

La mise en place d'un panel représentatif connecté à Internet mobile : bilan du pilote

ELIPSS

Stéphane Legleye^{1,2,3}, Anne Cornilleau⁴, Anne-Sophie Cousteaux⁴, Nirintsoa Razakamanana¹ et l'équipe ELIPSS

1. Institut national des études démographiques (INED), Paris (France)
2. Inserm, U669, Paris (France)
3. Univ Paris-Sud and Univ Paris Descartes, UMR-S0669, Paris (France)
4. Sciences Po, Centre de données socio-politiques, UMS828Paris (France)

L'équipe ELIPSS est composée de : Kevin Boudelle, AlinaDanciu, Emmanuelle Duwez, , Alexandre Mairot, Geneviève Michaud, Matthieu Olivier, Elodie Pétorin, NirintsoaRazakamana, Jérémy Richard, Patricia Sossa. Gabrielle Bouchet, Carmen Calandra, Adrien Ferreira, Daniele Guido, Adam Schaal ont également contribué à la mise en place du panel ELIPSS. Bernard de Clédats et Patricia Thauvin du service des enquêtes de l'INED participent ponctuellement au projet.Ce travail est collaboratif et reflète la participation de l'ensemble des membres du projet.

Domaine : Échantillonnage et collecte

Résumé

Internet est très utilisé par les sondeurs pour produire des enquêtes rapides et à bas coût. Pour cela, les instituts ont mis en place de larges panels d'internautes volontaires, qui peuvent s'y inscrire par exemple par des bannières postées sur des sites populaires. Ces panels sont donc constitués sans base de sondage ni procédure probabiliste, alors même que plus de 20% des ménages n'ont pas accès à Internet en 2013 [1].

ELIPSS (enquête longitudinale par Internet pour les sciences sociales) est un panel d'individus représentatif de la population des ménages ordinaires tiré aléatoirement dans le recensement, que l'on équipe gratuitement de tablettes tactiles connectées à Internet sans fil, en échange d'une participation mensuelle à 30 minutes de questionnaire. Ce panel fait partie d'un Equipement d'excellence *Données, infrastructures et méthodes en sciences humaines et sociales* (DIME-SHS) financé par l'ANR en 2010. ELIPSS est actuellement dans sa phase pilote (n=1000) et devrait atteindre la taille de 4000 individus fin 2015.

Le recrutement du pilote d'ELIPSS repose sur un échantillon de ménages tirés du recensement et fourni par l'INSEE (n=4500). Il a eu lieu en plusieurs phases (une phase de contact postal suivi d'un recrutement postal, Internet ou téléphonique conduit par l'INED, une phase de contact et de recrutement en face-à-face conduit par un institut de sondage). Dispositif d'enquête longitudinal original ouvert à la recherche, ELIPSS reste exposé à des biais de recrutement ainsi qu'à des risques d'attrition et de professionnalisation de ses membres qui pourraient en altérer la représentativité initiale.

Deux points sont à noter. Premièrement, dans la phase initiale conduite par l'INED, des incitations financières sous la forme de bons cadeaux ont été distribués aléatoirement aux ménages avec la lettre-annonce informant de l'enquête et donc indépendamment de leur participation. Deuxièmement, contrairement aux panels Internet existant en Europe, tous les panélistes reçoivent une tablette tactile et en particulier les individus dont le ménage n'était pas connecté à Internet au moment du recrutement.

Il s'agit ici de présenter rapidement le panel et le recrutement du pilote, la structure sociodémographique de l'échantillon de répondants pour insister sur la présentation de l'évolution de la participation au cours du temps, de la typologie des trajectoires de réponse de ses membres aux enquêtes administrées depuis 2013. Un regard particulier sera également porté au recrutement des personnes initialement non connectées à Internet et à l'utilité des bon-cadeaux.

Mots clefs : ELIPSS ; représentativité ; panel ; Internet.

Abstract

Internet is used by pollsters to produce quickly cheap surveys. For this purpose, survey agencies have built large panels of volunteers that are mostly recruited by banners posted on popular websites. These panels are thus non-probability samples whose representativeness is highly questionable: besides self-selection biases due to the recruitment process on internet, it has to be reminded that more than 20% of the French households do not have internet access.

ELIPSS (longitudinal survey no internet for social sciences) is a representative panel of individuals from ordinary households that is randomly drawn in the general population. The panelists receive a digital tablet connected to mobile internet 3G internet (whether they were equipped and connected to internet or not) to participate each month to a 30 minutes questionnaire. This panel is part of larger project devoted to social sciences that was elected to a national grant in 2010, DIME-SHS. ELIPSS is currently in its pilot phase with 1000 panelists and will reach its definitive size (n=4000) by the end of 2015 with the recruitment of 3000 new panelists using a random face to face procedure.

The recruitment of the pilot is based on a random sample of 4500 households. The fieldwork was divided in several phases. The primary contact was made by mail then phone contacts were made and a survey agency finally conducted a face to face survey to achieve the recruitment. The composition of each selected household was necessary to select randomly a potential panelist (two stage sample). Although ELIPSS is based on a random sample, there is still selection biases, attrition and a risk of professionalization of its members that may alter its representativeness.

Two points are noticeable. First, an experiment was conducted in a large part of the sample: a financial incentive was sent with the announcement letter to a random subsample of households, independently of the future participation of the household. Secondly, contrarily to the other similar panels that exist in Europe, all panelists received a digital tablet, whereas only the non-equipped households received a computer and a connection in the other representative panels using internet.

We present here briefly the pilot panel: the recruitment process, the initial representativeness and its evolution in time. We propose a typology of the trajectories of participation of the panelists based on the first 23 surveys administered since the beginning of the panel in 2013. We will also show the effect of the incentive and of the inclusion of non-equipped households on the recruitment.

Keywords: ELIPSS; panel; representativeness; Internet.

Présentation du projet

Le projet a vu le jour pour permettre aux chercheurs de mener des enquêtes probabilistes sur des thèmes qui ne sont que peu traités par la statistique publique française et d'offrir des facilités pour conduire des enquêtes longitudinales de grande ampleur, ordinairement hors d'atteinte de la recherche académique en sciences humaines et sociales. ELIPSS doit permettre de diminuer les coûts et le temps de collecte et de faciliter l'administration et le suivi des enquêtes.

