

D'UNE ENQUÊTE PONCTUELLE À UNE ENQUÊTE LONGITUDINALE : ÉTUDE DU COMPORTEMENT DES RÉPONDANTS EN EUROPE

Aurore Philibert¹, Nicolas Briant² & Anne Laferrère³

Résumé long – Résultats provisoires, ne pas citer.

¹ *Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny 75775 Paris Cedex 16*

aurore.philibert@dauphine.fr

² *Fondation Dauphine, Place du Maréchal de Lattre de Tassigny 75775 Paris Cedex 16*

nicolas.briant@fondation-dauphine.fr

³ *INSEE, 18 bd Adolphe Pinard 75675 Paris Cedex 14* anne.laferrere@insee.fr

Résumé. L'enquête SHARE (Survey on Health, Ageing and Retirement in Europe) étudie le vieillissement de la population en Europe en interrogeant tous les deux ans des personnes de 50 ans et plus dans une vingtaine de pays d'Europe. L'originalité de l'enquête tient à son caractère pluridisciplinaire, européen et surtout longitudinal. Conserver les répondants d'une vague à l'autre est alors primordial. Nous nous intéressons ici au comportement de réponse des enquêtés de SHARE face à des interrogations bisannuelles. Pour cela, nous comparons le taux de participation des individus interrogés aux quatre premières vagues en fonction du nombre de vagues consécutives de participation. L'analyse fait apparaître une augmentation de la participation en fonction du nombre de vagues d'interrogations, pour 9 des 11 pays considérés. Les individus participant à leur 2^{ème} interrogation apparaissent plus difficiles à interroger que lors de la 3^{ème} ou 4^{ème} interrogation. Le passage d'une enquête ponctuelle à une enquête longitudinale semble être une étape primordiale pour les répondants. Cet effet de seconde interrogation se retrouve dans la littérature, et l'enquête SHARE en confirme la robustesse. Dans un second temps, nous étudions les déterminants de l'attrition pour les quatre premières vagues de l'enquête.

Mots-clés. Enquête longitudinale, enquête européenne, pluridisciplinaire, vieillissement, taux de réponse, attrition

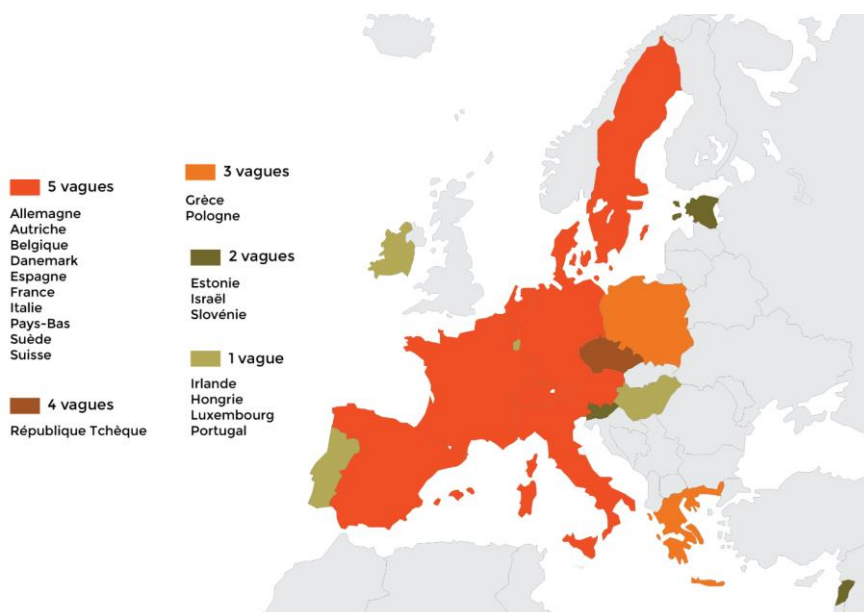
1 Introduction

L'enquête SHARE (Survey on Health, Ageing and Retirement in Europe) étudie le vieillissement de la population en Europe en interrogeant tous les deux ans des personnes de 50 ans et plus dans une vingtaine de pays d'Europe (<http://share.dauphine.fr/>). A travers des questions sur leur santé, leur bien-être, leur intégration dans leur famille, leur réseau social et leur situation économique, SHARE participe à la recherche sur le vieillissement afin de soutenir les chercheurs dans leur compréhension de ce processus.

