

ÉCHANTILLONNER DES ENQUÊTES MULTIMODE DANS FIDÉLI : RETOURS SUR L'EXPÉRIENCE DE L'ENQUÊTE FAMILLES ET EMPLOYEURS

Géraldine Charrance¹, Amandine Stephan² et Thomas Merly-Alpa³

¹ *Ined, France, geraldine.charrance@ined.fr*

² *Ined, France, amandine.stephan@ined.fr*

³ *Ined, France, thomas.merly-alpa@ined.fr*

Résumé. Après une première communication sur les coordonnées de contact présents dans Fidéli lors de la précédente édition du colloque francophone sur les sondages, nous vous proposons, cette fois-ci, de dresser le bilan de la mobilisation de ces coordonnées dans la conduite de l'enquête Familles et Employeurs, enquête de grande ampleur recourant à plusieurs protocoles : un protocole multimode séquentiel CATI-CAWI et un protocole monomode CAWI. Nous ferons, dans un premier temps, un état des lieux des informations issues de la base de sondage pour les personnes échantillonnées, puis nous reviendrons sur les procédures de nettoyage et d'enrichissement réalisées et les adaptations de protocoles mises en place pour tenir compte de la diversité des configurations. Enfin nous tenterons d'apprécier l'impact des coordonnées de contact et du protocole de relance mis en place sur la participation selon les différents protocoles.

Mots-clés. Base de sondage, enquête multimodes, protocoles, modes de collecte, coordonnées de contact, support de communication

Abstract. Following an initial presentation on the contact information in Fidéli at the previous edition of this conference, we now invite you to study the mobilization of these information during the data collection of the survey Familles et Employeurs, a large-scale survey using several protocols: a CATI-CATI sequential multimode protocol and a CAWI single-mode protocol. We will begin with an overview of the information available from the sampling frame for the people sampled, then look back at the cleaning and enrichment procedures carried out and the protocol adaptations implemented to take account of the diversity of configurations. Finally, we will attempt to assess the impact of contact information and the follow-up protocol used on participation according to the different protocols.

Keywords. Sampling frame, Mixed modes surveys, contact information, data collection modes,

1 Le point de départ d'une enquête : la base de sondage et ses qualités

La recherche d'une base de sondage adaptée constitue la première étape nécessaire à la réalisation d'une enquête. Une bonne base de sondage possède les propriétés suivantes : exhaustivité, sans double-compte, actualisée et contenant de l'information auxiliaire (caractéristiques des individus pouvant être mobilisées lors du tirage, pour la collecte et pour les redressements). Lorsqu'on parle d'informations auxiliaires pouvant être mobilisées pour la collecte, il est essentiellement question de coordonnées de contact. Pour des enquêtes en face-à-face, les enquêteurs devront disposer *a minima* de l'adresse de la personne à enquêter. Pour des enquêtes téléphoniques, il sera nécessaire de disposer d'un numéro de téléphone. Il est donc indispensable d'évaluer régulièrement les bases de sondages (ou les méthodes de constitution des échantillons) afin de juger de leur adéquation au contexte dans lesquelles elles sont mobilisées.

Aujourd'hui et depuis plusieurs décennies, les taux de réponse sont en baisse, et les enquêtes téléphoniques n'échappent pas à la règle. Parmi les causes de cette évolution, on note les sollicitations commerciales croissantes auxquelles les personnes sont confrontées, qui engendrent des stratégies de filtrage. Il semble nécessaire, pour pallier cette difficulté, de pouvoir informer les personnes de l'enquête en amont des appels, par un courrier, un SMS et/ou un mail.

Face à ces constats, Fidéli est apparu comme une source adaptée pour le tirage de l'échantillon de l'enquête Familles et Employeurs (FamEmp), enquête de grande ampleur du projet LifeObs¹. Dans cette présentation, nous ferons dans un premier temps un état des lieux des informations issues de la base de sondage et de l'enrichissement téléphoniques pour les personnes échantillonnées, puis nous décrirons le protocole d'animation de la collecte mobilisé pour l'enquête et les choix opérés, et enfin nous tenterons d'apprécier l'impact des informations de contact à disposition sur la participation selon les différents protocoles.

