Technical Paper

Sample attrition and distortion over the waves of the French Generations and Gender Survey

Arnaud Régnier-Loilier, Nelly Guisse

Institut national d'études démographiques

Abstract

Background and Objective

Attrition issue is crucial with longitudinal surveys: apart from the problem of a decrease in the sample size, which can compromise the strength of the statistical tests, attrition can also distort the initial sample structure and ultimately bias the results and their interpretation. Furthermore, because longitudinal surveys are becoming more common, better knowledge of the sample distortion over the waves can enable us to better define the sample draw. For instance, researchers can consider oversampling certain population categories for which the expected attrition is higher, or organize a more individualized follow-up between the waves.

Methods

Since France was one of the first countries to complete the data collection for the *Generations and Gender Survey*, we share our experience of the attrition between 2005 (1st wave) and 2011 (3rd wave).

Using descriptive analysis and modelisations (logit), we identify various factors linked to attrition: respondent's sociodemographic characteristics, residential status, respondent's health and sociability but also details on the interviewee's experience during the first wave (length of the questionnaire, refusal to answer certain questions, etc.).

Results

Aggregate attrition after 3 waves (2005-2011) was 43%, a rate similar to that found in other similar surveys in France. Attrition was highest between the first and the second waves (35 %).

Our study reveals that factors associated with attrition are overall consistent with those observed in other surveys. The comparison of the factors associated with attrition in the two inter-wave periods shows that most of the factors recur in both periods. Thus, the distortion of the sample structure increased over waves.

1. INTRODUCTION

Non-response is a major issue with questionnaire surveys, since participation predicts the representativity of the data. Non-response is even more problematic for longitudinal studies, because the risk of sample loss increases with the number of waves. Attrition is the continuous, selective erosion of the initial sample over the waves. The phenomenon is "continuous" insofar as the probability of an individual leaving the scope automatically increases with each wave, as the risks of geographical mobility and respondent fatigue compound the non-response factors encountered in cross-sectional surveys (Laurie, Smith and Scott 1999); it is "selective" because the probability of re-interviewing varies with the socio-demographic characteristics of the respondents and the conditions in which the previous questionnaires were administered. Apart from the obvious problem of a reduction in the sample size, which can compromise the robustness of the statistical tests, attrition can also distort the initial sample structure and ultimately bias the results and their interpretation (Razafindratsima and Kishimba 2004).

This issue of attrition is raised particularly in the *Generations and Gender Survey* currently being conducted in some 20 countries. The GGS involves questioning three times the same people, with a three-year interval between the waves, with the ultimate aim of performing international comparisons. Since France is one of the first countries to have completed data collection (the first wave took place in late 2005, and the last wave in late 2011), we propose to evaluate the sample loss observed in the survey.

After proposing a review of the literature and presenting the data, this study measures the attrition observed between the different waves of the French GGS, taking pains to distinguish between different reasons for leaving the sample: refusal, loss of contact, out of scope. In a second part, we characterise the attrition, by considering different categories of explanatory factors: socio-demographic variables, place of residence, state of health and sociability, experience of the previous interviews. Lastly, since the survey consists of three waves, we seek to see if the loss observed between the first two waves and the last two waves is attributable to the same factors. In other words, we want to see if the sample distortion is accentuated over the waves.

Attrition and the associated factors are a crucial issue downstream from data collection, because attrition is used to calculate adjustment variables (longitudinal weights). It is also very important to look at attrition because longitudinal surveys are becoming more common. Better knowledge of the sample distortion over the waves can enable us to better define the sample draw, by providing, for example, for oversampling of certain population categories for which the expected attrition is higher. Furthermore, better knowledge of the probability of individuals' leaving the panel allows for more individualised follow-up between the waves.

2. STATE OF THE ART

When an individual is included in a survey sample, his/her participation in the survey is conditional on two stages: contact (success or failure to contact the individual) and, if contact is established, the respondent's cooperation (agreement or refusal to respond). While earlier studies on attrition examined the phenomenon as a whole, most researchers now agree that the determinants of attrition differ depending on whether non-response is due to a loss of contact or to an explicit refusal to participate again in the survey (Groves and Couper, 1998; Watson and Wooden, 2009). The wealth of studies on sample loss in longitudinal surveys enables us to apprehend the phenomenon in all its complexity. Two categories of factors are usually advanced to understand attrition. On the one hand, there are the characteristics of the respondent, his/her household or place of residence. On the other hand, there are factors specific to the survey: the nature of the questionnaire but also the characteristics of the interviewer and the quality of the relationship that is established between him/her and the respondent, especially in connection with the fundamental issue of trust. For each category of factors, it is essential to distinguish between those that are linked to a loss of contact and those related to the refusal to cooperate, while bearing in mind that the same factor might influence both and or that it might have opposite effects on each phenomenon.

2.1. The socio-demographic determinants of establishing contact

Traditionally we observe that it is easier to establish contact with women, with old people and with parents of young children, since these individuals spend more time on average at home (Groves and Couper, 1998). Contact is established more frequently in households with more than one adult than with individuals who live alone, because it is more likely that at least one of the residents of the dwelling will be present when the interviewer visits. Conversely, people who live in high-density urban areas are harder to approach. Those areas have a higher percentage of tenants and of people living in apartment buildings that may be highly secure and hard to access (Uhrig, 2008; Blom, de Leeuw and Hox, 2011). People who live in large cities are also less likely to be at home because they spend less time there on average owing to longer travel times than in small towns or the countryside, but also because of a wider range of social and cultural activities on offer (Groves and Couper, 1998).

In longitudinal surveys, the probability of losing contact with a respondent increases as the risk of geographical mobility rises: that risk is higher among respondents who say they are considering moving house, who do not like the neighbourhood where they live, or who have already moved house in the years prior to the survey (Uhrig, 2008). Philippe Collomb (1979) followed by Benoît Riandey (1988) showed the decisive influence of the interviewers' determination to track down the respondents who have "disappeared". In the case of a change of address, the quality of the follow-up depends mainly on the commitment of the interviewers in what can turn into a real investigation.

2.2. The socio-demographic determinants of cooperation

Studies on attrition show also that once contact has been made between the respondent and the interviewer, the propensity to agree to participate varies with the socio-demographic characteristics of the target person.

Socio-economic level and education level can be decisive for the retention of the respondents (Uhrig, 2008), with the most educated being the most cooperative. The cognitive investment is sometimes greater for people with less education, who may be less comfortable answering a questionnaire. As well as being harder to contact, residents of large cities are less inclined to respond because of a diffuse sense of insecurity and lower social cohesion in high-density urban areas. A fear of the unknown also explains the lower rate of cooperation among old people (Holbrook, Green and Krosnick 2003; Uhrig, 2008). Foreigners also refuse to be re-interviewed more often (Watson and Wooden, 2009): they are less comfortable expressing themselves in French if it is not their native language and may also feel less concerned by a national survey. Furthermore, participation in social activities, involvement in politics or charity work are all factors that correlate positively with survey participation (Uhrig, 2008; Stoop, 2005; Lipps, 2007). Lastly, several authors have looked at

the link between life changes and response rate to surveys. Marital, occupational or geographical instability were found to be positively correlated with a refusal to participate. In addition to the automatic effect of a change of address, which can make it harder to contact people, people experiencing a change of circumstances are less likely to answer a survey that may be intrusive and time consuming (Voorpostel and Lipps, 2011). The effect is especially strong when the change is adverse, such as marital breakdown, a period of unemployment, or a drop in income (Fitzgerald, Gottschalk and Moffit, 1998).

2.3. Influence of questionnaire and interviewer characteristics on attrition

The conditions in which the survey is administered also influence the probability that an individual will agree to be re-interviewed. The theme of the survey, the quality of the previous wave and the characteristics of the interviewer may be determining factors in the risk that an individual will leave the sample.

Firstly, the people farthest from the topic of the survey may feel unconcerned and therefore less willing to spend time on the survey. A study of attrition in a French fertility intentions (*Intentions de fécondité*) survey (INED, 1998-2003) revealed a high probability of refusing to be recontacted among the respondents farthest from the questions related to having children, namely childless people and those who did not want to have children (Mazuy *et al*., 2005).

The amount of missing data in the previous waves can be an indicator of lower interest in the survey, or of an attitude of distrust by the respondent. Uhrig (2008), Lipps (2007) as well as Watson and Wooden (2009) showed that the respondents who refuse to fill out the questions in income are more likely to leave the sample in the next wave. More broadly, Loosvedt, Pickery and Billiet (2002) show that non-response to “sensitive” questions is related to lower cooperation in the next wave.

Similarly, the length of the questionnaire can predict the probability of response to the next waves. A shorter or longer than average questionnaire can increase the risk of refusal to participate in the next wave. A shorter questionnaire may reflect a lack of interest or distance from the theme of the survey; a longer questionnaire raises the opportunity cost by requiring a bigger commitment from the respondent (Watson and Wooden, 2009).

In another way, the securing of back-up contact persons at the end of a questionnaire can be decisive. Not only does this make it possible to find the respondent’s new contact details more easily in the event of a move, but, beyond that direct effect, the respondent’s agreement to provide such details can also be interpreted as an indicator of a mindset more open to the survey and therefore a greater propensity to agree to be followed up (Riandey, 1988; Laurie, Smith and Scott, 1999).

Lastly, some authors have shown that the interviewer himself/herself can have an impact on the response rate. Beyond being able to track down the respondent, the strategies used by the interviewer to convince a target person to respond are decisive. That impact depends on the interviewer's age and experience, which both correlate positively with the respondents' participation rate (Blom, de Leeuw and Hox, 2011), but also his/her personality, since the most open and extroverted obtain the best results (Jäckle *et al*., 2013).