Les personnes de 65 ans et plus représentaient, en 2013, 18% de la population de l'Europe des 27. Les projections de population d'Eurostat y prévoient une proportion de 30% de personnes de 65 ans et plus d'ici 2060 (Eurostat, 2014). SHARE s'inscrit dans ce contexte d'Europe vieillissante, véritable défi économique et social pour les pays concernés.

L'enquête a commencé en 2004 avec une première vague d'interrogation de plus de 25 000 individus dans 11 pays d'Europe dont plus de 3 000 en France. Puis, tous les deux ans, quatre autres vagues d'interrogation ont eu lieu. La dernière, en 2012-2013 (Vague 5) concerne près de 70 000 répondants dans 15 pays d'Europe dont plus de 4 000 en France. La prochaine vague (Vague 6) est en cours de préparation pour l'année 2015. Dix pays ont ainsi participé aux 5 vagues de SHARE, ce qui est le cas de la France tandis que 10 autres pays ont participé entre 1 et 4 vagues (Figure 1).

Figure 1: Participation des pays européens par nombre de vagues de participation



Cette enquête s'est fortement inspirée de deux précédentes études, l'enquête ELSA (English Longitudinal Survey on Ageing) en Angleterre qui existe depuis 1998 et l'enquête HRS (Health and Retirement Study) aux Etats-Unis qui existe depuis 1992.

L'originalité de l'enquête SHARE tient à quatre particularités. Tout d'abord, les données sont mises à disposition des chercheurs gratuitement, pour tous ceux qui en formulent la demande et justifient d'une affiliation scientifique¹, ce qui donne lieu à une production très riche². L'enquête a un caractère pluridisciplinaire puisque des questions en sociologie, économie et santé sont posées aux enquêtés. Des chercheurs issus de domaines divers comme l'épidémiologie ou la psychologie travaillent et utilisent ces données. L'enquête est internationale et concerne 19 pays européens plus Israël. Le questionnaire est identique pour tous les pays participants et généré en anglais ; chaque pays le traduisant dans sa (ou ses) langue(s). Enfin, et c'est ce qui nous intéresse ici, c'est une enquête longitudinale, puisqu'elle réinterroge les mêmes personnes tous les deux ans, ce qui permet d'étudier le vieillissement comme un processus et non comme un état stationnaire, d'approcher l'identification de causalités au-delà des corrélations ainsi que de tester les effets à long terme de certains comportements (fumer, boire,...) ou événements (divorce, perte de travail, retraite...).

C'est dans ce contexte particulier que nous nous intéressons au comportement de réponse des enquêtés de SHARE face à des interrogations bisannuelles. En effet l'analyse à partir des données longitudinales n'est valide que si l'attrition, c'est-à-dire la perte d'individus, de l'échantillon d'une vague à la suivante n'est pas trop forte. Par ailleurs cette attrition ne doit pas être liée aux phénomènes que l'on désire étudier. Dans la suite de ce document, nous décrivons tout d'abord dans la section 2 nos données et les taux de participation que nous avons utilisé, puis les premiers résultats obtenus dans la section 3. La section 4 présente l'analyse en cours sur les déterminants de l'attrition ; la section 5 conclut.

¹ Les données des quatre premières vagues de SHARE sont accessibles à l'adresse suivante : <http://www.share-project.org/data-access-documentation/research-data-center-data-access.html>.

² La liste des travaux réalisés à partir des données de SHARE est disponible sur le site de SHARE Europe : <http://www.share-project.org/publications.html>

2 Méthodologie

2.1. Source des données

Afin d'évaluer le suivi longitudinal des répondants dans chacun des pays au cours des vagues, nous avons créé un historique des participations de chaque individu ayant été interrogé au moins une fois au cours des 4 premières vagues de SHARE³. Les données de la vague 5 ne sont pas encore mises à disposition. Pour chaque vague, nous avons indiqué si la personne avait participé à un entretien classique (codé 1), un entretien de « fin-de-vie » (codé 2) ou n'avait pas participé (codé 0). Dans le cas où le répondant est décédé après une vague à laquelle il a participé, un entretien de « fin-de-vie » sur la dernière année de vie du répondant est conduit auprès d'une personne proche. Cette codification permet d'obtenir un nombre à 4 chiffres décrivant les différentes combinaisons possibles dans la participation individuelle. Par exemple le code 1102 décrit les individus entrés en vague 1, ayant participé à la vague 2 mais pas à la vague 3 et pour qui un entretien de fin-de-vie a été réalisé en vague 4.

