

L'ENQUETE ÉTUDE DES RELATIONS FAMILIALES ET INTERGÉNÉRATIONNELLES (ERFI 2) PREMIERE VAGUE : GESTION DU TERRAIN ET ENSEIGNEMENTS D'UNE COLLECTE MULTIMODE AVEC FIDELI

Milan Bouchet-Valat¹, Ruxandra Breda-Popa², Paul Cochet³, Efi Markou⁴, Laurent Toulemon⁵

¹ *Institut national d'études démographiques (Ined), France, milan.bouchet-valat@ined.fr*

² *Institut national d'études démographiques (Ined), France, ruxandra.popa@ined.fr*

³ *Institut national d'études démographiques (Ined), France, paul.cochet@ined.fr*

⁴ *Institut national d'études démographiques (Ined), France, markou@ined.fr*

⁵ *Institut national d'études démographiques (Ined), France, toulemon@ined.fr*

Résumé. L'Étude des relations familiales et intergénérationnelles 2 (Erfi 2) est la deuxième enquête française menée dans le cadre du Generations and Gender Programme (GGP), infrastructure de recherche transnationale qui vise à comprendre les dynamiques familiales et les parcours de vie des individus. Étude longitudinale, Erfi 2 est composée de trois vagues, la collecte de la première se terminant début juillet 2024. Cette collecte multimode concurrentielle (CAWI/CATI) s'étale sur un peu plus de cinq mois, selon un calendrier où l'exploitation se fait au fur et à mesure en quatre périodes distinctes. Ces quatre périodes sont comparables quant au protocole, la taille et la structure de leurs échantillons mais chacune a rencontré des contraintes spécifiques liées notamment au calendrier ou à l'organisation du travail sur la longue durée. Cette communication se propose de revenir sur les problèmes rencontrés, les ajustements qui ont été apportés et leur impact sur le taux de la participation et leurs éventuelles incidences sur la composition de l'échantillon répondant. Deux contraintes seront en particulier abordées : les vacances scolaires et les jours fériés ; et les variations du nombre d'enquêteurs.

Mots-clés. Enquêtes multimode ; CATI ; CAWI ; protocole de relance ; enquêteurs ; Fidéli-Nautile ; gestion de terrain

Abstract. Erfi 2 is the second French survey part of the Generations and Gender Programme (GGP), a transnational research infrastructure aimed at understanding family dynamics and individual life courses. A longitudinal study, Erfi 2 is made up of three waves, with the first wave ending in early July 2024. This competitive multi-mode collection (CAWI/CATI) lasted a little more than five months, according to a schedule in which batches are opened progressively in four distinct periods. The four periods were comparable in terms of protocol, sample size and structure, but each had its own specific constraints, notably in terms of scheduling and the organization of long-term work. This paper will review the problems encountered, the adjustments made and their impact on the participation rate, as well as their possible impact on the composition of the respondent sample. Two constraints in particular will be addressed: school and public holidays; and variations in the number of interviewers.

Keywords. Multimode surveys ; CATI; CAWI ; follow-up protocol ; interviewers ; Fidéli-Nautile ; field management

Le volet français du Generations and Gender Programme (GGP), l'Étude des relations familiales et intergénérationnelles (Erfi 2), conçue par l'Ined, se trouve actuellement en cours de collecte de la première vague. Cette collecte multimode concurrentielle (CAWI/CATI) s'étale sur un peu plus de cinq mois, selon un calendrier où l'exploitation de l'échantillon se fait au fur et à mesure en quatre périodes distinctes. Cette communication se propose de revenir sur les spécificités de chaque période de collecte, la gestion des contraintes rencontrées, les ajustements qui ont pu être apportés et leur impact sur le taux de participation et leurs éventuelles incidences sur la composition de l'échantillon répondant. Détaillant certains des ajustements du protocole initial, réalisés au fur et à mesure de la collecte des données, nous montrons que celle-ci, aussi préparée et testée qu'elle puisse être, n'est pas l'application rigide des règles fixées d'avance. Elle nécessite au contraire un suivi serré, tenant compte d'indicateurs différents, et des ajustements continus. Anticipés ou conçus suite aux aléas du terrain, ces ajustements peuvent éviter la dégradation de la qualité des données. C'est cet aspect que nous voudrions présenter lors de cette communication, une fois la collecte finie et les fichiers des données et des paradonnées préparés.