3. THE DATA

As part of the programme of comparative longitudinal Generations and Gender Surveys, France conducted the first wave of the *Etude des relations familiales et intergénérationnelles*[1] (Erfi-GGS) in autumn 2005. The same respondents were interviewed again three and six years later (in autumn 2008 and autumn 2011), although sample erosion occurred over the waves.

The survey is mainly concerned with family, and was conducted in France by the National Institute for Demographic Studies Institut national d'études démographiques, (INED) and the National Institute for Statistics and Economic Studies (Institut national de la statistique et des études économiques, INSEE). It contains detailed information about people's marital status and children, fertility intentions, the organisation of household tasks between the spouses, the occupational status of each spouse, their economic resources, their values and opinions, and intergenerational support (for more details on the Generations and Gender Survey, see Vikat *et al.*, 2007; Régnier-Loilier and Légaré, 2010; Sebille and Régnier-Loilier, 2007; Régnier-Loilier, Saboni and Valdès, 2011; Régnier-Loilier, 2012). The questionnaire is very similar, in both architecture and content, from one wave to another.

Given the central themes of the survey, particularly the study of inter-generational and gender relations, the respondents of the first wave were men and women aged 18-79. Since the broad scope and longitudinal ambition of the project entails a risk of sample attrition over the waves, the designers of the international survey recommended interviewing at least 10,000 people. The *Erfi-GGS* fulfilled that target, since 10,079 people answered the first questionnaire, out of 18,000 households drawn randomly from the 1999 French census.

The questionnaire, of an average duration of 60 minutes in each wave, was administered using computer-assisted personal interviews (CAPI). France's statistical authority (Conseil national de l'information statistique, CNIS) rated the survey as "of general interest and statistical quality" but not as "compulsory" (unlike some surveys like the population census and the employment surveys; an argument the interviewers use when they make contact in

[1] The French name for the Generations and Gender Survey. The name was changed in French because the term "gender" sounded too scientific and was unclear to the participants in the survey tests.

order to maximise the participation rate). Furthermore, no financial incentive was provided,[2] despite the sometimes positive role of this type of approach (see for example Olsen, Abelsen and Olsen, 2012), although its efficacy depends in reality on a whole set of factors, including type of gift (cash, lottery ticket, etc.) and survey administration method (Internet, face-to-face, telephone).

The survey was thus administered face-to-face by approximately 550 interviewers in the first wave (and by roughly 400 interviewers in the second and third waves, since there were fewer address files to process), across the whole of metropolitan France. As far as possible, the same interviewers were engaged for the different waves, since a positive role of continuity in face-to-face longitudinal surveys has been demonstrated (see for example Behr *et al.*, 2005; Nicoletti and Peracchi, 2005). Unfortunately, we do not know the percentage of interviewers who actually took part in more than one wave, since the survey did not collect any information on this.

After the first wave of the survey, only the people who gave their written consent by signing a form that provided a contact address and additional details, such as a telephone number and a back-up contact person so an interviewer could contact them three years later were eligible for the subsequent waves. The contact details were updated "manually" between the waves by INED's surveys department (cross-checking with administrative records was not permitted). The probability of losing track of people between the waves was much lower for respondents for whom we had contact details for at least one back-up contact person (although there was no difference between having one or two sets of contact details). Additionally, the number of telephone numbers provided by the respondent (up to three: home, work and mobile) was positively correlated with the probability of maintaining contact over the waves.[3]

In the end, 10,079 people responded to the first wave of the survey (2005), 6,534 to the second wave (2008) and 5,781 to the third wave (2011).

4. AIMS AND METHOD

This research pursues several aims. After outlining the strategies deployed to reduce sample attrition between the waves, the first aim is to measure the scale of attrition between each wave. We take care to distinguish between the different sources of attrition: refusal to take

[2] After collecting the data, the interviewer nevertheless gave the respondent a short (four-page) INED publication on a similar theme to those addressed in the survey to show to the respondent the way the responses were subsequently used. A token gift was also given: a pen with the survey logo in the first wave, a key ring in the second wave, and a cloth bag in the third wave.

[3] These results are not shown here but are available from the authors.

part in the next wave, loss of contact, out-of-scope[4] and the percentage of attrition ascribable to mortality.

The second aim is to identify various factors linked to attrition. By drawing on the existing literature on the subject and depending on the data available in the survey, we formulated various research hypotheses that we then tested, first descriptively then by constructing models in order to identify the impact of the various factors, all other things being equal (logit models). We considered different categories of factors. Firstly, the respondents' socio-demographic characteristics: gender, age, nationality, education level and family status. Secondly, their residential status: size of locality of residence, type of dwelling (house or apartment), occupancy status (owner, tenant, etc.) and intention of moving, with the hypothesis of higher or lower geographical mobility and differences in ease of access to the dwelling for the interviewer depending on these variables. Thirdly, factors related to the respondents' health and sociability, which can be approached in the survey through the frequency of exchanges of confidence with family and friends. Lastly, the specifics of the experience of the interview in the first wave formed a fourth category of factors: length of the questionnaire, attitude of distrust towards the questionnaire (measured by the refusal to answer certain questions) or interest in the study (measured by the wish to receive the initial results of the survey or not). In this section, our analysis focuses on the attrition observed between the first and last waves with a view to proposing a general evaluation of the survey.

In order to respond to the third aim of our study, we break down the sample attrition by comparing the explanatory factors behind attrition between the first two waves and the last two waves of the survey. This process enables us to see whether the same factors recur from one wave to another, thus reinforcing the distortion of the sample structure over the waves, or, conversely, whether a "selection effect" is revealed (attrition appears to be selective in the first inter-wave period but less or not at all in the second). However, while the comparison of significance thresholds between the two models is enlightening, it can be challenged because the two models are based on different numbers. Furthermore, even if a factor has the same type of influence in both models, its impact may not be of the same magnitude. Therefore a single model was constructed after stacking the samples from waves 1 and 2 (10,079 + 6,534) in the same file. For those 16,613 observations, we estimated the probability of having participated in the next wave depending on the same explanatory variables in the separate models but by controlling by the original sample ("wave 1/wave 2" dummy variable) and by including an interaction factor between the dummy variable and each of the explanatory dimensions. We were first interested in the result of those interactions: a statistically significant interaction indicates that the impact of the variable is different in the first and the second inter-wave periods; whereas a non-significant effect

[4] Respondents who had moved abroad or into an institution (e.g. retirement home) were not interviewed again.

indicates that the impact is roughly the same between the different waves. Here we advance the hypothesis of a smaller impact between waves 2 and 3 of the factors related to refusal to participate (the people most distrustful of the survey or who feel unconcerned by the theme – e.g. those who refused to answer certain questions or who did not wish to receive the results – would have left the sample at the end of wave 1), but a cumulative impact over the waves for the factors linked to loss of contact (i.e. geographical mobility, number of years in current residence, risk of leaving the scope: institutionalisation or death).

5. RESULTS

5.1. Response rate and attrition: overview

Figure 1 shows the size of the initial sample and the various sources of attrition (refusal, loss of contact, etc.) over the three waves of the survey. Of the 18,000 address files drawn randomly in 2005, 15% (or 2,688 address files) were unusable or were not used because the total number of expected respondents (10,000) had been reached; and 12% refused to participate (2,242). Other respondents were out-of-scope, impossible to trace (some of these are probably disguised refusals), long-term absentees or unable to respond (health problem, doesn't speak French); lastly, others went as far as a description of the household but the survey did not go further (refusal of the selected person to respond).

Of the 10,079 respondents to the first wave, 88% (9,099) agreed to be contacted again three years later for the second wave. A thank-you and reminder letter was sent to those who said they did not wish to continue. These respondents were asked if they "really" refused to be contacted again three years later: in the end, 150 agreed to be contacted. In addition to this first "reminder" letter, a more general procedure of follow-up and updating of addresses was conducted throughout the survey period (**Box**).

Box. Follow-up and updating of addresses between the waves

Between 2005 and 2011, each respondent received two letters per year (list of send-outs below), which sought to maintain contact with the respondent, to interest him/her in the purpose of the study and to update his/her contact details. Different types of letters were therefore sent out: thank-you letters for having participated in the previous wave, the initial results, greeting cards, and letters announcing the next wave. Here are the details:

- March 2006: three models of thank-you letter[5]: 1) a reminder for people who "refused" to continue to participate; 2) a thank you and a request for the details of a back-up contact person in case of a change of address, if we did not have these for the respondent; 3) a "simple" thank you for people who agreed to be contacted again and who provided contact details for at least one back-up contact person;

- December 2006: the initial survey results on the frequency of meetings between parents and children, sent out as a 4-page document;[6]

- June 2007: new results on the division of household tasks between spouses sent out as an 8-page document;[7]

- January 2008: New Year greetings card sent out with the survey logo together with a small sachet of flower seeds to plant at home;

- June 2008: a letter announcing the second wave of the survey, followed at the end of 2008 by notification of the interviewer's visit;

- March 2009: two models of thank-you letter: 1) if we did not have contact person's details for the respondent (approximately 50% of cases), the letter reminded the recipient of the importance of having these, his/ role and was accompanied by a form to fill out; 2) if we had details of at least one contact person, the send was a simple "thank-you" letter;

- December 2009: a letter accompanying the initial results based on the longitudinal data and showing the trend in the division of household tasks after the birth of a child;[8]

- June 2010: a letter and a brochure about the book Portraits de familles *(a collective book based on the data from the first wave and offering results on the different themes addressed in the survey) and a bookmark;*

- January 2011: a New Year card with the survey logo, and a sachet of cherry tomato seeds;

- June 2011: a letter announcing the survey, in two models: 1) one sent to the participants in the first two waves; 2) the other was sent to non-respondents to the second wave indicating how to contact INED if they did not wish to be contacted for the third wave.

