
UN PROTOCOLE INNOVANT, MULTIMODE EN DEUX PHASES, POUR MESURER LA VICTIMATION : RÉSULTATS ET PRINCIPAUX ENSEIGNEMENTS DE LA COLLECTE DE L'ENQUETE GENRE ET SECURITE (GENESE)

Hélène GUEDJ, Axelle QUIVIGER, Sandra ZILLONIZ (*), Tiaray RAZAFINDRANOVA (**)

(*) SSMSI, section Enquêtes statistiques

(**) SSMSI, section Méthodes statistiques et analyses conjoncturelles

helene.guedj@interieur.gouv.fr, axelle.quiviger@interieur.gouv.fr,
sandra.zilloniz@interieur.gouv.fr, tiaray.razafindravona@interieur.gouv.fr

Mots-clés : Protocole multimode, collecte, relance, internet, coordonnées de contact, parodonnées

Domaine concerné : 4. Collecte de données d'enquêtes / 4.3 Multimode

Résumé

Au sein du ministère de l'Intérieur, le Service statistique ministériel de la sécurité intérieure (SSMSI) est en charge du domaine couvrant la victimation, la délinquance et l'insécurité. Très sollicité sur la thématique des violences faites aux femmes, le SSMSI a obtenu en novembre 2019 un financement européen pour conduire en 2021, au niveau national, l'enquête Genre et Sécurité (Genese), portant sur les violences liées au genre et basée sur le questionnaire *Gender-Based Violence* développé par Eurostat. Cette première enquête en population générale conduite par le SSMSI a été conçue pour répondre à deux ambitions majeures : éclairer le débat public à l'échelle européenne avec des données inédites sur les parcours de violences sexistes et sexuelles ; réaliser une expérimentation multimode grandeur réelle pour mesurer les victimations et les perceptions en matière de sécurité dans le cadre de la refonte de l'enquête « Cadre de vie et sécurité » (CVS).

Pour répondre aux objectifs de l'enquête, le SSMSI a conçu, en lien avec des experts de l'Insee, un protocole multimode strictement nominatif en deux phases : d'abord une enquête filtre (internet/papier/téléphone) auprès d'un échantillon de 169 060 individus sélectionnés dans la base Fidéli puis une enquête de suivi (internet/téléphone) auprès d'un échantillon ciblé de 15 000 répondants de phase 1 dans lequel les victimes de violences physiques et sexuelles sont surreprésentées. En

phase 1 comme en phase 2, les stratégies de relances des non répondants prévoient des envois de lettres postales, de mails et des appels téléphoniques. Afin de réduire l'attrition entre les deux phases des profils les plus rares concernés par la thématique des violences liées au genre, une des originalités du dispositif était d'offrir la possibilité d'enchaîner les deux questionnaires et, en l'absence d'enchaînement, d'enclencher un protocole de relance personnalisé démarrant le lendemain de la complétion du questionnaire de phase 1. Le terrain de l'enquête Genese est, par ailleurs, complété d'un volet méthodologique en face-à-face auprès d'une sélection de 3 000 non répondants totaux de phase 1 résidant en Gironde et en Île de France.

L'étude propose une description détaillée du protocole principal de l'enquête Genese (hors volet face-à-face) et examine de manière approfondie la qualité des données de contact fournies dans la base de sondage (adresses postales, mails, numéros de téléphone), l'apport de l'enrichissement de coordonnées réalisé par le prestataire de collecte ainsi qu'une évaluation des différentes stratégies mises en place pour optimiser la participation. Au-delà des taux de réponse, l'étude propose également une mesure des gains de représentativité offerts par chacun des modes en phase 1 ainsi qu'une analyse de l'efficacité de l'articulation phase 1 / phase 2. Les résultats sont enrichis de l'exploitation de certaines paradoxes recueillies dans le dispositif concernant le comportement des enquêtés : refus exprimés, désinscriptions des mails, respect des consignes et modalités d'accès au questionnaire en ligne comme le type de support, ou les jours et heures de connexion.

Il ressort de ces travaux que les choix des concepteurs concernant notamment la nature et le rythme des relances ont permis d'atteindre des taux de réponse élevés (64 % en phase 1 et 68 % en phase 2), en particulier sur internet (80 % des répondants en phase 1 et 78 % en phase 2). Toutes choses égales par ailleurs, l'envoi soutenu de mails apparaît comme un levier important de participation, particulièrement en début de collecte. L'introduction de la relance téléphonique permet également de soutenir la participation, à la fois par téléphone mais aussi par internet, ce qui plaide en faveur de protocoles concurrentiels différés plutôt que séquentiels « purs ». Sur le plan de la représentativité, la collecte par téléphone a permis des gains importants sur les profils peu répondants en début de collecte, en particulier les populations socialement défavorisées. Enfin, l'articulation originale entre phase 1 et phase 2 s'est révélée pertinente : l'attrition des profils rares a été particulièrement maîtrisée grâce aux relances lissées et à l'enchaînement des questionnaires qui a été largement plébiscité par les individus éligibles.

Les conclusions de ces travaux ont permis de valider globalement le dispositif de l'enquête Genese et de conserver ce protocole comme cadre de référence pour la nouvelle enquête sur le vécu et le ressenti en matière de sécurité (enquête VRS) qui prend la suite de l'enquête CVS à partir de 2022. Elles peuvent également utilement servir les concepteurs d'enquête ménage dans le design de protocole multimode.

Abstract

Funded by EU and based-on the Gender-based violence questionnaire developed by Eurostat, the Genese (Genre et sécurité) survey has been conducted in 2021 by the French Ministerial statistical department for internal security (SSMSI) is an innovative two-phase protocol survey allowing identification of the target population. This study provides a detailed description of the Genese survey and assesses the efficacy of this protocol in terms of quality of contact data, reminder strategies, gains in terms of representativity provided by each mode in phase 1 and attrition between phase 1 and 2. The survey design, in particular the type and the frequency of reminders, has led to high response rates. The use of phone calls was found particularly relevant and led to substantial gains in representativity, especially among socially disadvantaged population. Finally, the attrition between phase 1 and 2 was found low, in particular among the population of interest composed by individuals who experienced gender-based violence.

1. Présentation de l'enquête Genese

1.1 Contexte, objectifs et calendrier

Au sein du ministère de l'Intérieur, le Service statistique ministériel de la sécurité intérieure (SSMSI) est en charge du domaine couvrant la victimation, la délinquance et l'insécurité. Depuis sa création en 2014 et jusqu'en 2021, il a assuré la co-maîtrise d'ouvrage de l'enquête « Cadre de vie et sécurité » (CVS) aux côtés de l'Insee et de l'Observatoire national de la délinquance et des réponses pénales (ONDRP, supprimé fin 2020). En avril 2018, l'Insee a annoncé au SSMSI et à l'ONDRP qu'en raison de ses fortes contraintes budgétaires, il ne serait plus en capacité de réaliser l'enquête CVS sous sa forme actuelle (rythme annuel, collecte face-à-face) à partir de 2022. Conformément aux engagements pris devant le Conseil national de l'information statistique (Cnis), le SSMSI a piloté dès le printemps 2019 avec l'appui de l'Insee le projet de refonte de l'enquête CVS, en explorant notamment les opportunités offertes par la combinaison de différents modes de collecte (internet, téléphone, papier, face-à-face, etc.) en cours de développement au sein du Service Statistique Public (SSP).

En parallèle, très sollicité sur la thématique des violences faites aux femmes, le SSMSI a répondu à un appel à projet européen et obtenu en novembre 2019 un financement pour conduire en 2021, au niveau national, l'enquête Genre et Sécurité (Genese), portant sur les violences liées au genre et basée sur le questionnaire *Gender-Based Violence* développé par Eurostat. Cette première enquête en population générale conduite par le SSMSI a ainsi été conçue pour répondre à deux ambitions majeures : d'une part, éclairer le débat public à l'échelle européenne avec des données inédites sur les violences sexistes et sexuelles, d'autre part, réaliser une expérimentation multimode grandeur réelle pour mesurer les victimations et les perceptions en matière de sécurité en vue de la refonte de l'enquête CVS à l'horizon 2022 [12].

Plus précisément, l'enquête Genese a été réalisée en vue de remplir les objectifs suivants : i/ estimer à l'échelle infranationale (notamment départementale) mais aussi sur des territoires spécifiques (Quartiers prioritaires de la ville QPV, Quartiers de reconquête républicaine QRR) le sentiment d'insécurité, les phénomènes délinquants observés dans le quartier, des indicateurs de victimation agrégés (« atteintes aux biens » et « atteintes à la personne » en 2020 et avant 2020) et des indicateurs sur le lien police/population ; ii/ examiner finement à l'échelle nationale un large champ d'atteintes détaillées¹ sur différentes périodes (en 2020 et avant 2020), mesurer les taux de plainte associés en 2020 et décrire le profil des victimes ; iii/ mesurer dans une approche genrée les violences sexistes et sexuelles subies à différentes périodes de la vie (12 derniers mois, 5 dernières années, vie entière) au global, par type de violences (violences physiques, violences sexuelles, violences psychologiques, harcèlement sexuel au travail et agressions répétées malveillantes) et par type d'auteur et caractériser la fréquence, le contexte et les conséquences de ces violences.

En janvier 2020, l'enquête Genese a obtenu le label d'intérêt général et de qualité statistique avec octroi du caractère obligatoire délivré par le Cnis. La collecte a été réalisée par un prestataire privé du 1^{er} mars au 4 juillet 2021. Préalablement à l'enquête grandeur réelle, un pilote auprès de 2 000 individus a été conduit à l'automne 2020.

¹ Cambriolages, tentatives de cambriolage, vols sans effraction et actes de vandalisme contre le logement, vols de voiture, vols de deux-roues à moteur et vols de vélo, autres vols avec ou sans violence, escroqueries bancaires, arnaques, violences physiques, viols, tentatives de viol, attouchements du sexe, agressions sexuelles, harcèlement sexuel et moral.

1.2 Le protocole de l'enquête

1.2.1 Développement du multimode et source Fidéli

Dans le cadre du paradigme de l'erreur d'enquête totale [1], le recours à un dispositif multimode, c'est-à-dire qui mixe les modes de collecte, peut s'envisager pour différentes raisons : améliorer la couverture, favoriser la participation à l'enquête et donc réduire l'erreur de non-réponse totale ou encore diminuer l'erreur de mesure. Le développement de ces dispositifs est également motivé par la maîtrise des coûts. Les modes peuvent par exemple être combinés dans l'objectif d'améliorer la qualité de l'enquête à coût constant ou pour réduire les coûts sans dégradation de la qualité [23]. Ainsi, la recherche de la combinaison pertinente entre les différents modes de collecte doit intégrer ces différentes dimensions dans l'analyse. Typiquement, la proposition séquentielle de différents modes de collecte se fait habituellement du mode le moins coûteux (internet) au mode le plus coûteux (respectivement papier, téléphone, face-à-face).

Le développement du recours au multimode dans les enquêtes ménages de l'Insee est devenu une nécessité qui s'inscrit dans les axes stratégiques d'Insee 2025 [13] comme dans ceux des principaux instituts nationaux de statistique étrangers. Cette orientation est rendue possible grâce à la richesse des informations de contact de la base de sondage développée par l'Insee qui s'appuie sur les fichiers démographiques sur les logements et les individus (Fidéli). Cette base est construite à partir des données de la Direction générale des finances publiques. Elle recoupe des données issues des déclarations de revenus, des fichiers de la taxe d'habitation et des propriétés bâties. Ces données sont aussi complétées par des données issues d'autres sources : les répertoires des communautés, des résidences hôtelières ainsi que des logements locatifs des bailleurs sociaux. Les données de Fidéli sont quasi-exhaustives (la sous-couverture a été estimée à environ 2 % : personnes en situation irrégulière et des cas très particuliers)². Les informations de contact disponibles dans la base de sondage permettent de construire des protocoles multimodes nominatifs : une année de naissance et une adresse postale sont systématiquement renseignées, un prénom et un nom de famille sont renseignés pour les personnes majeures dans plus de 98 % des cas³, un ou plusieurs numéros de téléphone dans plus de 75 % des cas et enfin une ou plusieurs adresses mail dans près de 70 % des cas. Les nombreuses informations auxiliaires disponibles, notamment sur les revenus et le logement, peuvent être mobilisées pour le contact, les traitements statistiques post-collecte et l'évaluation de l'efficacité du protocole. La table utilisée comme base de sondage de l'enquête Genese est composée de tous les individus présents dans le champ du dispositif Fidéli millésime 2019 (revenus de l'année 2018 et situation des occupants au 1^{er} janvier 2019), ayant 18 à 74 ans au 1^{er} janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire, soit environ 46 millions d'individus⁴.

² Les présentations d'un séminaire de l'Insee dédié à la source Fidéli et à la nouvelle base de sondage Nautile sont accessibles en ligne : <https://www.insee.fr/fr/information/4191244>

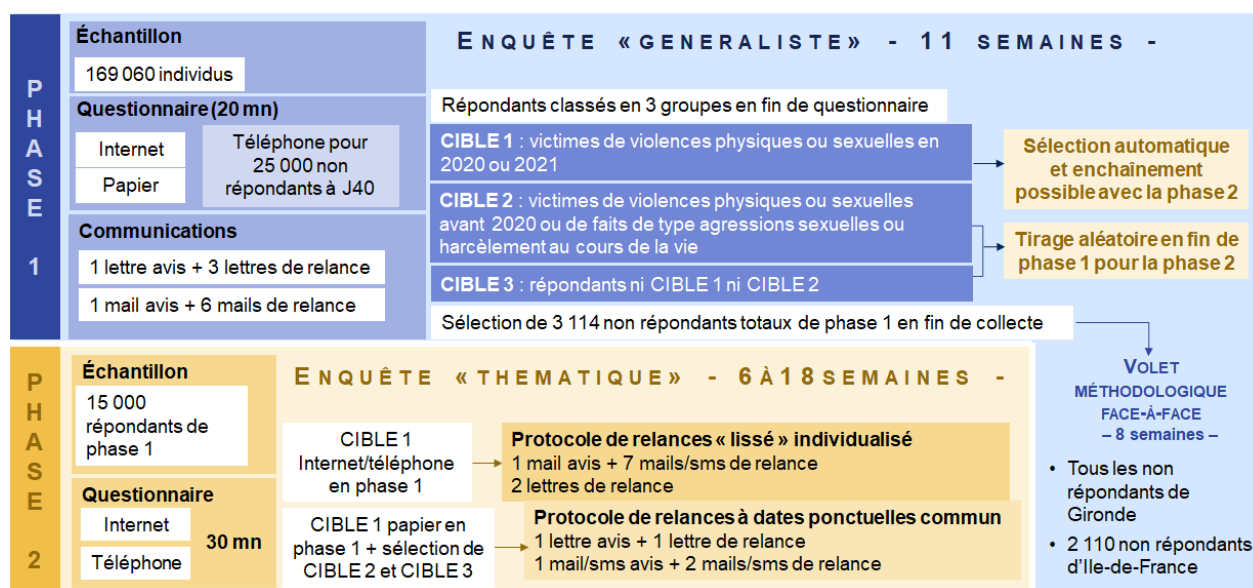
³ Plus de 99 % des individus majeurs sans prénom/nom renseignés sont des enfants à charge des référents fiscaux.

⁴ Les personnes identifiées comme vivant en communauté sont retirées de cette table. Cependant, ces dernières n'étant pas complètement identifiées par le dispositif Fidéli et par les recherches manuelles de la division Sondages de l'Insee, il restera des personnes hors-champ pour l'enquête dans la base de sondage.

1.2.2 Architecture générale de l'enquête Genese

Dans ce contexte stimulant et foisonnant, le SSMSI a préparé la refonte du protocole de l'enquête CVS réalisée en monomode face-à-face en visant une collecte multimode innovante. De manière générale, pour concevoir une enquête de victimation de qualité, le statisticien est confronté à deux difficultés majeures en amont de la conception du questionnaire. Il doit, d'une part, gérer la faible fréquence des atteintes mesurées – en particulier les violences les plus graves comme le viol⁵ – qui impose des tailles d'échantillons conséquentes. Il doit, d'autre part, prendre en compte la « sensibilité » de certaines victimations (violences sexuelles et domestiques) qui rend particulièrement délicat l'arbitrage entre les avantages et inconvénients des différents modes de collecte. Un échantillon trop faible conduit à des estimations dont la précision relative est médiocre et limite les possibilités de description des phénomènes étudiés. C'est un reproche qui a souvent été adressé à l'enquête CVS (qui en moyenne s'appuyait chaque année sur environ 15 000 ménages répondants). Afin de répondre à un coût soutenable aux besoins croissants de données localisées et plus précises en matière de victimation, tout en assurant une description détaillée des victimes et des atteintes subies, le SSMSI a conçu, en lien avec les experts de l'Insee, un protocole multimode en 2 phases de type enquête filtre/enquête de suivi (figure 1.1.1) en s'inspirant notamment d'enquêtes existantes qui ont fait leurs preuves, telles que l'enquête Vie quotidienne et santé (VQS)-Autonomie de la Drees. Ce protocole proposé pour l'enquête Genese s'inscrit dans les grands modèles promus par le programme multimode de l'Insee. Un protocole similaire a été mis en œuvre en 2020 pour l'enquête Epicov Drees-Inserm sur la diffusion du virus Sars-cov2 ainsi que les conséquences de l'épidémie sur le quotidien et la santé de la population.

Figure 1.1.1 – Architecture générale de l'enquête Genese



Source : SSMSI.

⁵ En France, d'après l'enquête CVS, chaque année environ 2,5 % des personnes âgées de 18 à 75 ans sont victimes de violences physiques ou sexuelles et 0,3 % sont victimes de viol (moyennes annuelles sur la période 2011-2017). Pour les taux de victimation annuels, voir [24].