2 L'enquête Famille et Employeurs, tirée dans Fidéli et reposant sur plusieurs protocoles de collecte

2.1 Thème, champ et protocole(s) de l'enquête Familles et Employeurs

L'enquête Famille et Employeurs est mise en place par une équipe de chercheurs de l'Ined en collaboration avec le service des enquêtes et des sondages de l'institut. Cette enquête, dont la première édition a eu lieu en 2004-2005, vise à faire un état des lieux des questions autour de l'articulation entre vie familiale et vie professionnelle tant du point de vue individuel que de celui des employeurs. Cette seconde édition de l'enquête, permettra de mesurer l'impact des changements économiques, sociaux, démographiques et politiques profonds intervenus depuis le début des années 2000 sur l'articulation des différents temps de vie. La dimension panel, permettra de prendre la mesure de ces équilibres et de leurs évolutions jusqu'en 2030.

¹ LifeObs, l'Observatoire français des parcours de vie, est une infrastructure de recherche qui vise à développer un programme d'enquêtes longitudinales et innovantes sur les comportements familiaux, à accroître la diffusion des données, et à former les utilisateurs. Cet observatoire multi site repose sur une coopération nationale entre des institutions clés dans le domaine des études sur les parcours de vie, la famille et la population : l'Ined (porteur du projet), les universités Paris Dauphine – PSL, de Bordeaux et de Strasbourg, l'Insee et la très grande infrastructure de recherche Progedo. Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du plan France 2030 portant la référence ANR-21-ESRE-0037.

L'objectif est d'interroger 30 350 femmes et hommes résidant en France métropolitaine, âgés de 20 à 65 ans au 1^{er} janvier 2024 afin d'explorer les dynamiques de conciliation au sein de sous-populations spécifiques (familles monoparentales, horaires atypiques, etc.) et d'anticiper l'attrition entre les vagues d'enquête. Pour atteindre cet objectif et pouvoir réaliser des analyses méthodologiques, un protocole multimode séquentiel CATI-CAWI (échantillon Multimode) et un protocole monomode CAWI (échantillon Méthodologique) ont été mis en place.

Tableau 1 : Taille des échantillons et nombre de répondants attendus

	Échantillon Fidéli	Nombre de répondants attendus
Échantillon Multimode	64 000	25 000 (20 000 CATI, 5 000 CAWI)
Échantillon Méthodologique	32 000	5 350
Total	96 000	30 350

Le tirage des échantillons a été réalisé dans la seconde version de la base Fidéli 2021. Les informations sur les personnes échantillonnées ont été actualisées en juin 2023 après le chargement du nouveau millésime des fichiers Fidéli dans l'application Nautile (Fidéli 2022), afin d'actualiser les coordonnées de contact des individus échantillonnés et d'identifier les personnes sortant du champ de l'enquête (déménagement hors France Métropolitaine, correction de l'année de naissance, personnes décédées).

Pour réaliser le tirage, la base de sondage a été partitionnée en 12 strates :

- la strate AEEH est composée des individus du champ de l'enquête repéré dans les données Filosofi 2020 comme percevant l'AEEH (Allocation d'éducation de l'enfant handicapé) ;
- les 11 strates D1 à D10 et D_NR regroupent les individus du champ de l'enquête selon leur décile de niveau de vie. Les individus pour lesquels le niveau de vie n'est pas renseigné dans Fidéli sont affectés à la strate D_NR.

2.2 État des lieux des informations issues de la base de sondage et de la procédure d'enrichissement

Grâce à Fidéli, on dispose, pour les personnes échantillonnées d'informations précieuses pour la collecte. Tout d'abord, on identifie clairement les personnes échantillonnées grâce à la présence de leur nom, prénom et date de naissance. Toutefois, pour certaines personnes (2,4% de l'échantillon FamEmp), on ne dispose que de l'année de naissance, les nom et prénom ne sont pas renseignés. Ce sont généralement des enfants à charge, d'où leur concentration parmi les plus jeunes (90% des personnes à l'identité inconnue appartiennent à la classe d'âge 18-24 ans). Dans ce cas, on dispose tout de même de l'identité du référent principal.