1 L'enquête Erfi 2

L'enquête Étude des relations familiales et intergénérationnelles (Erfi 2) fait partie de la deuxième édition du programme international d'enquêtes Generations and Gender Programme (GGP). Lancé à l'initiative de la Commission de l'ONU pour l'Europe au début des années 2000, ce projet a déjà réalisé un premier cycle d'enquêtes longitudinales dans une vingtaine de pays, dont la France (Erfi, 2005-2011 ; Régnier-Loilier, 2009, 2016). Pour connaître les changements dans les dynamiques familiales et les parcours de vie des individus survenus depuis lors, le Generations and Gender Programme a entamé en 2020 un nouveau cycle de collecte (GGP II), toujours sous la forme d'enquêtes longitudinales de 3 vagues espacées chacune de 3 ans. La nouvelle édition de l'enquête française (Erfi 2) est prise en charge par l'Institut national d'études démographiques (Ined). Après un premier test (fin 2021-début 2022) puis une répétition générale (automne 2023), l'Ined est en train d'achever la collecte des données de la première vague de l'enquête réelle. Deux autres vagues sont prévues en 2027 et 2030, réinterrogeant les mêmes personnes, ce qui permettra de mieux comprendre les liens entre les événements qui jalonnent les histoires familiales.

Erfi 2 vise à dresser un panorama détaillé des situations familiales, en mettant notamment l'accent sur les relations entre parents, enfants, petits-enfants et beaux-enfants. Elle est reconnue d'intérêt général et de qualité statistique par le Conseil national de l'information statistique (Cnis) mais sans être obligatoire. Il s'agit d'une enquête multimode concurrentielle (CAWI/CATI) par questionnaire auprès de personnes âgées de 18 à 79 ans, habitant en ménage ordinaire en France métropolitaine. Le questionnaire, identique pour les deux modes de collecte, est hébergé par le Central Hub GGP au NIDI (Netherlands Interdisciplinary Demographic Institute) à La Haye, qui coordonne l'ensemble du programme. Il dure en moyenne un peu plus d'une heure, portant sur la situation actuelle mais aussi sur l'histoire familiale et de couple des répondants, sur l'organisation du ménage, la santé et le bien-être, le travail et la situation financières, tout en incluant quelques questions sur les valeurs et les attitudes des personnes interrogées.

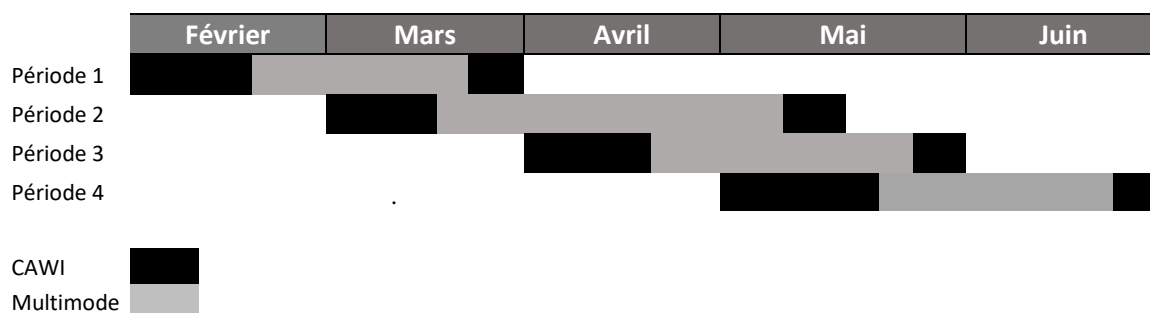
2 Échantillon et collecte de données sur quatre périodes

La taille de l'échantillon répondant attendue est de 10 500 individus avec deux objectifs en termes de surreprésentation. Tout d'abord nous avons cherché à surreprésenter les individus appartenant aux ménages situés aux déciles extrêmes de niveau de vie, ces déciles ayant des comportements démographiques qui se distinguent du reste de la population, et les déciles inférieurs ayant un taux de réponse plus faible. Deuxièmement, nous avons souhaité surreprésenter les couples de même sexe (CMS) afin d'obtenir diverses informations qui font aujourd'hui défaut dans la littérature au sujet de ces couples, notamment à propos de l'histoire conjugale ou encore sur les relations entretenues avec les parents.