La première phase consiste en une enquête de victimation « généraliste » courte (20 min en moyenne) basée sur le recueil des caractéristiques sociodémographiques, des questions de recensement des atteintes et des perceptions et opinions en matière de sécurité, sur le modèle de ce qui figure dans l'enquête CVS. Elle interroge un échantillon tiré dans Fidéli par l'Insee de 169 060 personnes âgées de 18 à 74 ans au 1^{er} janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire. La collecte, d'une durée de 11 semaines, est concurrentielle différée [23] : elle commence par internet puis prévoit une relance téléphonique auprès d'une sélection aléatoire de 25 000 non répondants et l'envoi d'un auto-questionnaire papier à différents stades de la collecte en fonction des coordonnées disponibles pour contacter l'enquêté. En plus d'internet, le papier s'est imposé comme un mode de relance rentable pour la phase 1, à l'image de ce qui est fait dans l'enquête sur les technologies de l'Information et de la Communication (TIC) de l'Insee, car le questionnaire est court et simple (peu de filtres). Le mode papier permet de réduire le recours à la relance téléphonique plus coûteuse et de toucher des populations moins connectées. Le volet papier dans l'enquête TIC représente 45 % des questionnaires complets et augmente significativement le taux de réponse global de l'enquête. L'objectif de participation à la phase 1, fixé à partir d'autres enquêtes multimode et expérimentations internet en cours au sein du SSP, était de 50 % : 35 % par internet, 10 % par papier et 5 % par téléphone.

Afin de cibler au mieux les victimes pour l'interrogation de phase 2, les répondants à l'enquête de phase 1 sont classés en trois groupes selon les atteintes qu'ils déclarent avoir subies :

- Cible 1 : victimes de violences physiques intentionnelles ou de violences sexuelles (viols, tentatives de viol ou attouchements du sexe) en 2021 ou 2020 ;
- Cible 2 : victimes de violences physiques ou sexuelles avant 2020 ou victimes de faits de type agressions sexuelles ou harcèlement au cours de la vie ;
- Cible 3 : répondants ni Cible 1 ni Cible 2.

La deuxième phase de l'enquête Genese consiste en une enquête de suivi thématique plus longue (30 min en moyenne) basée sur le questionnaire conçu par Eurostat qui permet un recueil d'informations très détaillées sur les violences sexistes et sexuelles subies tout au long de la vie. Elle est proposée par internet et par téléphone à un échantillon de 15 000 répondants de phase 1 dans lequel les victimes de violences physiques et sexuelles (Cible 1 et Cible 2) sont surreprésentées. Le mode papier n'est pas proposé en phase 2 car le questionnaire est complexe et très filtré et aussi parce que, contrairement à la phase 1, l'ensemble des non répondants avec téléphone est éligible en phase 2 à la relance téléphonique. Originalité du dispositif de l'enquête Genese, le protocole de phase 2 a été conçu de manière différenciée pour réduire l'attrition des populations particulièrement concernées par la thématique des violences liées au genre. Pour les répondants de phase 1 identifiés Cible 1 qui sont tous automatiquement sélectionnés pour la phase 2, l'enchaînement des deux questionnaires en une seule interrogation est proposé. Ainsi la collecte démarre à la validation du questionnaire de phase 1 et dure 18 semaines. Pour les répondants Cible 2 ou Cible 3 sélectionnés par tirage aléatoire pour la phase 2, la collecte dure 6 semaines et démarre une semaine après la clôture de la phase 1. Les objectifs de participation en phase 2 étaient fixés à 65 %, répartis à parts égales entre internet et téléphone.

Enfin, au terme de la phase 1, un volet méthodologique en face-à-face a été conduit sur 8 semaines du 1^{er} juin au 24 juillet auprès de plus de 3 000 personnes : tous les non répondants totaux de Gironde et une sélection aléatoire de non répondants totaux d'Île-de-France⁶.

⁶ Les résultats de ce volet ne sont pas présentés dans cette étude qui se concentre sur l'évaluation du protocole principal internet / papier / téléphone.

1.2.3 Schémas de contact et de relances

En phase 1, le schéma de contact et de relance prévoit l'envoi de relances par courrier et par mail ainsi qu'une relance téléphonique. Des traitements des informations de contact sont mis en œuvre en amont de la collecte : normalisation des adresses postales (traitement RNVP⁷) et enrichissement des coordonnées téléphoniques (cf. 2.1.1). Concernant les mails, étant donné la sensibilité des sujets abordés (notamment les violences domestiques), une règle a été définie pour s'assurer de leur caractère personnel afin de sécuriser au maximum ce mode de contact. Ainsi, les mails disponibles dans Fidéli sont qualifiés de « raisonnablement » personnels et utilisés uniquement si l'identité (nom/prénom) de l'individu est connue et à condition : i/ qu'il soit référent légal du foyer fiscal ou bien ii/ que l'adresse mail qui lui est rattachée dans la base soit unique au sein de son foyer fiscal s'il n'est pas le référent légal.

Le protocole de phase 1 distingue deux sous-échantillons en fonction des coordonnées exploitables après application de cette règle et enrichissement des coordonnées téléphoniques : le P1SS1 regroupe les personnes sans mail ni téléphone et le P1SS2 les personnes avec mail ou téléphone.

L'ensemble des individus sélectionnés reçoivent une lettre-avis nominative spécifiant clairement l'interdit du proxy (pour les quelques cas dont l'identité n'est pas connue, seule l'année de naissance est donnée) et les individus éligibles du P1SS2 reçoivent quelques jours plus tard un mail-avis nominatif. Afin de réduire le proxy, lors de la phase contact en ligne ou par téléphone, le caractère nominatif du tirage est rappelé et l'individu doit confirmer son nom/prénom et son année de naissance pour accéder au questionnaire.

Le schéma de contact et de relances de la phase 1 (annexe 1), qui se déroule sur 11 semaines, peut se décomposer en 4 périodes rythmées par l'envoi des relances postales toutes les 3 semaines environ :

- 1^{ère} – 3^e semaine (J0-J20) : le terrain est purement internet. Au cours de ces 3 semaines, 3 mails de relance sont envoyés aux individus éligibles dans le P1SS2 à J7, J14 et J20 ;
- 4^e – 6^e semaine (J20-J40) : une lettre de relance postale simple est envoyée aux non répondants du P1SS2, ainsi qu'un mail de relance à J33 aux non répondants éligibles. Les non répondants du P1SS1 reçoivent une lettre de relance accompagnée du questionnaire papier et de sa notice ;
- 7^e – 9^e semaine (J40-J60) : une lettre de relance postale simple est envoyée aux non répondants du P1SS1. Dans le P1SS2, 25 000 non répondants éligibles sont sélectionnés pour la relance téléphonique (P1SS2a) et en sont informés via une relance postale simple. Les non répondants non sélectionnés du P1SS2 (P1SS2b) reçoivent une relance postale accompagnée du questionnaire papier et de sa notice. Sur cette période, les non répondants éligibles du P1SS2a reçoivent un mail de relance à J45 et les non répondants éligibles du P1SS2b un mail de relance à J54 ;
- 10^e – 12^e semaine (J60-J80) : une relance postale accompagnée du questionnaire papier et de sa notice est envoyée aux non répondants sélectionnés pour la relance téléphonique du P1SS2a et une lettre de relance postale simple est envoyée aux autres non répondants. Un dernier mail de relance est envoyé à tous les non répondants éligibles à J69.

Dans le protocole, l'envoi du questionnaire papier a été programmé en fonction des coordonnées disponibles pour joindre l'enquêté afin de soutenir le plus longtemps possible la participation en ligne et de réduire ainsi le recours au mode papier dans les populations joignables par mail ou téléphone.

⁷ RNVP : restructuration, normalisation et validation postale. Ce traitement consiste à mettre au format et à corriger les adresses à partir du référentiel géographique construit par le Service National de l'Adresse.

La relance téléphonique a été proposée à un nombre fixé a priori de non répondants afin d'améliorer la participation de populations moins connectées tout en restant dans l'enveloppe budgétaire de la subvention européenne. La sélection par tirage aléatoire à la relance téléphonique des non répondants éligibles constitue en outre un cadre quasi-expérimental permettant de conduire des travaux d'expertise méthodologique sur les effets de mode⁸.

Les répondants en ligne ou par téléphone identifiés Cible 1 en phase 1 ont la possibilité d'enchaîner les deux questionnaires au cours de la même interrogation. Pour décider en connaissance de cause, ils sont informés de la durée des deux questionnaires avant leur démarrage respectif. S'ils enchaînent les questionnaires et complètent le questionnaire de phase 2, ils ne sont plus recontactés. Pour tous les autres sélectionnés pour la phase 2 n'ayant pas enchaîné, le schéma de contact et de relances de la phase 2 distingue quatre sous-échantillons selon le statut Cible, le mode de réponse en phase 1 et les coordonnées disponibles pour les recontacter (annexe 2). Le P2SS1 regroupe les répondants Cible 1 qui ont répondu par internet (dès le 1^{er} mars) et fourni un mail personnel en phase 1. Le P2SS2 regroupe les répondants Cible 1 qui ont répondu par téléphone à partir du démarrage de la relance téléphonique le 13 avril. Le P2SS3 regroupe les répondants Cible 2 ou 3 et Cible 1 papier ayant fourni un mail personnel en phase 1 ainsi que les répondants Cible 1 internet ou papier sans coordonnée. Enfin, le P2SS4 regroupe les répondants Cible 2 ou 3 sélectionnés et les Cible 1 papier sans mail personnel mais avec au moins un téléphone disponible.

Le schéma de contact et de relances de la phase 2 distingue 3 périodes :

- 12 premières semaines ou le terrain de phase 1 : dans le P2SS1 et le P2SS2, un protocole lissé personnalisé est mis en place : un mail ou SMS-avis est envoyé le lendemain de la validation du questionnaire de phase 1 puis des relances par mail ou SMS sont envoyées à intervalle régulier (au maximum 5) jusqu'à la clôture de la phase 1. Dans le P2SS1 seule la réponse par internet est proposée ;
- 13^e – 15^e semaine : une lettre-avis postale est envoyée aux individus du P2SS3 qui sont invités à répondre en ligne uniquement et aux individus du P2SS4 qui peuvent répondre en ligne ou par téléphone. Ces lettres sont doublées de l'envoi d'un mail-avis aux individus éligibles. En parallèle, une lettre de relance simple est envoyée aux individus du P2SS1 et du P2SS2, doublée d'un mail ou sms de relance à quelques jours d'intervalle. La relance téléphonique démarre dans le P2SS1 à la 13^e semaine. Selon les coordonnées disponibles, un mail ou un SMS de relance est envoyé aux non répondants du P2SS3 et du P2SS4 au milieu de cette période ;
- 16^e – 18^e semaine : une ultime lettre de relance postale simple est envoyée à l'ensemble des non répondants des 4 sous-échantillons, doublée d'un mail ou SMS de relance à quelques jours d'intervalle. La relance téléphonique démarre dans le P2SS3 à la 16^e semaine.

⁸ Une communication du SSMSI spécifiquement dédiée à l'évaluation des effets de mesure dans l'enquête Genese est également au programme des JMS 2022 [20].

2. Évaluation du protocole multimode de phase 1

Le succès d'une enquête multimode, principalement auto-administrée par internet, telle que Genese dépend fortement des informations disponibles au démarrage pour contacter les enquêtés et des stratégies de relances mises en œuvre [8]. Cette deuxième partie est ainsi consacrée à l'analyse de l'efficacité du protocole multimode internet/papier/téléphone en termes de participation. Elle examine : la qualité des données de contact fournies par l'Insee (adresses postales, mails, numéros de téléphone), l'apport de l'enrichissement des coordonnées réalisé par le prestataire de collecte, la participation des individus au cours de la collecte et l'efficacité des différents modes de relance proposés dans le dispositif.

2.1 Qualité des coordonnées de contact

2.1.1 Traitements sur les informations de contact

Les coordonnées présentes dans Fidéli sont fournies à l'administration fiscale par les déclarants, soit pour eux même lorsqu'ils sont référents, soit pour leur conjoint ou autres personnes à charge dans le cadre de déclarations communes pour les couples mariés ou pacsés. Néanmoins, en pratique, le référent remplit davantage de coordonnées pour lui-même que pour son conjoint ou les autres personnes du foyer, ce qui rend plus difficile le contact par mail ou téléphone d'un individu lorsque ce dernier n'est pas le référent légal⁹.

Pour améliorer la qualité du fichier de contact, le prestataire de collecte a procédé à une normalisation des adresses postales (essentiellement une correction orthographique du nom des communes ou des rues) et une recherche des personnes déménagées (environ 6 % des adresses) afin de réduire le nombre de plis non distribués. Il a également réalisé un enrichissement des coordonnées téléphoniques. Cette procédure a permis de renseigner une coordonnée téléphonique pour 36 % des individus qui n'en avaient pas, faisant passer la proportion d'individus joignables par téléphone de 77 % à 85 %, sachant que la quasi-totalité des numéros trouvés sont des numéros filaires (97 %). Parmi les 85 % de personnes joignables par téléphone (fixe ou mobile), la moitié possède un numéro de portable (58 %) : 32 % possèdent au moins un numéro de téléphone fixe et un numéro de téléphone portable, 27 % possèdent seulement un ou des numéros de téléphone fixe et 27 % possèdent un ou des numéros de téléphone portable.

Concernant les mails, aucun enrichissement n'a été réalisé. Avec l'application de la règle d'envoi des mails conservant uniquement les mails qualifiés de « raisonnablement » personnels (cf. 1.2.3), la proportion de personnes joignables par mail passe de 71 % dans le fichier à 56 %. Avec cette règle, les conjoints et les enfants majeurs à charge sont moins joignables par mail (respectivement 39 % et 26 %) que les référents légaux du foyer fiscal (77 %) et les autres personnes à charge (59 %).

Après enrichissement en numéros de téléphone et application de la règle d'envoi des mails, 10 % des personnes de l'échantillon n'ont ni mail « personnel » ni numéro de téléphone (fixe ou mobile). Elles constituent le P1SS1 (figure 2.1.1). Les autres personnes ayant au moins une coordonnée de contact mail ou téléphone constituent le P1SS2 et représentent 90 % de l'échantillon dont : 51 % avec un mail « personnel » et au moins un numéro de téléphone, 35 % sans mail « personnel » mais au moins un numéro de téléphone et enfin 5 % de personnes avec un mail « personnel » et pas de numéro de téléphone (figure 2.1.1).

⁹ Ces constats s'appuient sur des travaux non publiés sur la qualité des informations de contact dans Fidéli réalisés par Géraldine Charrance (Ined) et coordonnés par la Direction de la méthodologie et de la coopération statistique et internationale (DMCSI) de l'Insee.

Figure 2.1.1 – Coordonnées de contact avant et après traitements (en effectif et en % de l'échantillon total)

| | Avant traitements | | Après traitements | |
|---|-------------------|--------|-------------------|--------|
| P1SS1 – individus sans mail ni téléphone | 22 546 | 13,3 % | 16 051 | 9,5 % |
| P1SS2 – individus avec un mail ou un téléphone | 146 514 | 86,7 % | 153 009 | 90,5 % |
| <i>Dont :</i> | | | | |
| <i>individus avec mail et un téléphone</i> | 103 233 | 61,1 % | 85 917 | 50,8 % |
| <i>individus avec mail sans téléphone</i> | 16 001 | 9,5 % | 8 628 | 5,1 % |
| <i>individus sans mail et au moins un téléphone</i> | 27 280 | 16,1 % | 58 464 | 34,6 % |

Note : les traitements effectués sur les informations de contact sont : enrichissement des coordonnées téléphoniques et restriction de l'envoi de mail aux adresses qualifiées de « raisonnablement » personnelles.

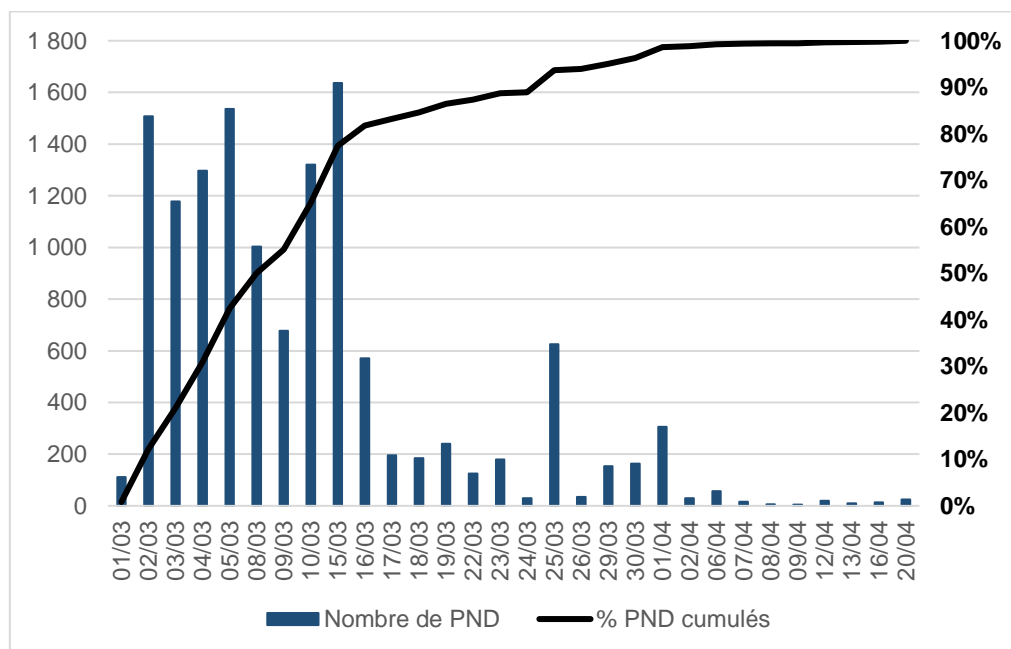
Source : enquête Genese, SSMSI.

2.1.2 Efficacité des informations de contact

Un suivi au fil de l'eau des plis non distribués (PND) a été réalisé pour la lettre-avis grâce au service Alliage de La Poste, ce qui a permis de les écarter des relances postales en cours de collecte.

La lettre-avis envoyée aux 169 060 personnes de l'échantillon a généré un taux de retour en PND de 8 % à la fin du terrain, les derniers retours étant arrivés près de deux mois après la date d'envoi (figure 2.1.2). Au moment de la première relance postale (base des non répondants au 22 mars), 87 % des plis non distribués avaient été reçus. Ces adresses ont ainsi pu être exclues, ce qui a permis d'éviter de nombreuses relances à tort très coûteuses sur des opérations mobilisant de larges échantillons.