Pour chaque pays, nous avons calculé le nombre de nouveaux individus à chaque vague (entrants et nouveaux conjoints), le nombre d'individus suivis d'une vague à l'autre et le nombre d'individus récupérés après un (ou plusieurs) échec(s) lors des vagues précédentes. Nous pouvons ensuite calculer le pourcentage d'individus conservés d'une vague à l'autre parmi ceux interrogés à la vague précédente. Seuls les individus pour lesquels un entretien de fin-de-vie a été réalisé n'ont pas été considérés pour les vagues suivantes.

Nous avons comparé les résultats des 10 pays ayant participé aux 4 premières vagues de SHARE : Autriche (AT), Belgique (Bf et Bn), Suisse (CH), Allemagne (DE), Danemark (DK), Espagne (ES), France (FR), Italie (IT), Pays-Bas (NL) et Suède (SE). En ce qui concerne la Belgique, la Flandre (Bf) et la Wallonie (Bn) possèdent chacune leur propre agence de collecte et leur propre réseau d'enquêteur, nous les comptabilisons donc ici comme deux pays distincts.

³ Cet article utilise les données de la vague 4 de SHARE version 1.1.1. du 28 mars 2013 (DOI: 10.6103/SHARE.w4.111), celles de la vague 1 et 2 de SHARE version 2.6.0 du 29 novembre 2013 (DOI: 10.6103/SHARE.w1.260 et 10.6103/SHARE.w2.260) et celles de la vague 3 SHARELIFE version 1 du 24 novembre 2010 (DOI: 10.6103/SHARE.w3.100). La collecte des données de SHARE a été principalement financée par la Commission européenne à travers le 5^{ème} Framework Programme (projet QLK6-CT-2001-00360 dans le programme thématique sur la qualité de vie), par le 6^{ème} Framework Programme (projets SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857 et SHARELIFE, CIT4-CT-2006-028812) et par le 7^{ème} Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 et SHARE M4, N° 261982). Un financement complémentaire provenant de l'U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 et OGHA 04-064) et du ministère allemand de l'éducation et de la recherche ainsi que diverses autres sources de financement nationales ont été vivement appréciées (voir <http://www.share-project.org/contact-organisation/funding.html> pour la liste complète des institutions de financement).

2.2. Taux de participation

Dans le cadre d'une enquête longitudinale, plusieurs taux de réponse peuvent être calculés. En se basant sur les travaux de Lynn (2005) présentant des taux de réponse conditionnels, nous définissons ici des taux de participation par vagues d'interrogations consécutives.

On définit la participation (ou l'interrogation) d'un individu à une vague donnée de SHARE quand cet individu était éligible et a répondu à l'enquête au cours du terrain de cette vague. Un entretien de fin de vie est considéré comme une participation. La participation à l'enquête SHARE n'a pas de caractère obligatoire, quel que soit le nombre de vagues auxquelles le répondant a déjà participé et quel que soit le pays.

Nous ne nous intéressons pas ici à la récupération des individus suite à un ou plusieurs échecs (trous de collecte).

Nous définissons donc trois taux de participations.

Tout d'abord, T4 le taux de participation à la 4^{ème} interrogation consécutive, qui est défini comme la part des individus ayant accepté de participer aux vagues 1, 2, 3 et 4 parmi les individus ayant participé aux vagues 1, 2 et 3 ; ce qui correspond à la formule suivante $T4 = \frac{N_{1111} + N_{1112}}{N_{1111} + N_{1112} + N_{1110}}$ avec par exemple N_{1111} le nombre d'individus ayant participé aux 4 premières vagues de SHARE. Ce taux est assimilable à un taux de réponse conditionnel longitudinal proposé par Lynn (2005) : $RR_{1234|123}$.