Ces diverses surreprésentations ont été rendues possibles par le tirage de l'échantillon dans la base de sondage Fidéli-Nautile qui offre des informations sur les déciles de niveau de vie et sur la composition du ménage, y compris le sexe du conjoint cohabitant (ou de la personne supposée telle). Afin de maximiser le nombre de couples de même sexe au sein de l'échantillon, d'autres ont pu être récupérés au moment de la disjonction de l'échantillon complémentaire de l'enquête Famille et Employeurs (FamEmp)¹ (Cochet et al., 2023).

La première vague de l'enquête Erfi 2 se déroule sur un peu plus de cinq mois, selon un calendrier où l'exploitation de l'échantillon s'est faite au fur et à mesure en quatre périodes de collecte distinctes qui se sont succédées en se recoupant partiellement de fin janvier à début juillet 2024. D'une durée globale allant de neuf à dix semaines, chacune des trois premières périodes a porté sur un échantillon d'environ 12 590 individus, contre 10 300 en quatrième période (Tableau 2). Malgré le caractère multimode concurrentiel CAWI/CATI de l'enquête, chaque début et fin de période recourt uniquement à une collecte par Internet, au moins deux semaines en amont et une semaine en aval des appels téléphoniques des enquêteurs (Tableau 1).

Tableau 1. Les quatre périodes de collecte de la vague 1



Le découpage en périodes a été initialement conçu pour mieux organiser le terrain téléphonique : il s'agissait de répartir de manière plus harmonieuse les effectifs d'enquêteurs et assurer la proximité entre les appels et l'invitation à répondre envoyée aux enquêtés. Il nous a aussi

¹ L'enquête FamEmp est une enquête portée par l'Ined portant sur les liens entre vie familiale et vie professionnelle. Cette enquête suivant un protocole et un calendrier proche de celui de l'enquête Erfi, les échantillons de ces deux enquêtes ont été tirés en même temps dans Fidéli.

permis d'intervenir en cours de collecte afin d'améliorer les résultats en fonction des enseignements du suivi de la progression du terrain. L'adaptation de la taille de l'échantillon mise en exploitation en est un exemple.

Avant le lancement de l'enquête, les fiches adresses étaient réparties aléatoirement par l'Insee en 21 lots de structure identique. Chaque lot comporte 2 000 individus afin de pouvoir ajuster au mieux la taille du sous-échantillon mis en collecte à chaque période. Les périodes 1 à 3 comportaient chacune six lots. Aux 12 000 individus contenus dans ces lots, nous avons ajouté à chacune de ces trois périodes un tiers des CMS captés via le complémentaire de l'enquête FamEmp.

La quatrième période de collecte quant à elle n'était constituée que de trois lots de manière à pouvoir y ajouter des fiches adresses tirées dans la réserve si le besoin de redresser l'échantillon apparaissait en cours de collecte. Or, les contrôles effectués sur l'échantillon répondant au cours de la collecte ont montré que les individus des déciles 4 à 10 avaient un taux de participation en deçà de ce qui était attendu. Ainsi, bien que l'objectif de 10 500 répondants soit atteignable avec l'échantillon principal seul, il a été décidé d'exploiter une partie de la réserve de l'enquête en y tirant des individus appartenant à ces déciles, quitte à obtenir plus de répondants que prévu à l'origine².

La répartition aléatoire des fiches adresses a notamment permis d'obtenir des périodes de collecte au sein desquelles les prévalences des différentes coordonnées disponibles dans la base sont comparables pour les périodes de collecte 1 à 3 (tableau 2). Pour ce qui est de la période 4, la prévalence est légèrement plus importante par rapport aux trois périodes précédentes. Cela s'explique par la plus grande proportion de personnes des déciles 4 à 10 durant cette période du fait du tirage de ces déciles dans la réserve : les coordonnées de contact dans Fidéli sont d'autant mieux renseignées que le décile de niveau de vie d'une personne est élevé (Charrance, 2024).