Figure 2.1.2 – Plis non distribués (PND) de la lettre-avis selon la date de retour



Note : la lettre-avis a été envoyée le 25 février.

Source : enquête Genese, SSMSI.

La part de PND est plus importante chez les plus jeunes (14 % pour les 18-39 ans), qui sont généralement plus mobiles, et chez les personnes seules (15 %). Les habitants en QPV (11 %) et QRR (12 %) sont également davantage concernés. La part de PND est ainsi plus élevée dans le P1SS1 (15 % vs 7 % dans le P1SS2) composé à 55 % d'individus de moins de 30 ans et à 46 % de personnes seules.

Concernant les mails-avis, 6 % n'ont pas pu être distribués (MND) : 80 % des cas correspondent à des adresses mail erronées, 14 % ont été classés en spam, enfin 6 % n'ont pas été distribués car la boîte mail du destinataire était pleine. La proportion de mails-avis non distribués est moins importante parmi les personnes les plus jeunes (3 % pour les 18-29 ans).

Dans le groupe des personnes relancées à J40 par téléphone, un protocole d'appels et de contacts a été mis en place par le prestataire de collecte pour permettre de gérer l'ensemble des numéros de téléphones disponibles, dans l'ordre de priorité défini conjointement avec l'équipe du SSMSI.

Sur les 39 200 numéros de téléphone disponibles chez les 25 000 individus¹⁰ : 79 % proviennent de Fidéli, 16 % de l'enrichissement, 4 % ont été récupérés par les enquêteurs téléphone durant la collecte, et moins de 1 % ont été renseignés par les enquêtés sur le site internet de l'enquête dans leur espace personnel.

Figure 2.1.3 – Efficacité des coordonnées téléphoniques en fonction de l'origine du numéro dans le P1SS2a

| | Ensemble | Fidéli | Enrichissement | Enquêteurs | Site internet |
|---|----------|--------|----------------|------------|---------------|
| Disponibles | 39 200 | 31 125 | 6 350 | 1 568 | 157 |
| Utilisés | 32 918 | 26 275 | 4 962 | 1 549 | 132 |
| Taux d'utilisation (utilisés/disponibles) | 84,0 % | 84,4 % | 78,1 % | 98,8 % | 84,1 % |
| Taux de faux numéro* | 25,3 % | 27,3 % | 19,8 % | 10,2 % | 6,3 % |
| Taux de hors cible* | 6,9 % | 7,2 % | 3,6 % | 10,8 % | 5,1 % |
| Taux d'injoignables* | 36,8 % | 34,6 % | 55,0 % | 19,6% | 9,2 % |
| Taux de refus/abandon* | 14,5 % | 14,8 % | 13,8 % | 13,0 % | 7,3 % |
| Taux de conversion ¹ | 16,4 % | 16,0 % | 8,1 % | 46,4 % | 65,6 % |

* en % des numéros utilisés ; ¹ Taux de conversion = nombre d'entretiens complets réalisés / nombre de numéros utilisés.

Source : enquête Genese, SSMSI.

27 % des numéros disponibles dans Fidéli utilisés durant la collecte sont des faux numéros (numéro d'entreprise, fax, modem, faux numéro nominatif), contre 20 % des numéros enrichis (figure 2.1.3). En revanche, les numéros utilisés issus de l'enrichissement sont plus souvent des numéros injoignables (numéros où il n'y a jamais eu de décroché) que ceux issus de Fidéli (55 % contre 35 %). Les numéros issus de Fidéli ont de fait un meilleur taux de conversion (16 %), que ceux issus de l'enrichissement (8 %). Ce constat est en lien avec les plus faibles performances des numéros de fixe en général (75 % des entretiens complets ont été réalisés sur un numéro de mobile). Or, moins de 3 % des numéros enrichis dans Genese correspondent à des numéros de mobile.

¹⁰ Les individus sélectionnés pour la relance téléphonique ont 1 à 5 numéros de téléphone renseignés, toutefois ces numéros correspondent aux numéros renseignés pour l'individu cible et pour les autres membres du foyer fiscal. Un même numéro peut ainsi être répété plusieurs fois et n'ont pas été dédoublonnés dans les résultats ici présentés.

Sur les 25 000 individus sélectionnés pour la relance téléphonique (P1SS2a), 12 741 (51 %) n'ont pas pu être joints par téléphone : l'ensemble des numéros disponibles pour ces individus étaient soit des faux numéros, des jamais décrochés (« injoignables ») ou bien n'ont pas été utilisés car l'enquête a répondu en ligne avant d'être appelé. Parmi eux, 5 002 ne présentaient que des faux numéros, soit 20 % du P1SS2a.

Il ressort de l'ensemble de ces constats sur les adresses mails et les numéros de téléphone que la multiplication des moyens de contact permet de réduire le nombre de personnes non contactées.

Au terme de la collecte, seuls 4 721 individus (3 %) n'ont pu être contactés par aucun des moyens mis en place, il s'agit de :

- 2 354 individus du P1SS1 qui n'ont pas reçu la lettre-avis (50 % des cas de « non joints ») ;
- 1 806 individus du P1SS2b non éligibles pour l'envoi de mails et classés PND pour la lettre-avis (38 %) ;
- 276 individus du P1SS2b éligibles pour l'envoi de mails et classés en PND pour la lettre-avis et en MND pour le mail-avis (6 %) ;
- 252 individus du P1SS2a non éligibles pour l'envoi de mails, classés en PND pour la lettre-avis et non joints par téléphone (6 %) ;
- 33 individus du P1SS2a éligibles pour l'envoi de mails, classés en PND pour la lettre-avis, en MND pour le mail-avis et non joints par téléphone (1 %).

2.2 Analyse de la participation

2.2.1 Participation globale à la fin du terrain

À l'issue des 11 semaines de terrain, sur les 169 060 individus de l'échantillon, 108 771 (64 %) ont été classés complets au sens des flux¹¹, 2 829 (2 %) ont commencé le questionnaire sans le compléter et 57 460 (34 %) sont non répondants.

La proportion d'incomplets rapportée aux effectifs de l'échantillon est plus élevée parmi les hommes, les jeunes adultes (18-39 ans), les personnes socialement plus défavorisées (niveau de vie du ménage faible, habitants des QPV ou en logement social) et les résidents des grandes agglomérations (≥ 200 000 habitants). Ainsi les populations plus difficiles à contacter sont aussi plus difficiles à faire adhérer à l'enquête [27].

Au total, la participation globale a dépassé de 14 points les objectifs assignés à la phase 1 de l'enquête, principalement en raison d'une réponse internet beaucoup plus élevée qu'attendue¹². De fait, 85 982 individus ont répondu par internet, ce qui représente plus de la moitié de l'échantillon (51 %). Les autres modes sont minoritaires : 17 382 ont répondu par papier (10 %) et 5 407 (3 %) par téléphone. La participation globale varie de 43 % dans le P1SS1 à 68 % parmi les individus du P1SS2 « avec mail », c'est-à-dire éligibles à l'envoi de mail. Dans ce dernier groupe, la participation internet atteint près de 58 % (figure 2.2.1). Dans le P1SS1, l'envoi du questionnaire papier contribue de moitié à la participation totale observée dans ce groupe.

¹¹ Le critère de complétude au sens des flux correspond à la règle qui met fin aux différentes relances au cours de la collecte : l'individu est compté comme complet si l'écran de fin du questionnaire est atteint pour les répondants en ligne ou par téléphone ou bien si le questionnaire papier a été retourné par courrier. Ce critère est distinct des règles de complétude mises en œuvre pour les traitements statistiques post-collecte.

¹² Dans l'expérimentation internet « Vols Violences et Sécurité » (VVS) menée en 2013, la participation avait atteint seulement 32 % [18].

Figure 2.2.1 – Participation globale par mode et par sous-échantillon

| | P1SS1 | P1SS2 | | | Ensemble |
|--|--------|-----------|-----------|----------|----------|
| | | sans mail | avec mail | Ensemble | |
| Total échantillon à J0 | 16 051 | 58 464 | 94 545 | 153 009 | 169 060 |
| Total répondants | 6 841 | 37 457 | 64 473 | 101 930 | 108 771 |
| Participation (en % de l'échantillon à J0) | 42,6 | 64,1 | 68,2 | 66,6 | 64,3 |
| dont : | | | | | |
| <i>internet</i> | 22,5 | 47,4 | 57,8 | 53,8 | 50,9 |
| <i>papier</i> | 20,1 | 12,6 | 7,2 | 9,2 | 10,3 |
| <i>téléphone</i> | NE | 4,1 | 3,2 | 3,5 | 3,2 |

NE : non éligibles.

Note : les « répondants » désignent les individus ayant complété le questionnaire au sens des flux.

Lecture : 50,9 % de l'échantillon total de Genese a répondu par internet à l'enquête.

Source : enquête Genese, SSMSI.

Globalement, le téléphone contribue de 3 points à la participation. Cela peut sembler faible mais la relance téléphonique n'est proposée qu'à 25 000 non répondants (P1SS2a) après environ 6 semaines de collecte (J40), soit 15 % de l'échantillon initial et 29 % des non répondants à J40.

Au total, 22 % des individus du P1SS2a ont répondu par téléphone à l'enquête. Ce taux de collecte téléphone auprès des non répondants sélectionnés s'est avéré inférieur aux objectifs (30 %). Il était attendu que les non répondants à ce stade de la collecte seraient plus difficiles à mobiliser, néanmoins le facteur le plus limitant du terrain téléphonique s'est révélé être la joignabilité des individus. Plus de 6 numéros sur 10 disponibles ont en effet été classés en faux numéros ou « injoignables » (cf. 2.1.2) malgré un protocole prévoyant de nombreuses tentatives sur des plages horaires variées et incluant au moins un appel sur un jour de week-end ainsi que le dépôt d'un message sur répondeur après 3 appels infructueux sur le 1^{er} numéro offrant la possibilité de laisser un message. Finalement, dans le P1SS2a, la participation atteint à la fin de la collecte 53 % des effectifs à J40 contre 35 % dans le P1SS2b et 36 % parmi les individus éligibles à la relance téléphonique dans le P1SS2b (figure 2.2.2). Le téléphone contribue fortement à cet écart mais il est également intéressant de noter que dans le P1SS2a la participation internet est plus importante que dans le P1SS2b : 23 % versus 16 %, soit +7 pts. A contrario, la réponse papier apparaît logiquement plus résiduelle dans le P1SS2a (9 % versus 19 %), le questionnaire papier n'étant envoyé dans ce groupe que lors de l'ultime relance à J60.

Figure 2.2.2 – Participation globale par mode dans le P1SS2a et le P1SS2b

| | P1SS2a | P1SS2b | P1SS2b éligibles* |
|---|--------|--------|-------------------|
| Total échantillon à J40 | 25 000 | 61 123 | 54 999 |
| Total répondants | 13 335 | 21 709 | 19 893 |
| Participation (en % de l'échantillon à J40) | 53,3 | 35,5 | 36,2 |
| dont : | | | |
| <i>internet</i> | 22,6 | 16,1 | 15,9 |
| <i>papier</i> | 9,1 | 19,4 | 20,3 |
| <i>téléphone</i> | 21,6 | NC | NC |

* éligibles à la relance téléphonique c'est-à-dire avec au moins un numéro de téléphone disponible. NC : non concernés. Note : les « répondants » désignent les individus ayant complété le questionnaire au sens des flux.

Lecture : 21,6 % du P1SS2a a répondu par téléphone à l'enquête.

Source : enquête Genese, SSMSI.

2.2.2 Avancée de l'enquête

La répartition des individus classés complets sur les 4 périodes de la phase 1 fait apparaître quelques constats intéressants. La première période (les 3 premières semaines) concentre 30 % de l'échantillon initial (figure 2.2.3). Les trois périodes suivantes regroupent entre 10 % et 12 % de l'échantillon. Le début de la collecte, exclusivement internet, est particulièrement « explosif » (36 %) dans le P1SS2 avec mail : l'écart est de minimum 25 points entre la première période et les suivantes. Ce constat reste vrai mais dans une moindre mesure dans le P1SS2 sans mail. Dans le P1SS1, la progression de l'enquête présente des écarts moins marqués, la première période n'enregistrant pas le score le plus élevé.

Une autre façon d'analyser la progression de la collecte consiste à regarder le nombre de personnes qui répondent au cours d'une période donnée rapporté au nombre de personnes non répondantes au début de la période afin de mesurer l'impact global des différentes communications (figure 2.2.3). Cette analyse montre que les relances en fin de collecte (J60-J80) mobilisent fortement les non répondants à J60 dans le P1SS2 et dans une moindre mesure dans le P1SS1. Les retours des questionnaires papier envoyés massivement à J40, la relance téléphonique dans le P1SS2a mais aussi possiblement l'annonce de la date de fin de collecte dans la relance de J60 expliquent sans doute cette mobilisation. Enfin, autre constat intéressant, passée la première période, l'envoi des mails semble avoir un impact plus limité dans le P1SS2.

Figure 2.2.3 – Répartition des participants par sous-échantillon et par période

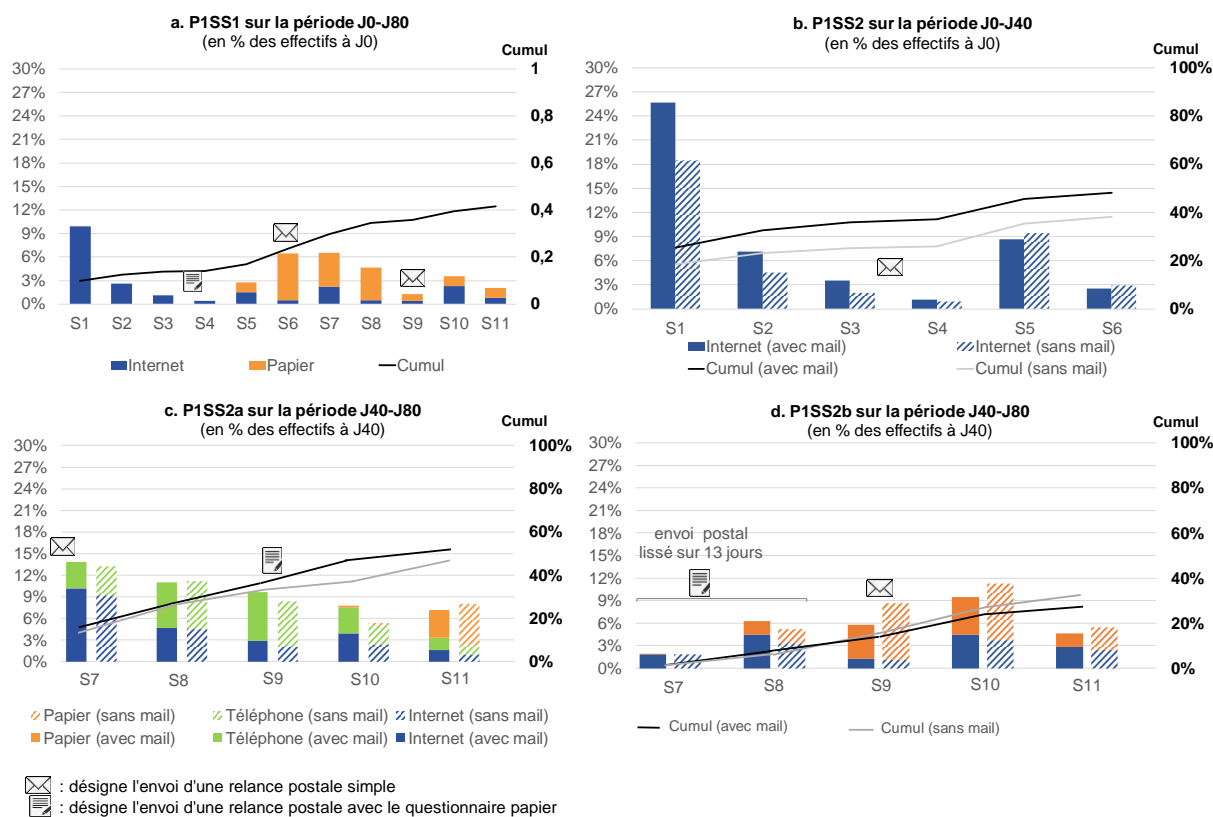
| Période | P1SS1 | | P1SS2 | | | | Ensemble | |
|---------|----------------------|-----------------------|----------------------|-----------------------|----------------------|-----------------------|----------------------|-----------------------|
| | en % N _{J0} | en % N _{JDP} | sans mail | | avec mail | | en % N _{J0} | en % N _{JDP} |
| | | | en % N _{J0} | en % N _{JDP} | en % N _{J0} | en % N _{JDP} | | |
| J0-J20 | 13,7 | 13,7 | 25,0 | 25,0 | 36,4 | 36,4 | 30,3 | 30,3 |
| J20-J40 | 7,2 | 8,4 | 12,1 | 16,1 | 11,4 | 17,9 | 11,2 | 16,1 |
| J40-J60 | 14,1 | 17,9 | 11,4 | 18,1 | 9,2 | 17,6 | 10,4 | 17,8 |
| J60-J80 | 7,5 | 11,6 | 15,6 | 30,2 | 11,2 | 26,1 | 12,4 | 25,7 |

Note : N_{J0} désigne l'effectif au démarrage de la collecte (J0) ; N_{JDP} désigne l'effectif de non répondants dans le groupe en début de période.

Lecture : parmi les individus du P1SS2 avec mail, la participation sur la période J40-J60 représente 9,2 % des effectifs à J0 et 17,6 % des effectifs de non répondants en début de période (J40).

