen de prévenir l'Ined si toutefois la personne refusait d'être sollicitée pour la troisième vague.

Dans chacun de ces courriers, il était rappelé à la personne l'importance de nous communiquer sa nouvelle adresse si toutefois elle avait déménagé ou envisageait de le faire : un coupon-réponse était systématiquement joint au courrier (remboursement des frais postaux sur demande), mais la personne pouvait aussi nous informer de son changement de situation par téléphone ou par mail (boîte électronique créée pour l'enquête).

Lorsque le courrier n'arrivait pas à destination (la personne avait déménagé sans faire suivre son courrier), il était alors automatiquement retourné au Service des enquêtes et sondages de l'Ined. Une procédure de recherche des nouvelles coordonnées du répondant était alors mise en œuvre : utilisation des informations disponibles sur la fiche de suivi (numéro de téléphone, contact auprès des personnes-relais, recherche dans l'annuaire). Si la recherche était infructueuse, le courrier suivant était cependant renvoyé à l'ancienne adresse. Au bout de trois retours de courrier avec la mention « n'habite plus l'adresse indiquée », la personne était sortie de l'échantillon pour la collecte de la deuxième vague. Pour la troisième vague en revanche, il avait été décidé d'envoyer dans tous les cas un enquêteur sur place : il lui était alors demandé de procéder à une recherche auprès du voisinage, de la mairie, etc. Notons également que pour la dernière vague de l'enquête, le fichier d'adresses avait été confié juste avant la collecte à la Poste afin de procéder à d'ultimes actualisations et corrections d'adresses (la Poste disposant d'un fichier de transfert d'adresse et de suivi du courrier des personnes en cas de déménagement).

(1) Ces remerciements intervenaient environ trois mois après la fin de la collecte, temps nécessaire au Service des enquêtes et sondages de l'Ined (SES) pour saisir l'intégralité des coordonnées des personnes (et éventuelles personnes-relais) ayant accepté que l'on reprenne contact avec elles trois ans plus tard.

(2) Régnier-Loilier A., 2006, « À quelle fréquence voit-on ses parents ? », *Population et sociétés*, 427. Le choix de cette thématique avait été guidé par le souhait d'intéresser la majorité des répondants à l'enquête (18-79 ans) en proposant un sujet où la plupart des personnes se sentiraient concernées (soit en tant que parent, soit en tant qu'enfant).

(3) Bauer D., 2007, « Entre maison, enfant(s) et travail : les diverses formes d'arrangement dans les couples », *Études et résultats*, Drees, 570.

(4) Régnier-Loilier A., 2009, « L'arrivée d'un enfant modifie-t-elle la répartition des tâches domestiques au sein du couple ? », *Population et sociétés*, 461.

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

			Taille de l'échantillon à différents temps de l'enquête		Type de sortie d'échantillon		Repêchage			
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		
Vague 1	Hors champ	Échantillon initial pour la vague 1 Logement vacant, détruit, impossible à identifier Fiche-adresse non traitée Échec Impossible à joindre, absence longue durée, etc. Personne inapte pour répondre (malade, non francophone, etc.) Réalisation partielle (TCM, abandon, etc.) Refus de participer Réussite <i>n</i> (enquêtes complètes V1)	18019							
					2430	13,5				
					258	1,4				
					1657	9,2				
					695	3,9				
				658	3,7					
				2242	12,4					
			10079	55,9						
Entre les vagues 1 et 2	Au terme de la vague 1	Refus de poursuivre en V2 Repêchage après courrier de relance Accepte le suivi (y compris après repêchage)			1130	11,2				
	Perte V1-V2	N'habite plus l'adresse indiquée (NPAI), refus, décès	9099				150	13,3		
					758	8,3				
Vague 2	Hors champ	Échantillon pour la vague 2 Décès, étranger, institution, accès adresse impossible Fiche-adresse non traitée Échec Impossible à joindre, absence longue durée, etc. Personne inapte pour répondre (malade, non francophone, etc.) Réalisation partielle (abandon en cours) Refus de participer Aucun habitant du ménage en 2008 ne correspond au répondant de 2005 Enquête réalisée en 2008 auprès d'une personne autre que le répondant de 2005 Réussite <i>n</i> (enquêtes complètes V2)	8341							
					165	2,0				
					45	0,5				
					547	6,6				
					190	2,3				
					8	0,1				
					794	9,5				
					16	0,2				
				42	0,5					
			6534	78,3						

Tableau 1. suite

			Taille de l'échantillon à différents temps de l'enquête		Type de sortie d'échantillon		Repêchage		
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Entre les vagues 2 et 3	Au terme de la vague 2	Refus de poursuivre en V3 Refus Vague 2, NPAI, etc. conservés dans le suivi Accepte le suivi au terme du questionnaire de la deuxième vague	6 296		238	3,6	1 274		
	Perte V2-V3	Refus exprimés, décès			48	0,8			
Vague 3		<i>Échantillon pour la vague 3 dont :</i>	7 522						
	Hors champ	<i>– n'a pas répondu à V2 (sous-échantillon 1)</i>	1 274	16,9					
		Échec	Décès, étranger, institution, accès adresse impossible Fiche-adresse non traitée			80	6,3		
	Réussite	Échec	Impossible à joindre, absence longue durée, etc. Personne inapte pour répondre (malade, non francophone, etc.)			23	1,8		
		Réussite	Réalisation partielle (abandon en cours)			357	28,0		
			Refus de participer			73	5,7		
		<i>n</i> (enquêtes complètes répondants V1 et V3)	345	27,3	4	0,3			
Hors champ	<i>– a pas répondu à V2 (sous-échantillon 2)</i>	6 248	83,1						
	Échec	Décès, étranger, institution, accès adresse impossible Fiche-adresse non traitée			150	2,4			
Réussite	Échec	Impossible à joindre, absence longue durée, etc. Personne inapte pour répondre (malade, non francophone, etc.)			36	0,6			
		Réalisation partielle (abandon en cours)			187	3,0			
	Réussite	Refus de participer			117	1,9			
		<i>n</i> (enquêtes complètes répondants V1, V2 et V3)	5 436	87,0	6	0,1			
	<i>n</i> (enquêtes complètes V3)	5 781	76,9	316	5,1				

Note : classification issue des données brutes de l'enquête ; regroupements opérés par les auteurs. 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). Source : Érfi-GGS123, Ined-Insee, 2005-2008-2011.

questionnaire de la deuxième vague ont été identifiées comme n'étant pas le même répondant qu'à la première vague (*proxy* comme le conjoint par exemple, ou personne sans lien avec le ménage) et ont donc été supprimées de la base de données de la deuxième vague (s'ajoute à cela quelques situations minoritaires : personnes inaptes à répondre, abandon en cours d'entretien). Celle-ci compte au final 6534 observations et, parmi elles, 96 % avaient accepté que l'on reprenne contact pour la troisième vague.