Tableau 2. Prévalence des coordonnées de contact selon les périodes de collecte

Période de collecte	Nombre de fiches adresses	Proportion avec un téléphone portable	Proportion avec un deuxième téléphone	Proportion avec une adresse e-mail
Période 1	12 596	58,1%	20,9%	68,2%
Période 2	12 587	57,5%	20,5%	68,1%
Période 3	12 593	56,6%	19,2%	67,5%
Période 4	10 311	58,9%	21,1%	70,9%

3 Description du protocole de relance d'une période

La collecte des questionnaires par téléphone a été prise en charge par l'institut Ipsos, les appels étant passés à partir du plateau téléphonique situé à Plérin (Côtes-d'Armor). Dans le cadre de cette enquête, Ipsos a mis en place un site internet où les personnes sélectionnées pouvaient s'identifier afin de mettre à jour leurs coordonnées, prendre rendez-vous pour ré-

² Cette augmentation de la taille de l'échantillon a été rendue possible budgétairement grâce à des coûts de collecte plus faibles qu'anticipé, du fait du fort taux de réponse en CAWI.

pondre par téléphone ou encore se connecter au questionnaire pour répondre en ligne. Une hotline a également fonctionné tout au long de l'enquête téléphonique. Toute l'équipe du terrain (enquêteurs, encadrants et hotliners) a été formée de manière extensive sur l'objectif de l'enquête, le protocole, le contact avec les enquêtés et le questionnaire, formation incluant nombre d'exercices pratiques et de simulations. Trois formations successives ont été organisées, regroupant de 10 à 18 enquêteurs chacune. Elles duraient 2,5 jours sauf pour la première (1 jour), destinée aux enquêteurs qui avaient déjà suivi la formation complète dans le cadre de la répétition générale.

Un protocole de contact (annonce de l'enquête et relances) des personnes invitées à participer a été mis en place, suivant le même schéma pour chaque période de collecte. Il était conçu de manière à tirer profit de la richesse de Fidéli en matière de coordonnées de contact : adresse postale, téléphone(s) – portable ou fixe, de la cible ou d'autres personnes de son ménage –, adresse(s) e-mail. Nous avons ainsi alterné les supports et les groupes visés par les relances ([Tableau 3](#)). Une lettre annonce a été envoyée à l'ensemble des participants en début de chaque période, suivie quelques jours plus tard d'un e-mail d'annonce, envoyé aux personnes pour lesquelles nous disposions des coordonnées e-mail³. Des relances par lettre, e-mail et SMS étaient envoyées à des dates définies, juste avant et pendant la phase d'appel téléphonique. Cependant, afin de ne pas déranger exagérément des personnes déjà bien sollicitées par l'enquête, certaines de ces relances s'adressaient à des groupes spécifiques. C'est par exemple le cas des e-mails M3 et M3bis, envoyés aux non répondants pour lesquels nous ne disposions d'aucun numéro de téléphone. Parallèlement aux relances à jours fixes, un message vocal était laissé sur les répondeurs des personnes n'ayant pas décroché au 2^e appel, et des messages SMS ou par e-mail étaient envoyés aux adresses n'ayant pas décroché au bout du 6^e et puis du 12^e appel.

3 Voir les proportions de chaque type de coordonnées par période au [Tableau 2](#).

Tableau 3. Planning-type d'une période de collecte de la vague 1

Calendrier	Type d'événement	Code
J-4	Envoi Lettre annonce	L0
J	Mail annonce	M0
J+10	Lettre de relance	L1
J+11	Relance mail	M1
J+12	SMS Annonce	S0
J+10 ; J+10 à 13 ; J+17 à 19	Formations enquêteurs	
J+13 à J+49 2 appels sans décroche 6 appels sans décroche 12 appels sans décroche	Appels Message sur répondeur Relance SMS ou mail Relance mail ou SMS	M2/S1 M2bis/S2
J+24	Relance mail à ceux qui n'ont pas de téléphone	M3
J+31	Relance mail à ceux qui n'ont pas de téléphone	M3bis
J+37	Lettre fin de collecte aux adresses sans tél ou avec tél mais sans e-mail	L2
J+44	Relance Mail et SMS avec date de fin de collecte	M4/S3
J+49	Fin du terrain CATI	
J+58	Fin du terrain CAWI	
J+65	Envoi des lettres de remerciement	