Les graphiques d'avancée d'enquête, distinguant les échantillons, l'envoi de mail et le mode de réponse, permettent d'illustrer et d'affiner ces premiers constats (figure 2.2.4). Ils font apparaître que l'envoi des relances postales a un effet important sur la participation – en particulier celle envoyée à J40 (S6 et S7 sur les graphiques). Dans le P1SS2, pendant la phase de collecte « pur internet » (six premières semaines, J0-J40), l'envoi de mail a un effet « brut » important au démarrage puis l'écart se maintient (figure 2.2.4b). Après J40, dans le P1SS2 (figure 2.2.4c), l'envoi de mail ne semble pas présenter d'effet « brut » significatif sur la participation. Dans le P1SS2b, l'effet des mails semble même épuisé auprès de la population avec mail puisque la population sans mail se mobilise davantage au cours des dernières semaines (figure 2.2.4d). Enfin, l'analyse hebdomadaire de la participation fait apparaître un pic de réponses par internet dans la semaine qui suit l'envoi de la relance postale à J40 bien plus élevé dans le P1SS2a que dans le P1SS2b (10 % vs 2 %). L'envoi à J40 de la relance postale accompagnée du questionnaire papier dans le P1SS2b s'est étalée sur près de 2 semaines en raison des volumes importants concernés, ce qui a pu diluer son effet et contribuer aux écarts observés en début de période. Néanmoins, l'écart de participation internet sur la période est important. Il est donc possible que, dans le P1SS2a, l'annonce du démarrage de la relance téléphonique ait provoqué un report vers le questionnaire en ligne afin d'échapper à un mode de collecte jugé plus intrusif.

Figure 2.2.4 – Participation hebdomadaire et cumulée par sous-échantillon et par mode



Source : enquête Genese, SSMSI.

Note : les questionnaires papier retournés après la fin des 11 semaines de la phase 1 ont été saisis jusqu'au 31 mai soit jusqu'en fin de semaine 12, ce qui explique les légers écarts sur le cumul de participation entre les graphiques de la figure 2.2.4 et la figure 2.2.5.

Source : enquête Genese, SSMSI.

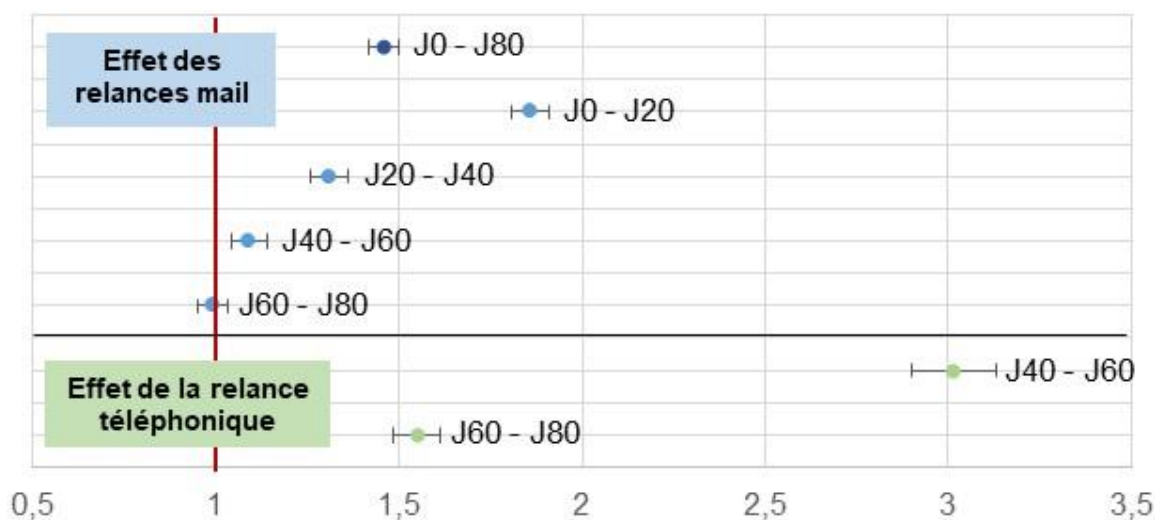
2.2.3 Effet des relances

La composition des sous-échantillons et du schéma de relances dépend des coordonnées de contact disponibles au démarrage de la collecte. La présence de ces informations (mail, téléphone) dépend elle-même des caractéristiques sociodémographiques des individus sélectionnés. Afin d'analyser l'effet des relances au cours des différentes périodes du terrain de phase 1, une analyse multivariée par régressions logistiques permettant de contrôler les caractéristiques des individus a été conduite. Les variables de contrôle introduites dans le modèle sont les informations disponibles dans la base de sondage : sexe, âge, type de ménage fiscal, statut dans le foyer fiscal (réfèrent, conjoint, enfant à charge ou autre personne à charge), décile de niveau de vie, la taille d'unité urbaine en 5 tranches, une indicatrice de résidence en Quartier prioritaire de la ville (QPV) et enfin la présence d'une coordonnée téléphonique. La propension des individus à répondre en fonction de leurs caractéristiques sociodémographiques fait l'objet d'un examen approfondi dans la partie 3. Concernant l'évaluation du protocole *stricto sensu*, les résultats de cette analyse permettent d'établir, conformément aux résultats obtenus par expérimentation randomisée lors du pilote de l'enquête, que l'envoi de mail a un effet significatif toutes choses égales par ailleurs sur la probabilité de participer au cours des 11 semaines (odds-ratio (OR) sur la période J0-J80 [IC95 %] : 1,46 [1,42 ; 1,50] ; figure 2.2.5 et annexe 3). En modélisant la probabilité de participer au cours des différentes périodes de la phase 1 rythmée par l'envoi des lettres de relance, l'analyse permet de confirmer un résultat apparent dans les graphiques

d'avancés d'enquête : l'effet des mails s'essouffle au cours de la collecte. Très élevé et significatif entre J0 et J20, il baisse graduellement au cours des semaines suivantes pour finir non significatif entre J60 et la fin du terrain. Cet effet n'est pas attribuable au nombre de relances programmées au cours de chaque période (3 entre J0 et J20, 2 entre J20 et J40 puis une entre J40 et J60 ainsi qu'entre J60 et J80). En effet, en conduisant la même analyse sur chaque sous-période définie par l'envoi des relances par mail, l'impact des mails apparaît croissant au début de collecte et maximal à la 2^e relance envoyée au bout de la 2^e semaine puis décroît régulièrement dès la 3^e relance (annexe 4). L'envoi rapproché de mails en début de collecte apparaît donc comme une stratégie importante pour booster la réponse internet. Des analyses complémentaires modélisant la probabilité de répondre au cours du temps pourraient être utilement mobilisées pour confirmer ces résultats en permettant de tester la significativité des écarts mesurés.

Un autre résultat intéressant est mis en lumière par cette analyse : toutes choses étant égales par ailleurs, l'effet de la relance téléphonique est particulièrement élevé entre J40 et J60 (3,01 [2,90 ; 3,13]) au démarrage des appels et ralentit fortement sur la période suivante au cours de laquelle l'effet apparaît moins élevé (1,55 [1,49 ; 1,62]). En modélisant la probabilité de répondre spécifiquement par internet sur ces deux périodes en contrôlant des mêmes variables, l'analyse confirme le constat dressé à partir des graphiques d'avancée d'enquête : toutes choses étant égales par ailleurs, la relance téléphonique booste significativement la réponse par internet sur la période J40-J60. En revanche, sur la période suivante, l'effet n'est plus significatif.

Figure 2.2.5 – Effet des relances par mail et par téléphone sur la probabilité de répondre au cours des différentes périodes du terrain de phase 1 dans le P1SS2



Note : 5 modèles de régression logistique ont été mis en œuvre afin d'évaluer l'effet de l'envoi de mails, de la présence d'une coordonnée téléphonique et l'effet de la relance téléphonique dans le P1SS2 au cours des différentes périodes de la phase 1. Le premier modèle « J0 – J80 » (N=142 101) mesure la probabilité de répondre entre J0 et la fin du terrain parmi les individus du P1SS2 ayant reçu la lettre-avis ; le deuxième modèle « J0-J20 » (N=142 101) mesure la probabilité de répondre entre J0 et J20 parmi les individus du P1SS2 ayant reçu la lettre-avis ; le troisième modèle « J20-J40 » (N=91 055) mesure la probabilité de répondre entre J20 et J40 parmi les individus du P1SS2 ayant reçu la première lettre de relance ; le quatrième modèle « J40-J60 » (N=73 905) mesure la probabilité de répondre entre J40 et J60 parmi les individus du P1SS2 ayant reçu la deuxième lettre de relance ; le cinquième modèle « J60-J80 » (N=55 635) mesure la probabilité de répondre entre J60 et J80 parmi les individus du P1SS2 ayant reçu la troisième lettre de relance. Les variables de contrôle introduites dans le modèle sont les informations disponibles dans l'échantillon (sexe, âge en 5 tranches, type de ménage en 5 catégories, statut fiscal en 4 catégories, décile de niveau de vie, TUU en 5 tranches, indicatrice de résidence en QPV et présence d'un

numéro de téléphone). Les données représentées sur le graphique correspondent aux odds-ratios et leurs intervalles de confiance à 95 % des indicatrices « envoi de mail » et « sélection pour la relance téléphonique » obtenus dans les 5 différents modèles.
Source : enquête Genese, SSMSI.

2.3 Analyse du comportement des enquêtés

2.3.1 Refus et désinscriptions

La participation à la phase 1 de l'enquête Genese a été particulièrement élevée, notamment sur internet, et le protocole mis en place, rythme et nature des relances, semblent avoir contribué à ce succès. L'envoi soutenu de mails, en particulier au démarrage de la collecte, et le nombre de relances postales combinés à la diversification des supports de collecte présentent-ils, en miroir, un effet délétère ? L'ensemble des non répondants à l'enquête Genese n'a pas été consulté et il n'est donc pas possible de conclure de manière contrôlée à la question de savoir si les stratégies de relance ont pu constituer un frein à la participation. Néanmoins, les données collectées concernant les refus exprimés auprès de la hotline ou lors de la relance téléphonique ainsi que les demandes de désinscriptions de l'envoi de mail apportent des éléments d'appréciation plutôt rassurants.

Lors du terrain téléphonique, alors qu'il s'agissait d'une relance auprès de personnes non répondantes après 6 semaines de collecte – soit après 2 lettres postales et 5 mails pour les éligibles - les refus fermes ont été relativement faibles : 10 % des appelés (2 553 individus) ont en effet décliné la participation à l'enquête. Les remontées des enquêteurs ont fait part d'une bonne acceptation du questionnaire de phase 1 dans l'ensemble. La réception de la lettre-avis décrivant l'enquête s'est révélée être un important facilitateur : le fait de ne pas l'avoir reçue pouvait susciter une méfiance de la part des enquêtés, en particulier chez les plus âgés qui se sentaient de manière générale moins concernés par le sujet.

Outre les refus exprimés lors du terrain téléphonique, 201 individus parmi les 10 990 ayant contacté la hotline au cours de la phase 1 (2 %) ont signifié leur refus de participer. Les caractéristiques des refusants via la hotline sont concordantes avec les remontées des enquêteurs téléphone : 38 % des refusants ont plus de 60 ans, 62 % ont plus de 50 ans. De manière générale, les principaux motifs de contact à la hotline ont été des demandes d'informations sur l'enquête (30 %) et des signalements sur l'impossibilité de répondre par internet (15 %).

Par ailleurs, dans chaque mail envoyé (avis et relance), un lien situé en bas du message offrait la possibilité au destinataire de demander la désinscription des envois de mail. Une fois confirmée, la demande de désinscription excluait automatiquement et immédiatement le demandeur de toute relance par mail. La désinscription des mails ne valait cependant pas refus, ce qui signifie que l'envoi de relances postales et la sélection pour la relance téléphonique restaient possibles. Sur la phase 1 de l'enquête Genese, 2 350 désinscriptions ont été enregistrées, ce qui représente 2,5 % des individus sollicités par mail. La moitié de ces désinscrits ont tout de même répondu à l'enquête, ce qui valide le choix de ne pas les considérer comme refusants. Le pourcentage de désinscriptions enregistré après chaque relance mail reste inférieur à 1 % à chaque vague de relance : les quatre premiers envois de mail ont provoqué un taux de désinscription d'environ 0,5 %, puis on observe une légère augmentation du taux de désinscription sur les quatre relances suivantes (entre 0,6 % et 0,8 %).

2.3.2 Auto-administration et respect des consignes

Le respect des consignes données aux enquêtés dans le cadre de dispositifs principalement auto-administrés est un enjeu particulièrement important. Tout au long de la collecte, deux consignes importantes étaient données aux enquêtés. La première concernait la désignation de la personne sélectionnée et l'interdiction du proxy¹³. La deuxième consigne spécifiait les supports à privilégier pour la réponse en ligne « un ordinateur ou une tablette ».

Concernant la consigne désignant la personne sélectionnée pour l'enquête, il n'est évidemment pas possible d'évaluer le taux de proxy avec certitude, néanmoins certains cas suspects peuvent être détectés en confrontant le sexe et la date de naissance connus et ceux déclarés. Les cas discordants sont extrêmement minoritaires : 380 individus sur les 108 771 répondants de phase 1 (<1 %) ont déclaré un sexe différent de celui enregistré dans Fidéli, et 15 individus sans sexe connu dans Fidéli n'ont pas déclaré de sexe dans le questionnaire (répondants papier). Par ailleurs, 457 individus ont déclaré un âge différent de celui connu dans Fidéli, et 720 n'ont pas répondu sur le questionnaire papier.

Sur le respect des supports à privilégier, les communications suggéraient « de préférence » de se connecter depuis un ordinateur ou une tablette car les questionnaires de l'enquête sont complexes et un certain nombre de questions étaient délibérément proposées sous forme de tableaux (en particulier le recensement des victimations d'une même famille d'atteintes). Néanmoins, le questionnaire internet a été développé dans l'optique d'être accessible sur tous les supports (questionnaire « *device agnostic* »), y compris smartphone (sans affichage de tableaux). Parmi les répondants internet, 19 % se sont connectés et ont complété le questionnaire sur un smartphone (75 % sur ordinateur et 6 % sur tablette). Moins de 1 % des répondants internet ont basculé d'un support vers l'autre. L'utilisation des smartphones est particulièrement importante parmi les 18-29 ans et les 30-39 ans (30 % dans les deux groupes).

2.3.3 Les usages de réponse sur internet

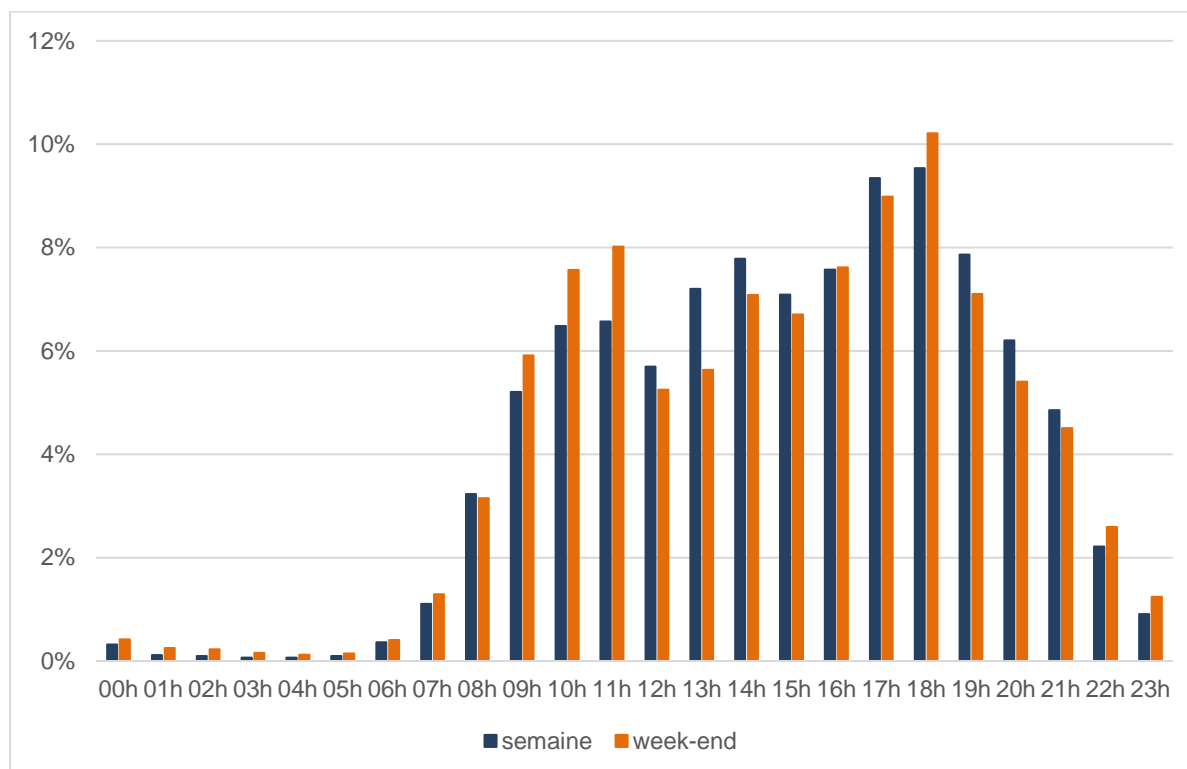
De nombreuses données sont collectées à partir du questionnaire en ligne permettant de renseigner les concepteurs sur les usages des enquêtés en matière de comportement de réponse par internet.

Lors du pilote de Genese, une question d'appréciation sur le mode de réponse montrait ainsi que la réponse en ligne donnait davantage satisfaction que la réponse par téléphone. Ainsi, 91 % des répondants internet avaient répondu concernant le mode de réponse « c'était très bien comme ça » contre 76 % des répondants téléphone. Par ailleurs, 24 % des répondants téléphone estimaient qu'ils auraient préféré répondre par internet contre seulement 2 % des répondants internet qui auraient préféré répondre par téléphone.

Concernant les heures de connexion préférées pour répondre par internet, les données de Genese confirment des résultats connus dans la littérature ([10], le créneau préféré est aussi 18h-19h). Un pic de réponses est enregistré en fin de journée : un tiers des individus qui ont répondu au questionnaire sur internet a répondu entre 17h et 21h, 10 % se sont connectés à 18h (figure 2.3.1). Au total, 24 % des connexions au questionnaire internet ont lieu en dehors des heures d'appels de la relance téléphone, à savoir le dimanche, le samedi entre 16h et 9h et entre 21h et 9h les autres jours.