Each of these letters reminded the recipients of the importance of letting us know their new address if they moved or intended to move: there was a change-of-address coupon to fill out with each letter (postage could be refunded on request) but it was also possible to inform us of a change of address by telephone or email (an email address was set up for the survey).

When the letter failed to reach the recipient (the person had moved without having the mail forwarded), it was returned to INED's surveys department. A procedure to track down the new address of the respondent was then implemented, using the information available on the follow-up sheet (telephone number, contact person) or the telephone directory. If the search was unfruitful, the next letter was nevertheless sent to the former address. After three returns of letters with "no longer at his address", the person was removed from the sample for the second wave. For the third wave, however, it was decided to send an interviewer to the former address anyway to make enquiries with neighbours, the local council, etc. Additionally, just before the data collection in the last wave of the survey, the French post office (*La Poste*) was commissioned to perform the final updates and corrections of addresses on the address files (*La Poste* has a file of address transfers and mail forwarding for people when they move).

[5] The thank-you letters were sent out approximately three months after the end of the data collection, which was the time the surveys department needed to enter the contact details for the respondents (and their back-up contact people) who agreed to be contacted again three years later.

[6] Régnier-Loilier A., 2006, "How often do adult children see their parents?", *Population & Societies*, 427.

[7] Bauer D., 2007, "Entre maison, enfant(s) et travail: les diverses formes d'arrangement dans les couples", *Études et résultats*, DREES, 570.

[8] Régnier-Loilier A., 2009, "Does the birth of a child change the division of household tasks between partners?", *Population & Societies*, 461.

A total of 758 people had been "lost" (including people who had died or were out-of-scope, i.e. people who had moved into an institution or abroad) between waves 1 and 2. The sample for wave 2 (2008) consequently consisted of 8,341 people.

Of these, 165 were out-of-scope when the interviewer came in 2008 (institutionalised, abroad or deceased); 794 refused to answer the questionnaire in 2008; and 547 could not be reached (long-term absence, moved house, etc.); and 42 people who answered the questionnaire in the second wave were identified as not having been the same respondent as in the first wave (either a proxy such as a spouse, or a person unconnected to the household) and were therefore deleted from the second-wave database. In the end, the database comprised 6,534 observations and, of these, 97% agreed to be contacted again for the third wave.

Between waves 2 and 3, 48 respondents announced their refusal to continue or a relative informed us of their death. For the third wave, 6,296 respondents from waves 1 and 2 were therefore eligible. To these were added a sample of 1,274 people who answered the first wave but not the second. These were mainly people who had been impossible to reach or who refused to answer even though they did not object at the end of the first wave to being contacted again by an interviewer. The decision to try to recontact these people for the third wave was motivated by the fact that the interviewers had indicated that these were not necessarily final refusals to participate in the survey but refusals to respond *at that time* due to particular circumstances (death of a close relative, not available, etc.). A letter was therefore sent to them before the third wave (in spring 2011) explaining that we had been unable to interview them in 2008 but would like to interview them again in 2011. If they did not wish to be interviewed again, they were asked to notify us by email, letter or telephone. In the end, of 7,522 addresses, 5,781 led to an interview in the third wave, with a success rate obviously much higher among people who responded to the first two waves (87%) than among those who had not been interviewed in 2008 (27%). Although the decision to recontacting people for the third wave who had been impossible to reach or who had refused to respond in the second wave had strong implications upstream[9] and downstream[10] of the data collection, this catch-up procedure made it possible to interview an additional 348 people, which was 6% of the longitudinal sample.

In the end, attrition was thus much higher between the first two waves (35%) than between the next two waves (17%), a typical result in panel surveys. This can be attributed partly to a

[9] Particularly the addition of a specific questionnaire linking their responses to the first wave. In the third wave, various questions sought to collect data on events that had occurred since the second wave, but for those who did not participate in the second wave, the retrospective questions had to refer to events that had occurred since the first wave.

[10] Two longitudinal weightings had to be calculated: one for mining data from all three waves, and the other for mining data from the first and third waves only.

selection effect. The people uninterested in the study or who found the questions too intrusive left at the end of the first wave (almost 11% of the 2005 respondents refused to be contacted again for the next wave, compared with only 3% of the respondents to the second wave). Moreover, the logistical problems encountered between the first two waves did not recur subsequently.[11] Over the whole survey period (2005-2011), attrition was 43%,[12] which in was attributable roughly half of the time to refusals (to continue after the first or the second wave, or to respond when the interviewer came), and half of the time to a loss of contact, people impossible to reach, out-of-scope (institutionalised or abroad) or dead. On the latter point, we do not know the exact number of respondents who died between 2005 and 2011 but our estimate is 430,[13] accounting for 10% of the total attrition.

The rate of attrition observed in the Erfi-GGS is finally fairly similar to that observed in other panel surveys after six years: 42% for the European Community Household Panel - HCHP (1994-2001) and 45% for the European Union Statistics on Income and Living Conditions - EU-SILC (2004-…). However, in those two surveys, which are annual, the rate of attrition after three years was much lower (approximately 25% versus 35%), seeming to indicate better short-term retention when the interval between waves is short. It is worth noting, however, that attrition is influenced by various factors, such as whether proxies are allowed or whether the survey is compulsory or not. On the latter point, sample loss in the French version of SILC[14] was low in the first four waves when the survey was compulsory but considerably increased when the survey subsequently became optional. Similarly, the attrition observed in a French survey on health and occupational pathways (*Santé et itinéraire professionnel* - SIP) was 19% (compared with 35% for the Erfi-GGS after three years), a difference largely explained by the fact that the SIP survey was compulsory but also by the announcement right from the first wave that the survey would consist of two waves. The individuals who agreed to take part in the first wave knew they were committing to respond to the second wave. This was reflected in very few refusals to be contacted again for the second wave, compared with almost 10% for the Erfi-GGS (Mermilliod, 2012). The procedure of announcing a second wave as soon as respondents were contacted for the first

[11] There was a combination of incidents after the first wave. The follow-up forms were filled out manually by the interviewer after the questionnaire. Some were incorrectly filled out, making it impossible to contact the respondent for the second wave. Other interviewers' follow-up forms were lost. And lastly, because some individual identifiers were incorrectly entered, some questionnaires could not be matched to the respondents' contact details, so some observations had to be left out of the second-wave sample. The experience of these mistakes led to more care in the third wave (follow-up forms were pre-printed with the respondent's contact details, which only had to be changed if they were incorrect; and every follow-up form was photocopied before it was posted to the department that would enter the contact details).

[12] If we consider the respondents to waves 1 and 3, regardless of whether they responded to wave 2; however 46% of the respondents in 2005 did not participate in all three waves (the either participated in wave 1 only, or in waves 1 and 2, or in waves 1 and 3).

[13] Author's estimate based on life tables by five-year age groups applied to the 2005 sample structure.

[14] The French version of the EU-SILC is called *les statistiques sur les revenus et les conditions de vie (SRCV).*

wave of the SIP survey might have caused a selection bias, however, with perhaps more refusals to participate in the survey before the first wave than in the Erfi-GGS.

Figure 1. Breakdown of the response rate and attrition rate between the three waves of the Érfi-GGS

			Sample size at various points in survey		Reason left sample		Catch-up	
			n	%	n	%	n	%
WAVE 1		**Initial sample for wave 1**	**18019**					
	Out of scope	*Home vacant, demolished, impossible to find*			2430	13,5		
		Address file not processed			258	1,4		
	Failure	*Impossible to reach, long-term absence, etc.*			1657	9,2		
		Person unable to respond (ill, doesn't speek French)			695	3,9		
		Partial response (basic household data only, dropped out during interview, etc.)			658	3,7		
		Refused to participate			2242	12,4		
	Success	**n (complete surveys W1)**	**10079**	55,9				
BETWEEN WAVES 1 & 2	At end of wave 1	*Refused to continue in W2*			1130	11,2		
		Retained after reminder letter					150	13,3
		Agreed to continue	9099					
	Loss W1-W2	*No longer lives at address provided, refusal, death*			758	8,3		
WAVE 2		**Sample for Wave 2**	**8341**					
	Out of scope	*Death, abroad, institutionalised, impossible to access home*			165	2,0		
		Address file not processed			45	0,5		
	Failure	*Impossible to reach, long-term absence, etc.*			547	6,6		
		Person unable to respond (ill, doesn't speak French)			190	2,3		
		Partial response (dropped out during interview)			8	0,1		
		Refused to participate			794	9,5		
		No resident of the household in 2008 matches the 2005 respondent			16	0,2		
		Person interviewed in 2008 not the 2005 respondent			42	0,5		
	Success	**n (complete sureys W2)**	**6534**	78,3				
BETWEEN WAVES 2 & 3	At end of wave 2	*Refused to continue in W3*			238	3,6		
		Refused in wave 2, no longer living at address, etc. retained					1274	
		Agreed to continue	6296					
	Loss W2-W3	*Refusal, death*			48	0,8		
	Sub-sample 1	Sample for W3 that did not respond to W2	1274	16,9				
	Sub-sample 2	Sample for W3 that did respond to W2	6248	83,1				
		Sample for wave 3 (TOTAL)	**7522**	100,0				
Sub-sample 1 WAVE 3	Out of scope	*Death, abroad, institutionalised, impossible to access home*	80	6,3				
		Address file not processed	23	1,8				
	Failure	*Impossible to reach, long-term absence, etc.*	354	27,8				
		Person unable to respond (ill, doesn't speak French)	73	5,7				
		Partial response (dropped out during interview)	4	0,3				
		Refused to participate	392	30,8				
	Success	**n (complete surveys respondents W1 and W3)**	**348**	27,3				
Sub-sample 2 WAVE 3	Out of scope	*Death, abroad, institutionalised, impossible to access home*	150	2,4				
		Address file not processed	36	0,6				
	Failure	*Impossible to reach, long-term absence, etc.*	190	3,0				
		Person unable to respond (ill, doesn't speak French)	117	1,9				
		Partial response (dropped out during interview)	6	0,1				
		Refused to participate	316	5,1				
	Success	**n (complete surveys respondents W1, W2 and W3)**	**5433**	87,0				
Total WAVE 3	Success	**n (complete surveys W3)**	**5781**	76,9				