ELIPSS fait partie d'un Equipement d'excellence Données, infrastructures et méthodes en sciences humaines et sociales (DIME-SHS) financé par l'ANR de 2010 à 2019 (ANR-10-EQPX-19-01). Cet Equipement dirigé par Laurent Lesnard au Centre de données sociopolitiques (CDSP) de Sciences Po comprend également l'Ined (Institut national d'études démographiques), le Groupe des écoles nationales d'économie et statistiques (GENES), l'Université Paris-Descartes (UP-5), Telecom Paris Tech, le département recherche et développement d'Electricité de France (EDF R&D), l'École des hautes études en sciences sociales (EHESS).

ELIPSS est un projet unique en son genre de panel à partir d'un échantillon probabiliste (représentatif) dont les membres sont équipés d'une tablette tactile connectée à Internet, en échange de leur participation à une interrogation mensuelle de 30 minutes au maximum. Ce dispositif est entièrement dédié à la recherche : les enquêtes sont proposées par des équipes de chercheurs des sciences humaines et sociales et sélectionnées par un comité scientifique et technique sur leur qualité méthodologique et scientifique.

Si ELIPSS est unique en son genre en raison de l'utilisation de tablettes connectées, il n'est pas pionnier dans le domaine des panels Internet représentatifs en population générale. En effet, des dispositifs similaires utilisant les ordinateurs personnels au sein des ménages ordinaires existent aux Pays-Bas (LISS Panel : [http://www.lissdata.nl/\[2\]](http://www.lissdata.nl/)) aux USA (Knowledge Panel : <http://www.gfk.com/us/Solutions/consumer-panels/Pages/GfK-KnowledgePanel.aspx>) et en Allemagne (GIP http://reforms.uni-mannheim.de/internet_panel/home[3], GESIS Panel : <http://www.gesis.org/en/services/data-collection/gesis-panel/>). Toutefois, ces panels n'équipent en matériel que les ménages qui en sont dépourvus.

Le pilote ELIPSS

La première phase du panel est sa phase pilote qui s'achèvera fin 2015. Elle avait pour ambition de tester les procédures d'échantillonnage, de contact et de recrutement des panélistes, et définir les procédures de gestion du panel et de production des enquêtes et de développer les outils nécessaires. Il s'agit également de produire des estimations de l'attrition du panel pour préparer le plus précisément possible le cas échéant le passage à un panel à 4000 membres. Echantillonnage, critères d'éligibilité

- Echantillonnage, critères d'éligibilité

Le panel ELIPSS est un panel aléatoire d'individus résidant en ménages ordinaires au sens de l'INSEE. Sont exclus les personnes sans domicile, les mineurs et les plus de 75 ans, les personnes résidant dans des collectivités (prisons, maisons de retraite, résidences étudiantes, etc.), mais aussi celles dans des logements précaires non considérés comme des logements au sens de l'INSEE et celles ne maîtrisant pas suffisamment la langue française pour répondre à des questionnaires auto-administrés.

La base de sondage est constituée des logements recensés en 2011, 2010 et 2009 (ces deux dernières années étant très minoritaires (3% des adresses)). Dans cette base, sont renseignés, pour chaque logement, son adresse, le nom et le prénom de la personne de référence du ménage.

A partir de cette base de sondage, un échantillon de 4500 logements a été tiré par l'INSEE par sondage stratifié sur les critères suivants:

- les zones d'action enquêteur (ZAE).
- les typologies de communes (rural/urbain, taille des communes).

Au sein de chaque ménage sélectionné par l'INSEE, une seule personne devait être sélectionnée aléatoirement. Le sondage est donc stratifié à deux degrés (ménages puis individus). Le poids de sondage est obtenu à partir de la probabilité de sélection du logement (fournie par l'INSEE) et de la probabilité de sélection d'un individu au sein de son logement (fournie au cours du protocole).

L'échantillon initial fourni par l'INSEE a été subdivisé en une partie principale (3500 logements) et deux réserves, à utiliser si l'exploitation de la partie principale ne permettait pas d'atteindre l'objectif initial de 1500 panélistes.

- Le recrutement du pilote

Afin de tester différents modes de contact, le recrutement du pilote a été réalisé selon une procédure séquentielle, du mode le moins coûteux au plus onéreux : courrier postal, puis téléphone et enfin face-à-face. Lors de la première phase de recrutement réalisée par l'équipe ELIPSS entre juin et septembre 2012, la partie principale de l'échantillon INSEE, soit 3500 logements, a été exploitée. Le protocole a reposé sur l'envoi de courriers d'invitation à participer au panel, à destination de la personne de référence des ménages. Deux types de relance ont ensuite été réalisés: relance téléphonique dans le cas où un numéro était connu (45% des cas) et relance postale sinon. Il est à noter qu'au total, sur les 4500 ménages, seules 89 adresses ont été jugées inexploitable.

Il a ensuite été fait appel à l'institut de sondage TNS-Sofres pour la seconde phase du recrutement. Celle-ci a porté sur une large partie des ménages non contactés ou qui avaient refusé de répondre lors de la première phase et sur les échantillons de réserve. Elle a eu lieu pour partie par téléphone, pour partie en face-à-face (ce mode étant obligatoire pour les ménages de l'échantillon principal finalement traités par l'institut). Au total, 3629 ménages étaient concernés dont 2642 adresses de l'échantillon principal.

Au sein de chaque logement, un individu parmi les éligibles était sélectionné par tirage aléatoire simple. Si la personne sélectionnée acceptait de participer, elle devait impérativement signer une convention précisant le cadre du prêt et les conditions d'utilisation de la tablette au sein du panel. A la réception et l'activation de la tablette par le panéliste, une formation téléphonique à l'utilisation de la tablette lui était proposée et fournie gratuitement via l'opérateur de téléphonie mobile.