¹³ Pour rappel, le tirage est nominatif et précise les nom/prénom de l'individu sélectionné pour plus de 98 % de l'échantillon, sinon la date de naissance. La phrase expliquant la règle indiquait « Vous faites partie des personnes sélectionnées aléatoirement pour participer à l'enquête. **Personne ne peut répondre au questionnaire à votre place.** Le respect de cette consigne est primordial pour la fiabilité des résultats. » dans les courriers postaux et « Ce lien est personnel et unique, personne ne peut répondre au questionnaire à votre place. » dans les mails.

Figure 2.3.1 – Heure de connexion au questionnaire internet



Champ : 78 165 répondants internet en une seule connexion.

Source : enquête Genese, SSMSI.

3. Analyse de la représentativité

Le taux de participation est un indicateur nécessaire mais insuffisant pour mesurer l'impact potentiel du biais de non-réponse à une enquête. On trouve dans la littérature des études montrant que certains protocoles de collecte sont susceptibles d'améliorer le taux de participation mais dans le même temps conduisent à une amplification du biais de non-réponse [14].

En effet, le biais de non-réponse des estimateurs est déterminé par deux facteurs [4] :

- le taux de réponse à l'enquête : plus il est faible, plus l'impact potentiel du biais de non-réponse est élevé ;
- la divergence en moyenne des profils entre répondants et non répondants par rapport à la variable d'intérêt cible : plus ils diffèrent, plus le biais sera large.

La représentativité des répondants à l'enquête Genese est ici étudiée, au sens de la similarité des caractéristiques des répondants par rapport aux individus enquêtés initialement. Cette analyse s'appuie sur les variables auxiliaires disponibles et ne préjuge pas entièrement du potentiel biais de non-réponse : ce biais dépend également des corrélations entre variables auxiliaires et variables d'intérêt et des éventuels mécanismes de non-réponse non ignorable. Toutefois, cette analyse de la représentativité, menée au regard du protocole de l'enquête, permet d'appréhender en première approximation les apports éventuels des différents modes de collecte sur cette dimension.

Les différences standardisées sont utilisées pour mesurer la représentativité des répondants par rapport aux individus enquêtés. Cette mesure s'affranchit de la taille de l'échantillon et permet de quantifier la distance de deux échantillons, variable par variable.

3.1 Évolution de la représentativité au cours de la collecte

Selon l'avancée de la collecte aux dates clés de l'enquête, à savoir à J20 (au moment où le papier est introduit dans le P1SS1), à J40 (au moment où le téléphone est introduit dans le P1SS2a et le papier dans le P1SS2b), à J60 (au moment où le papier est introduit dans le P1SS2a), et à la fin de la collecte, à J80, la répartition des caractéristiques des répondants présente des différences en comparaison à la composition de l'échantillon initial (figure 3.1.1).

Figure 3.1.1 – Structure de l'échantillon et des répondants à date (en %)⁽¹⁾

| | Echantillon | Répondants à J20 | Répondants à J40 | Répondants à J60 | Répondants à J80 |
|--------------------------------------|-------------|------------------|------------------|------------------|------------------|
| <i>Effectif</i> | 169 060 | 51 247 | 70 249 | 87 872 | 108 771 |
| Sexe | | | | | |
| Femme | 49,9 | 48,6 | 49,5 | 50,2 | 50,8 |
| Homme | 50,1 | 51,4 | 50,5 | 49,8 | 49,2 |
| Age | | | | | |
| 18-29 ans | 19,5 | 16,9 | 17,0 | 17,2 | 17 |
| 30-39 ans | 17,9 | 17,3 | 16,9 | 16,8 | 16,6 |
| 40-49 ans | 18,4 | 20,1 | 19,9 | 19,4 | 19,1 |
| 50-59 ans | 18,9 | 20,2 | 20,5 | 20,3 | 20,1 |
| 60-74 ans | 25,4 | 25,5 | 25,7 | 26,3 | 27,2 |
| Type de commune⁽²⁾ | | | | | |
| Rurale | 22,2 | 22,3 | 22,7 | 23,1 | 23,6 |
| Urbaine | 77,8 | 77,7 | 77,3 | 76,9 | 76,4 |
| Résidence en QPV | | | | | |
| QPV | 7,1 | 4,0 | 4,4 | 4,8 | 5,2 |
| Hors QPV | 92,9 | 96,0 | 95,6 | 95,2 | 94,8 |
| Type de ménage (fiscal) | | | | | |
| Personne seule | 18,7 | 15,2 | 14,0 | 15,3 | 15,9 |
| Couple sans enfant | 23,5 | 26,2 | 25,9 | 25,8 | 25,6 |
| Couple avec enfants | 40,1 | 44,4 | 44,4 | 43,8 | 43,1 |
| Famille monoparentale | 10,9 | 8,9 | 9,2 | 9,4 | 9,6 |
| Autre type de ménage | 6,8 | 5,3 | 5,5 | 5,6 | 5,8 |
| Décile de niveau de vie | | | | | |
| d01 | 8,9 | 5,4 | 5,8 | 6,2 | 6,6 |
| d02 | 8,7 | 5,6 | 6,1 | 6,6 | 7,2 |
| d03 | 8,7 | 6,6 | 7,0 | 7,3 | 7,8 |
| d04 | 8,8 | 7,4 | 7,7 | 8,1 | 8,4 |
| d05 | 9,2 | 8,7 | 8,9 | 9,1 | 9,4 |
| d06 | 9,5 | 9,8 | 9,9 | 10,0 | 10,0 |
| d07 | 10,1 | 11,7 | 11,6 | 11,4 | 11,2 |
| d08 | 10,2 | 12,9 | 12,6 | 12,2 | 11,8 |
| d09 | 10,3 | 14,2 | 13,5 | 12,9 | 12,2 |
| d10 | 10,5 | 15,0 | 14,2 | 13,3 | 12,5 |
| NA | 5,1 | 2,6 | 2,6 | 2,7 | 2,8 |

Notes : ⁽¹⁾ Les structures de l'échantillon et des répondants sont pondérées avec le poids de tirage ; ⁽²⁾ Le type de commune est décomposé en communes rurales (communes n'appartenant pas à une unité urbaine) et urbaines (communes appartenant à une unité urbaine de plus de 2 000 habitants).

Source : enquête Genese, SSMSI.

La proportion d'hommes et de femmes répondants tout au long du terrain est relativement stable, avec une légère surreprésentation des femmes en fin de collecte.

Les 60-74 ans sont très légèrement surreprésentés au début du terrain, à J20, et leur surreprésentation s'accroît avec le temps.

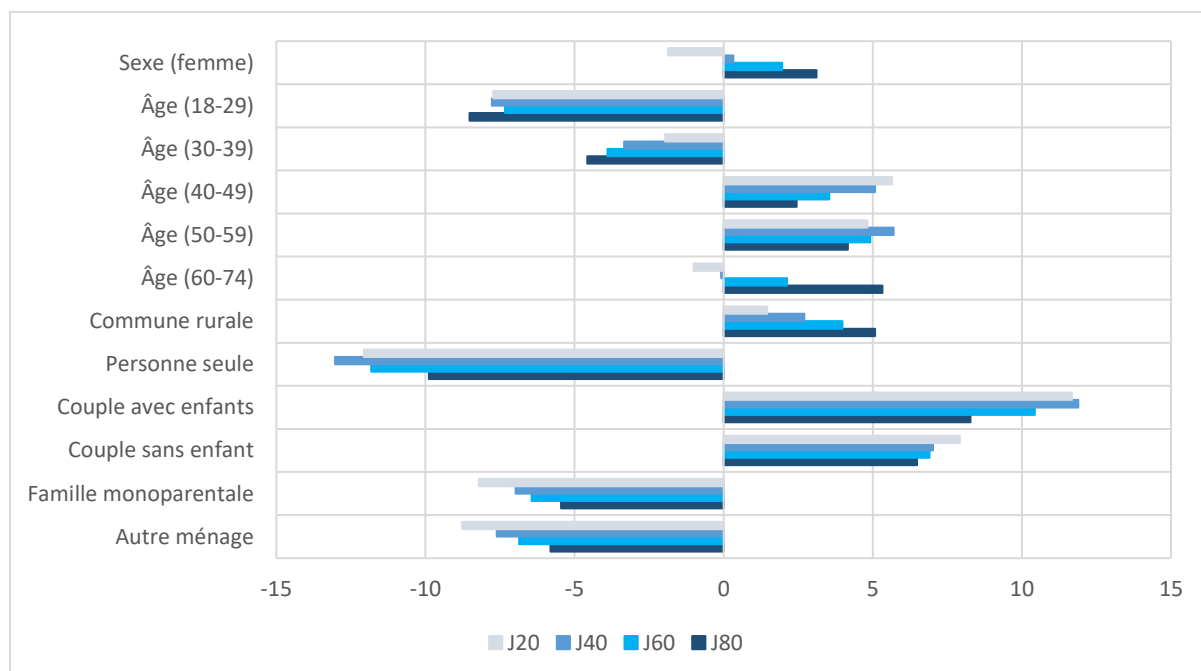
Les habitants en Quartier Prioritaire de la Ville (QPV), de même que les déciles de niveau de vie les plus faibles, sont sous-représentés à J20 : leurs proportions parmi les répondants sont croissantes au fur et à mesure du terrain, bien qu'ils restent sous-représentés en fin de collecte par rapport à la composition initiale de l'échantillon. Les ménages composés de personnes seules, familles monoparentales et autres types de ménage connaissent le même schéma d'évolution.

Les différences standardisées sont mesurées à chaque étape de la collecte, en comparaison avec l'échantillon initialement tiré. La différence standardisée mesure la différence des moyennes d'une variable dans deux groupes, rapportée à une mesure de dispersion (écart-type intragroupe). Pour une variable dichotomique et deux groupes, à savoir l'échantillon (contrôle) et les répondants à date (traitement) :

$$d = 100 \times \frac{\hat{p}_{\text{traitement}} - \hat{p}_{\text{contrôle}}}{\sqrt{\frac{\hat{p}_{\text{traitement}}(1 - \hat{p}_{\text{contrôle}}) + \hat{p}_{\text{contrôle}}(1 - \hat{p}_{\text{traitement}})}{2}}}$$

Les différences standardisées sont ainsi mobilisées dans la littérature sur l'évaluation causale des effets de traitements pour vérifier les propriétés équilibrantes des échantillons d'observations appariées. Dans ce contexte particulier, d'après Austin et Stuart (2015) [2], une différence standardisée inférieure à 10 est négligeable ; plus elle s'élève, plus elle signe un déséquilibre préoccupant.

Figure 3.1.2 – Différences standardisées selon l'avancée de la collecte



Lecture : les femmes répondantes à J20 ont une différence standardisée de -1,9 point par rapport à la structure de l'échantillon tiré.

Source : enquête Genese, SSMSI.

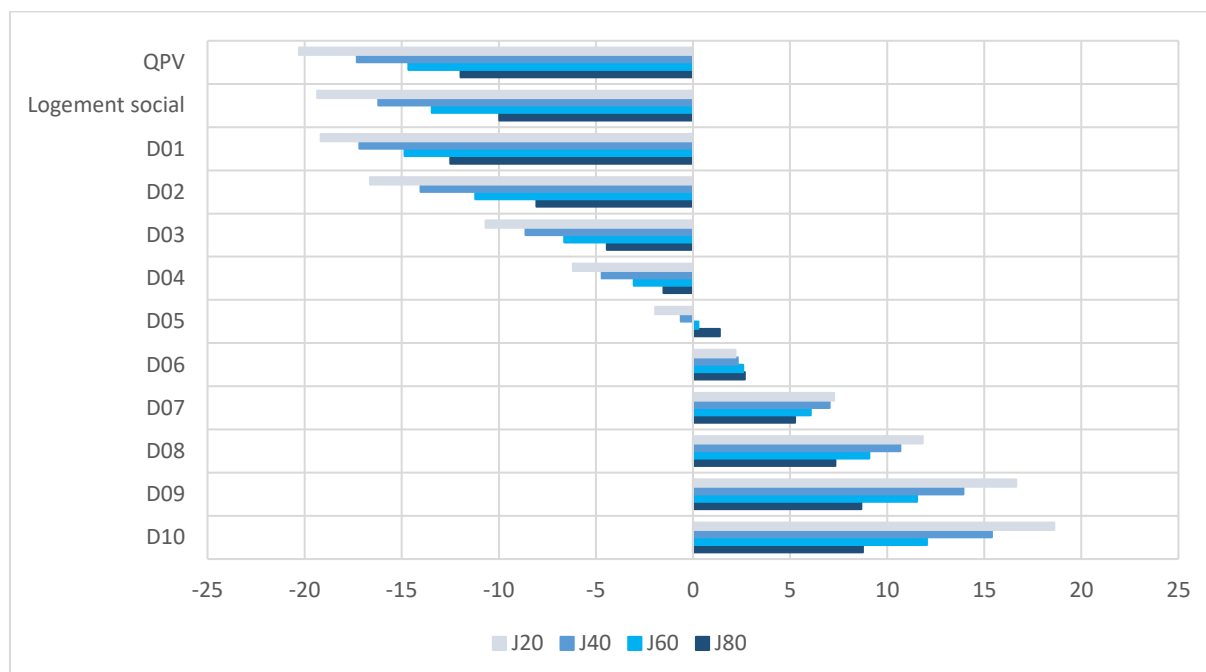
Si les femmes sont légèrement sous-représentées parmi les répondants à J20 (-1,9), elles finissent légèrement surreprésentées à la fin de la collecte à J80 (3,1), mais ces différences restent peu importantes (figure 3.1.2). De la même manière, les moins de 40 ans et les habitants de communes urbaines sont sous-représentés tout au long de la collecte avec un écart faible par rapport à l'échantillon initial.

En début de collecte, les ménages de personnes seules et les couples avec enfants sont respectivement sous-représentés et surreprésentés à des niveaux assez marqués (-12,1 et 11,7), mais ces différences s'atténuent au cours de la collecte pour redescendre sous les 10 points en fin de collecte.

Les populations les moins favorisées sont globalement sous-représentées parmi les répondants, illustrant le lien assez classique entre revenus/conditions de vie et propension à répondre (figure 3.1.3). Mais l'évolution des différences standardisées souligne un résultat assez remarquable : sur ces dimensions, les structures se rapprochent de l'échantillon initial au fur et à mesure de l'avancée de la collecte. Les différences standardisées sont néanmoins assez marquées pour les habitants en QPV, en logement social ou encore dans les déciles de niveau de vie les plus bas, que les enquêtes ont historiquement du mal à aller chercher, en particulier dans les collectes où le mode de réponse principal est internet [3 ; 9].

Ainsi, malgré une amélioration au fil du terrain, en fin de collecte, les habitants en QPV et le décile le plus pauvre restent à des niveaux de sous-représentation importants (-12,0 et -12,5). En revanche, l'avancée de la collecte permet de se rapprocher de la structure de l'échantillon pour les autres déciles de niveau de vie et la résidence en logement social.

Figure 3.1.3 - Différences standardisées selon l'avancée de la collecte



Lecture : les résidents en QPV répondants à J20 ont une différence standardisée de -20,3 points par rapport à la structure de l'échantillon tiré.

Source : enquête Genese, SSMSI.

La représentativité des répondants aux différentes étapes de la collecte peut également être appréhendée par les coefficients de variation des propensions à répondre. Ces indicateurs sont liés à la littérature sur les indicateurs de représentativité et en particulier la théorie sur les R-indicateurs [22].

Le R-indicateur est une mesure du manque d'association entre réponse et variables auxiliaires, ou dit autrement, de la similarité entre répondants et population enquêtée [16]. Il s'exprime comme :

$$R(\rho) = 1 - 2 S(\rho)$$

avec :

$$S(\rho) = \sqrt{\frac{1}{N-1} \sum \omega_i (\rho_i - \bar{\rho})^2}$$

où N est la taille de la population, ω_i le poids de sondage, ρ_i la propension individuelle à répondre et $\bar{\rho}$ la propension moyenne (pondérée) à répondre.

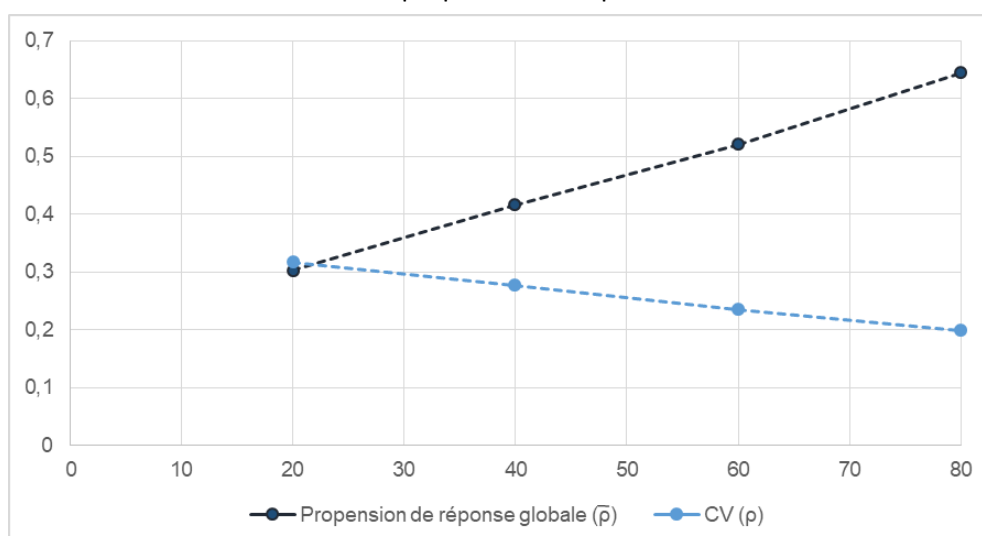
Empiriquement, les propensions individuelles à répondre étant inconnues, elles sont estimées par un modèle (typiquement par régression logistique) qui fait intervenir des variables auxiliaires¹⁴. Le R-indicateur global est compris entre 0 et 1, 1 signifiant une réponse parfaitement représentative, et 0 signifiant une déviation maximale de la réponse représentative. Néanmoins, cet indicateur présente une principale propriété non désirable (trajectoire en fonction de l'avancée de la collecte « en U » qui reflète non pas uniquement la représentativité intrinsèque mais la faible dispersion des propensions à répondre lorsque les taux de réponse sont faibles). Ainsi, l'utilisation des coefficients de variations des propensions à répondre sont préférées dans certaines situations [17] :

$$CV(\rho) = \frac{S(\rho)}{\bar{\rho}}$$

Le coefficient de variation estimé diminue aux différents jalons temporels de la collecte, indiquant une représentativité globale qui s'améliore au cours du temps (figure 3.1.4).