Entre les vagues 2 et 3, 48 personnes ont annoncé leur refus de poursuivre ou un proche nous a annoncé leur décès. Pour la troisième vague, 6296 répondants aux vagues 1 et 2 étaient donc éligibles. À ceux-ci a été ajouté un échantillon de 1274 personnes ayant répondu à la première vague mais pas à la deuxième. Il s'agit principalement de personnes impossibles à joindre ou ayant refusé de répondre alors qu'elles ne s'étaient pas opposées au terme de la première vague à ce qu'un enquêteur reprenne contact avec elles. Le choix de tenter de recontacter ces personnes pour la troisième vague était motivé par le fait que les enquêteurs indiquaient qu'il ne s'agissait pas nécessairement de refus définitifs de participer à l'enquête, mais de refus de répondre à cette période-là, en raison de circonstances particulières (décès d'un proche, faible disponibilité, etc.). Un courrier leur avait donc 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. Si elles ne le souhaitaient pas, elles pouvaient le faire savoir par mail, courrier ou téléphone. Au final, sur les 7522 adresses, 5781 ont donné lieu à un entretien lors de la troisième 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 %). Bien que le choix de recontacter pour la troisième vague des personnes impossibles à joindre, ou ayant refusé de répondre lors de la deuxième vague ait eu des implications fortes en amont de la collecte<sup>(8)</sup> mais aussi en aval<sup>(9)</sup>, cette procédure de repêchage a ainsi permis d'interroger 345 personnes supplémentaires, soit 6 % de l'échantillon longitudinal.

Au final, l'attrition entre les deux premières vagues a donc été bien plus élevée (35 %) qu'entre les deux suivantes (17 %<sup>(10)</sup>), résultat classique dans les enquêtes panel en France. Cela tient en partie à un effet de sélection. Les personnes peu intéressées par l'étude ou ayant trouvé le questionnement trop indiscret sont sorties de l'échantillon dès la fin de la première vague (près de 11 % des répondants de 2005 ont refusé que l'on reprenne contact avec eux pour la vague suivante contre seulement 4 % des répondants à la deuxième

(8) Notamment la mise en place d'un questionnaire spécifique permettant de faire la jonction avec la première vague. En effet, lors de la troisième vague, un ensemble de questions visait à collecter les événements survenus depuis la deuxième vague mais, pour les non-participants à la deuxième vague, le questionnement rétrospectif a dû être étendu aux événements survenus depuis la première vague.

(9) Construction de deux pondérations longitudinales, l'une pour les exploitations mobilisant des informations issues des trois vagues, l'autre pour celles ne mobilisant que la première et la dernière vague.

(10) Taux calculé sans tenir compte des 345 personnes interrogées en vagues 1 et 3, mais qui n'avaient pas répondu à la deuxième vague.

vague). En outre, les difficultés logistiques rencontrées entre les deux premières vagues ne se sont pas reposées ensuite<sup>(11)</sup>. Sur l'ensemble de la période d'enquête (2005–2011), l'attrition s'élève ainsi à 43 %<sup>(12)</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 (départs en institution, 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>(13)</sup>, soit 10 % de l'attrition totale.

Le taux d'attrition observé dans Érfi est au final assez semblable à celui observé dans d'autres enquêtes panel après 6 ans : 59 % des répondants à la première vague du Panel européen des ménages ont répondu à l'ensemble des vagues de l'enquête entre 1994 et 2001 (Lhommeau, 2003) et 55 % pour les Statistiques sur les revenus et les conditions de vie (SRCV<sup>(14)</sup>, 2004–... ; Burricand et Lorgnet, 2014) contre 54 % des répondants à l'enquête Érfi<sup>(15)</sup>. Le taux d'attrition de ces deux enquêtes, qui sont annuelles, étaient cependant plus faible à trois ans (environ 25 % contre 35 %) dans Érfi, indiquant un meilleur suivi à court terme lorsque les vagues sont rapprochées. Toutefois, le taux de participation se dégrade ensuite, peut-être en raison de la charge que représente le fait de répondre chaque année à un questionnaire. Notons cependant qu'un certain nombre de facteurs influent sur l'attrition, comme l'autorisation de *proxy* 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'idées, l'attrition observée dans l'enquête Santé et itinéraire professionnel (SIP) était de 19 % (contre 35 % pour Érfi 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 allait compter deux vagues. Ainsi, ceux qui acceptaient de prendre part à la première vague savaient qu'ils s'engageaient à répondre

(11) Plusieurs situations se sont en réalité cumulées suite à la première vague. Les fiches de suivi étaient complétées manuellement par l'enquêteur au terme du questionnaire, or certaines avaient été mal renseignées (imprécisions) et n'ont pas permis de retrouver le répondant au moment de la deuxième vague. En outre, pour quelques enquêteurs, les fiches de suivi ont été perdues. Enfin, des erreurs dans la saisie de l'identifiant individuel ont rendu l'appariement impossible entre le questionnaire et les coordonnées des personnes, conduisant à évincer quelques observations lors de la constitution de l'échantillon de la deuxième vague. L'expérience de ces « ratés » a conduit à plus de vigilance pour la troisième vague (pré-impression des fiches de suivi avec les coordonnées des personnes, à modifier uniquement en cas d'erreurs, photocopie systématique des fiches de suivi avant envoi par la Poste au service en charge de la saisie des adresses).

(12) 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).

(13) Estimation de l'auteur à partir des tables de mortalité par groupes d'âges quinquennaux appliquées à la structure de l'échantillon de 2005.

(14) L'enquête SRCV est la partie française du système communautaire EU-SILC : European union-Statistics on income and living conditions.

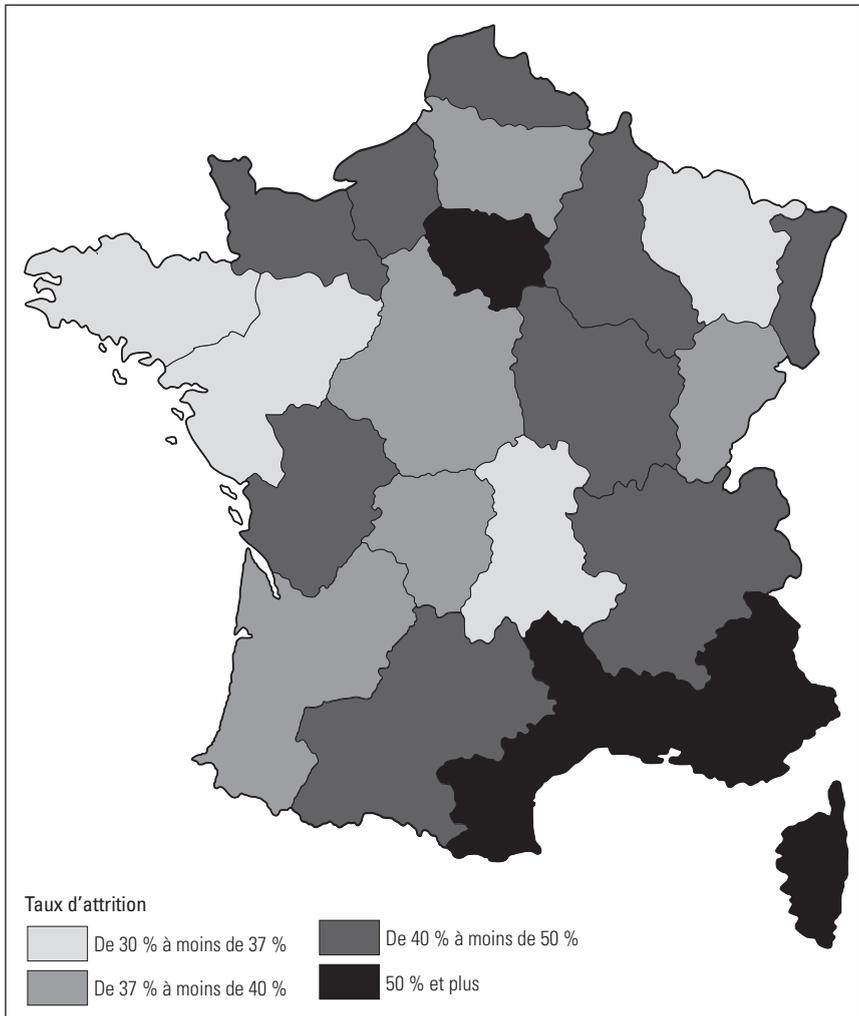
(15) 5436 répondants parmi les 10079 personnes interrogées en vague 1 ont participé aux trois vagues.