4 L'adaptation du protocole aux spécificités du terrain

Cette collecte longue de plus de 5 mois soumet les échantillons ouverts progressivement dans le temps à des contraintes spécifiques, liées au calendrier ou aux événements du terrain. Des adaptations ont été nécessaires. Grâce aux indicateurs de suivi fournis par Ipsos de manière régulière – le remplissage des questionnaires en particulier était suivi jour par jour, voire heure par heure –, nous avons pu adapter les protocoles-type de relances à la progression du terrain et des appels afin d'intervenir au plus tôt pour limiter l'impact de certains événements imprévisibles.

Les vacances scolaires et les jours fériés constituaient pour le travail du terrain des moments spécifiques car nous craignons des ralentissements de la collecte. Ils nécessitaient des ajustements propres à chaque période. Nous avons par exemple avancé le lancement de la collecte par Internet pour la période 4 (P4) afin que la lettre annonce arrive bien avant les grands ponts du 1^{er} puis des 8 et 9 mai, et que les enquêtés disposent d'un temps de réponse suffisant avant leur éventuel départ. Nous avons de surcroît retardé d'une semaine le lancement des appels de la P4 pour permettre à la P3, qui avait aussi souffert de ces ponts et d'un déficit d'enquêteurs disponibles, de se terminer dans des bonnes conditions.

Le suivi du déroulement de la collecte nous a aussi amené à d'autres infléchissements du protocole de contact, sans toutefois trop dévier de la règle d'une exploitation homogène des différents lots qui composent l'échantillon. Dans le suivi des envois des e-mails de la P1 et P2, nous avons observé que la relance M3bis (envoyée uniquement aux adresses pour lesquelles nous ne disposions pas d'un numéro de téléphone, et auprès desquelles nous cherchions à mobiliser davantage les autres supports de relance) concernait un très faible pourcentage d'adresses, quelques centaines tout au plus. Nous avons donc décidé d'élargir son périmètre, contactant toutes les adresses actives pour lesquelles un e-mail était disponible. Les relances

M3bis des P3 et P4 ont couvert ce nouveau périmètre – avec de résultats bien meilleurs. La P3 a par ailleurs bénéficié d'un autre élargissement, de la L2 cette fois (la lettre annonçant la date de fin de collecte aux adresses, ici aussi, sans téléphone ou avec téléphone mais sans e-mail – celles avec e-mail ayant reçu un message électronique au lieu de la lettre). Suite à une erreur de filtrage avant l'impression et de la mise sous pli, toutes les adresses actives ont reçu cette lettre de relance, ce qui s'est avéré bénéfique. On peut considérer que l'élargissement des périmètres de ces deux relances a permis de redresser le taux de participation de la P3, qui avait initialement pâti d'un nombre insuffisant d'enquêteurs (voir infra).

C'est pour améliorer davantage le taux de participation tout en incluant des répondants tardifs et éventuellement atypiques que nous avons décidé au cours de la quatrième période, de solliciter à nouveau par e-mail et par un petit nombre d'appels téléphoniques l'ensemble des non répondants (et non refusants) des trois premières périodes de collecte. À l'heure actuelle, cette dernière relance n'est pas encore réalisée.