Ensuite, nous définissons T3 le taux de participation à la 3^{ème} interrogation consécutive. Cette troisième interrogation peut avoir eu lieu en vague 3 ou 4. Si elle a lieu en vague 3, T3 peut être assimilable au taux de réponse conditionnel $RR_{123|12}$, et il s'agit de la part des individus entrés en vague 1 ayant participé aux vagues 1, 2, 3 qu'ils aient ou non participé en vague 4 ($N_{1110} + N_{1111} + N_{1112} + N_{1120}$) parmi les individus ayant participé aux vagues 1 et 2 (quelle que soit leur participation en vague 3 et 4) ($N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1100} + N_{1101} + N_{1102}$). Si elle a lieu en vague 4, T3 est assimilable au taux de réponse conditionnel $RR_{234|23}$, il s'agit de la part des individus entrés en vague 2 ayant participé aux vagues 2, 3, 4 ($N_{0111} + N_{0112}$) parmi les individus ayant participé aux vagues 2 et 3 (quelle que soit leur participation en vague 4) ($N_{0111} + N_{0112} + N_{0110}$). Le taux de participation T3 est une combinaison de ces deux taux de réponse et se définit grâce à la formule suivante :

$$T3 = \frac{N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{0111} + N_{0112}}{N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1100} + N_{1101} + N_{1102} + N_{0111} + N_{0112} + N_{0110}}$$

avec par exemple N_{0111} le nombre d'individus ayant participé aux vagues 2, 3 et 4 de SHARE.

Enfin nous définissons T2 le taux de participation à la 2^{ème} interrogation consécutive. Cette deuxième interrogation peut avoir lieu en vague 2, 3 ou 4. Si elle a lieu en vague 2, T2 est assimilable au taux de réponse conditionnel $R_{12|1}$, c'est à dire la part des individus entrés en vague 1 ayant participé aux vagues 1 et 2 (quelle que soit leur participation aux vagues 3 et 4) ($N_{1100} + N_{1101} + N_{1102} + N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1200}$) parmi les individus ayant participé à la vague 1 (quelle que soit leur participation aux vagues 2, 3 et 4) ($N_{1100} + N_{1101} + N_{1102} + N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1200} + N_{1000} + N_{1001} + N_{1002} + N_{1010} + N_{1011} + N_{1012} + N_{1020}$). Si la deuxième interrogation a lieu en vague 3, T2 est assimilable au taux de réponse conditionnel $R_{23|2}$ comme la part des individus entrés en vague 2, ayant participé aux vagues 2 et 3 (quelle que soit leur participation en vague 4) ($N_{0110} + N_{0111} + N_{0112} + N_{0120}$) parmi les individus ayant participé à la vague 2 (quelle que soit leur participation en vague 3 et 4) ($N_{0110} + N_{0111} + N_{0112} + N_{0120} + N_{0100} + N_{0101} + N_{0102}$). Si la deuxième interrogation a lieu en

vague 4, T2 est assimilable au taux de réponse conditionnel $R_{3|3}$ comme la part des individus entrés en vague 3, ayant participé aux vagues 3 et 4 ($N_{0011} + N_{0012}$) parmi les individus ayant participé à la vague 3 (quelle que soit leur participation en vague 4) ($N_{0011} + N_{0012} + N_{0010}$).

Le taux de participation T2 est une combinaison de ces trois taux de réponse conditionnel et se définit grâce à la formule suivante :

$$T2 = \frac{N_{1100} + N_{1101} + N_{1102} + N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1200} + N_{0110} + N_{0111} + N_{0011} + N_{0012} + N_{0112} + N_{0120}}{N_{0011} + N_{0012} + N_{0010} + N_{1100} + N_{1101} + N_{1102} + N_{1110} + N_{1111} + N_{1112} + N_{1120} + N_{1200} + N_{1000} + N_{1001} + N_{1002} + N_{1010} + N_{1011} + N_{1012} + N_{1020} + N_{0110} + N_{0111} + N_{0112} + N_{0120} + N_{0100} + N_{0101} + N_{0102}}$$

avec par exemple N_{1200} le nombre d'individus ayant participé à la première vague de SHARE puis ayant eu un entretien de fin de vie en vague 2.