à la seconde vague. Ceci s'est traduit par très peu de refus d'être recontacté pour la seconde vague contre près de 10 % dans Érfi (Mermilliod, 2012). Cette procédure d'annonce d'une deuxième vague dès la prise de contact lors de la première vague de l'enquête SIP a cependant pu occasionner un biais de sélection à l'entrée avec davantage de refus de participer à l'enquête dès la première vague que dans Érfi.

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

Au-delà du taux d'attrition dans son ensemble, nous avons cherché à repérer si certaines caractéristiques favorisaient ou au contraire venaient

**Figure 1. Taux d'attrition entre 2005 et 2011 par région**



Source: Érfi-GGS123, Ined-Insee, 2005-2008-2011.

limiter la déperdition de l'échantillon. De manière descriptive d'abord, on se focalise sur la répartition géographique de l'attrition et sur l'effet de l'âge et du sexe du répondant.

a. Géographie de l'attrition

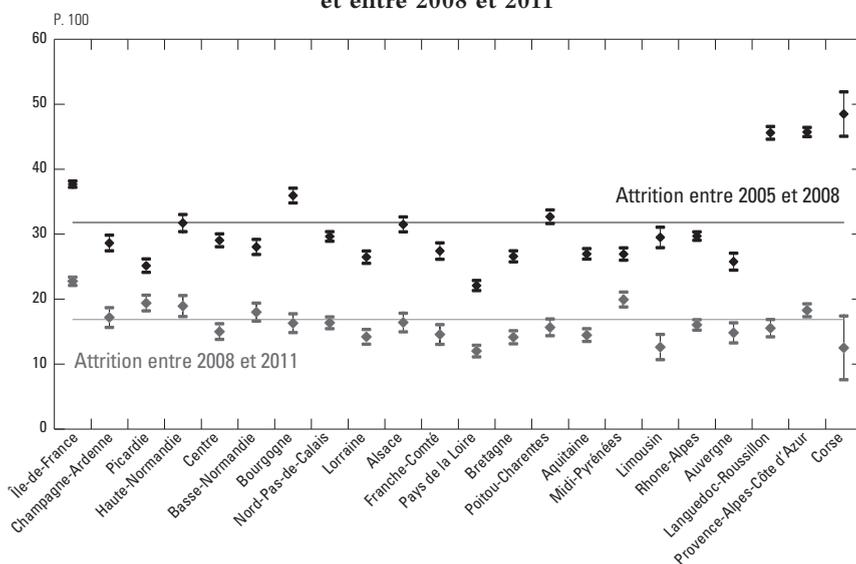
L'attrition a été plus ou moins marquée selon les différentes régions (figure 1), avec au final une déperdition plus forte de l'échantillon dans les régions méditerranéennes et en Île-de-France (attrition supérieure à 50 %). Mais c'est surtout entre la première et la deuxième vague que les disparités régionales se sont jouées (figure 2). En effet, entre 2008 et 2011, les différences interrégionales sont plus faibles et certaines régions qui avaient subi 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

■ 67

**Figure 2. Taux d'attrition par région entre 2005 et 2008 et entre 2008 et 2011**



Note: intervalles de confiance à 95 %. Les lignes horizontales continues matérialisent les taux d'attrition moyen (France entière). Source: Érfi-GGS123, Ined-Insee, 2005-2008-2011.

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 % des répondants des villes de plus de 100 000 habitants avaient l'intention ferme<sup>(16)</sup> de déménager dans les trois ans contre moins de 8 % des habitants de 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 (tableau 2)<sup>(17)</sup>: les résidences secondaires y étant plus nombreuses, l'attrition pourrait s'expliquer par des difficultés accrues à rentrer en contact avec le ménage en raison d'absences de longue durée; la population est également plus âgée, ceci pouvant expliquer davantage de départs en institution et de décès.

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

68 ■

À la fois entre les deux premières vagues et entre les vagues 2 et 3, l'attrition est bien plus élevée aux âges extrêmes, avant 30 ans et au-delà de 60 ans (figure 3). 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; 2011). 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

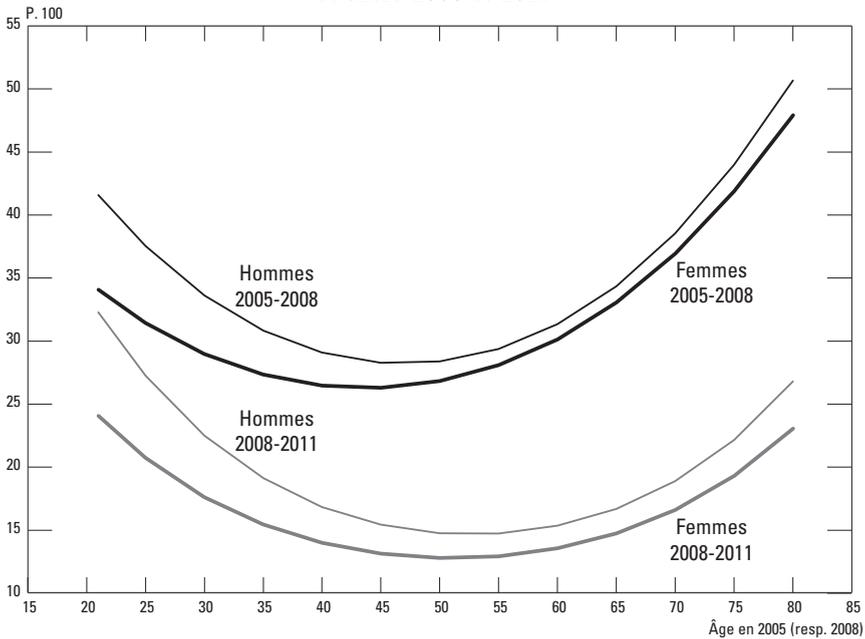
(16) Réponse « oui ».

(17) Le tableau 2 présente les résultats d'une régression logistique modélisant la probabilité que l'attrition observée entre les vagues 1 et 3 provienne d'un « refus » plutôt que d'autres motifs (perte de vue du répondant, décès, etc.).