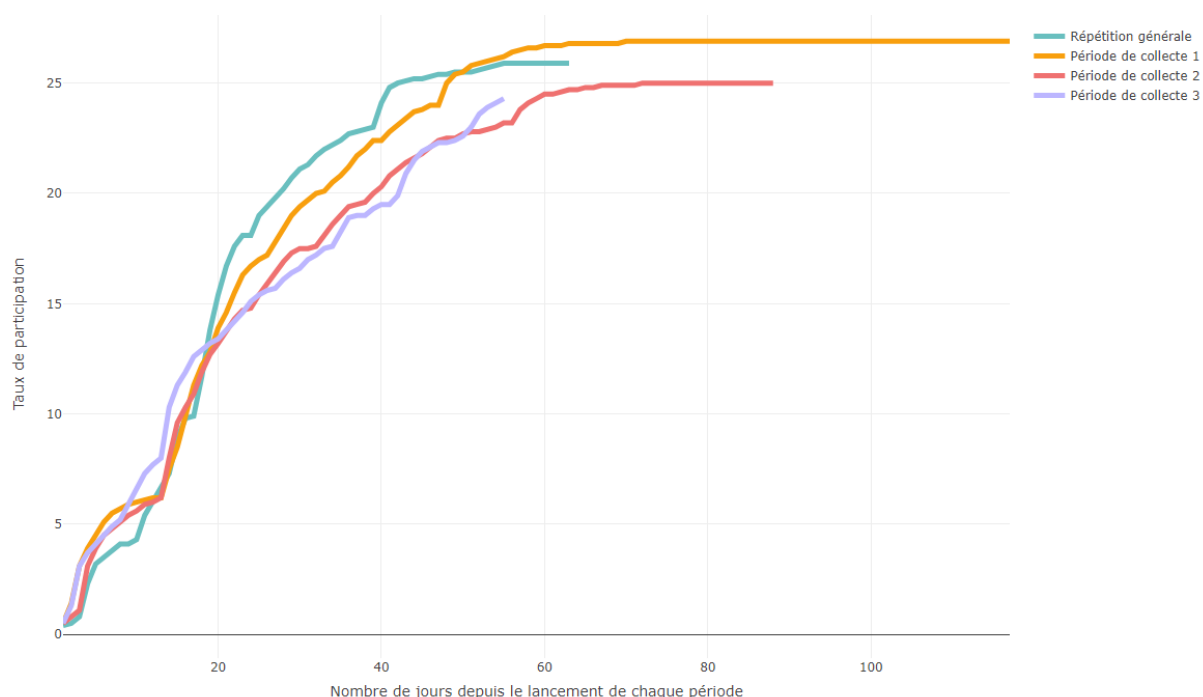
Malgré une répétition générale et une préparation importante de l'enquête, certains événements inopinés sont inévitables pendant le déroulement du terrain et nécessitent une attention constante de la part de l'équipe. Durant la première période par exemple, le terrain nous a alerté sur l'épuisement accéléré des adresses, événement relié à un blocage inattendu des appels fait par l'automate. Ce blocage s'est avéré être la conséquence d'une des règles du protocole. Une reprogrammation rapide a été nécessaire pour reprendre le déroulement normal du terrain.

Dans le même ordre d'évènements, certains e-mails ont nécessité un envoi rectificatif comme lors du lancement de la quatrième période de collecte. Le premier e-mail envoyé à l'ensemble des individus pour lesquels nous disposions d'une adresse électronique manquait d'indiquer la personne à laquelle l'enquête s'adressait. Cet envoi incomplet a été redoublé d'un deuxième e-mail correct très rapidement pour éviter à la fois une incompréhension du message qui leur était adressé, un risque de confusion sur le destinataire et une perte de confiance.

5 Focus sur les périodes de vacances et la variation du nombre d'enquêteurs

Les quatre périodes constituent autant de collectes des données indépendantes et comparables entre elles. Parties de la même enquête (même questionnaire et mêmes conditions de réalisation), elles utilisent des échantillons et des protocoles similaires tout en variant sur certains détails. Les courbes du taux de participation, tout en étant proches, diffèrent selon la période de collecte ([Figure 1](#)). Dans quelle mesure pouvons-nous attribuer ces différences aux ajustements du protocole de collecte ?