Résultats globaux

- Recrutement dans le panel

Au terme de l'exploitation complète des 4500 adresses de l'échantillon fourni par l'INSEE, 1039 individus ont accepté de faire partie du panel et ont signé la convention. Cependant, seuls 925 individus ont répondu en avril 2013 à la première partie de l'enquête annuelle dans laquelle il s'agit de collecter les informations socio-démographiques des panélistes. En effet, une partie est entrée dans le panel après l'enquête annuelle, et une autre partie n'a jamais utilisé sa tablette pour répondre aux enquêtes.

Le recours à un sondage à deux degrés (ménages puis individus) implique de fournir des indicateurs de participation au niveau ménage et individu. Suivant les recommandations fournies par l'American Association for Public Opinion Research (AAPOR : www.aapor.org) et notamment celles adaptées aux panels internet par Callegaro[4], il faut distinguer plusieurs issues à chaque niveau : contact, refus, acceptation (dite aussi participation).

Notons que l'éligibilité est parfois délicate voire impossible à déterminer (dans les cas où aucun contact avec des membres du ménage n'a eu lieu) : pour tous ces cas, on applique le taux d'éligibilité observé sur les unités pour lesquelles l'éligibilité est connue.

Pour la première phase, 1349 ménages avaient accepté de répondre au questionnaire ; pour la seconde, 1181 individus ont accepté de participer à ce stade au sein de ces ménages ; finalement, 1039 panélistes furent effectivement recrutés et activés après réception de leur tablette.

Avec ces conventions, le taux de panélistes recrutés et activés est de 27%, le taux de refus de 43% et le taux de non contact de 31%. Le taux de recrutement est à peine plus élevé (30%) pour la partie du panel n'ayant fait l'objet que d'un recrutement en face-à-face (et marginalement au téléphone) par l'institut de sondage.

Il est à noter que 131 panélistes (12.6% du total) ne disposaient pas de connexion Internet à leur foyer avant leur inclusion dans le panel. Ce taux est plus bas que ce qui est constaté dans la dernière enquête de l'INSEE [1], ce qui souligne un biais de sélection en faveur des ménages équipés.

- L'envoi de bons cadeaux durant le recrutement

Les lettres-annonces envoyées à l'échantillon principal contenaient 2000 bons-cadeaux d'une valeur de 10€ attribués au hasard. L'objectif de cette expérimentation était de vérifier si cette incitation financière non conditionnée à la participation pouvait se révéler utile pour augmenter le taux de participation, comme cela a régulièrement été observé dans la littérature scientifique internationale [5].

Les résultats montrent un effet très important au niveau de l'acceptation ménage ($ORa=1,90$ [1,54-2,34]) et individu ($ORa^1=1,41$ [1,19-1,68]) ainsi qu'un effet neutre sur la fidélité comme le montre la figure 2. Une étude détaillée a montré que l'effet des bons-cadeaux est identique suivant l'âge et le diplôme de la personne de référence, le statut d'occupation, le type de logement etc. Cela laisse ouverte la perspective de donner un bon en cas de difficulté ou pour certaines sous-population mal représentées dans le panel de répondants (ménages plutôt vieux, par exemple) lors d'un recrutement futur pour étendre la taille du panel.

- Utilité de recruter des personnes non connectées

L'équipement Internet du ménage préalablement au recrutement dans le panel était demandé lors de la phase de contact ménage, puis au panéliste lors de la première enquête. En combinant les deux informations, nous trouvons que 98 panélistes n'avaient pas accès à Internet avant ELIPSS [6]. La proportion correspondante, 9,4% est inférieure à la proportion réelle de non-équipés dans la population française, qui s'élevait à 14,5% en 2013 d'après l'enquête TIC [1], ce qui souligne un biais de sélection. Ce biais peut être en tout ou partie directement lié à l'équipement et la maîtrise d'Internet, donc au projet ELIPSS, mais aussi refléter un biais de sélection classique des enquêtes en population générale, lié au diplôme, la profession et catégorie sociale, etc. Ces individus sont en effet différents des autres du point de vue du statut d'occupation du logement, du statut d'activité professionnelle, de l'âge, du statut marital et du nombre de personnes dans le foyer, de la taille de l'Unité urbaine de résidence, du diplôme etc. (certaines caractéristiques, comme le diplôme, étant plus proches du RRP pour ces individus que pour le reste de l'échantillon des répondants). Toutefois, leur faible nombre n'a qu'une influence marginale sur la représentativité de l'échantillon de répondants.

- Evolution de la participation aux enquêtes

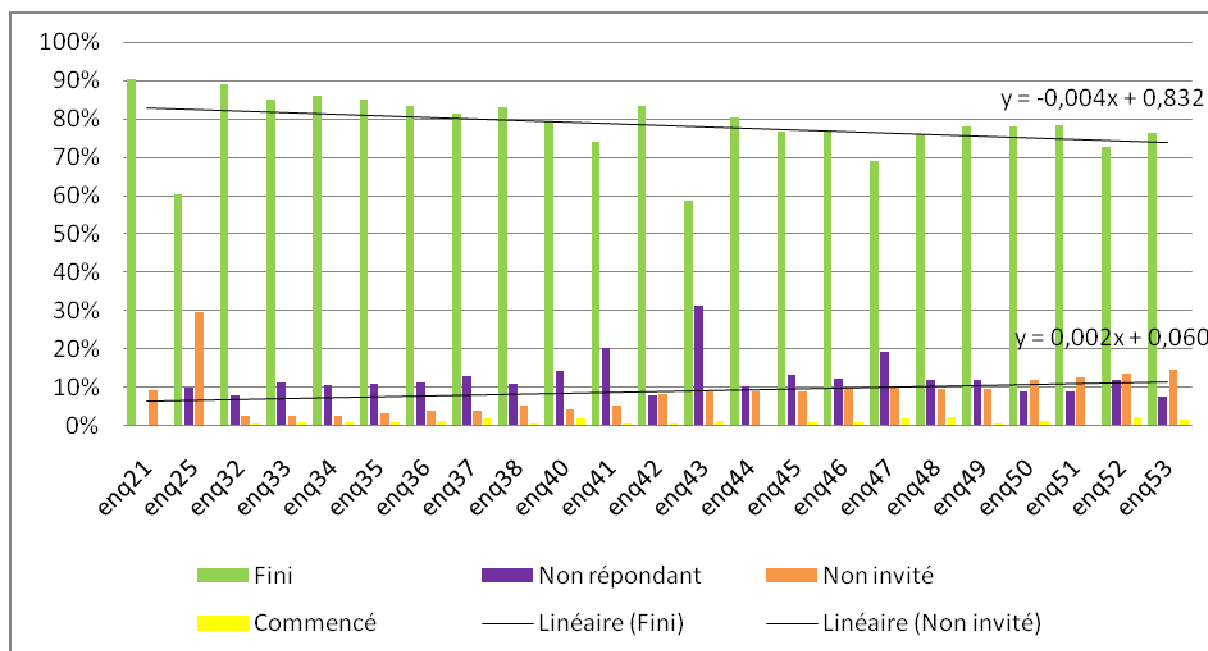
A l'heure actuelle, 26 enquêtes ont été administrées et 23 sont achevées². En moyenne, les 1039 panélistes ont participé à 18,0 enquêtes (moyenne brute du nombre de questionnaires achevés sur l'ensemble des enquêtes). L'évolution de la participation au cours de ces enquêtes achevées est la suivante (figure 1). Elle montre clairement une baisse continue de la participation au fil du temps, corolaire d'une augmentation des non-réponses et des non-invitations (exclusions volontaires ou non, problèmes techniques ou indisponibilités).