Au-delà des traits d'identité, pour chaque personne échantillonnée, on peut disposer :

- de l'adresse postale (systématique) ;
- des numéros de téléphone (fixe et mobile) de la personne échantillonnée ;
- de l'adresse mail de la personne échantillonnée ;
- des numéros de téléphone des deux référents du logement (référent principal et conjoint) ;
- de l'adresse mail de chacun des deux référents.

Fidéli, permet donc de disposer de nombreuses coordonnées de contact. Cependant, dans les faits, tous les champs ne sont pas renseignés ou certaines coordonnées peuvent être en doublon (même numéro renseigné dans plusieurs champs par exemple). Afin d'optimiser l'information avant la collecte, une étape de normalisation des champs destinés aux coordonnées de contact et une étape d'enrichissement sont réalisées par le prestataire en charge de la collecte.

L'étape de normalisation repose sur deux opérations : un traitement « Restructuration Normalisation Validation Postale » (RNVP) afin d'effectuer des contrôles et corrections sur les adresses postales présentes dans le fichier et une homogénéisation des formats des numéros de téléphone et des adresses mails, avec dédoublonnages vertical (entre les personnes de l'échantillon) et horizontal (dédoublonnage des numéros pour une même personne).

Une fois les différents champs dédiés aux coordonnées de contact normalisés, des opérations d'enrichissement sont réalisées à partir de sources externes :

- Une recherche des déménagés a été effectuée afin d'obtenir une potentielle nouvelle adresse postale (nouvelle adresse en métropole pour un peu moins de 5% de l'échantillon et repérage d'un déménagement hors du champ géographique pour 0,17% de l'échantillon).
- Un enrichissement téléphonique a été réalisé sur les individus soumis à une exploitation CATI (échantillon Multimode). Cette opération a permis d'augmenter la part de fiches avec coordonnées téléphoniques de 5,4 points, on est passé de 85,1% à 90,5%. L'enrichissement permet surtout de récupérer des numéros de fixes, ils représentent 90% des nouveaux numéros.

Parmi les individus soumis à une exploitation Multimode, on dispose de coordonnées téléphoniques pour 90,5% et coordonnées téléphoniques propres à la cible pour 76,5% d'entre eux.. On compte en moyenne 1,9 numéro par fiche sur l'échantillon Multimode (après dédoublonnage).

Pour l'échantillon Méthodologique, c'est surtout la présence de numéros de mobile qui nous intéresse puisque les personnes ne seront pas appelées mais seront éventuellement averties et relancées pour participer à l'enquête par l'envoi de SMS. On dispose d'au moins un numéro de mobile pour 80,9% de l'échantillon et du mobile de la cible pour 60,9% de l'échantillon.

Tableau 2 : Prévalences des coordonnées téléphoniques par échantillon et au total

	Multimode	Méthodologique	Total
Fidéli et enrichissement			
N° de tél.	90,5	85,2	88,4
N° de tél. cible	76,5	64,3	28,3
N° de mobile	82,2	80,9	81,7
N° de mobile de la cible	63,7	60,9	62,6
Fidéli			
N° de téléphone	85,1	85,2	85,1
N° de tél. de la cible	64,8	64,3	64,6
N° de mobile	80,7	80,9	80,8

N° de mobile de la cible	61,3	60,9	61,1
Enrichissement			
N° de tél. issu(s) de l'enrichissement (diff. des n° Fidéli)	30,6	-	18,4
Fixe issu de l'enrichissement (diff. Fidéli)	27,8	-	16,7
Mobile issu de l'enrichissement (diff. Fidéli)	4,3	-	2,6

La part de fiches comptant au moins une adresse mail est légèrement inférieure (84,6%), et celles des fiches avec l'adresse mail de la cible est de 72,2%. Pour 12,4% de l'échantillon, on ne dispose donc que de l'adresse mail d'un tiers. Parmi les fiches avec une adresse mail pour la cible, dans 80% des cas, l'adresse mail est unique, c'est-à-dire non répliqué dans le champ dédié à une autre personne du ménage. Pour les 20% restants, une même adresse mail est indiquée pour plusieurs personnes, ce qui laisse planer le doute sur le fait que ce soit l'adresse mail personnelle de la cible.