**Tableau 2. Facteurs associés au risque d'avoir refusé de répondre à la vague 3 versus ne pas y avoir participé pour une autre raison (modèle logit, coefficients  $\beta$  et seuil de significativité)**

Constante	0,34*		
<b>Habitat</b>			
<i>Taille de la localité de résidence</i>			
Ville de 100 000 habitants et plus	- 0,09		
Ville de 5 000 à 99 999 habitants	Réf.		
Ville de moins de 5 000 habitants	- 0,06		
Commune rurale	- 0,01		
<i>Région</i>			
Île-de-France	0,28*		
Bassin parisien	0,22		
Nord	Réf.		
Est	- 0,09		
Ouest	0,01		
Sud-Ouest	0,15		
Centre-Est	- 0,04		
Méditerranée	- 0,34**		
<i>Statut d'occupation du logement</i>			
Propriétaire, usfruitier	Réf.		
Accédant à la propriété	0,09		
Locataire, logé gratuitement	- 0,26***		
<i>Type de logement</i>			
Maison	Réf.		
Appartement, autre	- 0,01		
<i>Intention de déménager dans les 3 ans</i>			
Non, certainement pas	Réf.		
Ne sait pas, oui probablement, non probablement pas	- 0,10		
Oui, certainement	- 0,46***		
<i>Type de ménage</i>			
Couple sans enfant	Réf.		
Couple avec enfant(s) de plus de 3 ans	- 0,01		
Couple avec enfant(s) de moins de 3 ans	- 0,18		
Seul avec enfant(s) de plus de 3 ans	- 0,23		
Seul avec enfant(s) de moins de 3 ans	- 0,32		
Ménage de plusieurs personnes	0,08		
Seul	- 0,25***		
<b>Caractéristiques sociodémographiques</b>			
<i>Sexe</i>			
Homme	Réf.		
Femme	0,23***		
<i>Âge</i>			
Âge	0,03		
Âge <sup>2</sup>	- 0,42***		
<i>Niveau de diplôme</i>			
Aucun diplôme	- 0,04		
Diplôme inférieur au Bac	Réf.		
Diplôme de niveau Bac et plus	- 0,24***		
<i>Nationalité</i>			
Française	Réf.		
Étrangère	0,03		
<i>Catégorie socioprofessionnelle</i>			
Agriculteur	0,44		
Artisan, commerçant et chef d'entreprise	- 0,12		
Cadre et profession intellectuelle supérieure	- 0,36**		
Profession intermédiaire	- 0,03		
Employé	Réf.		
Ouvrier	- 0,01		
Chômeur	- 0,01		
Retraité	- 0,33**		
Au foyer	- 0,15		
Étudiant	- 0,09		
Autre inactif	- 0,56		
<b>Santé perçue, sociabilité</b>			
<i>État de santé perçue</i>			
Très bon, bon	Réf.		
Moyen, mauvais	- 0,06		
Très mauvais	- 0,77***		
<i>A écouté quelqu'un parler de sa vie</i>			
Oui	Réf.		
Non	0,16**		
<b>Posture par rapport à l'enquête</b>			
<i>Réponse donnée concernant le revenu du ménage</i>			
Réponse	Réf.		
Ne sait pas	- 0,16		
Refus de répondre	0,60		
<i>Signature de l'autorisation express ?</i>			
Signée	Réf.		
Refus de signer	0,91***		
Non concerné	0,91***		
<i>Souhaite recevoir les résultats de l'enquête ?</i>			
Oui	Réf.		
Non	1,40***		
<i>Durée du questionnaire de 2005</i>			
20 à 44 min	- 0,04		
45 min à moins de 1 h	Réf.		
1 h à moins de 1 h et quart	0,07		
1 h et quart à moins de 2 h et demie	0,08		
Non communiqué	0,03		
<i>Champ</i> : tous les non-répondants de la troisième vague (2011). <i>Lecture</i> : 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 refusé de participer à une vague. <i>Légende</i> : Réf.: situation de référence; *** significatif à 1%; ** significatif à 5%; * significatif à 10%; <i>Source</i> : Érfi-GGS123, Ined-Insee, 2005-2008-2011.			

**Figure 3. Taux d'attrition par sexe et âge entre 2005 et 2008 et entre 2008 et 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 « résumant bien » les valeurs observées pour chaque âge (non présentées ici), en gommant les variations aléatoires.

Source: Érfi-GGS123, Ined-Insee, 2005-2008-2011.

70 ■

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.

### **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. Seuls trois modèles sont ici présentés (tableau 3). Les deux premiers prennent en compte la catégorie socioprofessionnelle de la personne et se distinguent par l'inclusion ou non de la durée du questionnaire en première vague (distinction faisant émerger un lien entre la composition du ménage et la

durée de l'entrevue). Le troisième modèle reprend les mêmes caractéristiques que le deuxième, mais la catégorie socioprofessionnelle est remplacée par la situation détaillée d'activité (en tenant notamment compte du temps de travail afin de tester l'hypothèse d'un impact de la disponibilité sur l'attrition)<sup>(18)</sup>. Le choix des variables retenues fait suite aux hypothèses construites sur la base de la revue de la littérature présentée précédemment. Il s'agit des caractéristiques des personnes en 2005 (première vague). Quatre grands groupes de facteurs sont considérés : l'habitat, les caractéristiques socio-démographiques, la santé perçue et la sociabilité de la personne et, enfin, sa posture par rapport à l'enquête.

### a. 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 où elle tient davantage à d'autres motifs (tableau 2). Les locataires, personnes vivant en

**Tableau 3. Modélisation de la probabilité d'avoir répondu à la troisième vague de l'enquête Érfi (paramètres estimés de la régression logistique)**

	Modèle 1	Modèle 2	Modèle 3
Constante	0,96 ***	0,95 ***	0,95 ***
<b>Habitat</b>			
<i>Taille de la localité de résidence</i>			
Ville de 100 000 habitants et plus	- 0,01	- 0,01	0,00
Ville de 5 000 à 99 999 habitants	<i>Réf.</i>	<i>Réf.</i>	<i>Réf.</i>
Ville de moins de 5 000 habitants	0,33 ***	0,34 ***	0,34 ***
Commune rurale	0,00	0,00	0,02
<i>Région</i>			
Île-de-France	- 0,51 ***	- 0,49 ***	- 0,48 ***
Bassin parisien	- 0,06	- 0,05	- 0,05
Nord	<i>Réf.</i>	<i>Réf.</i>	<i>Réf.</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 ***
<i>Statut d'occupation du logement</i>			
Propriétaire, usufruitier	<i>Réf.</i>	<i>Réf.</i>	<i>Réf.</i>
Accédant à la propriété	0,14 **	0,12 *	0,11
Locataire, logé gratuitement	- 0,12 *	- 0,12 *	- 0,13 *

(18) Ces deux variables ne sont pas intégrées dans un même modèle en raison de leur forte colinéarité.

Tableau 3. suite

	Modèle 1	Modèle 2	Modèle 3
<i>Type de logement</i>			
Maison	Réf.	Réf.	Réf.
Appartement, autre	- 0,18 ***	- 0,17 ***	- 0,17 ***
<i>Intention de déménager dans les 3 ans</i>			
Non, certainement pas	Réf.	Réf.	Réf.
Ne sait pas, oui probablement, non probablement pas	- 0,05	- 0,07	- 0,07
Oui, certainement	- 0,36 ***	- 0,37 ***	- 0,37 ***
<i>Type de ménage</i>			
Couple sans enfant	Réf.	Réf.	Réf.
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>Caractéristiques socio-démographiques</b>			
<i>Sexe</i>			
Homme	Réf.	Réf.	Réf.
Femme	0,14 ***	0,15 ***	0,15 ***
<i>Âge</i>			
Âge	- 0,14 *	- 0,17 **	- 0,15 *
Âge <sup>2</sup>	- 0,63 ***	- 0,63 ***	- 0,64 ***
<i>Niveau de diplôme</i>			
Aucun diplôme	- 0,23 ***	- 0,23 ***	- 0,23 ***
Diplôme inférieur au Bac	Réf.	Réf.	Réf.
Diplôme de niveau Bac et plus	0,29 ***	0,29 ***	0,31 ***
<i>Nationalité</i>			
Française	Réf.	Réf.	Réf.
Etrangère	- 0,58 ***	- 0,60 ***	- 0,60 ***
<i>Catégorie socioprofessionnelle</i>			
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	
Employé	Réf.	Réf.	
Ouvrier	- 0,05	- 0,06	
Chômeur	- 0,36 ***	- 0,36 ***	
Retraité	0,08	0,09	
Au foyer	- 0,08	- 0,08	
Étudiant	- 0,10	- 0,07	
Autre inactif	- 0,44 ***	- 0,44 ***	