Les panélistes sont en principe invités à répondre à chaque enquête mais ils peuvent avoir un problème de tablette (dysfonctionnement, désactivation de la part des gestionnaires de panel) ou bien être indisponibles (vacances, maladie, etc.). Ils peuvent également ne pas répondre malgré les relances. La désactivation définitive et donc la sortie du panel intervient après des non réponses immotivées trop fréquentes (généralement quatre successives).

L'attrition est donc relativement modérée, puisqu'elle est d'environ 15 points en 23 enquêtes successives administrées sur près de deux ans, soit 141 sorties du panel. Un peu plus de la moitié (54%) de ces sorties sont des exclusions pour non-réponses successives, les autres sont des sorties volontaires ou des décès. Au total, 925 individus ont répondu à l'enquête annuelle de 2013 (enquête n°32) et 793 ont participé à la vingt-troisième (enquête n°53).

¹ Dans les deux analyses, les variables d'ajustement sont : âge du chef de ménage, diplôme du chef de ménage, taille d'unité urbaine, ZEAT, commune en Zone urbaine sensible (ZUS), nombre de pièces du logement, HLM
² 21 : Première enquête ELIPSS ; 25 : Votre quartier / enjeux politiques ; 32 : Vos conditions de vie ; 33 : Pratiques, croyances et opinions ; 34 : Loisirs culturels, médias et Internet ; 35 : Fécondité, contraception, sexualité ; 36 : Les grands sujets de société et la vie citoyenne ; 37 : Environnement & Énergie ; 38 : Le couple, les enfants et la famille ; 40 : Enquête Santé, Travail et Environnement ; 41 : Les relations entre générations ; 42 : Vos conditions de vie en 2014 ; 43 : Les élections municipales ; 44 : Questions générales sur la société française ; 45 : Après les élections municipales ; 46 : Quelques questions sur l'Europe ; 47 : Les élections européennes ; 48 : Après les élections européennes ; 49 : Votre opinion sur les familles ; 50 : Le jeu des métiers ; 51 : Vos goûts et vos préférences ; 52 : Le jeu des portraits ; 53 : Modes de vie et société.

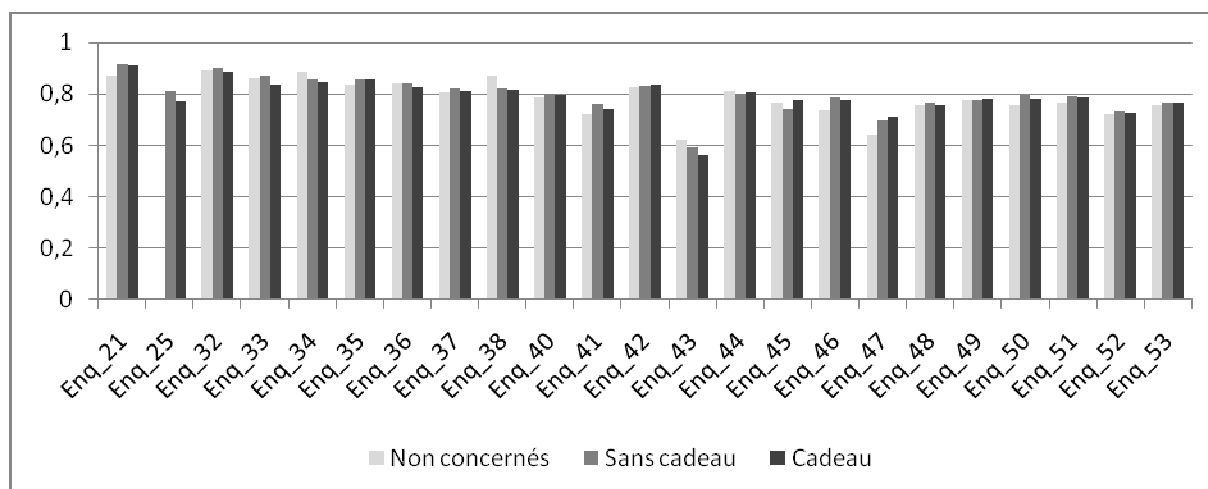
Figure 1 : Evolution de la participation dans ELIPSS.



Note : les statuts enregistrés à chaque enquête sont : questionnaire complet (fini), commencé non fini (commencé), non invité (exclu ou non), non répondant.