Tableau 3 : Prévalences des mails par échantillon et au total

	Échantillon Multimode	Échantillon Méthodologique	Échantillon FamEmp
Mail(s) issu(s) de Fidéli	84,7	84,4	84,6
Mail de la cible issu de Fidéli	72,4	71,9	72,2
Mail de la cible unique	58,1	58	58,1

À noter, qu'il est extrêmement rare de disposer des coordonnées personnelles des personnes dont l'identité nous est inconnue : on dispose d'un numéro de téléphone ou d'une adresse mail renseigné dans un des champs dédiés à la cible dans seulement 2% des cas. Toutefois, pour 78,8% d'entre eux, on dispose d'au moins un numéro de téléphone appartenant à l'un des référents et pour 69,3% d'une adresse mail. Il est donc possible de mobiliser ces canaux de communication, afin de les « atteindre » par l'intermédiaire d'un tiers.

3 Un protocole d'animation de la collecte s'appuyant sur la richesse des coordonnées disponibles

3.1 Protocole d'animation de la collecte de l'enquête FamEmp

Le protocole de contact (annonce de l'enquête et animation de la collecte par des relances) a été conçu pour mettre à profit la diversité et la richesse des coordonnées de contact issues de Fidéli et de l'enrichissement. Pour être plus précis, ce sont trois protocoles de contact qui ont été conçus : un pour l'échantillon Méthodologique et deux pour l'échantillon Multimode. En effet, malgré l'enrichissement téléphonique, nous ne disposons d'aucune coordonnée téléphonique pour environ 10% de l'échantillon Multimode. Les individus faisant partie de l'échantillon Multimode sans coordonnées téléphoniques avaient la possibilité de répondre à l'enquête par Internet ou de communiquer à la hotline un numéro de téléphone pour être rappelé et interrogé par un enquêteur.

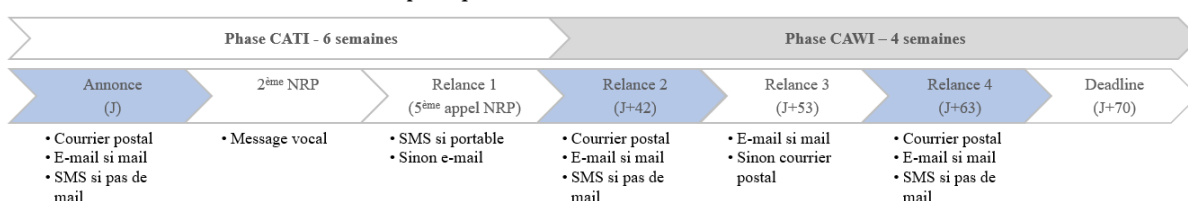
L'ensemble des modes de contact disponibles ont été mobilisés pour animer la collecte :

courriers, mails, coordonnées téléphoniques (fixe et mobile) pour l'exploitation CATI, coordonnées mobiles pour les SMS.

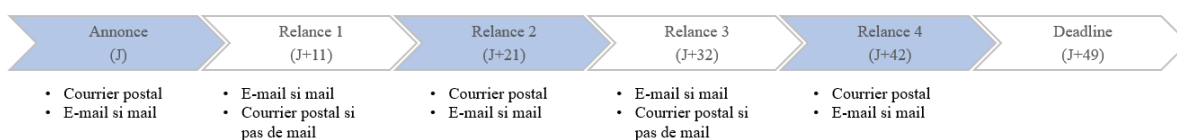
Les courriers postaux ont été mobilisés pour les lettres annonce de tous les échantillons et une partie des relances afin de limiter le coût financier et écologique de l'enquête tout en essayant de maximiser le taux de participation. Plusieurs modes de contact ont pu être mobilisés pour un même envoi, tout en mobilisant deux canaux de communication maximum par relance. Le choix des modes de contact utilisés est intrinsèquement lié au mode de collecte de l'échantillon concerné. Lors de l'exploitation CAWI, la priorité était ainsi donnée aux mails et les SMS étaient utilisés seulement en l'absence d'adresse mail. Les figures 1 à 3 décrivent le protocole d'animation de la collecte de chaque échantillon ainsi que les modes de contact utilisés.