Tableau 3. suite

	Modèle 1	Modèle 2	Modèle 3
<i>Situation d'activité</i>			
Indépendants, chef d'entreprise			- 0,13
Salarié (temps plein, 40h et plus par semaine)			0,06
Salarié (temps plein, moins de 40h par semaine)			Réf.
Salarié (temps partiel)			0,01
au foyer			- 0,08
Chômeur			- 0,35 ***
Retraité			0,07
Etudiant			- 0,05
Autre inactif			- 0,44 ***
<b>Santé perçue, sociabilité</b>			
<i>État de santé perçue</i>			
Très bon, bon	Réf.	Réf.	Réf.
Moyen, mauvais	- 0,09	- 0,09 *	- 0,09 *
Très mauvais	- 0,47 **	- 0,51 **	- 0,51 **
<i>A écouté quelqu'un parler de sa vie</i>			
Oui	Réf.	Réf.	Réf.
Non	- 0,19 ***	- 0,17 ***	- 0,16 ***
<b>Posture par rapport à l'enquête</b>			
<i>Réponse donnée concernant le revenu du ménage</i>			
Réponse	Réf.	Réf.	Réf.
Ne sait pas	- 0,31 **	- 0,29 **	- 0,28 **
Refus de répondre	- 0,57 ***	- 0,56 ***	- 0,57 ***
<i>Signature de l'autorisation express ?</i>			
Signée	Réf.	Réf.	Réf.
Refus de signer	- 1,07 ***	- 1,05 ***	- 1,05 ***
Non concerné	- 0,72 ***	- 0,69 ***	- 0,69 ***
<i>Souhaite recevoir les résultats de l'enquête ?</i>			
Oui	Réf.	Réf.	Réf.
Non	- 1,30 ***	- 1,27 ***	- 1,27 ***
<i>Durée du questionnaire de 2005</i>			
20 à 44 min		- 0,24 ***	- 0,24 ***
45 min à moins de 1 h		Réf.	Réf.
1 h à 1 h et quart		- 0,01	- 0,01
1 h et quart à moins de 2 h et demie		0,16 **	0,16 **
Non communiqué		- 0,14	- 0,14
<b>Qualité du modèle</b>			
Likelihood Ratio	***	***	***
R <sup>2</sup>	0,16	0,17	0,17
% de paires concordantes	69,3	69,6	69,6
<i>Champ</i> : tous les répondants de la première vague (en 2005). <i>Lecture</i> : 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 3 <sup>e</sup> vague. <i>Légende</i> : Réf.: situation de référence; ***significatif à 1%; **significatif à 5%; *significatif à 10%; <i>Source</i> : Érfi-GGS123, Ined-Insee, 2005-2008-2011.			

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 résidents de communes de moins de 5 000 habitants ont davantage pris part à la troisième vague. C'est en particulier le cas des agriculteurs pour qui l'attrition est bien plus faible, ceci s'expliquant par une moindre mobilité de ces populations et par un accès au logement plus simple pour l'enquêteur que dans les grandes villes.

### **b. Les caractéristiques sociodémographiques**

Les hommes, déjà sous-représentés dès la première vague de l'enquête<sup>(19)</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. Watson et Wooden (2009) ont relevé un refus plus fréquent de la part des étrangers d'être réinterrogés, ce qui ne se retrouve pas ici. Leur moindre participation à la troisième vague tient autant à d'autres facteurs qu'à un refus explicite (tableau 2).

74 ■

### **c. Santé perçue, sociabilité**

Les personnes s'estimant en mauvaise santé, vivant seules 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 (tableau 2). 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>(20)</sup> ont ainsi moins souvent pris part à la troisième vague de l'enquête, avec des refus plus fréquents de poursuivre l'étude (tableau 2).

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

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

(20) 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 ? ».

parfois perçue comme une occasion d'échanger avec l'extérieur. Toutefois, l'inclusion d'indicateurs de cet ordre<sup>(21)</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).

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

Les études méthodologiques mettent l'accent sur un lien entre l'attrition et la posture du répondant par rapport à l'enquête, son intérêt pour la thématique abordée dans le questionnaire 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 Érfi : 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>(22)</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 de poursuivre l'étude sans doute parce qu'elles ont estimé que le questionnaire était trop indiscret. De la même manière, 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 1 h 15 et 2 h 30 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 d'Érfi. 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

(21) 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) ? ».

(22) Le questionnaire de 2005 comportait deux questions sensibles au sens de la loi informatique et libertés, 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 Sebillé et Régnier-Loilier, 2007.

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. Le fait d'intégrer la durée du questionnaire (modèle 2, tableau 3) annule en partie l'effet de la variable « type de ménage » observé dans le modèle 1 : 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.

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 participer aux vagues suivantes. À cette fin, le modèle 3 (tableau 3) intègre une variable de situation d'activité tenant compte du statut d'activité (salaarié 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 ?*

76 ■

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 ; 2011) 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 la structure de l'échantillon entre les vagues 2 et 3 (les mêmes facteurs explicatifs de l'attrition se rejouent à chaque vague) ou si l'attrition a été sélective entre les deux premières vagues, mais qu'elle a été plus aléatoire entre les vagues 2 et 3.

Dit autrement, on peut faire l'hypothèse que certains facteurs ont pu jouer sur l'attrition entre les vagues 1 et 2, mais n'ont plus joué ensuite entre les vagues 2 et 3 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>(23)</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 calculées sur le même principe que le modèle 2 du tableau 3, mesurant la probabilité de suivi entre 2005 et 2008 et entre 2008 et 2011 (tableau 4)<sup>(24)</sup>. La comparaison des  $R^2$ , 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 « 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>(25)</sup> ou encore s'estimant en mauvaise santé ; le suivi est en revanche meilleur chez les agriculteurs, les habitants de communes comptant moins de 5000 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 rejouent 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 montrant plus de refus de leur part lié à 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 diplômés, 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

(23) 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.

(24) 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).

(25) 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) ; en revanche, 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).