¹⁴ La propension à répondre est modélisée par le sexe, l'âge, le décile de niveau de vie, le type de commune, la résidence en QPV et le type de ménage au sens fiscal.

Figure 3.1.4 – Coefficient de variation des propensions à répondre au cours de la collecte



Lecture : à J20, la propension de réponse globale est de 0,32 et le coefficient de variation de la propension à répondre est de 0,30.

Source : enquête Genese, SSMSI.

Globalement, l'adjonction des différents modes de réponse au fil de la collecte, outre les gains en précision induits par la participation accrue, semble également contribuer à améliorer la représentation des sous-populations les plus défavorisées sans nuire trop sensiblement à la représentativité sur les autres dimensions. Ces constats peuvent être affinés par l'analyse sur les différents sous-échantillons constitués au cours de la collecte.

3.2 Analyse de la représentativité entre les différents sous-échantillons

La représentativité des répondants est maintenant analysée selon le protocole de collecte qui leur a été attribué.

Le P1SS1, qui comprend les individus avec seulement une adresse postale comme moyen de contact, a une structure qui se différencie fortement de l'échantillon global tiré. Il s'agit d'une population plus pauvre, moins souvent en couple, et qui habite plus fréquemment en QPV. Les catégories d'âge extrêmes (plus jeunes et plus âgés) y sont surreprésentées (figure 3.2.1).

Le P1SS2a et le P1SS2b sont les non répondants à J40 qui sont tous éligibles à la relance téléphonique, mais dont seule une partie (P1SS2a) a été sélectionnée pour y participer. Ces deux échantillons sont très similaires à J40 car le tirage de l'échantillon relancé par téléphone a été pseudo-stratifié par tri sur différentes variables sociodémographiques. Ces deux échantillons se différencient logiquement de l'échantillon initial par une plus forte sous-représentation des catégories qui ont moins participé avant J40, en particulier les sous-populations plus défavorisées.

Figure 3.2.1 – Structure des sous-échantillons (en %)

| | Echantillon | P1SS1 J0 | P1SS2a J40 | P1SS2b éligibles J40 |
|--------------------------------|--------------------|---------------------|-----------------------|---------------------------------|
| <i>Effectif</i> | 169 060 | 16 051 | 25 000 | 54 999 |
| Sexe | | | | |
| Femme | 49,9 | 53,5 | 50,2 | 50,3 |
| Homme | 50,1 | 46,5 | 49,8 | 49,7 |
| Age | | | | |
| 18-29 ans | 19,5 | 27,1 | 20,4 | 20,6 |
| 30-39 ans | 17,9 | 9,6 | 19,8 | 19,6 |
| 40-49 ans | 18,4 | 12,9 | 18,0 | 17,8 |
| 50-59 ans | 18,9 | 18,4 | 17,6 | 17,8 |
| 60-74 ans | 25,4 | 32,0 | 24,2 | 24,2 |
| Type de commune | | | | |
| Rurale | 22,2 | 19,6 | 22,3 | 22,5 |
| Urbaine | 77,8 | 80,4 | 77,7 | 77,5 |
| Résidence en QPV | | | | |
| QPV | 7,1 | 11,7 | 8,6 | 8,7 |
| Hors QPV | 92,9 | 88,3 | 91,4 | 91,3 |
| Type de ménage | | | | |
| Personne seule | 18,7 | 31,8 | 18,5 | 18,3 |
| Couple sans enfant | 23,5 | 22,1 | 21,9 | 21,9 |
| Couple avec enfants | 40,1 | 26,0 | 39,7 | 39,8 |
| Famille monoparentale | 10,9 | 12,6 | 12,0 | 12,0 |
| Autre type de ménage | 6,8 | 7,5 | 7,9 | 7,9 |
| Décile de niveau de vie | | | | |
| d01 | 8,9 | 15,7 | 10,4 | 10,3 |
| d02 | 8,7 | 13,4 | 10,3 | 10,2 |
| d03 | 8,7 | 10,8 | 9,9 | 9,9 |
| d04 | 8,8 | 9,2 | 9,7 | 9,6 |
| d05 | 9,2 | 7,8 | 9,6 | 9,8 |
| d06 | 9,5 | 6,9 | 9,8 | 9,7 |
| d07 | 10,1 | 6,0 | 9,5 | 9,5 |
| d08 | 10,2 | 4,9 | 9,1 | 9,1 |
| d09 | 10,3 | 3,9 | 8,6 | 8,7 |
| d10 | 10,5 | 3,7 | 8,5 | 8,4 |
| NA | 5,1 | 17,7 | 4,5 | 4,7 |

Source : enquête Genese, SSMSI.

La structure des répondants du sous-échantillon P1SS1 s'éloigne assez nettement de sa composition initiale (figure 3.2.2 et annexe 5 pour l'ensemble des variables) : si certaines variables améliorent leur représentativité entre J20 et J80 (sexe, type de commune, résidence en QPV), de nombreuses variables présentent des différences standardisées de plus de 10 points. En fin de collecte, sont particulièrement sous-représentés les ménages de personnes seules (-26,2) et les 18-29 ans (-21,4), alors que ces derniers sont surreprésentés (6,1) à J20 quand seul internet est proposé comme mode de réponse.

Le caractère aléatoire de la sélection à la relance téléphonique permet une comparaison assez directe des différences standardisées relatives aux sous-échantillons P1SS2a et P1SS2b.

Les différences standardisées en fin de collecte sont globalement beaucoup moins marquées pour les répondants du P1SS2a que pour les répondants du P1SS2b, indiquant une meilleure représentativité des répondants, en particulier sur les dimensions de sexe, âge, type de commune, résidence en QPV ou encore niveaux de vie. Dans le P1SS2a, seuls les habitants en QPV et le décile le plus pauvre dépassent la barre des 10 points en absolu en fin de collecte, avec tout de même une amélioration entre J60 et J80 pour ces variables.

Le constat est plus mitigé pour le P1SS2b des éligibles à la relance téléphonique mais non sélectionnés : la sous-représentation des moins de 30 ans s'aggrave en cours de collecte, de même que la surreprésentation des plus de 60 ans. Les autres variables suivent globalement l'évolution constatée dans le P1SS2a, mais à des niveaux de différences plus élevés, dont beaucoup restent supérieurs à 10 points en fin de collecte.

Figure 3.2.2 – Différences standardisées selon le sous-échantillon et l'avancée de la collecte

| | P1SS1 | | P1SS2a | | P1SS2b éligible | |
|--------------------------------------|-------|-------|--------|-------|-----------------|-------|
| | J20 | J80 | J60 | J80 | J60 | J80 |
| Sexe (femme) | 15,1 | 11,8 | 7,4 | 5,5 | 7,8 | 9,2 |
| Age (18 – 29 ans) | 6,1 | -21,4 | -4,6 | -4,9 | -5,4 | -16,3 |
| Age (60 – 74 ans) | -25,1 | 13,6 | 2,2 | 3,8 | 9,9 | 22,0 |
| Commune (rurale) | 15,3 | 14,7 | 8,1 | 7,1 | 11,1 | 11,7 |
| QPV (appartenance) | -24,3 | -12,6 | -13,2 | -11,7 | -18,7 | -14,2 |
| Type de ménage (personne seule) | -46,3 | -26,2 | -12,5 | -6,2 | -14,0 | -8,2 |
| Type de ménage (couple avec enfants) | 44,6 | 17,7 | 12,4 | 7,4 | 13,6 | 6,3 |
| Décile de niveau de vie (d01) | -10,4 | -10,6 | -13,9 | -12,4 | -19,7 | -15,3 |
| Décile de niveau de vie (d10) | 27,7 | 13,4 | 6,7 | 5,5 | 13,8 | 7,1 |

Lecture : Dans le P1SS1, les femmes répondantes à J20 ont une différence standardisée de 15,1 points par rapport à la structure du P1SS1.

Source : enquête Genese, SSMSI.

Ces différents résultats suggèrent que la possibilité offerte de répondre par questionnaire papier peut tendre à surreprésenter les plus âgés parmi les répondants au détriment des plus jeunes.

Proposer le téléphone comme mode de réponse permet en revanche d'assurer clairement une meilleure représentativité des répondants par rapport à l'échantillon initial sur les principales variables sociodémographiques. Le coefficient de variation des propensions à répondre est ainsi nettement moins important pour les répondants du P1SS2a (0,17) que pour les répondants du P1SS2b éligibles (0,30).

Ainsi, le protocole multimode implémenté pour l'enquête Genese, en plus de favoriser un taux de participation élevé, permet de ramener à des niveaux satisfaisants sur le temps long des profils qui sont peu répondants en début de collecte, en particulier les catégories les plus défavorisées de la population (habitants en QPV, déciles de niveau de vie les plus faibles, habitants en logement social), et permet globalement d'améliorer la représentativité par rapport à l'échantillon initial sur la plupart des variables sociodémographiques.

4. Efficacité de l'articulation phase 1 – phase 2

La seconde phase de l'enquête, qui est une phase de ré-interrogation, présente deux principaux enjeux : assurer une bonne participation et une attrition non sélective, en particulier pour permettre de bien représenter les Cible 1 aux profils rares.

Afin d'assurer une bonne participation, le protocole de la phase 2 est adapté au statut Cible du répondant et aux coordonnées disponibles (annexe 2). Son efficacité tient à l'adhésion des Cible 1 éligibles à l'enchaînement proposé, sinon à un impact bénéfique du protocole lissé personnalisé pour relancer leur participation, et à la performance du dispositif pour faire participer au mieux les individus en fonction des coordonnées disponibles.

Pour se prémunir d'une déformation importante des répondants par rapport à l'échantillon de personnes sélectionnées pour la phase 2 : i/ le tirage en fin de phase 1 rééquilibre la structure de l'échantillon des répondants Cible 2 et 3 sélectionnés pour participer à la phase 2 ; ii/ le protocole permet de contacter les répondants immédiatement (Cible 1) ou peu de temps après leur participation à la phase 1 (Cible 2 et 3) afin de limiter au maximum le taux d'injoignables ou de démenagés ; iii/ le téléphone est proposé à tous, soit dès le début de la collecte, soit à mi-parcours, ce qui représente une alternative pour les personnes ne pouvant répondre sur internet (le mode de réponse papier n'est pas proposé en phase 2).

4.1 Règles d'inclusion et coordonnées de contact

La phase 2 de l'enquête a concerné 15 000 répondants à la phase 1 : 1 781 individus du groupe Cible 1, 10 001 individus du groupe Cible 2 et 3 218 individus du groupe Cible 3. Les personnes Cible 1 ont été systématiquement sélectionnées¹⁵, les personnes Cible 2 ont été surreprésentées (elles représentent 67 % des sélectionnées pour la phase 2 alors que leur poids parmi les répondants de la phase 1 est de 38 %) et les personnes Cible 3 ont été tirées aléatoirement pour compléter l'échantillon (21 % des sélectionnés contre 60 % parmi les répondants de phase 1).

Les coordonnées de contact privilégiées pour informer les enquêtés et/ou les relancer pour participer en phase 2 sont les coordonnées laissées par les répondants à la fin de la phase 1. Un module à la fin du questionnaire de phase 1, informe le répondant de l'existence d'une enquête complémentaire (sans annoncer les sujets exacts qui seront abordés¹⁶) à laquelle il a été sélectionné (Cible 1) ou pour laquelle il est susceptible d'être recontacté (Cible 2 et 3) et lui demande de renseigner une adresse mail et un numéro de téléphone personnels (mobile de préférence). Les coordonnées téléphoniques laissées viennent compléter celles pouvant déjà être disponibles dans le fichier de contact afin d'améliorer la joignabilité du répondant par téléphone. À l'inverse, en raison de la sensibilité des sujets abordés en phase 2, seule l'adresse mail personnelle renseignée dans ce module est conservée pour contacter l'enquêté afin de cibler au mieux la personne sélectionnée. Au-delà de disposer de nouvelles coordonnées, le fait de renseigner des coordonnées peut s'interpréter comme un signal en faveur d'une réinterrogation et donc d'une meilleure participation.

Au total, parmi l'ensemble des répondants de phase 1, 29 % ont communiqué un numéro de téléphone (fixe ou mobile) et une adresse mail personnels, 24 % seulement une adresse mail, 6 % uniquement un

¹⁵ Le taux de sondage est de 97 % car la sélection pour la phase 2 a exclu quelques répondants papier Cible 1 dont l'identité (sexe et date de naissance) ne correspondait pas aux informations de contact du fichier d'origine.

¹⁶ Voici ce qui s'affiche à l'écran : « L'enquête sur le vécu et les opinions en matière de sécurité comporte un second volet qui permet d'aborder plus en détail certaines atteintes à la personne qui ont été évoquées dans ce questionnaire. »

numéro de téléphone et 41 % n'ont laissé ni coordonnées mail ni coordonnées téléphoniques (figure 4.1.1). Le fait de laisser des coordonnées dans ce module, toutes choses étant égales par ailleurs (sexe, âge, localisation, type de ménage contrôlés), est fortement lié au fait d'avoir déclaré des atteintes liées au genre dans le questionnaire de phase 1 : les Cible 2, ayant déclaré une violence physique ou sexuelle avant 2020 ou des faits d'agressions sexuelles ou de harcèlement au cours de leur vie, ont ainsi plus souvent laissé un mail ou un téléphone que les autres cibles. Le moment et le mode de collecte ont également une influence. Les répondants tardifs (après J40), relancés à plusieurs reprises avant de participer, laissent moins leurs coordonnées (mail ou téléphone) que ceux qui ont complété leur questionnaire au début de la collecte (avant J20). Par ailleurs, les répondants par internet donnent moins leurs coordonnées que les répondants par téléphone ou par papier. Enfin, des coordonnées sont plus souvent laissées lorsque des coordonnées étaient déjà disponibles dans la base de sondage.

Figure 4.1.1 – Coordonnées de contact laissées par les répondants pour être recontactés pour la phase 2 (en %)

| | Répondants de phase 1 | Répondants de phase 1 sélectionnés pour la phase 2 |
|---|-----------------------|--|
| A laissé une adresse mail et un numéro de téléphone (fixe et/ou mobile) | 28,7 | 33,5 |
| A laissé seulement une adresse mail | 24,2 | 28,0 |
| A laissé seulement un numéro de téléphone (fixe et/ou mobile) | 6,3 | 4,3 |
| N'a laissé ni adresse mail ni numéro de téléphone | 40,8 | 34,2 |

Source : enquête Genese, SSMSI.

En lien avec ces résultats, les répondants Cible 1 et 2 étant surreprésentées dans l'échantillon de la phase 2, on dispose plus souvent de coordonnées en fin de phase 1 pour eux que pour l'ensemble des répondants de phase 1 (66 % contre 59 %) : 34 % ont communiqué un numéro de téléphone (fixe ou mobile) et une adresse mail, 28 % seulement une adresse mail, 4 % seulement un numéro de téléphone.

Finalement, en cumulant les informations téléphoniques laissées en fin de phase 1 et celles disponibles dans la base de sondage, pour 96 % des répondants de phase 1 sélectionnés au moins une coordonnée mail ou téléphonique est disponible sachant que pour la majorité des personnes (59 % des cas) on dispose à la fois d'un mail personnel et d'au moins un numéro de téléphone (figure 4.1.2).

Figure 4.1.2 – Ensemble des coordonnées de contacts utilisées pour la phase 2 (en %)

| | Répondants de phase 1 | Répondants de phase 1 sélectionnés pour la phase 2 |
|---|-----------------------|--|
| Mail et au moins un numéro de téléphone | 50,8 | 59,3 |
| Mail et pas de téléphone | 2,1 | 2,3 |
| Pas de mail et au moins un téléphone | 41,9 | 34,3 |
| Pas de mail et pas de téléphone | 5,2 | 4,2 |

Source : enquête Genese, SSMSI.

Les personnes ayant déjà été contactées pour participer à la phase 1 de l'enquête et de nouvelles coordonnées ayant été fournies par de nombreux répondants en plus de celles déjà disponibles dans les fichiers, la part de personnes n'ayant jamais reçu de sollicitation lors de la phase 2 (ni lettre-avis, ni mail-avis, ni contact téléphonique avec un enquêteur) est relativement faible. Au démarrage de la collecte pour l'ensemble de l'échantillon (semaine 12), seul 1 % de l'échantillon n'a reçu ni lettre-avis, ni mail-avis. Parmi ces personnes, 33 % a répondu au questionnaire par téléphone.

4.2 Analyse de la participation

4.2.1 Bilan de la participation

La phase 2 a permis de collecter 10 225 questionnaires complets : 7 920 par internet et 2 305 par téléphone, soit un taux de collecte de 68 % (figure 4.2.1). La participation internet est de 53 % et de 22 % par téléphone. L'objectif initialement fixé a été atteint et dépassé de 3 points grâce à une bonne participation par internet qui compense une participation par téléphone en-deçà des attentes.

La part de questionnaires incomplets représente 4 % de l'ensemble des personnes interrogées. Elle est plus importante par internet qu'au téléphone, la présence de l'enquêteur incitant le répondant à aller au bout du questionnaire. La moitié des abandons sont des connexions aux premières pages du questionnaire, les autres s'arrêtant surtout dans les premières parties. Les personnes qui commencent le questionnaire sans aller au bout sont davantage des répondants Cible 1, qui ont un peu plus souvent participé par téléphone en phase 1.

La participation est liée aux coordonnées mail et téléphoniques présentes pour joindre la personne à enquêter (les adresses ont été éprouvées lors de la phase 1, le taux de PND est ici inférieur à 1 %). Elle est plus élevée pour les personnes ayant laissé un numéro de téléphone et un mail personnel en fin de phase 1 : 75 % contre 38 % pour les personnes sans adresse mail personnelle ni téléphone.