Figure 1 : Animation de la collecte selon l'échantillon

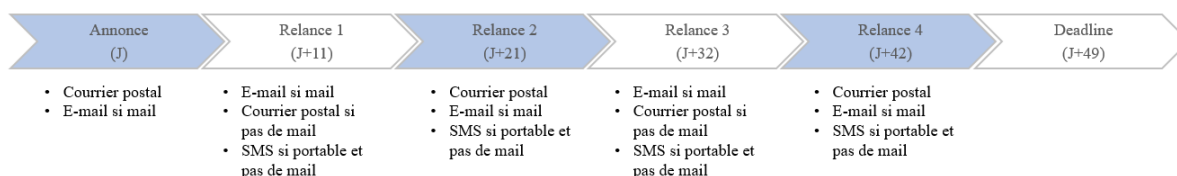
Échantillon Multimode avec coordonnées téléphoniques



Échantillon Multimode sans coordonnées téléphoniques (CAWI exclusif)



Échantillon Méthodologique (CAWI exclusif)



En bleu : relance avec courrier postal pour tous

3.2 Choix opérés sur l'utilisation des coordonnées compte tenu du protocole

Compte tenu de la multiplicité des adresses mails ou des numéros de mobile disponibles pour un individu, il est nécessaire d'opérer des choix méthodologiques afin d'identifier les coordonnées les plus pertinentes à mobiliser pour contacter la cible.

Pour l'envoi des courriers postaux, on dispose, pour chaque personne échantillonnée, de l'adresse de son logement, c'est donc cette adresse qui sera mobilisée pour l'envoi de courriers. Pour les individus dont l'identité est inconnue, nous avons fait le choix d'adresser le courrier au référent principal du logement, en mentionnant l'année de naissance de la personne échantillonnée résidant dans le ménage afin qu'il puisse lui transmettre.

Le mail contient les informations pour se connecter à l'espace enquêté en ligne (et au questionnaire pour les individus soumis au CAWI), c'est pourquoi les adresses mails n'appartenant pas à la cible ou répliquées dans plusieurs champs de la base ont été écartées. Ainsi, pour l'envoi des emails, seule l'adresse mail de la cible, non dupliquée dans un autre

champ, a été mobilisée (sur l'échantillon, 58,1% des fiches présentent une adresse mail unique pour la cible). Cela peut être l'adresse mail dans Fidéli ou une adresse mail récupérée sur le terrain (site internet, hotline, contact CATI) ; dans ce dernier cas, celle-ci devient prioritaire pour les envois.

En l'absence d'une adresse mail personnelle de l'enquêté, on lui préférera l'envoi d'un SMS le cas échéant, celui-ci ne donnant pas les identifiants de connexion il peut donc être réceptionné par une autre personne du logement. Pour l'envoi des SMS, la priorité a été donnée au numéro de mobile récupéré sur le terrain. En l'absence de mobile terrain, les SMS sont envoyés sur le mobile de la cible issu de Fidéli, puis celui du référent 1 et enfin celui du référent 2 en fonction de leur disponibilité.

Concernant le protocole d'appels de l'échantillon Multimode avec téléphone, le choix a été fait de mobiliser l'ensemble des numéros de téléphone disponibles (Fidéli, enrichissement, collectés sur le terrain) afin de maximiser les chances de contacter les individus. Les numéros de téléphone présents en base avant le lancement sont mobilisés dans l'ordre suivant : mobile de la cible, fixe de la cible, fixe du référent 1, mobile du référent 1, fixe du référent 2, mobile du référent 2, mobile issu de l'enrichissement, et enfin fixe issu de l'enrichissement. Lorsqu'un numéro de téléphone est collecté sur le terrain, celui-ci devient prioritaire.