**Tableau 4. 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		Suivi V1-V2	Suivi V2-V3
Constante	1,32 ***	2,04 ***	<i>Nationalité</i>		
<b>Habitat</b>			Française	Réf.	Réf.
<i>Taille de la localité de résidence</i>			Étrangère	-0,42 ***	-0,49 ***
Ville de 100 000 habitants et plus	-0,01	-0,01	<i>Catégorie socioprofessionnelle</i>		
Ville de 5 000 à 99 999 habitants	Réf.	Réf.	Agriculteur	-0,02	0,37
Ville de moins de 5 000 habitants	0,33 ***	0,34 ***	Artisan, commerçant et chef d'entreprise	-0,20	0,05
Commune rurale	0,00	0,00	Cadre et profession intellectuelle supérieure	0,13	0,10
<i>Région</i>			Profession intermédiaire	0,00	0,19
Île-de-France	-0,51 ***	-0,49 ***	Employé	Réf.	Réf.
Bassin parisien	-0,06	-0,05	Ouvrier	-0,06	-0,08
Nord	Réf.	Réf.	Chômeur	-0,21 **	-0,24
Est	0,05	0,04	Retraité	0,00	0,29 *
Ouest	0,07	0,08	Au foyer	-0,28 ***	0,07
Sud-Ouest	-0,02	-0,03	Étudiant	-0,04	-0,36
Centre-Est	0,06	0,05	Autre inactif	-0,23 *	-0,75 ***
Méditerranée	-0,58 ***	-0,59 ***	<b>Santé perçue, sociabilité</b>		
<i>Statut d'occupation du logement</i>			<i>État de santé perçue</i>		
Propriétaire, usufruitier	Réf.	Réf.	Très bon, bon	Réf.	Réf.
Accédant à la propriété	0,14 **	0,12 *	Moyen, mauvais	-0,05	-0,21 **
Locataire, logé gratuitement	-0,12 *	-0,12 *	Très mauvais	-0,63 ***	-1,26 ***
<i>Type de logement</i>			<i>A écouté quelqu'un parler de sa vie</i>		
Maison	Réf.	Réf.	Oui	Réf.	Réf.
Appartement, autre	-0,18 ***	-0,17 ***	Non	-0,19 ***	-0,14 *
<i>Intention de déménager dans les 3 ans</i>			<b>Posture par rapport à l'enquête</b>		
Non, certainement pas	Réf.	Réf.	<i>Réponse donnée concernant le revenu du ménage</i>		
Ne sait pas, oui probablement, non probablement pas	-0,05	-0,07	Réponse	Réf.	Réf.
Oui, certainement	-0,36 ***	-0,37 ***	Ne sait pas	-0,37 ***	-0,30 *
<i>Type de ménage</i>			Refus de répondre	-0,63 ***	-0,81 ***
Couple sans enfant	Réf.	Réf.	<i>Souhaite recevoir les résultats de l'enquête ?</i>		
Couple avec enfant(s) de plus de 3 ans	0,05	0,04	Oui	Réf.	Réf.
Couple avec enfant(s) de moins de 3 ans	-0,07	-0,09	Non	-1,62 ***	-1,30 ***
Seul avec enfant(s) de plus de 3 ans	-0,02	0,00	<i>Durée du questionnaire de 2005</i>		
Seul avec enfant(s) de moins de 3 ans	-0,54 **	-0,53 **	20 à 44 min	-0,21 ***	-0,21 **
Ménage de plusieurs personnes	-0,17	-0,13	45 min à moins de 1 h	Réf.	Réf.
Seul	-0,12 **	-0,08	1 h à moins de 1 h et quart	0,09	0,02
			1 h et quart à moins de 2 h et demie	0,19 ***	-0,15
			Non communiqué	-0,04	-0,51 ***
<b>Caractéristiques sociodémographiques</b>			<b>Qualité du modèle</b>		
<i>Sexe</i>			Likelihood Ratio	***	***
Homme	Réf.	Réf.	R <sup>2</sup>	0,15	0,10
Femme	0,07	0,21 ***	% de paires concordantes	68,8	67,4
<i>Âge</i>					
Âge	-0,06	0,14			
Âge <sup>2</sup>	-0,47 ***	-0,86 ***			
<i>Niveau de diplôme</i>					
Aucun diplôme	-0,27 ***	-0,05			
Diplôme inférieur au Bac	Réf.	Réf.			
Diplôme de niveau Bac et plus	0,23 ***	0,35 ***			

*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. *Légende*: Réf.: situation de référence; \*\*\*significatif à 1%; \*\*significatif à 5%; \*significatif à 10%. *Source*: Érfi-GGS123, Ined-Insee, 2005-2008-2011.

perdre le contact<sup>(26)</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 rejouent 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 significative entre V1 et V2 et entre V2 et V3. L'effet d'une même variable peut en effet jouer dans le même sens ou être différent, mais sans que l'on sache 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>(27)</sup> sur la base des modèles présentés en tableau 4, 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 (tableau 5) confirme un effet significativement différent pour les non diplômés et en région 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).

■ 79

## Conclusion

L'attrition dans les enquêtes *Generations and Gender Survey* est une préoccupation de premier ordre, à plusieurs titres. En premier lieu, outre des considérations financières qui ont empêché certains pays de conduire les différentes vagues de l'enquête, l'attrition trop élevée entre les deux premières vagues de l'enquête en a conduit d'autres à renoncer à poursuivre l'enquête (c'est notamment le cas de l'Allemagne, où seul un tiers des personnes interrogées lors de la première vague ont participé à la deuxième : voir également chapitre 1). Une déperdition trop élevée peut ainsi venir compromettre en partie l'objectif comparatif du projet. En deuxième lieu, l'attrition détermine plus généralement la qualité des données et, dans le contexte d'une opération internationale, il importe que celle-ci soit précisément documentée. L'étude préliminaire de l'attrition dans le volet français de l'enquête a permis de

(26) 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).

(27) 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.

**Tableau 5. Modélisation de la probabilité d'avoir répondu à la vague suivante et facteurs d'interaction avec la vague**

	Paramètre estimé du facteur	Paramètre estimé de l'interaction du facteur avec la variable «Vague d'observation»
Constante	1,44	
<b>Vague d'observation</b>		
Première vague (2005)	<i>Réf.</i>	
Deuxième vague (2008)	0,60**	
<b>Habitat</b>		
<i>Taille de la localité de résidence</i>		
Ville de 100 000 habitants et plus	- 0,11*	0,19 *
Ville de 5 000 à 99 999 habitants	<i>Réf.</i>	<i>Réf.</i>
Ville de moins de 5 000 habitants	0,33	0,09
Commune rurale	0,05	0,01
<i>Région</i>		
Île-de-France	- 0,46***	- 0,17
Bassin parisien	- 0,07	- 0,09
Nord	<i>Réf.</i>	<i>Réf.</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 ***
<i>Statut d'occupation du logement</i>		
Propriétaire, usufruitier	<i>Réf.</i>	<i>Réf.</i>
Accédant à la propriété	0,20***	- 0,13
Locataire, logé gratuitement	- 0,17***	0,03
<i>Intention de déménager dans les 3 ans</i>		
Non, certainement pas	<i>Réf.</i>	<i>Réf.</i>
Ne sait pas, oui probablement, non probablement pas	- 0,05	-0,08
Oui, certainement	- 0,41***	0,10
<i>Type de ménage</i>		
Couple sans enfant	<i>Réf.</i>	<i>Réf.</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>Caractéristiques sociodémographiques</b>		
<i>Sexe</i>		
Homme	<i>Réf.</i>	<i>Réf.</i>
Femme	0,10*	0,11
<i>Âge</i>		
Âge	- 0,09	0,23
Âge <sup>2</sup>	- 0,44***	- 0,42 **
<i>Niveau de diplôme</i>		
Aucun diplôme	- 0,30***	0,25 *
Diplôme inférieur au Bac	<i>Réf.</i>	<i>Réf.</i>
Diplôme de niveau Bac et plus	0,20***	0,15