Nous avons vérifié que l'envoi d'un bon cadeau ne modifiait pas la participation aux enquêtes. Parmi les 1039 panélistes, 797 avaient été concernés par l'envoi potentiel d'un bon cadeau (509 en ont reçu un, 288 non) et 242 n'étaient pas concernés par cette expérimentation. L'envoi des bon-cadeaux n'a eu aucun rôle sur la participation aux enquêtes successives comme le montre la Figure 2 présentant le taux de questionnaires achevés (finis) suivant l'envoi ou non d'un bon cadeau (si l'on excepte l'enquête 25, enquête proposée aux seuls panélistes de l'échantillon principal ayant pu bénéficier d'un bon cadeau). Au total, les 509 récipiendaires d'un bon cadeau ont participé à 18,2 enquêtes sur les 23 étudiées ici, contre 18,3 pour les 288 non récipiendaires et 17,3 pour les 242 non concernés (les chiffres correspondants sont 17,4, 17,5 et 17,3 si l'enquête 25 est exclue). Aucun de ces écarts n'est significatif.

Figure 2 : Participation aux enquêtes suivant l'envoi préalable d'un bon cadeau (%) [champ =ensemble des 1039 panélistes, dont 509 récipiendaires d'un bon cadeau, 288 non récipiendaires et 242 non concernés]

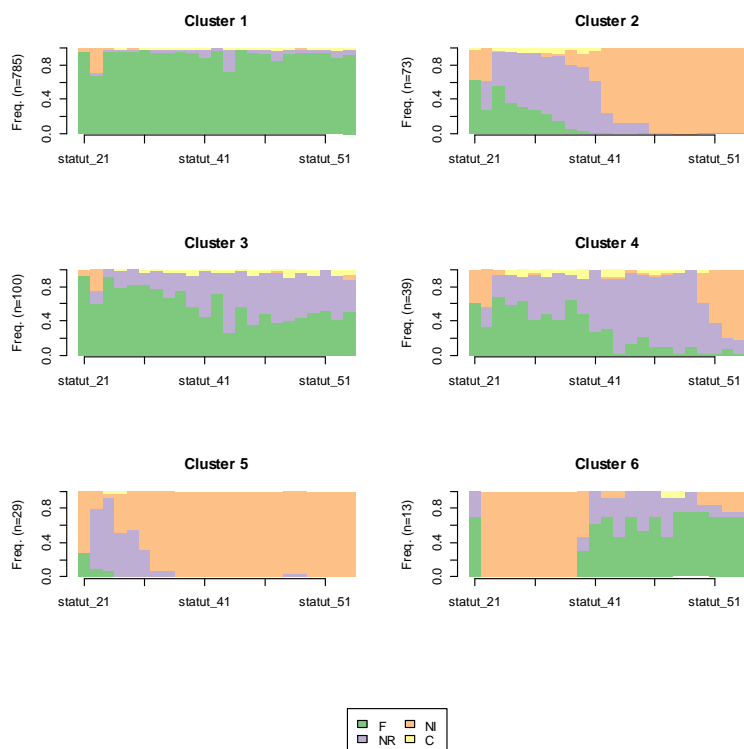


- Différents types de participants

Toutefois, cette évolution globale masque des comportements très différents au niveau individuel. Ainsi, en utilisant les données individuelles de participation à chacune de ces neuf enquêtes, et en procédant à une analyse de séquences (Optimal Matching : [7]) suivie d'une classification ascendante hiérarchique, nous pouvons observer 6 profils de panélistes distincts (figure 2). Ce travail est encore préliminaire (des analyses de robustesse doivent être menées). Le principal groupe est constitué de 785 panélistes (76%) qui ont participé à toutes les enquêtes. Le second est constitué de 73 individus (7%), dont la participation a baissé très rapidement. Le troisième groupe de 100 individus (10%) comprend des personnes qui participent moyennement mais qui ne sont pas exclus, malgré une tendance nette à la baisse. Le quatrième (n=39, 4%) comprend des personnes dont la participation baisse continuellement. Le groupe 5 (n=29, 3%) est celui des exclus précoces tandis que le groupe 6 (n=13, 1%) est celui des recrutés tardifs, qui participent assez fortement aux enquêtes depuis leur arrivée. Globalement, les groupes 1, 3 et 6 comprennent 86% d'individus participant systématiquement ou très souvent aux enquêtes.

On peut qualifier les panélistes du groupe 1 et les ménages dont ils sont issus. Ce sont des hommes à 52.6% (contre 48.4% dans le reste des panélistes), de nationalité française (92.0% vs 84.6%, $p < 0.001$), diplômés du Bac (64.3% vs 53.9%, $p = 0.003$), mariés (48.6% vs 26.9%, $p < 0.001$), résidant moins souvent en HLM (13.1% vs 24.0%, $p < 0.001$), en ZUS (4.2% vs 9.8%), et habitant plus souvent un logement d'au moins 4 pièces (68.3% vs 58.7%, $p = 0.005$), ou d'au moins 5 pièces (45.5% vs 31.1%, $p < 0.001$), sans différence entre le nombre de personnes ou d'éligibles résidant dans les ménages (2.60 vs 2.63, ns ; 1.83 vs 1.84, ns). Signalons que ces individus fidèles n'ont pas plus que les autres été destinataires d'un bon cadeau (parmi les ménages concernés).

Figure 3 : Typologie de la participation dans ELIPSS



F=fini (questionnaire complet) ; C=commencé non validé
 NI=non invité (exclu temporaire ou définitif) ; NR=Non réponse

- Qualité de l'échantillon de répondants à la première enquête annuelle de 2013

Sur le plan des variables individuelles, l'échantillon apparaît assez peu déformé pour ce qui est de l'enquête d'inclusion, traitant des conditions de vie, comme le montre le tableau 1. Les populations suivantes apparaissent sous-représentées : les 18/24 ans, les 65/75 ans, les personnes nées à l'étranger, les personnes titulaires d'un diplôme inférieur au baccalauréat, ou encore les personnes résidant dans la région parisienne, bassin parisien, le Nord et l'Est. Les populations les plus surreprésentées sont celles des français de naissance, des très diplômés, et les personnes résidant plutôt dans les régions Centre, Sud et Ouest. On constate donc un léger biais de sélection en faveur des personnes les plus aisées, biais classique dans les enquêtes en population générale.