3 Résultats

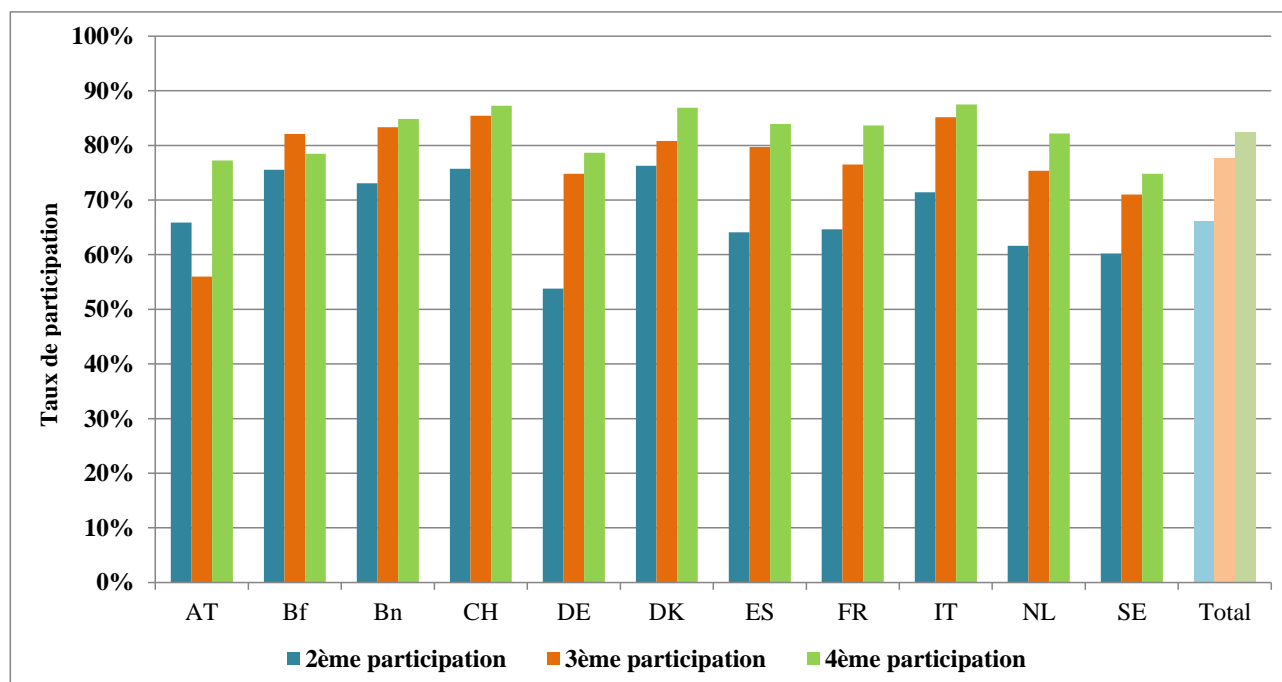
Le taux de participation individuelle augmente en fonction du nombre de vagues de participations consécutives (Figure 2).

Cette augmentation est visible en moyenne et se retrouve pour 9 des 11 pays. En moyenne, 66% des individus sont conservés à la 2^{ème} participation, 78% à la 3^{ème} et 82% à la 4^{ème}. L'écart est de 12 points entre la 2^{ème} et la 3^{ème} participation et passe à 4 points entre la 3^{ème} et la 4^{ème} participation.

L'Autriche voit son taux de participation diminuer de 10 points entre la 2^{ème} et la 3^{ème} participation avant de remonter de 19 points. La Flandre subit une augmentation de son taux de participation de 6 points entre la 2^{ème} et la 3^{ème} interrogation avant de diminuer de 4 points entre la 3^{ème} et la 4^{ème}.

La France présente des résultats très proches de la moyenne avec une participation à la 2^{ème} interrogation de 65%, de 77% à la 3^{ème} et de 84% à la 4^{ème}.