Figure 1. Progression du taux de participation par période de collecte



Deux déterminants du taux de participation et de la composition de l'échantillon répondant retiennent particulièrement notre attention : d'une part, les vacances scolaires ; d'autre part, les heures travaillées par les enquêteurs. Dans une collecte CAWI/CATI concurrentielle, nous pouvons minimiser l'effet de ce second facteur en nous concentrant uniquement sur les questionnaires remplis en ligne suite à une relance par lettre ou par e-mail. Les analyses successives de la progression de la collecte CAWI suite à l'envoi d'un support de contact et selon sa place dans le protocole peuvent fournir des indications utiles.

À partir des données du suivi de la collecte, nous pouvons formuler l'hypothèse que les vacances scolaires ralentissent peu le remplissage sur internet lorsque les enquêtés sont contactés par e-mail. En revanche ceux-ci répondent moins bien pendant les périodes non travaillées de plus courte durée, comme les ponts ou les longs weekends. Néanmoins, selon la même hypothèse, si les enquêtés sont contactés par courrier postal pendant les vacances scolaires, ils répondent moins bien dans la semaine qui suit le contact.

Nous entendons mettre à l'épreuve ces hypothèses avec les données et parodonnées finales. Nous affinerons aussi les résultats en analysant l'impact des vacances sur la composition même de l'échantillon répondant et sur la répartition des réponses dans le temps : si les enquêtés contactés par lettre répondent moins pendant les vacances scolaires, est-ce que cela est confirmé par le ralentissement des répondants habitant dans les zones en vacances ? Et est-ce que les non-répondants de ces zones viennent remplir leur questionnaire une fois les vacances finies, ou sont-ils davantage perdus pour la collecte ?

Par ailleurs, au total 40 enquêteurs ont été formés pour réaliser l'enquête téléphonique. Ce dispositif a été estimé par Ipsos en fonction de la taille de l'échantillon, de l'évaluation du

nombre d'heures de production CATI et de l'étalement du terrain. Si la présence des enquêteurs a été plutôt conséquente durant les deux premières semaines d'enquête CATI de la première période (P1) et moindre par la suite, certaines variations sont observées entre les différentes périodes. En effet, pour la première période en sortie de formation, les enquêteurs ont été plus fortement mobilisés sur le terrain CATI, puis leur nombre a été ajusté pour s'adapter aux besoins du terrain. Pour les périodes suivantes, la présence des enquêteurs CATI au démarrage est moins marquée et une augmentation de leur effectif apparaît plus tardivement.

Nos premières observations ainsi que les retours des enquêteurs laissent penser que la première période de collecte, bénéficiant d'un nombre important d'enquêteurs au démarrage du terrain CATI et des appels proches de la réception des documents de communication sur l'enquête, a reçu un meilleur accueil et a conduit à un meilleur taux de participation. À l'inverse, la réduction des effectifs des enquêteurs durant les deux premières semaines de démarrage de la P2 et de la P3, et leur relative augmentation en fin de période, semble avoir amené des appels plus éloignés de la réception des relances, et donc un rattrapage plus difficile des taux de réponse. La collecte Erfi 2 par périodes successives permettra de tirer quelques enseignements sur le rôle de la distribution du nombre d'enquêteurs et des heures de travail sur le plateau téléphonique, en comparant l'évolution du taux de réponse selon les différentes périodes.

Bibliographie

- Charrance, G. (2024). *La disponibilité des coordonnées de contact dans Fidéli-Nautile : Quels enseignements pour les protocoles de collecte*. Documents de travail Insee. <https://www.insee.fr/fr/statistiques/8183708>
- Cochet, P., Charrance, G., Merly-Alpa, T., & Toulemon, L. (2023). *Surreprésenter les populations rares grâce au tirage simultané de plusieurs enquêtes*. 12e colloque francophone sur les sondages, Aubervilliers. https://drive.google.com/file/d/1nc8LTVZkJnHH2wtOt-vfhgNn14_BTqfhB/view?usp=drivesdk&usp=embed_facebook
- Régnier-Loilier, A. (dir.) (2009). *Portraits de familles : L'enquête Étude des relations familiales et intergénérationnelles*. Ined éditions.
- Régnier-Loilier, A. (dir.) (2016). *Parcours de familles : L'enquête Étude des relations familiales et intergénérationnelles*. Ined éditions.