**Tableau 5. suite**

	Paramètre estimé du facteur	Paramètre estimé de l'interaction du facteur avec la variable «Vague d'observation»
<i>Nationalité</i>		
Française	<i>Réf.</i>	<i>Réf.</i>
Étrangère	- 0,47***	- 0,03
<i>Catégorie socioprofessionnelle</i>		
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
Employé	<i>Réf.</i>	<i>Réf.</i>
Ouvrier	- 0,04	- 0,04
Chômeur	- 0,20	- 0,05
Retraité	0,00	0,28
Au foyer	- 0,19*	0,26
Étudiant	- 0,04	- 0,32
Autre inactif	- 0,21	- 0,54 **
<b>Santé perçue, sociabilité</b>		
<i>État de santé perçue</i>		
Très bon, bon	<i>Réf.</i>	<i>Réf.</i>
Moyen, mauvais	- 0,05	- 0,16
Très mauvais	- 0,71***	- 0,54
<i>A écouté quelqu'un parler de sa vie</i>		
Oui	<i>Réf.</i>	<i>Réf.</i>
Non	- 0,17***	0,03
<b>Posture par rapport à l'enquête</b>		
<i>Réponse donnée concernant le revenu du ménage</i>		
Réponse	<i>Réf.</i>	<i>Réf.</i>
Ne sait pas	- 0,43***	0,12
Refus de répondre	- 0,67***	- 0,14
<i>Souhaite recevoir les résultats de l'enquête ?</i>		
Oui	<i>Réf.</i>	<i>Réf.</i>
Non	- 1,70***	0,40 **
<i>Durée du questionnaire de 2005</i>		
20 à 44 min	- 0,25***	0,04
45 min à moins de 1 h	<i>Réf.</i>	<i>Réf.</i>
1 h à moins de 1 h et quart	0,04	- 0,03
1 h et quart à moins de 2 h et demie	0,21***	- 0,35 **
Non communiqué	- 0,11	- 0,40 **
<b>Qualité du modèle</b>		
Likelihood Ratio	***	
R <sup>2</sup>	0,17	
% de paires concordantes	71,30	

*Champ* : tous les répondants de la première vague (2005) et de la deuxième vague (2008). *Lecture* : la probabilité de répondre à la vague suivante est significativement moindre en région Méditerranée (paramètre négatif et statistiquement significatif). Toutefois, l'effet de ce facteur est significativement différent entre les vagues 1-2 et entre les vagues 2-3 dans le sens d'une meilleure participation en vague 3 qu'en vague 2 (paramètre positif et statistiquement significatif). *Légende* : *Réf.* : situation de référence ; \*\*\*significatif à 1 % ; \*\*significatif à 5 % ; \*significatif à 10 %. *Source* : Érfi-GGS123, Ined-Insee, 2005-2008-2011.

construire de manière raisonnée une variable de pondération longitudinale, mais il importera, dans une perspective de comparaisons internationales, de réfléchir à la construction d'une pondération qui soit relativement homogène entre les différents pays. Enfin, l'ampleur de l'attrition et la déformation de l'échantillon initial peut avoir des implications au niveau des études statistiques qui seront ensuite conduites. Outre une moindre précision statistique des estimateurs à mesure que la taille d'un échantillon diminue, une attrition trop sélective peut induire des biais sur les variables d'intérêt<sup>(28)</sup>. Si ce chapitre met en lumière un certain nombre de variables liées à une plus forte attrition, il ne met pas l'accent sur les variables d'intérêt<sup>(29)</sup> ni ne vise à estimer ces éventuels biais. L'enquête étant généraliste, les variables d'intérêts sont en effet très nombreuses et il importe à chaque chercheur de s'assurer que l'attrition n'affecte pas ses résultats, une fois la variable de pondération appliquée.

Au terme des trois vagues de l'enquête Érfi (2005–2011), l'attrition atteint 43 %, taux semblable à ceux relevés dans d'autres enquêtes du même type en France. Comme dans la plupart des enquêtes longitudinales, c'est entre la première et la deuxième vague que l'attrition a été la plus forte (35 %)<sup>(30)</sup>. Au-delà de ce taux, notre étude révèle que les facteurs qui y sont associés sont globalement conformes à ceux observés dans d'autres enquêtes. Sur le plan des caractéristiques sociodémographiques, on trouve d'abord un effet significatif du sexe, de l'âge, du niveau de diplôme, de la nationalité. Sur ce dernier point cependant, il est intéressant de souligner que l'attrition ne tient pas pour les étrangers à des refus plus fréquents de participer aux vagues suivantes<sup>(31)</sup>, mais à d'autres raisons (perte de contact notamment). Le lieu de vie des personnes joue également fortement, avec une attrition plus forte dans les grandes villes, en Île-de-France, chez les locataires ou encore chez les personnes qui avaient l'intention de déménager. La posture du répondant par rapport à l'étude joue également un rôle important : une moindre coopération lors du premier questionnaire (refus de répondre à certaines questions) et un moindre intérêt pour l'étude (refus de recevoir les premiers résultats) s'accompagne d'une sortie plus fréquente de l'échantillon au fil des vagues. On notera cependant que l'on ne trouve pas d'effet négatif d'une durée élevée de questionnaire, au contraire : un questionnaire long s'accompagne d'un meilleur taux de suivi et inversement. L'architecture du questionnaire peut en partie expliquer cela, les personnes plus enclines à ne pas

82 ■

(28) Pour exemple, dans l'enquête française Intentions de fécondité (Ined, 1998–2003), l'attrition élevée au terme des trois vagues d'enquête (70 %) a conduit à une déformation significative de la structure de l'échantillon mais aussi entraîné une modification significative du lien entre les variables d'intérêt (l'intention d'avoir un enfant, notamment) et ses facteurs explicatifs (Mazuy *et al.*, 2005).

(29) À cet égard, voir par exemple l'analyse réalisée par Isabelle Buber-Ennser (2014) à partir des deux premières vagues de l'enquête GGS autrichienne, où elle étudie le lien entre l'attrition et les variables relatives à la fécondité.

(30) En regard avec d'autres pays ayant réalisé les deux premières vagues de l'enquête GGS, le taux d'attrition observé en France entre les vagues 1 et 2 se situe dans la moyenne (elle était par exemple de 17 % en Géorgie et 22 % en Autriche, mais atteint 68 % en Allemagne ou en République tchèque).

(31) Par exemple, en raison d'un moindre intérêt pour une étude nationale ou de difficultés accrues de compréhension. Mais il est à noter que seuls des francophones étaient interrogés lors de la première vague.

se sentir concernées par une enquête sur la famille se voyant interrogées sur beaucoup moins de questions que les personnes en couple, avec enfants, etc. Enfin, le fait d'être moins disposé à des échanges de confidences avec son entourage mais aussi, sans surprise, un mauvais état de santé s'accompagnent d'un moins bon suivi (risque plus fort d'être indisposé pour répondre, de se retrouver en institution, et risque supérieur de décéder entre les vagues). Toutefois, suite au retour des enquêteurs lors des bilans de collecte, nous faisons l'hypothèse d'un meilleur suivi des personnes se sentant plus isolées, certaines voyant dans la venue de l'enquêteur une occasion d'échanger. Celle-ci ne se vérifie pas sur la base des indicateurs ici utilisés.

La mise en regard des facteurs associés à l'attrition entre les deux premières vagues et entre les deux suivantes montre que la plupart des facteurs se rejouent d'une vague à l'autre. Nous postulons notamment que les facteurs liés à une plus forte mobilité géographique ou à une prise de contact plus difficile avec le répondant avaient tout lieu de se rejouer d'une vague à l'autre, mais que cela devait être moins le cas pour les facteurs liés à une posture de refus en raison d'un effet de sélection: les personnes méfiantes ou peu intéressées pour répondre seraient sorties dès la fin de la première vague. En réalité, presque tous les effets qui pèsent sur l'attrition entre les deux premières vagues se rejouent entre les deux vagues suivantes, avec une intensité comparable (facteur d'interaction non significatif). Seuls les non diplômés, les habitants de la région méditerranéenne et les personnes ne souhaitant pas recevoir les résultats, dont la probabilité d'attrition est plus forte (paramètre estimé de signe négatif), ont mieux participé à la vague 3 qu'ils ne l'avaient fait lors de la vague 2 (facteur d'interaction positif et significatif). Cela pourrait traduire un effet de sélection mais, au final, la déformation de la structure de l'échantillon s'est amplifiée au fil des vagues, comme en atteste la dispersion plus élevée de la variable de pondération calculée suite à la troisième vague (par rapport à celle issue de la vague 2).