Figure 2: Taux de participation individuelle en fonction du nombre d'interrogations consécutives



L'analyse des taux de participation par nombre d'interrogations consécutives dans le cadre de l'enquête européenne et longitudinale SHARE nous montre que les individus interrogés pour la seconde fois apparaissent plus difficiles à conserver que lors d'une 3^{ème} ou 4^{ème} interrogation consécutives. Une fois l'étape de deux interrogations effectuée, la 3^{ème} et 4^{ème} semblent plus facilement acceptées par les enquêtés. Le passage d'une enquête ponctuelle à une enquête longitudinale semble être une étape importante pour les répondants. Cependant, d'autres facteurs peuvent jouer : type de questionnaire (par exemple la vague 3 de SHARE a été très différente), choix de l'agence de collecte (par exemple un changement d'agence a eu lieu en Autriche entre les vagues 2 et 3), contexte local (par exemple certains pays comme l'Allemagne ne peuvent légalement tenter de réinterroger un individu qui a exprimé son désir de ne plus participer), ou le fait que les répondants aient été informés ou non du caractère longitudinal de l'enquête dès leur première interrogation.

Nous retrouvons cette évolution de participation au fil des vagues pour l'enquête longitudinale annuelle Statistiques sur les Revenus et les Conditions de Vie des ménages (SRCV). Burricand et Lorgnet (2012) ont étudié, sur une base de 16 000 logements français, l'évolution du taux de réponse sur 8 vagues. Après une chute de 10 points, due au passage obligatoire à non obligatoire de l'enquête, le taux de réponse remonte de 4 points entre la 2^{ème} et la 3^{ème} interrogation puis de 1 point entre la 3^{ème} et la 4^{ème} interrogation. Une hypothèse proposée par Burricand et Lorgnet (2012) est que les individus qui restent après cette seconde interrogation sont les individus les plus « motivés et intéressés ».

Régnier-Loilier et Guisse (2011) présentent pour l'enquête Étude des relations familiales et intergénérationnelles (Érifi), basé sur une interrogation tous les trois ans, le même type de résultat sur une base de 10 000 individus et trois interrogations à intervalle de trois ans chacune. Le taux d'attrition est de 35% à la 2^{ème} interrogation pour 17% à la 3^{ème}.⁴

Ce résultat, grâce à l'originalité de l'enquête SHARE, présente cet effet 2^{ème} interrogation comme un phénomène robuste puisque nous le retrouvons pour 9 des 11 pays ayant effectué les 4 vagues de SHARE.

4 Les déterminants de l'attrition

Réduire l'attrition dans un panel est crucial pour la validité des analyses. Par exemple, une attrition de 30% a priori faible conduit à perdre les trois-quarts d'un échantillon au bout de 4 vagues, si aucun individu n'est récupéré après une ou plusieurs absences de participation. Par ailleurs l'attrition n'est en général pas exogène aux phénomènes étudiés, ce qui rend problématique certaines études (Davezies et D'Haultfœuille, 2013). Comment étudier par exemple la mobilité résidentielle si on perd les répondants quand ils déménagent, ou les conséquences du divorce si on ne peut suivre les deux ex-conjoints?

L'attrition touche plus particulièrement les personnes âgées (Bhamra et al., 2008 ; Chatfield et al., 2005) ce qui rend indispensable la compréhension des déterminants de l'attrition dans le cadre d'enquête longitudinale sur ces personnes comme c'est le cas pour l'enquête SHARE.

Nous présentons ici une analyse des déterminants de l'attrition. Tout d'abord nous utilisons des variables de niveau individuel récoltées lors de la vague d'entrée, telles l'âge, le genre, le niveau d'éducation, la santé du répondant (physique et mental), la composition de son ménage, sa profession, son revenu, son patrimoine, son type de logement et sa localisation géographique.

⁴ Voir aussi Radler et Ryff (2010) sur l'enquête américaine MIDUS (Midlife in the United States) et Laurie et al. (1999) sur l'enquête BHPS (British Household Panel Study).