Source: INED, INSEE, Erfi-GGS1-3, 2005-2011
Note 1: classification based on the raw survey data; regrouped by the author.
Note 2: of the 5,781 respondents of 2011, 8 were deleted because they did not match the panel individuals (most certainly different people from the previous waves)

5.2. Factors linked to attrition: descriptive elements

Beyond the overall rate of attrition, we sought to identify if some characteristics favoured or mitigated sample loss. First descriptively, we focus on the geographical breakdown of attrition and on the impact of the respondent's age and gender.

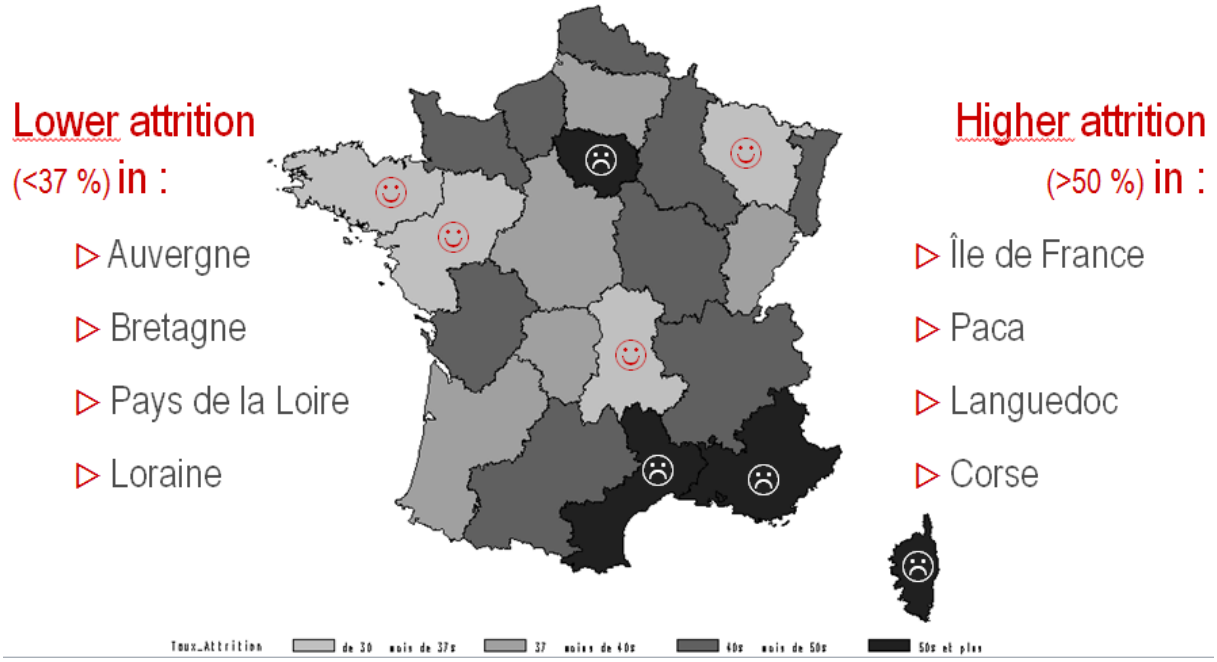
5.2.1. The geography of attrition

There was a regional variation in attrition (**Figure 2**), with higher cumulative sample loss in the Mediterranean and Île-de-France regions (attrition above 50%). But regional disparities were most pronounced between the first and second waves (**Figure 3**). The inter-regional differences were smaller between 2008 and 2011, and some regions, like Corsica and Languedoc-Roussillon, that had recorded strong losses between 2005 and 2008 even caught

up some of their lag with a slightly higher-than-average participation rate in the third wave. Conversely, Ile-de-France (Greater Paris), which had already experienced higher attrition in 2008, widened its gap by recording the lowest retention rate in 2011.

Conversely, respondents from Pays de la Loire, the region with the highest retention between the first two waves, confirmed its higher "loyalty" than other regions in 2011, as it was the region with the lowest attrition (31%).

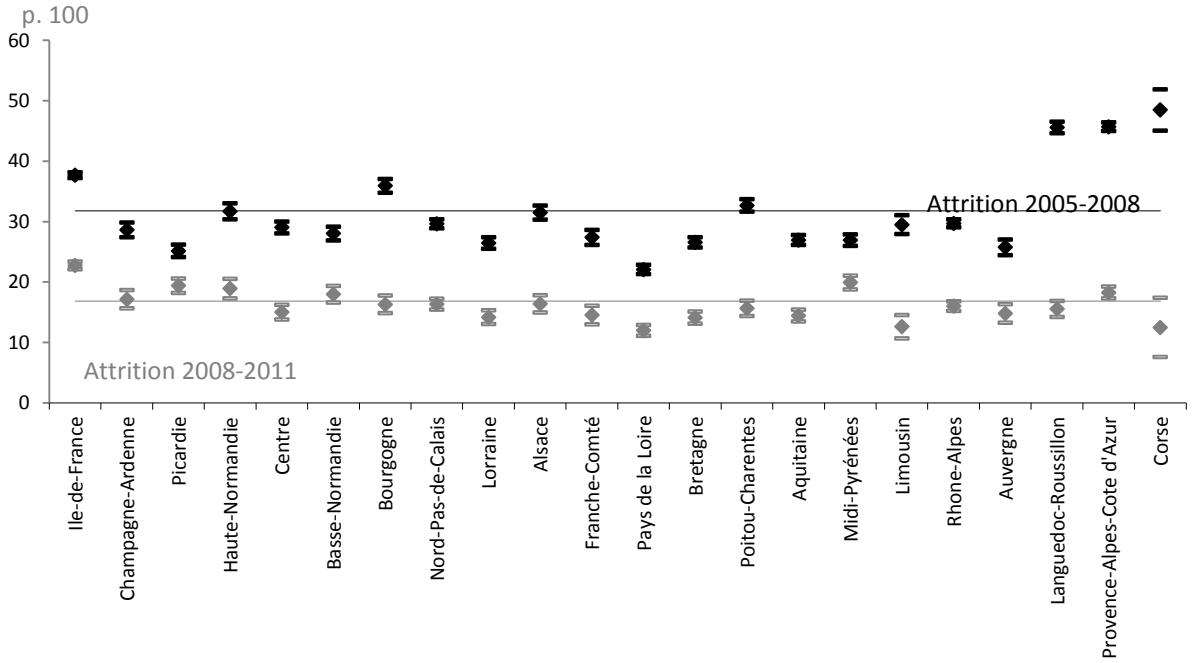
Figure 2. Rate of attrition between 2005 and 2011 by region



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

These regional differences can be linked to urban density, with lower attrition in small towns and rural areas (33% in municipalities with a population under 5,000; 39% in rural municipalities). Most residents of large cities, especially Paris and the Mediterranean cities, live in apartments (in 2005, 73% of the Érfi respondents in Paris lived in apartments and more than 50% of residents of cities with a population of at least 100,000, compared with only 5% of residents of rural municipalities). Apartments are often harder for interviewers to access than houses (door codes at the building entrance; people harder to reach). These highly urban populations are also more geographically mobile; they more frequently declared an intention to move within the next three years (20% firm intentions to move within three years in 2005 among respondents in cities with a population over 100,000 compared with less than 8% in rural municipalities).

Figure 3. Rate of attrition between 2005 and 2011 and between 2008 and 2011 by region



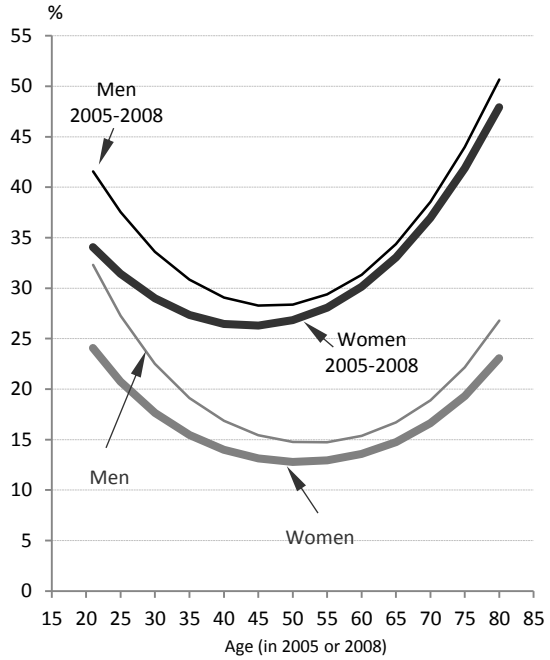
Source: INED-INSEE, Erfi-GGS1-3, 2005-2011
Note: 95% confidence intervals. The continuous horizontal lines represent the average rate of attrition (whole of France)

However, higher attrition can be attributed to more frequent refusals (more distrust, less availability) in Île-de-France but more to other reasons in the Mediterranean region (**Figure 8**): second homes are more common there, so attrition could be explained by greater difficulty of establishing contact with the household because of long-term absences; the population is also older there, which could explain more institutionalisation and deaths.