Tableau 1 : structure de l'échantillon de répondants ELIPSS

Variable	Modalité	INSEE RRP ¹	ELIPSS répondants enquête 1	
		18-75 ans	Non Pondéré	Pondéré ²
		%		
Sexe	Hommes	48.6	48.0	48.6
	Femmes	51.4	52.0 ns	51.4 ns
age	18-24 ans	12.0	8.2	12.0
	25-34 ans	17.7	19.1	17.7
	35-44 ans	20.0	25.7	20.0
	45-54 ans	19.6	21.8	19.6
	55-64 ans	18.1	16.5	18.1
	65-75 ans	12.6	8.8 ***	12.6 ns
Taille ménage	1	16.7	25.9	19.5
	2	34.9	25.8	25.9
	3	20.1	17.9	22.3
	4	18.0	20.6	23.7
	5+	10.4	9.9 ***	8.7 ***
Statut marital	Marié/en couple	51.1	44.9	49.6
	Seul	36.8	40.1	38.2
	Autres	12.1	15.1 ***	12.2 ns
Nationalité	Français de naissance	88.2	90.9	88.2
	Français par acquisition	5.4	5.2	5.4
	Etranger	6.4	3.9	6.4 ns
Statut d'activité professionnelle	Occupe un emploi	59.7	64.9 ***	55.8
	Etudiant(e) ou en stage	4.6	6.0	8.7
	Chômeur	7.8	7.5	8.3
	Retraité(e)	20.1	15.2	19.6
	Autre situation	8.0	6.5 ***	7.6 ***
Statut d'occupation du logement	Propriétaire (ou copropriétaire)	60.5	58.5	59.0
	Locataire ou Sous-locataire	37.2	35.3	32.6
	Occupant à titre gratuit	2.3	4.8	6.5
	NSP ou NVPR	---	1.4 ***	1.8 ***
Diplôme	Aucun/CEP/BEPC	28.4	17.6	28.4
	CAP/BEP	24.9	20.9	24.9
	Bac_bac+2	32.3	36.5	32.3
	Bac+3 et plus	14.4	24.9 ***	14.4 ns

¹ : Recensement rénové de la population

² : La stratégie de redressement est présentée dans le paragraphe éponyme. *, **, ***, ns : valeur-test du test du K_{hi}² de Pearson respectivement <0.05, <0.01, <0.001 et non significative.

- Attrition : sorties volontaires et exclusions

A la fin de la 23^{ème} enquête, soit un peu plus de deux ans, 159 individus sont définitivement sortis du panel : 6 pour décès, 89 pour exclusion, 3 pour perte de la tablette ou impossibilité de joindre le panéliste, 61 pour une sortie volontaire. Soit un taux d'attrition de 15%, 56% étant dus à des exclusions suite à non-respect de la convention.

- Qualité de l'échantillon de répondants au cours du temps

Il est possible de mesurer l'évolution de la représentativité tout au long des enquêtes. Qu'on l'examine à partir des données de la base de sondage (données du RRP, figure 3) ou bien des données individuelles déclarées par les panélistes (figure 4), les deux graphiques montrent une grande stabilité au cours du temps une fois passé le recrutement c'est-à-dire les étapes « acceptation ménage » (« men ») puis « acceptation individu » « ind » (activation du panéliste et de sa tablette). Seul le diplôme se dégrade un peu, les moins diplômés étant plus facilement non répondants ou exclus. Au total, la représentativité se dégrade donc légèrement, surtout pour le diplôme : le biais de sélection en faveur de ménages plutôt favorisés s'accroît donc légèrement au fil des enquêtes (figure 6). Toutefois, les graphiques 5 et 6 soulignent une relative stabilité des indicateurs de représentativité depuis l'enquête 48 (voir les encadrés sur les graphiques).

Figure 4 : représentativité du panel à partir des données logement (RRP) : données pondérées par le poids de sondage, non redressées

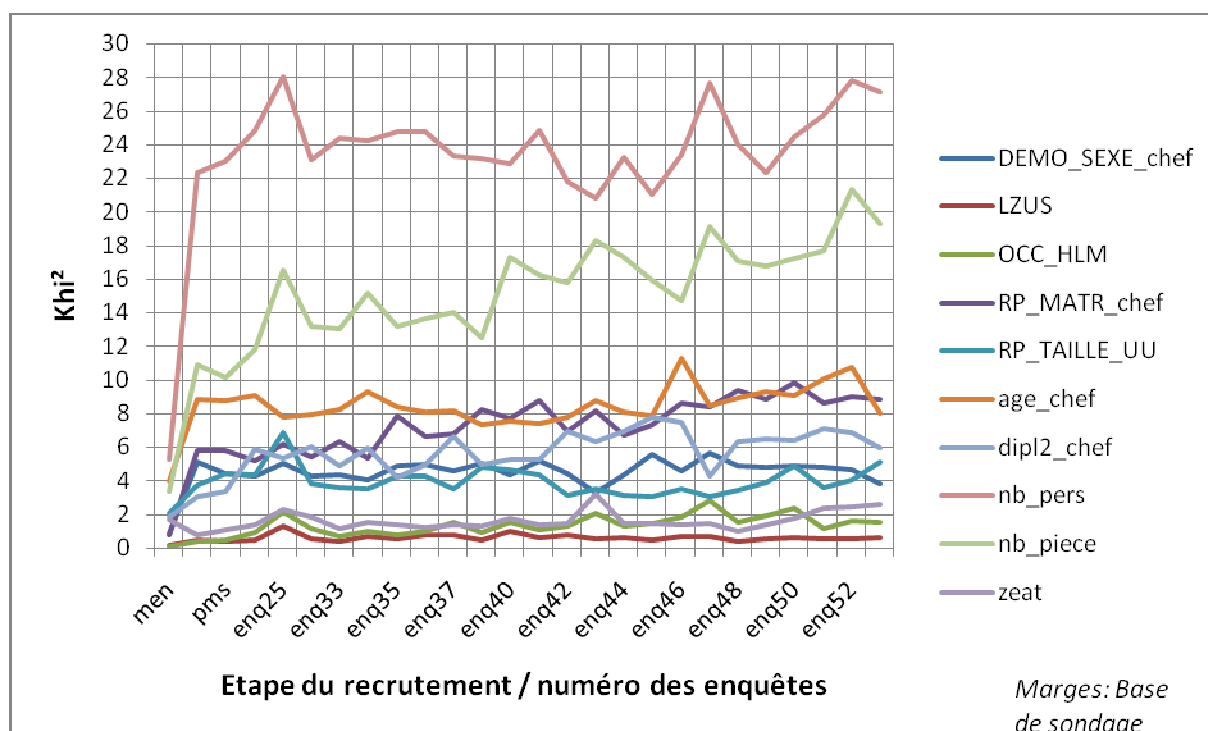


Figure 5 : représentativité du panel à partir des données individuelles des panélistes : données pondérées par le poids de sondage, non redressées