SHARE permet aussi de récolter l'opinion de l'enquêteur sur le déroulement de l'entretien. Nous introduisons aussi des variables indicatrices pour chacun des onze pays considérés. Nous estimons des modèles probit de la forme suivante :

$$Prob(\text{non participation en vague } V / \text{participation en } V-1) = f(X_{VE}, S_{VE}, C, N)$$

X_{VE} : caractéristiques du répondant et de son ménage en vague d'entrée VE

S_{VE} : opinion de l'enquêteur sur l'entretien en vague d'entrée VE

C : Indicatrice du pays

N : nombre de vagues précédentes consécutives de participation à l'enquête, vague d'entrée

Le modèle sera estimé globalement, puis selon les effets trouvés, par vague d'entrée.

Dans cette partie nous ne prendrons pas en compte les 785 individus entrés en vague 3 ($N_{0011} + N_{0012} + N_{0010}$). Cette vague portait en effet sur l'histoire de vie des répondants, ce qui implique que la majorité des questions étaient différentes des autres vagues et que les informations utilisées au niveau de la vague d'entrée étaient manquantes.

Nous en sommes à ce jour à l'étape de constitution des variables, nous n'avons pas encore de résultats disponibles. Cependant des déterminants de l'attrition ont déjà été étudiés sur d'autres enquêtes longitudinales sur des personnes âgées et seront donc à comparer avec nos résultats sur l'enquête SHARE. Chatfield et al. (2005) et Bhamra et al. (2008) présentent ainsi une revue de la littérature des déterminants de l'attrition dans le cadre d'enquêtes longitudinales effectuées auprès de personnes âgées. Mein et al. (2012) examine, quant à lui, l'attrition dans le cadre de l'enquête Whitehall II.

5. Conclusion

Au vu de ces résultats, nous pouvons nous interroger sur un possible effet de nombre de participation *seuil*, dû à une lassitude de la part des répondants par exemple, à partir duquel le taux de réponse chuterait ou stagnerait. Il serait alors intéressant d'étudier l'existence et les caractéristiques de ce seuil quand le nombre de vagues de SHARE le permettra.

Des conclusions concernant l'analyse des déterminants de l'attrition dans l'enquête SHARE seront apportées une fois les résultats obtenus.

Bibliographie

- Burricand, C. et Lorgnet, J-P. (2012), L'attrition dans SRCV : Déterminants et effets de l'attrition sur la mesure des variables, *Journées de Méthodologie statistique*, http://jms.insee.fr/files/documents/2012/866_2-JMS2012_S05-1_BURRICAND-ACTE.PDF
- Bhamra, S., Tinker, A., Mein, G., Ashcroft, R., Askham, J., (2008) The retention of older people in longitudinal studies: A review of the literature, *Quality in Ageing and Older Adults*, 9 (4), 27-35.
- Chatfield, M., Brayne, C. and Matthews, F. (2005) A systematic review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *Journal of Clinical Epidemiology*, 58, 13-19.
- Davezies, L. and D'Haultfoeuille X. (2013), Endogenous attrition in panels, http://www.crest.fr/ckfinder/userfiles/files/Pageperso/xdhaultfoeuille/panel_attrition.pdf
- Eurostat (2014), http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main_tables
- Laurie, H., Smith, R., Scott, L. (1999), Strategies for reducing nonresponse in a longitudinal panel survey, *Journal of official statistics*, 15(2), 269-282.
- Lynn P. (2005) Outcome categories and definitions of response rates for panel surveys and their surveys involving multiple data collection events from the same units. Unpublished manuscript. *University of Essex, Colchester*.
<http://www.nonresponse.org/uploadi/editor/1129102463Outcome%20categories%20same%20units.pdf>
- Mein, G., Johal, S., Grant, R.L., Seale, C., Ashcroft, R., and Tinker, A. (2012) Predictors of two forms of attrition in a longitudinal health study involving ageing participants: an analysis based on the Whitehall II study. *BMC Medical Research Methodology*, 12:164.
- Radler, B.T. and Ryff, C.D. (2010), Who participates ? Longitudinal retention in the MIDUS National Study of Health and well-being, *Journal of aging and health*, 22(3), 307-331.
- Régnier-Loilier, A. et Guisse, N. (2011), La déperdition au fil des trois vagues de l'enquête longitudinale Érfi, http://sondages2012.ensai.fr/wp-content/uploads/2011/01/Article_Regnier_La-d%C3%A9perdition-au-fil-des-trois-vagues.pdf