5.2.2. Higher attrition at the extreme ages and large gender differences at young ages

In both inter-wave periods, attrition is much higher at extreme ages, i.e. before 30 and after 60 (**Figure 4**). For the youngest, attrition can be interpreted as a higher probability of mobility: moving out of their parents' home, or moving away to attend university, to form a union or after the birth of a child, or a change of employment status. Loss of contact among old people can also be attributed to moving, e.g. after retirement, but it is also explained by a higher propensity to refuse to continue the study after the first wave (Régnier-Loilier, 2009; 2011). In this group, there is a higher probability of leaving the scope (institutionalisation, e.g. in a retirement home) and of death.

Figure 4. Rate of attrition by gender and age between 2005 and 2008 and between 2008 and 2011



Source: INED-INSEE, Erfi-GGS1-3, 2005-2011

Note: The age is that in 2005 for observation of attrition between 2005 and 2008; and that in 2008 for observation of attrition between 2008 and 2011. The series are smoothed by models including the continuous age (in relation to the average age of the sample) and its square. These curves "sum up" the values observed for each age (not shown here), by erasing random variations.

A more detailed comparison of the rates of attrition between waves 1 and 2 and between waves 2 and 3 reveals a trend inversion between the youngest and the oldest respondents. The oldest took part in the second wave less often but responded to the third wave slightly more often. This is partly attributable to a selection effect of the oldest after the first wave: as explained earlier, there were far more refusals in this group, but after the selection was performed, they do not leave the sample any more frequently than younger people.

Lastly, regardless of the inter-wave period observed (waves 1 and 2 or waves 2 and 3), attrition is higher among men, but the gap is most pronounced before age 50: young men participated much less in the next wave than young women.

5.3. Factors in attrition: "all other things being equal"

Attrition can be attributed to a combination of effects. For example, there is a strong link between the size of the town/city, the type of dwelling, and an intention to move. In order to measure the "net" effect of the various factors on the probability of attrition, a series of models was constructed that play with different variables. Only three models are presented here (**Figure 5**). The variables were chosen in light of the hypotheses made on the basis of the existing literature. These are the individual characteristics in 2005 (first wave). Four major categories of factors are considered.

5.3.1. Dwelling

All other things being equal – particularly the type of dwelling, the size of the town/city and an intention to move within three years – we find a specific effect for Île-de-France and the Mediterranean regions, with lower participation in wave 3. The effect is different, however, with attrition in Île-de-France attributable more to refusals, and in the Mediterranean region mainly to other reasons (**Figure 8**). Tenants, people living in apartments or who expressed their intention to move within three years were also more often lost between 2005 and 2011, mainly due to more frequent loss of contact (does not live at the address provided, impossible to reach). Conversely, home-buyers and residents of municipalities with a population of less than 5,000 took part in the third wave more often. That was particularly demonstrated by farmers, among whom attrition is much lower, because of the lower mobility of this population and easier access to the dwelling for the interviewer than in large cities.

5.3.2. Socio-demographic characteristics

Attrition was higher among men, already under-represented in the first wave of the survey,[15] confirming the descriptive results (see above); this is also true of the youngest and oldest people. Their lower participation is not attributable to more frequent refusals to participate but more to other reasons (impossible to reach, moved without leaving a forwarding address, institutionalised, etc.). The least educated, the unemployed and foreigners also participated less in the third wave of the survey. Watson and Wooden (2009) identified a more frequent refusal to be re-interviewed among foreigners, but this is not found here. The lower participation of foreigners in the third wave is as much attributable to other factors as to an explicit refusal (**Figure 8**).

5.3.3. Perceived health and sociability

People who consider themselves to be in poor health, who live alone or who are single parents of a small child (aged under 3) in 2005 answered the questionnaire less often in 2011; this was not due to more frequent refusals but to other reasons (**Figure 8**). We can assume more frequent institutionalisation or deaths among people in poor health, and we might assume it is more difficult for the interviewer to contact people who live alone (the more people there are in the dwelling, the more likely the interviewer is of finding someone at home when he/she visits).

[15] The lower participation of men in surveys is "typical" (less available and/or less interested in responding, harder to reach, etc.).

"Sociability" also plays a significant role: people who reported not having exchanged confidences with family or friends in the past 12 months[16] participated less in the third wave of the survey, with more frequent refusals to continue the study (**Figure 8**).

Conversely, based on feedback from interviewers at test or data collection evaluations, we made the hypothesis of higher participation among lonely people, since the interviewer's visit can be perceived as an opportunity to talk to someone. However, the inclusion of that type of indicator[17] in the model does not show a significant influence on the probability of participation in the third wave of the survey (results not shown here).

5.3.4. Attitude to the survey

Methodological studies emphasise a link between attrition and the respondent's attitude to the survey, interest in the theme of the questionnaire and experience of the interview. While we have little information about this, some indicators nevertheless enable us to test this hypothesis for the Erfi-GGS: a refusal to respond to certain questions (e.g. income) or a refusal for the response to "sensitive" questions (religion or civil partnership) to be recorded;[18] not wishing to receive the survey results; or the length of the interview in 2005.

An attitude of distrust towards the collection of data, e.g. refusing to let the interviewer keep the responses to the sensitive questions, leads to more frequent refusal to participate in the next waves of the survey. The same is true of respondents who refused to respond to certain questions (e.g. monthly income). The highest attrition stems from a refusal to continue the study (**Figure 8**), probably because these people felt the questionnaire was too intrusive. Similarly, a refusal to receive the initial results of the survey expresses a lower interest in the theme of the study and leads, unsurprisingly, to more refusals to continue and, ultimately, to much higher attrition.

Furthermore, although we were expecting higher attrition among people whose first interview was significantly longer than average, due to a possible fatigue effect, the reverse was in fact observed: respondents whose interview lasted between 75 minutes and 150 minutes in the first wave, were more likely to participate in the third wave. That result, which seems to contradict our preconceived ideas (we usually limit interviews to 60 minutes

[16] The title of the question was "In the last 12 months, has anyone talked to you about their life or their feelings?"

[17] Several questions referred to loneliness. For example: "I'm going to read you some sentences. Can you tell me for each one how accurately it describes your life at the moment. "You don't feel you have anyone close to you" or "Can you tell me how often you have felt the following in the past week: lonely, etc.".

[18] The 2005 questionnaire included two sensitive questions as defined by the French Data Protection Act, which implied warning the respondents about the two questions and asking them at the end of the interview whether they agreed to their responses being recorded; if so, the respondent had to sign a consent form, with his/her name, confirming that he/she agreed that his/her responses to those two questions would be recorded. For a more detailed explanation, see Sebille and Régnier-Loilier, 2007.

of questioning, fearing that the respondent will lose concentration or become weary), has an easy interpretation in the case of the Erfi-GGS. Through filters, the variance in interview length is high, and the interview tended to be longer when the respondent's personal situation matched the theme of the survey. These people (who had a long interview) probably felt more interested in and concerned by the survey, which explains their higher participation in the subsequent waves. Factoring in the length of the interview (Model 2, **Figure 5**) partly cancels out the influence of the "household type" variable observed in Model 1: living alone and childless in the household reduces the probability of responding to the following waves if we leave out the length of the interview, but that effect disappears when the length of the interview is taken into account.

Apart from the length of the interview, the daily availability of the people can influence retention. In particular, we make the hypothesis that having a full-time job with long working hours could lead to a lower propensity to participate in the subsequent waves. For that purpose, Model 3 (**Figure 5**) incorporates an "employment status" variable that takes account of whether the person is employed or not and the person's weekly working hours in the first wave (this variable is a substitute for socio-occupational category). We found that neither working part-time nor working more than 40 hours per week had an impact.

5.4. Do the successive waves show an accentuation of the sample distortion or a selection effect?

Independently of the factors of attrition observed between waves 1 and 3, we can investigate the distortion of the sample over the waves. Between waves 1 and 2, attrition was not random (Régnier-Loilier, 2009; 2011) and between waves 1 and 3, it was not random either (see previous section). This raises the question of whether there was an accentuation of the distortion caused by attrition between waves 2 and 3 (the same explanatory factors of attrition are present in every wave) or whether attrition was selective between the first two waves but more random between waves 2 and 3.

In other words, we can make the hypothesis that some factors may have influenced attrition between waves 1 and 2 but did not have an impact between waves 2 and 3 because of a selection effect. For example, people with little interest in the theme of the survey or who found the questions too intrusive refused to continue after the first wave; the sample in the second wave was therefore selective, with the people remaining in the sample being the most enthusiastic and interested in continuing. Conversely, we can assume that other variables, related for example to the probability of mobility or being out-of-scope (death, institutionalisation) would have the same influence in both inter-wave periods.

Based on some indicators, it seems that the two effects are combined. The attrition observed between waves 1 and 2 is much higher than that observed between waves 2 and 3, which supports a selection effect; conversely, the scatter of the weighting variable is greater for wave 3 than for wave 2,[19] indicating rather an accentuation of the distortion in the sample structure.