Figure 4.2.1 – Participation et complétude des questionnaires en phase 2

| | Nombre de répondants CAWI | Nombre de répondants CATI | Nombre de répondants total |
|--|---------------------------|---------------------------|----------------------------|
| Individus sollicités | 15 000 | 10 622* | 15 000 |
| Questionnaires complets | 7 920 | 2 305 | 10 225 |
| Questionnaires incomplets | 458 | 191 | 649 |
| Taux de participation (complets/ sollicités) | 52,8% | 21,7% | 68,2% |

* les individus sollicités par téléphone sont les personnes pour lesquelles au moins un numéro de téléphone est disponible et ne faisant pas partie du sous échantillon de personnes qui ont répondu en ligne à l'enquête avant que le terrain téléphonique ne démarre pour eux (avant le 18/06).

Source : enquête Genese, SSMSI.

Le protocole mis en place a permis de bien surreprésenter le groupe Cible 1 dans l'échantillon de phase 2. Ces individus ont mieux participé que les autres, ce qui conduit mécaniquement à une accentuation de leur surreprésentation parmi les répondants (figure 4.2.2).

Figure 4.2.2 – Participation globale selon la cible

| | Nombre de répondants en phase 1 | Nombre de répondants de phase 1 sélectionnés pour la phase 2 | Nombre de répondants en phase 2 | Taux de réponse en phase 2 (en %) |
|----------|---------------------------------|--|---------------------------------|-----------------------------------|
| Cible 1 | 1 840 | 1 781 | 1 388 | 77,9 |
| Cible 2 | 41 082 | 10 001 | 6 847 | 68,5 |
| Cible 3 | 64 927 | 3 218 | 1 990 | 61,8 |
| Ensemble | 107 849 | 15 000 | 10 225 | 68,2 |

Source : enquête Genese, SSMSI.

En fonction du sous-échantillon, le protocole de contact et de relance est adapté et tire parti au maximum des coordonnées de contact disponibles. Au terme de la collecte, 88 % des personnes du P2SS1, 70 % de celles du P2SS3, 62 % au sein du P2SS2 et enfin, 61 % parmi celles du P2SS4 ont répondu à la phase 2 de l'enquête (figure 4.2.3). La participation est plus élevée sur internet qu'au téléphone pour tous les groupes, sauf pour le P2SS2 où sont regroupés des individus ayant répondu par téléphone en phase 1. Les Cible 2, qui ont mieux participé que les Cible 3 à l'enquête (figure 4.2.2), représentent une part plus élevée des individus dans le P2SS3 que dans le P2SS4 (79 % contre 58 %), ce qui peut expliquer l'écart global de participation entre ces sous-échantillons.

Figure 4.2.3 – Participation globale par mode et sous-échantillon du protocole (en %)

| | P2SS1 (N = 932) | P2SS2 (N = 137) | P2SS3 (N = 8 920) | P2SS4 (N = 5 011) |
|-----------------------|--------------------|--------------------|----------------------|----------------------|
| Répondants CAWI | 82,5 | 2,9 | 60,4 | 35,2 |
| Répondants CATI | 5,6 | 59,1 | 10,0 | 25,6 |
| Participation globale | 88,1 | 62,0 | 70,4 | 60,8 |

Source : enquête Genese, SSMSI.

4.2.2 Participation selon l'avancée de la collecte

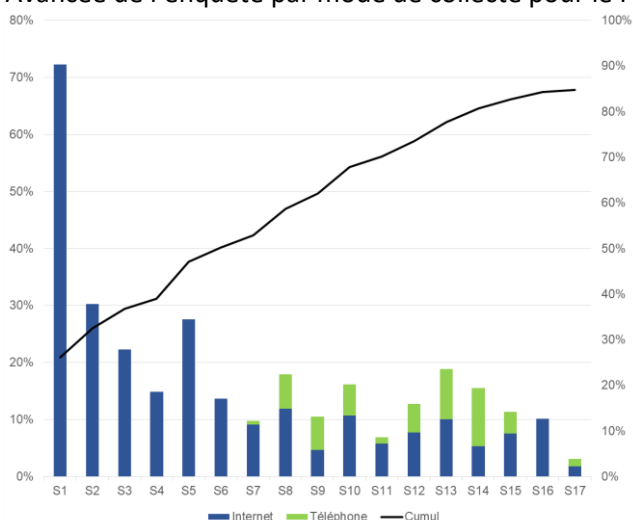
Rapportée aux répondants éligibles chaque semaine, la participation du P2SS1&2 (cumul de P2SS1 et P2SS2) est particulièrement importante au démarrage de la collecte, ce qui valide le protocole permettant l'enchaînement des deux phases (figure 4.2.4). La participation atteint un pic lors de la première semaine de collecte avec plus de 70 % de participation en phase 2 parmi les répondants identifiés Cible 1 en phase 1. Le taux cumulé de participation atteint 50 % au bout de 6 semaines de collecte et 70 % à la 11^e semaine avant l'envoi de la lettre de relance. La poursuite de la collecte (semaines 12 à 17) permet d'augmenter le taux global de participation de 70 % à 85 %, avec le maintien d'un taux de participation hebdomadaire relativement stable lors de cette dernière période, soutenu par la réception de la lettre de relance (semaines 12 et 13) et la relance téléphonique pour tous (effet sur les semaines 13 et 14).

La participation du P2SS3 est relativement stable au cours des 4 premières semaines de collecte avec un taux entre 11 et 14 % de personnes répondantes par semaine (figure 4.2.5). À la fin de la 4^e semaine (semaine 15), le taux de participation global cumulé est de 49 %. La participation est soutenue par le téléphone lors de la 5^e semaine et baisse clairement lors de la 6^e semaine (S17). Néanmoins, la courbe du taux de participation cumulé n'atteint pas véritablement un plateau en fin de collecte, ce qui laisse

penser que la poursuite de la collecte aurait possiblement permis d'obtenir un taux de participation global plus élevé que le 70 % enregistré en fin de semaine 17.

L'avancée de la collecte dans le P2SS4 est plus élevée que dans le P2SS3 pendant les deux premières semaines de la collecte avec en première semaine (S12) une bonne participation internet et en deuxième semaine un pic de participation cumulé internet/téléphone avec le mode téléphone comme mode majoritaire (figure 4.2.6). Ensuite la participation décroît, avec une participation sous la barre des 10 % à partir de la 4^e semaine (S15).

Figure 4.2.4 – Avancée de l'enquête par mode de collecte pour le P2SS1&2 (en %)



Lecture : au cours de la 8^e semaine de collecte, 12 % des répondants de phase 1 Cible 1 (éligibles à l'enchaînement à cette date n'ayant pas encore complété leur questionnaire de phase 2) ont complété un questionnaire en ligne et 6 % par téléphone (échelle de gauche). À cette date, le taux cumulé de participation est de 59 % (échelle de droite).

Figure 4.2.5 – Avancée de l'enquête par mode de collecte pour le P2SS3 (en %)

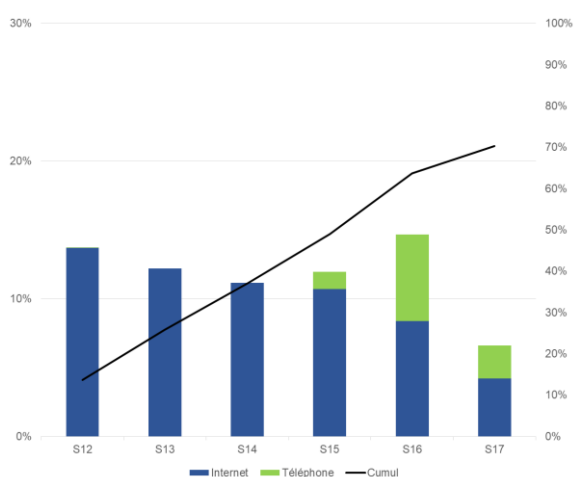
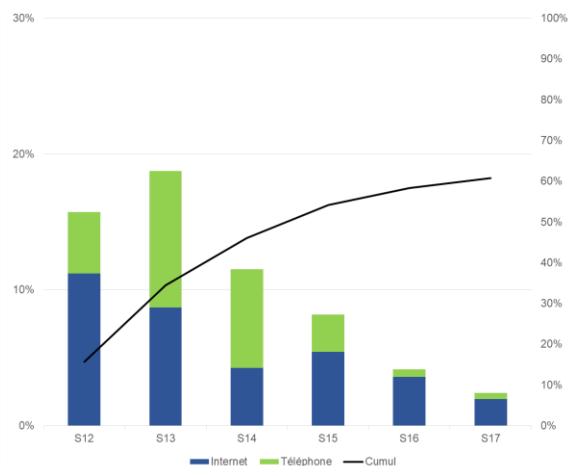


Figure 4.2.6 – Avancée de l'enquête par mode de collecte pour le P2SS4 (en %)



Lecture (figure 4.3.5) : au cours de la 16^e semaine de collecte, 8 % des individus du P2SS3 ont complété leur questionnaire en ligne (CAWI) et 6 % par téléphone (CATI) (échelle de gauche). A cette date, le taux cumulé de participation est de 64 % (échelle de droite).

Source : enquête Genese, SSMSI.

4.3 Enchaînement et protocole lissé

Le dispositif en place permet aux répondants particulièrement concernés par la thématique de l'enquête, les Cible 1, ayant répondu par internet ou téléphone à la phase 1, d'enchaîner les deux interrogations phase 1 puis phase 2. La participation de ces personnes pouvant enchaîner est meilleure que celle des Cible 1 ne pouvant pas enchaîner (répondants papier), avec une participation globale de 81 % contre 51 % (figure 4.3.1). Parmi les personnes pouvant enchaîner, un peu plus de la moitié (53 %) a accepté de le faire et s'est connectée directement au questionnaire de phase 2. Parmi elles, presque 100 % ont complété entièrement et validé le questionnaire de phase 2, quasi-toutes (96 %) en une seule connexion. 97 % étaient des répondants par internet en phase 1 et ont de fait poursuivi sur internet.

Pour les personnes qui n'ont pas enchaîné alors que c'était possible et pour lesquelles une coordonnée mail ou téléphonique est disponible, des relances selon un protocole lissé ont été envoyées en fonction des coordonnées disponibles et de la date de réponse à la phase 1 de l'enquête. La participation de ces personnes est plus élevée que celle obtenue pour les personnes pouvant enchaîner mais n'ayant pas été relancées pour le faire (62 % contre 40 %). Mécaniquement, l'enchaînement réduit le délai de réponse entre les phases 1 et 2 : Ce temps est de 7,5 jours en moyenne parmi les répondants Cible 1 pouvant enchaîner dans le protocole lissé contre 53,5 jours pour les répondants Cible 1 qui démarrent la collecte de la phase 2 à partir de la semaine 12.

Figure 4.3.1 – Possibilités d'enchaînement, relances et participation des répondants Cible 1 (en %)

| | Répartition parmi les Cible 1 | Participation globale | Participation internet | Participation téléphone |
|---|-------------------------------|-----------------------|------------------------|-------------------------|
| Ensemble des Cible 1 pouvant enchaîner | 90,6 | 80,7 | 70,0 | 10,7 |
| Cible 1 pouvant enchaîner et ayant accepté de le faire | 53,0 | 100,0 | 96,6 | 3,4 |
| Cible 1 pouvant enchaîner et n'ayant pas enchaîné | 37,7 | 53,7 | 32,6 | 21,0 |
| <i>Dont personnes dans le protocole lissé (avec coordonnées mail ou téléphonique)</i> | <i>63,0</i> | <i>61,5</i> | <i>37,6</i> | <i>23,9</i> |
| Cible 1 ne pouvant pas enchaîner (répondants papier) | 9,4 | 50,9 | 18,6 | 32,3 |

Champ : répondants Cible 1.

Source : enquête Genese, SSMSI.

4.4 Analyse de l'attrition

Avec près d'un tiers des répondants de phase 1 n'ayant pas complété le questionnaire de la phase 2, le taux d'attrition observé dans Genese est assez semblable à celui de l'enquête Care-ménages 2015 qui a réinterrogé en face-à-face avec le réseau d'enquêteurs de l'Insee les personnes âgées de 60 ans ou plus ayant répondu à l'enquête VQS 2014-2015 [21]. Il est, en revanche, supérieur à celui constaté entre la vague 1 et la vague 2 d'Epicov qui se sont déroulées à six mois d'intervalle (attrition de 20 %)¹⁷[26].

L'analyse des caractéristiques des répondants montre une bonne représentation des femmes, surreprésentées dans l'échantillon de répondants sélectionnés pour la phase 2, et une faible attrition entre la phase 1 et la phase 2 selon l'âge, l'habitat (type et localisation), le type de ménage, le niveau de diplôme, la nationalité, l'état de santé perçu, le niveau de vie et le fait d'exercer ou non un emploi.

¹⁷ Un kit d'auto-prélèvement sanguin a été envoyé aux répondants volontaires afin de détecter la présence d'anticorps dirigés contre le SARS-CoV-2.

Pour autant, certains répondants de phase 1, en particulier les répondants papier en phase 1, les répondants tardifs en phase 1 (après J41), les Cible 3 et les répondants n'ayant pas laissé de coordonnées en fin de phase 1 ont moins bien répondu à la phase 2 (figure 4.4.1).

Pour mieux identifier les effets qui se conjuguent et mesurer un effet « propre » des différents facteurs sur la probabilité de répondre et d'avoir complété son questionnaire de phase 2, la participation à la phase 2 a été modélisée par régression logistique (figure 4.4.2 et annexe 6). La littérature sur l'attrition s'accorde à dire que les déterminants diffèrent selon que la non-réponse provient d'une perte de contact ou d'un refus à participer de nouveau à l'enquête [11 ; 25]. S'agissant de la phase 2 de l'enquête, qui suit dans un temps court la phase 1, on peut raisonnablement considérer que la non-participation se traduit le plus souvent par un refus de l'enquête. Cinq grands groupes de facteurs sont considérés : l'habitat, les caractéristiques sociodémographiques, la santé perçue, les types de relances envoyées et, enfin, la « posture » de l'enquêté par rapport à l'enquête de phase 1. Il ressort de cette analyse que les facteurs liés à la collecte (coordonnées disponibles, mode de réponse, temps de passation en phase 1 et période de collecte en phase 1) et le statut Cible du répondant sont particulièrement importants pour expliquer la participation en phase 2.

Ainsi, le fait d'être Cible 1 augmente fortement la probabilité de participer à l'enquête. Les répondants repérés comme étant les plus concernés par la thématique de l'enquête lors de la phase 1 ont mieux participé que les autres à la seconde phase. En conduisant la même analyse mais en introduisant le fait d'avoir bénéficié ou non de l'enchaînement et du protocole lissé croisé avec la cible à la place de la cible seule (non présentée ici), l'effet est significatif et positif pour les Cible 1 pouvant enchaîner du protocole lissé relativement à tous les autres et notamment aux Cible 1 ne faisant pas partie de ce protocole (soit par manque de coordonnées disponibles, soit parce que c'était des répondants papier en phase 1). Ce résultat confirme que la possibilité d'enchaîner et le protocole de contact et de relance lissé adapté aux coordonnées disponibles ont permis d'avoir une moindre attrition des répondants Cible 1.

Les relances mail et téléphoniques effectuées (en lien avec les coordonnées disponibles) permettent également de faire davantage participer les enquêtés : le fait d'être destinataire d'un mail et d'être contacté par téléphone pour remplir le questionnaire avec un enquêteur, jouent chacun positivement sur la probabilité de répondre à l'enquête. Le fait d'avoir laissé un téléphone personnel en fin de phase 1 favorise par ailleurs davantage la participation que le fait de disposer d'un autre type de téléphone.

Le fait de répondre à la phase 2 est également expliqué par le fait d'avoir adhéré/ d'avoir répondu tôt à l'enquête de la phase 1 (avant J20 relativement aux autres périodes), indépendamment de la Cible (et de l'enchaînement possible pour certains répondants). On aurait pu penser que l'attrition serait plus importante à mesure du temps écoulé mais ce n'est pas ce qui ressort des résultats de cette analyse, les plus « motivés » par l'enquête en phase 1 ayant également mieux participé en phase 2. Les répondants sont préparés par la procédure d'annonce de la seconde phase de l'enquête à être recontactés dans un temps relativement court, cela a pu avoir pour effet de limiter l'attrition et de favoriser leur participation en phase 2.

Les répondants papier en phase 1 ont moins souvent répondu en phase 2 que les répondants internet et téléphone de la phase 1. Les répondants ont très majoritairement conservé le mode de réponse de la phase 1 pour la réponse en phase 2. Ce résultat sur l'inertie du mode de collecte est retrouvé également dans la vague 2 d'Epicov¹⁸. Les répondants internet en phase 1, majoritaires parmi les répondants à la phase 1, ont très majoritairement répondu à la phase 2 par internet (84 % des

¹⁸ Ces constats s'appuient sur des travaux non publiés portant sur la collecte de l'enquête Epicov. Ils ont été présentés par Guillaume Bagein (Drees) lors d'une réunion du groupe de travail Méthodologie d'enquête du SSMSI dans le cadre de la refonte de l'enquête CVS.

répondants internet de phase 1) alors que les répondants par téléphone ou papier en phase 1 ont majoritairement répondu à la phase 2 par téléphone (respectivement 82 % des répondants téléphone et 66 % des répondants papier).

Par ailleurs, on aurait également pu s'attendre à une attrition plus forte chez les personnes ayant eu un temps de passation plus long en phase 1. Or, c'est l'inverse que l'on observe (ce qui n'est pas contradictoire avec les travaux sur l'attrition portant sur des enquêtes par vagues [19]) : les enquêtés qui ont mis moins de 20 minutes à remplir leur questionnaire de phase 1 ont moins souvent répondu à celui de la phase 2 que ceux qui ont mis 20 minutes et plus. Par le jeu des filtres, le questionnaire de phase 1 est plus ou moins long et ce résultat peut s'interpréter, au-delà du temps de passation uniquement, par le fait que ces personnes (pour lesquelles le temps de passation est plus court) ont pu se sentir moins concernées et moins intéressées par l'enquête de phase 1 et expliqué qu'elles n'aient pas souhaité participer à la phase 2.