Nous regrettons de n'avoir pu creuser davantage certaines hypothèses, notamment en lien avec les conditions de passation: quel effet du sexe et de l'âge de l'enquêteur lors d'une étude dont les rapports de genre et les effets de générations sont au cœur du questionnement? Malheureusement, ces informations ne sont pas disponibles dans l'enquête. Il importera, pour l'avenir, de mieux anticiper les informations à collecter en vue d'études méthodologiques. L'étude des facteurs liés à l'attrition et la mesure de leur impact entre les différentes vagues répond en effet à des objectifs méthodologiques plus généraux. En amont de la conception d'enquête notamment, une meilleure connaissance de ces facteurs peut éclairer les concepteurs à la fois dans la définition de l'échantillonnage (certaines catégories de personnes, plus sujettes à la déperdition, pourraient être surreprésentées dans le tirage initial afin de permettre de disposer d'effectifs suffisants pour les études en longitudinal), le choix des enquêteurs (sexe ou âge, selon la thématique de l'enquête), mais aussi dans les stratégies à mettre en œuvre pour suivre l'échantillon entre les vagues (un suivi plus ou moins rapproché pourrait

être envisagé selon la probabilité d'attrition). En aval de la collecte, le repérage des facteurs liés à la déperdition de l'échantillon permet de mettre en œuvre de manière raisonnée la construction d'une variable de pondération visant à corriger la déformation de l'échantillon. En outre, répondre à la question de savoir si les mêmes facteurs d'attrition se rejouent entre les vagues a pour objectif de mettre en garde les utilisateurs des données sur les risques de biais d'analyses en longitudinal. Même si la construction de variables de redressement est toujours possible, une déformation excessive de la structure initiale de l'échantillon au fil des vagues doit inciter le chercheur à s'assurer que celle-ci n'affecte pas ses résultats.

## ■ Références bibliographiques

BEHR A., BELLGARDT, E. RENDTEL U., 2005, "Extent and determinants of panel attrition in the European Community household panel", *European Sociological Review*, 21(5), p. 489-512.

BLOM A., DE LEEUW E., HOX J., 2011, "Interviewer effects on nonresponse in the European Social Survey", *Journal of Official Statistics*, 27(2), p. 359-377.

BREUIL-GENIER P., VALDELIÈVRE H., 2001, « Le Panel européen : l'intérêt d'un panel d'individus », *Économie et statistique*, 349-350, p. 17-40.

BUBER-ENNSER I., 2014, "Attrition in the Austrian Generations and Gender Survey: is there a bias by fertility-relevant aspects?", *Demographic Research*, 31(16), p. 459-496.

BURRICAND C., LORGNET J.-P., 2014, « L'attrition dans SRCV : déterminants et effets sur la mesure des variables monétaires », *Économie et statistique*, 469-470, p. 19-35.

CHARRANCE G., RAZAFINDRSTIMA N., 2012, « Les pondérations de la vague 3 de l'enquête Érfi », in Régnier-Loilier A. et Guisse N., *Dictionnaire des codes. Troisième vague de l'enquête Érfi*, document de travail interne, p. 17-35; nesstar.ined.fr/other\_docs/IE0228\_C.pdf

COLLOMB P., 1979, « Les émigrants de l'Ouest audois dix-neuf ans après. Examen des méthodes de collecte et d'analyse », *Population*, 34(1), p. 65-90.

FITZGERALD J., GOTTSCHALK P., MOFFIT R., 1998, "An analysis of sample attrition in panel data: the Michigan panel study of income dynamics", *The Journal of Human Resources*, University of Wisconsin Press, 33(2), p. 251-299.

GROVES R., COUPER M., 1998, *Nonresponse in Household Interview Surveys*, Wiley series in probability and statistics.

HOLBROOK A., GREEN M., KROSNIK J., 2003, "Telephone versus face-to-face Interviewing of national probability samples with long questionnaires. Comparisons of respondent satisficing and social desirability response bias", *Public Opinion Quarterly*, 67, p. 79-125.

JÄCKLE A., LYNN P., SINIBALDI J., TIPPING S., 2013, "The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys", *Survey Research Methods*, 7(1), p. 1-15.

- LAURIE H., SMITH R., SCOTT L., 1999, "Strategies for reducing non-response in a longitudinal panel survey", *Journal of Official Statistics*, Vol. 15, p. 269-282.
- LHOMMEAU B., 2003, « Trajectoires passées par un emploi à bas salaire. Une étude à partir du panel européen des ménages », Dares, Document d'études, 78
- LIPPS O., 2007, *Attrition in the Swiss Household Panel*, Methoden-DatenAnalysen, 1(1), p. 45-68. [http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.1\\_Heft\\_1/MDA1\\_Lips.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.1_Heft_1/MDA1_Lips.pdf)
- LOOSVEDT G., PICKERY J., BILLIET J., 2002, "Item nonresponse as a predictor of unit nonresponse in a panel survey", *Journal of Official Statistics*, 18(4), p. 545-557.
- MAZUY M., RAZAFINDRATSIMA N., LA ROCHEBROCHARD, É. de, 2005, *Déperdition dans l'enquête « Intentions de fécondité »*, Ined, Document de travail n° 129.
- 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, 37, Drees.
- NICOLETTI, C., PERACCHI, F., 2005, "Survey response and survey characteristics: microlevel evidence from the European Community household panel", *Journal of the Royal Statistical Society, Series A*, vol. 168, part 4, p. 763-781.
- 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., 2009, « La déperdition dans les enquêtes longitudinales : l'exemple de l'enquête Érfi », *Courrier des statistiques*, 128, p. 55-62.
- RÉGNIER-LOILIER A., 2011, « Attrition entre les deux premières vagues de l'Étude des relations familiales et intergénérationnelles », in Tremblay M.-É., Lavallée P., El haj Tirari M. (dir.), *Pratiques et méthodes de sondage*, Paris, Dunod, p. 254-258.
- RIANDEY B., 1988, « La qualité du suivi des échantillons dans les enquêtes démographiques : un bilan », *Population*, 43(4-5), p. 829-854.
- STOOP I., 2005, *The hunt for the last respondent. Nonresponse in sample survey*, The Hague, Social and Cultural Planning of the Netherlands.
- UHRIG N., 2008, *The Nature and Causes of Attrition in the British Household Panel Survey*, Institute for social and economic research, ISER Working Paper series, n° 2008-05.
- VOORPOSTEL M., LIPPS O., 2011, "Attrition in the Swiss household panel: is change associated with drop-out?" *Journal of Official Statistics*, 27(2), p. 301-318.
- WATSON N., WOODEN M., 2009, "Identifying factors affecting longitudinal survey response", in Lynn P. (dir.), *Methodology of Longitudinal Surveys*, Chichester, John Wiley and Sons; DOI:10.1002/9780470743874.ch10.