In order to refine the analysis, we first compared two models based on the same principle as Model 2 in **Figure 5**, measuring the probability of retention between 2005 and 2008 and between 2008 and 2011 (**Figure 6**).[20] The comparison of R^2, traditionally low in social science, nevertheless shows lower quality of the "W2 W3 retention" model. Although the models are not directly comparable (notably because the sample sizes are different), this seems to support a selection effect, at least on the basis of the variables used in the model: they explain the retention between W2 and W3 "less" than the retention between W1 and W2. However, a comparison of the significant modalities shows few differences, with most of the factors linked to attrition between 2005 and 2008 having an impact between 2008 and 2011.

Whatever the inter-wave period, retention was lower in Île-de-France, among people who intended to move, at the extreme ages (young people are more mobile, while old people have a strong propensity to move out-of-scope), among respondents who refused to answer certain questions (especially on household income), among respondents who refused to receive the initial results[21] or who considered themselves to be in poor health; and conversely retention was higher among farmers, residents of municipalities with a population of under 5,000, and among the most educated people.

Other variables for which we assumed a selection effect (and therefore a smaller impact between 2008 and 2011) come into play again, however: these are gender and nationality. Regarding nationality, that result is unsurprising in the light of the finding described above: in the Erfi-GGS, attrition among foreigners cannot be attributed more to more refusals than average, unlike the results of other research that show more refusals by foreigners, linked to lower interest in a national survey.

Some factors cease to have an influence between waves 2 and 3: low educational level, Mediterranean region, and tenants (regarding tenants, there may be a selection effect: the

[19] The comparison was done after normalising the weights for 2008 and 2011 to make them comparable, independently of the very different magnitude of the attrition between 2005 and 2008 and between 2008 and 2011.

[20] Unless otherwise indicated in the table, the characteristics are those observed in 2005 for the first model (retention W1 – W2) and in 2008 for the second model (W2 – W3).

[21] A net selection effect appears with the wish to receive the initial survey results: at the end of the interview in the first wave (2005), 7.7% of respondents did not wish to receive the initial results, compared with 3.1% after the second wave (2008); conversely, the percentage of people who refused to fill out the household income range was the same in every wave, and very low (1.5%) compared with other surveys (e.g. Trajectories and Origins, INED-INSEE).

most mobile tenants moved between the first two waves and were more often lost; those who remained are the least mobile, reducing the risk of losing contact[22]). Lastly, the length of the interview no longer has an impact on retention between 2008 and 2011, but it should be noted that this is the interview length observed in 2005 in the two models (since interview length in 2008 was not available).

In the end, therefore, most factors related to higher attrition between the first two waves recur between the next two waves, accentuating the distortion of the sample structure.

A second method was tested to see whether the difference in the impact of the same factor was significantly different between W1 and W2 and between W2 and W3. The same variable can have the same or a different impact but without us knowing whether the difference is significant. To determine this, the samples of 2005 and 2008 were "stacked" into the same file, with the inclusion of a variable to show whether it was the sample of the first or second wave (wave dummy variable). We then assessed the probability of participating in the next wave[23] on the basis of the models shown in **Figure 6**, by including, in addition to the wave dummy variable an interaction factor between it and each of the variables in the model. A statistically significant interaction indicates that the effect of the variable is different between waves 1 and 2 and between waves 2 and 3; a non-significant effect indicates that the effect is overall of the same size (either the variable had no impact between waves 1 and 2 or between waves 2 and 3, or it was significant between waves 1 and 2 and significant between waves 2 and 3). This model (**Figure 7**) confirms a significantly different impact of low educational level and Mediterranean region (higher attrition in wave 2, which disappears in wave 3), age, length of interview and wishing to receive the initials results (smaller impact in wave 3, supporting a selection effect over the waves for that variable).

[22] In fact, tenants had been living in their dwellings for an average of 8 years in 2005 (median: 4 years) compared with an average of 9 years in 2008 (median: 5 years).

[23] For the sample in wave 1, having responded to wave 2; for the sample in wave 2, having responded to wave 3.

6. CONCLUSION AND DISCUSSION

Attrition in the Generations and Gender Survey is a major concern, for several reasons. Firstly, apart from financial considerations that prevented some countries from administering the different waves of the survey, a too high rate of attrition between the first two waves of the survey led others to withdraw from the study (notably Germany, where only one-third of the people interviewed in the first wave participated in the second wave). Excessive sample loss can also compromise the comparative aim of the project. Secondly, attrition determines the quality of the data more broadly and, in an international survey, it is important for attrition to be precisely documented. The preliminary study of attrition in the French survey enabled us to construct a longitudinal weighting variable based on reasoning, but it will be important for the purposes of international comparisons, to consider constructing a weighting that is relatively homogeneous between the participating countries. Lastly, the magnitude of the attrition and the distortion of the initial sample can have implications for the subsequent statistical studies. Apart from reduced statistical accuracy as the sample size decreases, excessively selective attrition can induce biases on the variables of interest.[24]

Since France was one of the first countries to complete the data collection for the Generations and Gender Survey, we would like to share our experience. Cumulative attrition after the three waves of the survey (2005-2011) was 43%, a rate similar to that found in other similar surveys in France. As in most longitudinal surveys, it is between the first and the second waves that attrition was the highest (35%). Beyond that figure, our study reveals that the factors associated with attrition are consistent on the whole with those observed in other surveys. In terms of socio-demographic characteristics, we first find a significant effect of gender, age, education level and nationality. Regarding nationality, however, it should be noted that attrition among foreigners does not stem from more frequent refusals to participate in the subsequent waves[25] but from other reasons (particularly loss of contact). Place of residence also has a strong impact, with higher attrition in large cities, in Île-de-France, among tenants and among people who intended to move. The respondent's attitude to the study also plays an important role: less cooperation during the first interview (refusal to answer certain questions) and less interest in the study (refusal to receive the initial results) correlate with a higher rate of leaving the sample between the waves. Interestingly,

[24] For example, in the French survey on fertility intentions (INED, 1998-2003), the high rate of attrition after the three waves of the survey (70%) caused significant distortion of the sample structure but also led to a significant change to the relationship between the variables of interest (especially the intention to have a child) and its explanatory factors (Mazuy *et al.*, 2005).

[25] For example, because of less interest in a national study or more difficulty understanding the questions. It should be noted, however, that only respondents who spoke French were interviewed in the first wave.

we did not find a negative impact for a long interview: on the contrary, a long interview was associated with higher retention and vice versa. This can be attributed in part to the architecture of the questionnaire, since the people most likely to feel unconcerned by a survey about family are asked far fewer questions than people in a union, with children, etc. Lastly, being less inclined to confide in family and friends and, unsurprisingly, poor health are also associated with lower retention (higher risk of being unable to respond, to be in an institution or of dying between the waves). However, based on the feedback from interviewers during evaluations of the data collection, we made the hypothesis of higher retention of respondents who felt isolated, since some might see the interviewer's visit as an opportunity to communicate. This was not verified.

The comparison of the factors associated with attrition in the two inter-wave periods shows that most of the factors recur in both periods. We postulated that the factors related to higher geographical mobility or more difficulty making contact with the respondent were highly likely to recur from one wave to another, but that this should be less the case for factors related to an attitude of rejection due to a selection effect: people who are wary of or uninterested in responding would have left at the end of the first wave. In fact, almost all the factors that influenced attrition between the first two waves also have an impact between the next two waves, with a similar intensity (non-significant interaction factor). Only people with a low education level, residents of the Mediterranean region and people who did not wish to receive the results, who have a higher probability of attrition (an estimated parameter with a negative sign), participated more in wave 3 than in wave 2 (positive and significant interaction factor). That might reflect a selection effect but the distortion of the sample structure increased over the waves, as attested by the greater dispersion of the weighting variable calculated after the third wave (compared with that calculated on wave 2).

We regret that we were unable to investigate some hypotheses more thoroughly, especially those related to the interview conditions: what is the effect of the interviewer's gender and age in a survey of gender relations and generational effects? Unfortunately, that information is not available in the survey. It will be important in future to better anticipate the type of data to be collected, with a view to conducting methodological studies. Indeed, the study of factors linked to attrition and the measurement of their impact between the different waves corresponds to a more general methodological objective. Upstream of the survey design, better knowledge of these factors can help the designers to define the sampling (certain categories of people, more vulnerable to attrition, could be over-represented in the initial random draw in order to have a sufficiently large sample for longitudinal studies), to choose the interviewers (gender or age, depending on the theme of the survey) and to adopt

strategies to retain the sample between the waves (smaller or larger inter-wave intervals could be considered depending on the probability of attrition). Downstream of the data collection, identifying the factors linked to sample loss makes it possible to construct a weighting variable, based on reasoning, to correct the sample distortion. Furthermore, answering the question of whether the same factors of attrition recur between the waves alerts data users to the risks of analysis bias in longitudinal surveys. Even if it is always possible to construct adjustment variables, excessive distortion of the initial sample structure over the waves should encourage researchers to ensure that this does not affect their results.

Acknowledgements

The authors would like to thank Laurent Toulemon (INED, France) for his advice and suggestions for the analysis, and Catriona Dutreuilh for the English proof-reading.

The Erfi-GGS was funded by INED, INSEE, ANR, CNAF, CNAV, COR, DARES and DREES.

References

Behr, A., Bellgardt, E. and Rendtel, U. (2005), "Extent and Determinants of Panel Attrition in the European Community Household Panel", *European Sociological Review*, 21(5): 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): 359–377.

Collomb P., 1979, "Les émigrants de l'Ouest audois dix-neuf ans après. Examen des méthodes de collecte et d'analyse", *Population*, (34)1: 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): 251-299.