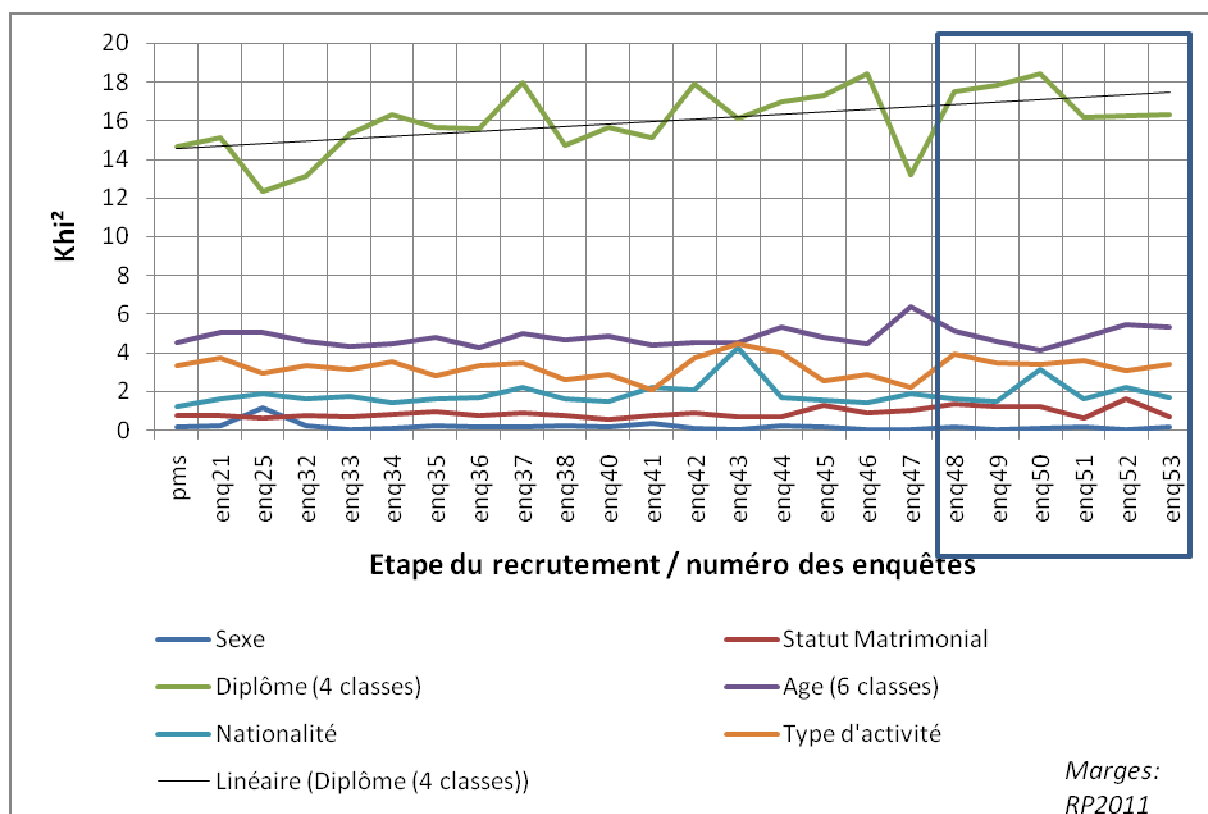
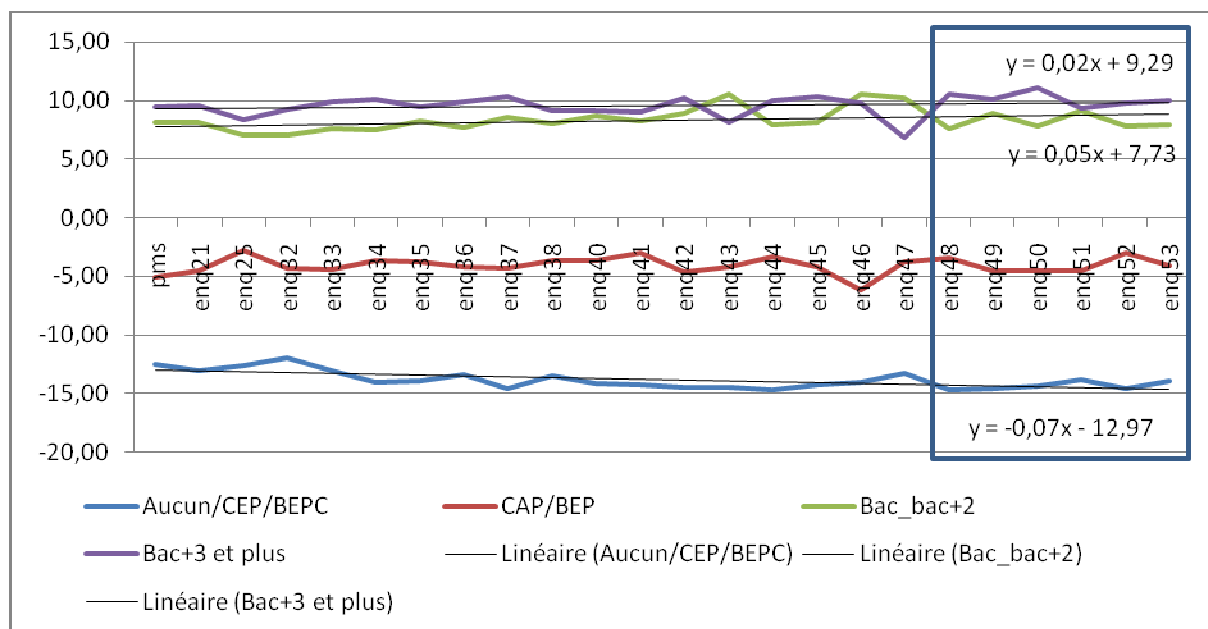