4 Impact de l'utilisation des différentes coordonnées en termes de participation

La collecte de l'enquête a débuté en janvier 2024. L'exploitation de l'échantillon méthodologique est terminée. L'exploitation de l'échantillon Multimode est en cours de collecte et devrait se terminer à la fin de l'été. Les premiers résultats présentés ici sont donc provisoires. Il apparaît toutefois déjà que la participation varie très nettement en fonction des coordonnées disponibles : le taux de réponse parmi les personnes avec coordonnées téléphoniques est supérieur de 60% à celui des personnes sans coordonnées téléphoniques, sur les deux protocoles (multimode séquentiel CATI-CAWI ou monomode CAWI). La présence du mail de la cible est également associée à une participation plus importante : le taux de participation des personnes dont on dispose de l'email personnel est supérieur de 24% sur le protocole multimode et de 19% sur le protocole CAWI exclusif à celui des personnes sans adresse mail mobilisable pour le terrain.

Tableau 4 : Taux de participation en fonction des coordonnées disponibles

	Ech. Multimode (lots terminés)	Ech. Méthodologique
Numéro de téléphone		
Oui	42,6	48,0
Non	27,1	30,0
Mobile		
Oui	43,8	48,3
Non	28,4	32,7
Mail de la cible (unique)		
Oui	44,6	48,6
Non	36,1	40,9

Ces résultats peuvent s'expliquer par deux effets :

- Les individus pour qui l'on dispose de coordonnées de contact (et les ayant donc fournis à l'administration fiscale) sont plus enclins à participer aux enquêtes (la disponibilité des coordonnées de contact, tout comme la participation, sont liés aux profils sociodémographiques des individus (Charrance, 2023)) ;
- Le protocole de relance mis en place, mobilisant mails et SMS, a un effet bénéfique sur la participation.

5 Prochains travaux

Nous pouvons d'ores et déjà affirmer que la pertinence de Fidéli pour échantillonner des enquêtes multimodes est incontestable. La réussite d'une enquête multimode, et a fortiori d'une enquête exclusivement conduite sur Internet, repose sur l'animation de la collecte, c'est-à-dire un protocole d'annonce et de relances étoffé et mobilisant plusieurs canaux de communication. Nous avons pu constater que l'envoi de SMS ou de mails en plus des courriers postaux permet un fort gain de participation sur un protocole CAWI exclusif. Ces résultats constituent de premières explorations, mais de nombreuses autres analyses pourront également être investiguées sur la base de cette enquête de grande ampleur afin d'améliorer le protocole pour la prochaine vague.

Nous présenterons en novembre les résultats actualisés avec la fin de la collecte. Nous déterminerons notamment si ces deux effets se conjuguent pour expliquer la meilleure participation sur la population avec coordonnées de contact. Ayant recours à un protocole multimode séquentiel, nous pourrons étudier l'impact de ces coordonnées en fonction du mode de collecte (CATI ou CAWI), en séparant les échantillons ainsi que les phases de collecte. Nous étudierons également si le fait de ne disposer d'aucune coordonnée personnelle pour la cible est associée à une moindre participation, à caractéristiques sociodémographiques égales.

Bibliographie

Bondon M., Bouchet-Vala M., Breda-Popa R., Charrance G., Cochet P., Markou E. et Toulemon L. (2023), L'enquête test Etude des relations familiales et intergénérationnelles (Erfi) 2.0 : Enseignements d'une collecte multimode téléphone/internet avec et sans incitations financières, communication proposée pour le 12ème colloque international francophone sur les Sondages, 22 au 24 mars 2023, Aubervilliers

Bondon M., Charrance G., Cochet P. et Leduc A. (2022), Apprendre des paradonnées pour améliorer les protocoles de collecte : l'exemple d'EpiCov, communication au cours des Journées de Méthodologie Statistique, 29 au 31 mars 2022, Paris

Charrance G., Merly-Alpa, (2023) Les coordonnées de contact dans Fidéli-Nautile : Quels enseignements en tirer pour la conduite d'enquêtes ?, communication proposée pour le 12ème colloque international francophone sur les Sondages, 22 au 24 mars 2023, Aubervilliers

Lollivier S., Lamarche P. (2021), Fidéli, l'intégration des sources fiscales dans les données sociales, Courrier des statistiques N6, juin 2021, INSEE