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

Holbrook A., Green M., Krosnick 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: 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): 1-15.

Laurie H., Smith R., Scott L., 1999, "Strategies for Reducing Non-response in a Longitudianl Panel Survey", *Journal of Official Statistics*, Vol. 15: 269-282.

Lipps O., 2007, *Attrition in the Swiss Household Panel*, Methoden – Daten – Analysen, 1(1): 45-68. 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): 545-557.

Mazuy M., Razafindratsima N., La Rochebrochard, É. de, 2005, *Déperdition dans l'enquête "Intentions de fécondité",* Documents de travail 129, Ined, 36 pages.

Mermilliod C., 2012, *L'impact de la mise en place d'un suivi d'adresse entre les deux vagues de l'enquête santé et itinéraire professionnel (SIP)*, Document de travail, Série Sources et méthodes, 37, Drees, 26 pages.

Nicoletti, C. and 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: 763-781.

Olsen F., Abelsen B., Olsen J. A., "Improving response rate and quality of survey data with a scratch lottery ticket incentive", *BMC Medical Research Methodology*, 15-52: www.biomedcentral.com/1471-2288/12/52

Razafindratsima N., Kishimba N., 2004, "La déperdition dans la cohorte Cocon entre 2000 et 2002", *Population*, 59 (3-4): 419-448.

Régnier-Loilier A., 2009, "La déperdition dans les enquêtes longitudinales: l'exemple de l'enquête Érfi", *Courrier des statistiques*, 128: 55-62.

Régnier-Loilier A., 2011, “Attrition entre les deux premières vagues de l’Étude des relations familiales et intergénérationnelles ”, in Tremblay M.-É., Lavallée P. et El haj Tirari M. (dir), *Pratiques et méthodes de sondage*, Dunod: 254-258.

Régnier-Loilier A., 2012, *Présentation, questionnaire et documentation de l’Étude des relations familiales et intergénérationnelles (Érfi-GGS3)*, Documents de travail 187, Ined, 327 pages.

Régnier-Loilier A., Légaré J., 2010, “Generations and Gender Programme. Présentation et perspectives”, *Revue de l’Observatoire franco-québécois de la santé et de la solidarité*, 2: 129-135.

Régnier-Loilier A., Saboni L., Valdes B., 2011, *Presentation and modifications to the GGS questionnaire in France (wave 2)*, Documents de travail 173, Ined, 146 pages.

Riandey B., 1988, “La qualité du suivi des échantillons dan les enquêtes démographiques: un bilan”, *Population*, 43(4-5): 829-854.

Sebille P., Régnier-Loilier A., 2007, *Modifications to the Generations and Gender Surveys questionnaire in France (wave 1)*, Documents de travail 144, Ined, 191 pages.

Stoop I., 2005, *The hunt for the last respondent. Nonresponse in sample survey*, Social and Cultural Planning of the Netherlands, The Hagues, ma, 339 pages.

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, 85 pages.

Vikat *et al.*, 2007, “Generations and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course”, *Demographic research*, 17 (14), p. 389-440.

Voorpostel M., Lipps O., 2011, “Attrition in the Swiss Household Panel: Is Change Associated with Drop-out?” *Journal of Official Statistics*, 27(2): 301-318.

Watson N., Wooden M., 2009, “Identifying Factors Affecting Longitudinal Survey Response”, in Lynn P. (dir.), *Methodology of Longitudinal Surveys*, Chichester, UK: John Wiley & Sons, chapitre 10.

Figure 5. Model of the probability of having replied to the third Érfi survey wave (estimated parameters of the logistic regression)

	MODEL 1	MODEL 2	MODEL 3
Constant	0,96 ***	0,95 ***	0,95 ***
Gender			
Male	*ref*	*ref*	*ref*
Female	0,14 ***	0,15 ***	0,15 ***
Age			
Age	-0,14 *	-0,17 **	-0,15 *
Age ²	-0,63 ***	-0,63 ***	-0,64 ***
Education level			
No educational qualification	-0,23 ***	-0,23 ***	-0,23 ***
Lower secondary education	*ref*	*ref*	*ref*
Secondary and higher education	0,29 ***	0,29 ***	0,31 ***
Socio-occupational category			
Farmer	0,40 *	0,37 *	
Self-employed: business, trade or crafts	-0,17	-0,16	
Manager, professional or higher-level intellectual occupation	0,09	0,09	
Intermediate occupation	-0,01	-0,01	
Clerical/sales	*ref*	*ref*	
Manual worker	-0,05	-0,06	
Unemployed	-0,36 ***	-0,36 ***	
Retired	0,08	0,09	
Homemaker	-0,08	-0,08	
Student	-0,10	-0,07	
Other economically inactive	-0,44 ***	-0,44 ***	
Employment status			
Self-employed			-0,13
Employee (full-time, 40+ hours a week)			0,06
Employee (full-time, less than 40 hours a week)			*ref*
Employee (part-time)			0,01
Homemaker			-0,08
Unemployed			-0,35 ***
Retired			0,07
Student			-0,05
Other economically inactive			-0,44 ***
Size of locality of residence			
City with pop. 100,000 or more	-0,01	-0,01	0,00
Town with pop. 5,000-99,999	*ref*	*ref*	*ref*
Town with pop. under 5,000	0,33 ***	0,34 ***	0,34 ***
Rural municipality	0,00	0,00	0,02
Region			
Ile-de-France	-0,51 ***	-0,49 ***	-0,48 ***
Parisian Basin	-0,06	-0,05	-0,05
North	*ref*	*ref*	*ref*
East	0,05	0,04	0,04
West	0,07	0,08	0,08
South-West	-0,02	-0,03	-0,02
Centre-East	0,06	0,05	0,05
Mediterranean	-0,58 ***	-0,59 ***	-0,58 ***

	MODEL 1	MODEL 2	MODEL 3
Nationality			
French	*ref*	*ref*	*ref*
Foreign	-0,58 ***	-0,60 ***	-0,60 ***
Household income question			
Responded	*ref*	*ref*	*ref*
Didn't know	-0,31 **	-0,29 **	-0,28 **
Refused to respond	-0,57 ***	-0,56 ***	-0,57 ***
Dwelling occupancy status			
Outright owner, life tenant	*ref*	*ref*	*ref*
Home-buyer (paying off a home loan)	0,14 **	0,12 *	0,11
Tenant, guest	-0,12 *	-0,12 *	-0,13 *
Intention to move within 3 years			
No, definitely not	*ref*	*ref*	*ref*
Don't know, probably, probably not	-0,05	-0,07	-0,07
Yes, definitely	-0,36 ***	-0,37 ***	-0,37 ***
Perceived health			
Very good, good	*ref*	*ref*	*ref*
Average, poor	-0,09	-0,09 *	-0,09 *
Very poor	-0,47 **	-0,51 **	-0,51 **
Type of dwelling			
House	*ref*	*ref*	*ref*
Apartment, other	-0,18 ***	-0,17 ***	-0,17 ***
Wishes to receive the survey results?			
Yes	*ref*	*ref*	*ref*
No	-1,30 ***	-1,27 ***	-1,27 ***
Signed consent form?			
Signed	*ref*	*ref*	*ref*
Refused to sign	-1,07 ***	-1,05 ***	-1,05 ***
Not concerned	-0,72 ***	-0,69 ***	-0,69 ***
Listened to someone talk about his/her life			
Yes	*ref*	*ref*	*ref*
No	-0,19 ***	-0,17 ***	-0,16 ***
Type of household			
Couple without children	*ref*	*ref*	*ref*
Couple with children aged over 3	0,05	0,04	0,04
Couple with children aged under 3	-0,07	-0,09	-0,09
Alone with children aged over 3	-0,02	0,00	0,00
Alone with children aged under 3	-0,54 **	-0,53 **	-0,52 **
Household of several people	-0,17	-0,13	-0,12
Alone	-0,12 **	-0,08	-0,08
Length of interview in 2005			
20-44 minutes		-0,24 ***	-0,24 ***
45-59 minutes		*ref*	*ref*
60-75 minutes		-0,01	-0,01
75-150 minutes		0,16 **	0,16 **
Not indicated		-0,14	-0,14

MODEL QUALITY			
Likelihood Ratio	<.0001	<.0001	<.0001
R^2	0.1635	0.1670	0.1666
Percent Concordant	69.3	69.6	69.6

Source: Érfi-GGS, INED-INSEE, 2005-2011
Scope: all respondents to the first wave (in 2005).
Key: a statistically significant, positive coefficient (negative resp.), indicates a factor that increases (incomplete resp.) the probability of refusing to participate in the third wave in 2011.
***: significant at 1%; **: significant at 5%; *: significant at 10%; *ref*: reference category.