Figure 6 : Evolution de la représentativité de la variable diplôme (%)



o Redressement

Nous avons pris le parti, dans un premier temps, de ne proposer que des pondérations transversales pour chaque enquête, afin de retenir une méthode simple et reproductible. La méthode repose sur une correction de la non-réponse totale puis un calage. Les données auxiliaires ELIPSS sont très nombreuses : on peut compter les

variables de la base de sondage du recensement et les parodonnées (présence d'un bon cadeau, mode de recrutement, nombre de tentatives de contact). C'est toutefois au niveau ménage (n=1349) que la distorsion de l'échantillon est la plus grande comme le montre le graphique 4C'est en effet lors de cette phase exigeant un retour actif du chef de ménage que la plus grande partie des refus et non-réponse sont enregistrées. S'il est possible de modéliser chaque phase du recrutement, depuis l'acceptation ménage, l'acceptation individu (n=1182) puis l'activation (n=1039) et enfin la participation à chaque enquête. Après plusieurs tentatives et examen de la dispersion des poids finaux, il a donc été décidé de corriger de la non-réponse au niveau ménage puis de procéder à un calage. Les variables utilisées pour la correction de la non-réponse sont au nombre de 39. Les plus influentes sont : le mode de recrutement, la présence d'une personne de plus de 60 ans au sein du foyer, la taille de l'unité urbaine, le nombre d'actifs au sein du ménage, l'ancienneté de la construction du logement, le nombre de pièces, le nombre de diplômés du supérieur dans le ménage, le nombre de relances téléphoniques durant la collecte, le revenu fiscal moyen de la commune. Toutes ces variables sont issues de la base de sondage tirée du RRP, sauf deux parodonnées (nombre de relances téléphoniques et mode de recrutement). L'âge du chef de ménage et la ZEAT n'ont pas été introduites car elles déformaient trop la structure des poids avant calage. Cette modélisation a été effectuée par la méthode du boosting sous le package R *gbm*, qui est une méthode itérative permettant de sélectionner au mieux les variables prédictives dans un modèle en s'affranchissant des problèmes de multicollinéarité. A l'issue de celle-ci des groupes homogènes de réponse ont été constitués par classification automatique, au nombre de 9, afin d'assurer un nombre suffisant de participants dans chaque groupe (n=45 minimum).

Pour la phase deux du redressement, un calage sur marge a été effectué pour la participation à chaque enquête. La population cible est celle du recensement de la population, dont est issu l'échantillon initial fourni par l'Insee. Le choix des variables de calage a été guidé par la taille restreinte de l'échantillon et le contexte du panel ELIPSS. L'âge, le sexe, le niveau de diplôme et la nationalité sont des choix classiques auxquelles la ZEAT a été ajoutée. Les marges ont été calculées à partir du recensement de la population 2011. Ces pondérations sont livrées avec les données de chaque enquête et détaillées dans la documentation afférente.

La pondération longitudinale est pour l'instant proposée pour une seule enquête, en concertation avec l'équipe de recherche en fonction de leurs besoins d'analyse.

Discussion et perspectives

Le panel ELIPSS est innovant à plus d'un titre : par sa nature représentative, son caractère d'infrastructure ouverte dédiée à la recherche en sciences humaines et sociales, et l'usage de l'Internet mobile. Sur ce dernier point, il se distingue de tous les panels représentatifs connectés existant à ce jour.

Le pilote a également montré que le recrutement séquentiel en plusieurs modes (courrier, téléphone et face-à-face) était complexe à gérer, nécessitant de nombreux échanges avec la société de sondage et la gestion de nombreux fichiers. Le prochain recrutement de 3000 nouveaux membres pour atteindre la taille finale de 4000 panélistes nécessitera un effort considérable. Conformément à ce qui se dégage de l'analyse des résultats des différentes étapes du recrutement du pilote, celui du futur panel sera confié à un institut de sondage privé pour être effectué en face-à-face lorsque cela est possible, afin de maximiser les contacts et la participation tout en accélérant le processus de recrutement (moins d'échanges de courrier, démonstration de la tablette et signature immédiate de la convention, etc.)

Un soin particulier est actuellement accordé à la définition des stratégies de relance et continueront d'être affinées dans le panel final. Celles-ci pourront s'appuyer sur l'analyse des parcours de participation des enquêtés qui sera complétée et enrichie au fur et à mesure.

Le panel pilote était pour l'instant réservée aux chercheurs des institutions membres de DIME-SHS. L'ouverture du panel en 2015 aux équipes de recherche externes nécessitera également la définition d'une politique tarifaire spécifique qui reste à imaginer, afin de parvenir à l'autonomie financière, comme le stipule le cahier des charges des équipements d'excellence.

Références

1. Gombault, V., *L'internet de plus en plus prisé, l'internaute de plus en plus mobile*. Insee Première, 2013(1452): p. 1-4.
2. Scherpenzeel, A.C. and M. Das, "True" Longitudinal and Probability-Based Internet Panels: Evidence From the Netherlands., in *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, M. Das, P. Ester, and L. Kaczmirek, Editors. 2011, Routledge: London. p. 77-104.
3. Blom, A.G., C. Gathmann, and U. Krieger, *Setting Up an Online Panel Representative of the General Population: The German Internet Panel*. Field Methods, 2013. **27**(4).
4. Callegaro, M. and C. Disogra, *Computing response metrics for online panels*. Public Opinion Quarterly, 2008. **72**(5): p. 1008-1032.
5. Edwards, P., et al., *Methods to increase response to postal and electronic questionnaires*. Cochrane Database of Systematic Reviews, 2009(3).
6. Revilla, M., et al., *What is the gain in a probability-based online panel to provide Internet access to sampling units that did not have access before?* Social Science Computer Review, 2014. **Submitted**.
7. Gabadinho, A., et al., *Analyzing and Visualizing State Sequences in R with TraMineR*. Journal of Statistical Software, 2011. **40**(4): p. 1–37.