Figure 6. Model of the probability of having replied to the second wave ("follow-up W1-W2) / to the third wave (follow-up V2-V3) of the Érfi survey

	Retent W1-W2	Retent W2-W3
Constant	1,32 ***	2,04 ***
Gender		
Male	*ref*	*ref*
Female	0,07	0,21 ***
Age		
Age	-0,06	0,14
Age ²	-0,47 ***	-0,86 ***
Education level		
No educational qualification	-0,27 ***	-0,05
Lower secondary education	*ref*	*ref*
Secondary and higher education	0,23 ***	0,35 ***
Socio-occupational category		
Farmer	-0,02	0,37
Self-employed: business, trade or crafts	-0,20	0,05
Manager, professional or higher-level intellectual occupation	0,13	0,10
Intermediate occupation	0,00	0,19
Clerical/sales	*ref*	*ref*
Manual worker	-0,06	-0,08
Unemployed	-0,21 **	-0,24
Retired	0,00	0,29 *
Homemaker	-0,28 ***	0,07
Student	-0,04	-0,36
Other economically inactive	-0,23 *	-0,75 ***
Size of locality of residence		
City with pop. 100,000 or more	-0,08	0,09
Town with pop. 5,000-99,999	*ref*	*ref*
Town with pop. under 5,000	0,33 ***	0,42 ***
Rural municipality	0,07	0,06
Region		
Ile-de-France	-0,50 ***	-0,63 ***
Parisian Basin	-0,10	-0,16
North	*ref*	*ref*
East	-0,08	0,03
West	0,01	0,09
South-West	-0,11	-0,06
Centre-East	-0,08	-0,07
Mediterranean	-0,72 ***	-0,20

	Retent W1-W2	Retent W2-W3
Nationality		
French	*ref*	*ref*
Foreign	-0,42 ***	-0,49 ***
Household income question		
Responded	*ref*	*ref*
Didn't know	-0,37 ***	-0,30 *
Refused to respond	-0,63 ***	-0,81 ***
Dwelling occupancy status		
Outright owner, life tenant	*ref*	*ref*
Home-buyer (paying off a home loan)	0,13 *	0,07
Tenant, guest	-0,25 ***	-0,14
Intention to move within 3 years		
No, definitely not	*ref*	*ref*
Don't know, probably, probably not	-0,04	-0,13
Yes, definitely	-0,40 ***	-0,31 ***
Perceived health		
Very good, good	*ref*	*ref*
Average, poor	-0,05	-0,21 **
Very poor	-0,63 ***	-1,26 ***
Type of dwelling		
House	*ref*	*ref*
Apartment, other	-1,62 ***	-1,30 ***
Listened to someone talk about his/her life		
Yes	*ref*	*ref*
No	-0,19 ***	-0,14 *
Type of household		
Couple without children	*ref*	*ref*
Couple with children aged over 3	0,00	0,04
Couple with children aged under 3	0,09	0,14
Alone with children aged over 3	0,01	-0,20
Alone with children aged under 3	-0,10	-0,25
Household of several people	-0,05	0,16
Alone	-0,08	-0,05
Length of interview in 2005		
20-44 minutes	-0,21 ***	-0,21 **
45-59 minutes	*ref*	*ref*
60-75 minutes	0,09	0,02
75-150 minutes	0,19 ***	-0,15
Not indicated	-0,04	-0,51 ***

MODEL QUALITY		
Likelihood Ratio	<.0001	<.0001
R^2	0.1507	0.0992
Percent Concordant	68.8	67.4

Source: Érfi-GGS, INED-INSEE, 2005-2008-2011
Scope: all respondents to the first wave (in 2005) for the 'retention W1-W2' model; all respondents to the second wave (in 2008) for the 'retention W2-W3' model.
Key: a statistically significant, positive coefficient (negative resp.), indicates a factor
that increases (incomplete resp.) the probability of responding to the next wave.
***: significant at 1%; **: significant at 5%; *: significant at 10%; ref: reference category.

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

	Estimated parameter of the factor		Estimated parameter of the interaction between the factor and the "wave" variable	
Constant	1,4421			
Observation wave				
First wave (2005)				
Second wave (2008)	0,601	**		
Gender				
Male				
Female	0,10	*	0,11	
Age				
Age	-0,09		0,23	
Age ²	-0,44	***	-0,42	**
Education level				
No educational qualification	-0,30	***	0,25	*
Lower secondary education				
Secondary and higher education	0,20	***	0,15	
Socio-occupational category				
Farmer	0,16		0,21	
Self-employed: business, trade or crafts	-0,16		0,22	
Manager, professional or higher-level intellectual occupation	0,08		0,02	
Intermediate occupation	0,03		0,16	
Clerical/sales				
Manual worker	-0,04		-0,04	
Unemployed	-0,20		-0,05	
Retired	0,00		0,28	
Homemaker	-0,19	*	0,26	
Student	-0,04		-0,32	
Other economically inactive	-0,21		-0,54	**
Size of locality of residence				
City with pop. 100,000 or more	-0,11	*	0,19	*
Town with pop. 5,000-99,999				
Town with pop. under 5,000	0,33		0,09	
Rural municipality	0,05		0,01	
Region				
Ile-de-France	-0,46	***	-0,17	
Parisian Basin	-0,07		-0,09	
North				
East	-0,04		0,07	
West	0,01		0,08	
South-West	-0,05		-0,02	
Centre-East	0,00		-0,07	
Mediterranean	-0,75	***	0,56	***

	Estimated parameter of the factor		Estimated parameter of the interaction between the factor and the "wave" variable	
Nationality				
French				
Foreign	-0,47	***	-0,03	
Household income question				
Responded				
Didn't know	-0,43	***	0,12	
Refused to respond	-0,67	***	-0,14	
Dwelling occupancy status				
Outright owner, life tenant				
Home-buyer (paying off a home loan)	0,20	***	-0,13	
Tenant, guest	-0,17	***	0,03	
Intention to move within 3 years				
No, definitely not				
Don't know, probably, probably not	-0,05		-0,08	
Yes, definitely	-0,41	***	0,10	
Perceived health				
Very good, good				
Average, poor	-0,05		-0,16	
Very poor	-0,71	***	-0,54	
Wishes to receive the survey results?				
Yes				
No	-1,70	***	0,40	**
Listened to someone talk about his/her life				
Yes				
No	-0,17	***	0,03	
Type of household				
Couple without children				
Couple with children aged over 3	0,03		0,01	
Couple with children aged under 3	0,01		0,12	
Alone with children aged over 3	-0,01		-0,18	
Alone with children aged under 3	-0,35		0,10	
Household of several people	0,01		0,15	
Alone	-0,11	*	0,05	
Length of interview in 2005				
20-44 minutes	-0,25	***	0,04	
45-59 minutes				
60-75 minutes	0,04		-0,03	
75-150 minutes	0,21	***	-0,35	**
Not indicated	-0,11		-0,40	**

MODEL QUALITY	
Likelihood Ratio	<.0001
R^2	0.1748
Percent Concordant	71.3

Source: Érfi-GGS, INED-INSEE, 2005-2008-2011
Scope: all respondents to the first wave (2005) and the second wave (2008).
***: significant at 1%; **: significant at 5%; *: significant at 10%; ref: reference category.
Key: the probability of responding to the next wave is significantly lower in the Mediterranean region (a negative, statistically significant parameter). However, the impact of the factor is significantly different between waves 1-2 and between waves 2-3, with higher participation in wave 3 than in wave 2 (a positive, statistically significant parameter)

30

Figure 8. Probability of having refused to reply versus having not taken part in the third wave for another reason by respondent's characteristics in 2005

	PROBABILITY OF REFUSAL
Constant	0,34 *
Gender	
Male	*ref*
Female	0,23 ***
Age	
Age	0,03
Age ²	-0,42 ***
Education level	
No educational qualification	-0,04
Primary or lower secondary education	*ref*
Secondary or higher education	-0,24 ***
Socio-occupational category	
Farmer	0,44
Self employed: business, trade or crafts	-0,12
Manager, professional or higher-level intellectual occupation	-0,36 **
Intermediate occupation	-0,03
Clerical/sales	*ref*
Manual worker	-0,01
Unemployed	-0,01
Retired	-0,33 **
Homemaker	-0,15
Student	-0,09
Other economically inactive	-0,56 ***
Size of locality of residence	
City with pop. 100,000 or more	-0,09
Town with pop. 5,000-99,999	*ref*
Town with pop. under 5,000	-0,06
Rural municipality	-0,01
Region	
Ile-de-France	0,28 *
Parisian Basin	0,22
North	*ref*
East	-0,09
West	0,01
South-West	0,15
Centre-East	-0,04
Mediterranean	-0,34 **
Nationality	
French	*ref*
Foreign	0,03
Household income question	
Responded	*ref*
Didn't know	-0,16
Refused to respond	0,60 **
Dwelling occupancy status	
Outright owner, life tenant	*ref*
Home-buyer (paying off a home loan)	0,09
Tenant, guest	-0,26 ***
Intention to move within 3 years	
No, definitely not	*ref*
Don't know, probably, probably not	-0,10
Yes, definitely	-0,46 ***
Perceived health	
Very good or good	*ref*
Average, poor	-0,06
Very poor	-0,77 ***
Type of dwelling	
House	*ref*
Apartment, other	-0,01
Wishes to receive the survey results?	
Yes	*ref*
No	1,40 ***
Signed consent form?	
Signed	*ref*
Refused to sign	0,91 ***
Not concerned	0,91 ***
Listened to someone talk about his/her life	
Yes	*ref*
No	0,16 **
Type of household	
Couple without children	*ref*
Couple with children aged over 3	-0,01
Couple with children aged under 3	-0,18
Alone with children aged over 3	-0,23
Alone with children aged under 3	-0,32
Household of several people	0,08
Alone	-0,25 ***
Length of interview in 2005	
20-44 minutes	-0,04
45-59 minutes	*ref*
60-75 minutes	0,07
75-150 minutes	0,08
Not indicated	0,03

Source: Érfi-GGS, INED-INSEE, 2005-2008-2011
Scope: all non-respondents to the third wave (2011).
Key: a statistically significant, positive coefficient (negative resp.) indicates a factor that increases (incomplete resp.) the probability of refusing to participate in a wave.
*** : significant at 1%; **: significant at 5%; *: significant at 10%; *ref*: reference category