Concernant les caractéristiques des individus, les personnes les plus jeunes (18-29 ans et 30-39 ans) ont moins bien participé que les 40-49 ans ; à l'inverse, les personnes les plus âgées (50-59 ans et 60-74 ans) ont mieux participé à l'enquête. Les personnes peu diplômées, les étrangers, les personnes avec un niveau de vie dans les premiers déciles de la distribution, les personnes en famille monoparentale ou faisant partie d'un type de ménage autre qu'un couple ou une personne seule ont également moins souvent répondu à l'enquête. Il en est de même pour les personnes ayant déclaré une mauvaise santé lors de la phase 1. Ces résultats sont retrouvés dans la littérature sur la participation : les plus diplômés sont plus souvent coopérants, l'effort étant plus important pour les moins diplômés, les étrangers refusent plus souvent d'être interrogés pour des raisons de difficultés à s'exprimer dans la langue ou bien parce qu'ils se sentent moins concernés par une enquête nationale [19].

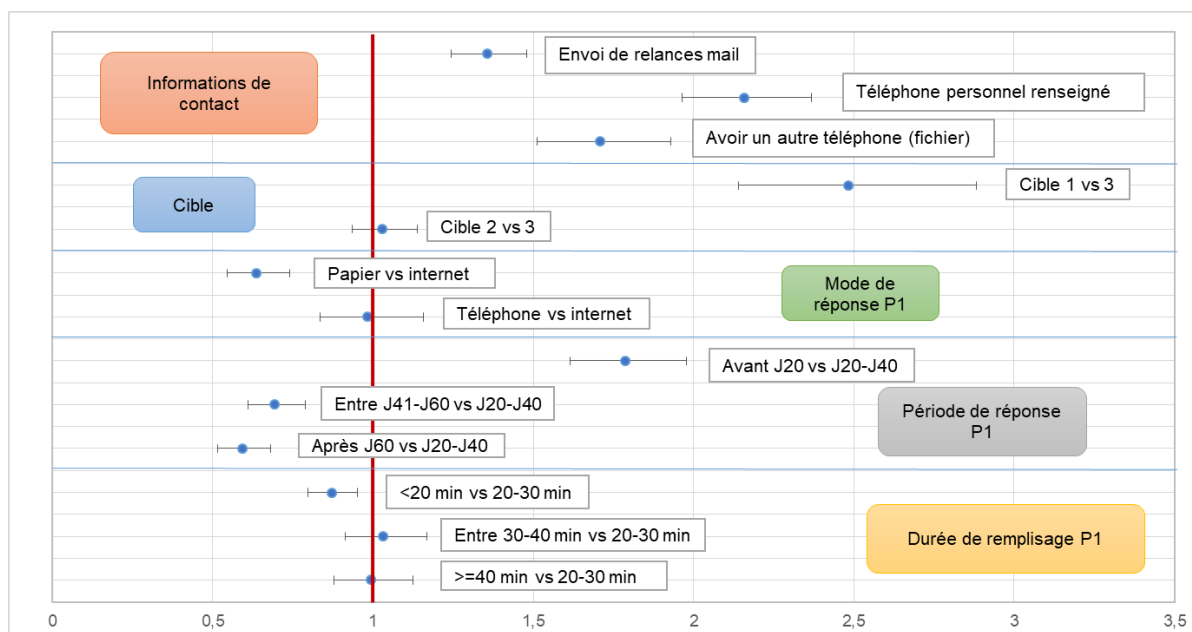
Figure 4.4.1 – Répartition des personnes sélectionnées et répondantes à l'enquête de suivi (phase 2) en fonction du sexe, de l'âge des enquêtés, du mode, de la période de collecte et des coordonnées laissées en fin de phase 1 (effectifs et %)

| | Répondants à la phase 1 sélectionnés pour la phase 2 | | Répondants à la phase 2 | |
|-----------------------------------|--|------|-------------------------|------|
| | Effectif | % | Effectif | % |
| Sexe du répondant | | | | |
| Homme | 3 525 | 23,5 | 2 420 | 23,7 |
| Femme | 11 475 | 76,5 | 7 805 | 76,3 |
| Âge du répondant | | | | |
| 18-29 ans | 3 325 | 22,2 | 2 190 | 21,4 |
| 30-39 ans | 2 916 | 19,4 | 1 877 | 18,4 |
| 40-49 ans | 3 006 | 20,0 | 2 074 | 20,3 |
| 50-59 ans | 2 778 | 18,5 | 1 947 | 19,0 |
| 60-74 ans | 2 975 | 19,8 | 2 137 | 20,9 |
| Mode de réponse en phase 1 | | | | |
| CAWI | 12 658 | 84,4 | 9 073 | 88,7 |
| CATI | 958 | 6,4 | 508 | 5,0 |
| Papier | 1 384 | 9,2 | 644 | 6,3 |

| Période de collecte en phase 1 | | | | |
|---|-------|------|-------|------|
| Avant J20 | 7 629 | 50,9 | 5 989 | 58,6 |
| Entre J20 et J40 | 2 718 | 18,1 | 1 786 | 17,5 |
| Entre J41 et J60 | 2 440 | 16,3 | 1 347 | 13,2 |
| Après J60 | 2 213 | 14,8 | 1 103 | 10,8 |
| A laissé des coordonnées en fin de phase 1 | | | | |
| A laissé une adresse mail et un numéro de téléphone (fixe et/ou mobile) | 5 031 | 33,5 | 4 032 | 39,4 |
| A laissé seulement une adresse mail | 4 206 | 28,0 | 2 831 | 27,7 |
| A laissé seulement un numéro de téléphone (fixe et/ou mobile) | 641 | 4,3 | 452 | 4,4 |
| N'a laissé ni adresse mail ni numéro de téléphone | 5 122 | 34,2 | 2 910 | 28,5 |

Source : enquête Genese, SSMSI.

Figure 4.4.2 – Effet des variables liées à la collecte sur la réponse à l'enquête de suivi (phase 2) (odds-ratio, modèle de régression logistique)



Note : le modèle complet est présent en annexe avec l'ensemble des variables auxiliaires prises en compte, à savoir en plus des variables présentées ici : le sexe, l'âge, le type de ménage, le niveau de diplôme le plus haut atteint, la nationalité, le décile de niveau de vie, le type de logement, le fait d'habiter en Ile-de-France, le fait de vivre dans une commune rurale, en QPV, en QRR, l'état de santé perçu.

Lecture : un odds-ratio supérieur à 1 (resp. inférieur à 1), statistiquement significatif à 10 %, indique que l'on est en présence d'un facteur qui augmente (resp. réduit) la probabilité d'avoir répondu à la phase 2 de l'enquête. Pour les répondants Cible 1, qui par ailleurs présentent l'ensemble des caractéristiques prises en référence, la chance relative d'avoir répondu à l'enquête est 2,5 fois plus élevée que pour les répondants Cible 3 (IC95% : 2,1 ; 2,9)].

Source : enquête Genese, SSMSI.

Conclusion

Comme cela a été rappelé au début de cet article, le SSMSI a conçu l'enquête Genese en réponse à un appel à projet européen dans le but d'éclairer le débat public sur les questions de violence liée au genre et de réaliser une expérimentation multimode grandeur réelle pour mesurer les victimations et les perceptions en matière de sécurité en vue de la refonte de l'enquête CVS.

Les objectifs assignés à l'enquête en termes de participation ont été largement dépassés pour un coût inférieur à celui budgété, notamment grâce au succès du volet internet. En phase 1, sur les 169 060 individus âgés de 18 à 74 ans résidant en France métropolitaine sollicités, près de 109 000 (64 %, +14 pts par rapport aux objectifs) ont répondu dont 51 % par Internet, 3 % par téléphone et 10 % par papier. Le protocole prévoyait une lettre-avis et trois relances postales envoyées à intervalles réguliers ainsi que l'envoi d'un mail-avis et de six relances mail auprès des individus éligibles (56 % de l'échantillon initial), dont trois au cours des 3 premières semaines. Trois semaines après le démarrage, le questionnaire papier a été envoyé aux non répondants sans mail ni téléphone (10 % de l'échantillon initial). Après 6 semaines, 25 000 non répondants avec téléphone disponible (85 % de l'échantillon initial après enrichissement) ont été sélectionnés pour la relance téléphonique. Les non répondants avec mail et/ou téléphone non sélectionnés pour la relance téléphonique ont été destinataires du questionnaire papier. Enfin, au bout de 9 semaines, le questionnaire papier a été envoyé aux non répondants de la relance téléphonique. Les analyses multivariées conduites montrent que l'envoi de mails a un effet propre significatif sur la probabilité de participer en phase 1 et que cet effet tend à décroître au fil des périodes de la collecte. La relance téléphonique contribue également fortement toutes choses égales par ailleurs à la participation dans le groupe concerné mais pas seulement via le mode téléphone. En effet, au moment de l'annonce de l'introduction du téléphone, un pic de réponses par internet a été enregistré dans le groupe concerné. Ce résultat plaide en faveur de la conception de protocoles multimode concurrentiel différé qui permettent de relancer les non répondants sur des modes de plus en plus coûteux tout en maintenant ouverts la réponse par des modes moins coûteux, en particulier le mode internet. Au-delà du taux de réponse, l'étude montre également l'importance de l'apport du téléphone en termes de représentativité des répondants de phase 1 par rapport à l'échantillon initial sur la plupart des variables sociodémographiques, en particulier les marqueurs de situations sociales plus défavorisées (habitants en QPV, déciles de niveau de vie plus faibles, habitants en logement social).

Concernant la phase 2 de l'enquête, les objectifs de participation ont été atteints : 10 225 individus parmi les 15 000 sélectionnés (68 %) ont répondu à l'enquête de phase 2 dont 53 % par internet et 15 % par téléphone. L'enquête de phase 1 a permis d'identifier plus de 1 800 victimes de violences physiques ou sexuelles en 2020 ou 2021 et de les sélectionner automatiquement pour la phase 2 en leur proposant d'enchaîner les deux questionnaires (hors réponse papier), ce que plus de la majorité d'entre elles ont fait. Les autres ont été relancées de manière personnalisée à intervalles réguliers en fonction de leur date de complétion de la phase 1. Au final, près de 1 400 individus de ce groupe (78 %) ont participé à la phase 2. Les individus avec des profils moins rares sélectionnés pour la phase 2 non soumis à l'enchaînement des questionnaires ni au protocole lissé ont un peu moins bien participé (67 %). Ces résultats permettent d'envisager des analyses inédites grâce à un échantillon particulièrement important de profils rares en matière de violence liée au genre. Dans la mesure où les phénomènes étudiés dans les enquêtes de victimation concernent souvent de faibles proportions d'individus dans la population, les conclusions de ces analyses valident globalement pour la refonte de l'enquête CVS le principe d'une articulation en deux phases multimode de type enquête filtre / enquête de suivi et le protocole différencié selon le niveau de qualification pour la phase 2 des répondants de phase 1.

En l'absence de cadre expérimental pour procéder à des évaluations rigoureuses, certains aspects du protocole n'ont pas pu être évalués dans cette étude. Il n'a par exemple pas été possible de dégager un effet propre du protocole mis en place en phase 2 pour déterminer la meilleure des stratégies entre l'introduction du téléphone dès le démarrage ou à mi-parcours de la phase 2. La place du mode papier dans le dispositif n'a pas non plus été évaluée. Il serait en effet intéressant d'évaluer de manière distincte les deux effets possibles du mode papier : il peut d'un côté susciter l'envie de participer auprès de populations qui ne répondraient pas à l'enquête sans son introduction mais il peut également siphonner la réponse internet en détournant des individus qui de toutes façons répondraient à l'enquête mais sur un autre mode si le papier n'était pas introduit.

De nombreux travaux d'expertise méthodologique pourront être initiés ou approfondis à partir des données des premières éditions de l'enquête Vécu et ressenti en matière de sécurité (enquête VRS) qui prend la suite de l'enquête CVS à partir de 2022.

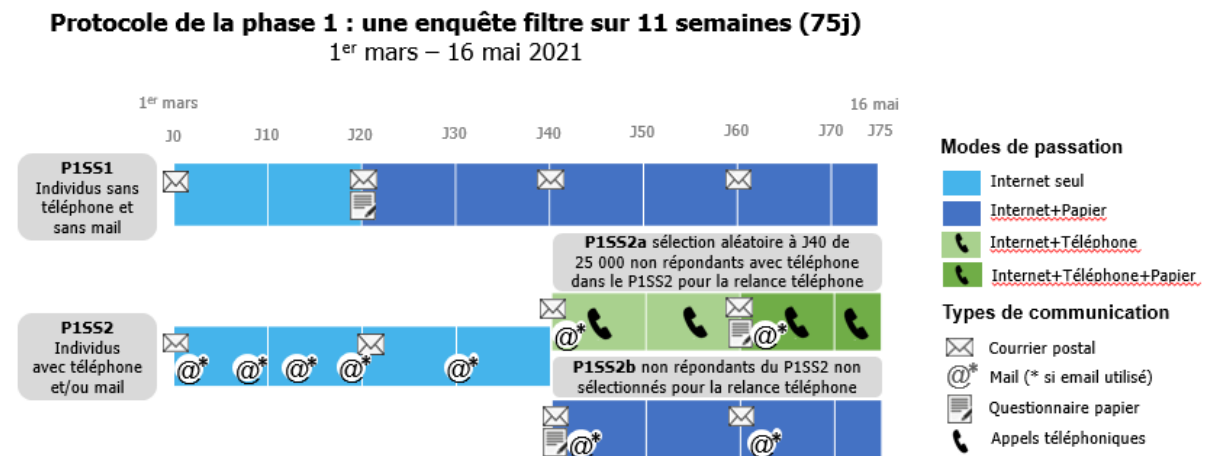
Bibliographie

- [1] Andersen, R., Kasper, J. D., & Frankel, M. R., *Total survey error*, Jossey-Bass, 1979.
- [2] Austin, P. C., & Stuart, E. A., « Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies », *Statistics in medicine*, vol.34, n°28, pp.3661-3679, 2015.
- [3] Barret, C., Cissé, M., Gaubert, E., Mazari, Z., & Olaria, M., « Efficacité d'un protocole multimode (téléphone et internet) », in *XIIIèmes Journées de Méthodologie Statistique de l'INSEE*, 2018.
- [4] Bethlehem, J. G., « Reduction of nonresponse bias through regression estimation », *Journal of Official Statistics*, vol.4, n°3, pp.251-260, 1988.
- [5] Castell, L., Clerc, M., & Legleye, S., « Victimations déclarées et effets de mode : l'expérimentation CVS Panel », in *11^e Colloque International Francophone sur les Sondages*, 2021.
- [6] De Heij, V., Schouten, B., & Schlomo, N., « Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation », *RISQ manual 2.1*, 2015.
- [7] De Leeuw, E. D., « Mixed-mode: Past, present, and future », In *Survey Research Methods*, vol.12, n°2, pp.75-89, 2018.
- [8] Dillman, D.A., « The promise and challenge of pushing respondents to the Web in mixed-mode surveys », *Survey Methodology, Statistics Canada*, 2017.
- [9] Friquet, S., « Les effets de la collecte par internet sur les taux de réponse et sur les résultats d'une enquête auprès d'un public précaire. Résultats à partir d'une expérimentation aléatoire », in *XIIIèmes Journées de Méthodologie Statistique de l'INSEE*, 2018.
- [10] Garner M., « Le projet MUSE : 5 ans d'expérimentations pour préparer l'introduction d'Internet dans l'enquête Emploi », *Insee, Document de travail, F1907*, 2019.
- [11] Groves R., & Couper M., *Nonresponse in Household Interview Surveys*, Wiley series in probability and statistics, 1998.
- [12] Guedj, H., Razafindranovona, T., & Zilloniz, S., « Une enquête multimode annuelle pour mesurer la victimation en France à compter de 2022 : chronique d'une refonte en 4 actes », *Statéco*, n°166, 2022.
- [13] Insee, « Horizon 2025 : une stratégie ambitieuse pour l'Insee », rapport stratégique de l'Insee 2016 – 2025, 2016.
- [14] Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S., « Consequences of reducing nonresponse in a national telephone survey », *Public opinion quarterly*, vol.64, n°2, pp.125-148, 2000.
- [15] Luiten, A., & Schouten, B., « Tailored fieldwork design to increase representative household survey response: an experiment in the Survey of Consumer Satisfaction », *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol.176, n°1, pp.169-189, 2013.
- [16] Merly-Alpa, T., & Rebecq, A., « L'utilisation des R-indicateurs pour prioriser la collecte des enquête ménages: une application à l'enquête Patrimoine 2010 », in *XIIèmes Journées de Méthodologie Statistique de l'INSEE*, 2015.
- [17] Moore, J. C., Durrant, G. B., & Smith, P. W., « Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice », *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol.181, n°1, pp.229-248, 2018.
- [18] Razafindranovona, T., « Exploitation de l'enquête expérimentale Vols, violence et sécurité », *Insee, Document de travail de la DMCSI, M2016-03*, 2016.

- [19] Régnier-Loilier A., & Guisse N., « Attrition et déformation de l'échantillon au fil des vagues de l'enquête Érfi » in Régnier-Loilier A. (dir.), *Parcours de familles. Enquête Étude des relations familiales et intergénérationnelles*, Collection « Grandes Enquêtes », Ined, 430 p., 2016.
- [20] Rosenzweig, A., Razafindranovona, T., Guedj, H., Quiviger A., & Zilloniz, S., « Effets de mode relatifs aux victimations et aux perceptions en matière de sécurité : estimations à partir de l'enquête Genre et Sécurité (Genese) », in *XIVèmes Journées de Méthodologie Statistique de l'Insee*, 2022.
- [21] Roy D., « Le dispositif d'enquêtes « Care » », in *Colloque de la Chaire Economie de la dépendance des personnes âgées*, 2017.
- [22] Schouten, B., Cobben, F., & Bethlehem, J., « Indicators for the representativeness of survey response », *Survey Methodology*, vol.35, n°1, pp.101-113, 2009.
- [23] Schouten, B., van den Brakel, J., Buelens, B., Giesen, D., Luiten, A., & Meertens, V., « Mixed-Mode Official Surveys: Design and Analysis », *Chapman and Hall/CRC*, 2021.
- [24] SSMSI, Rapport d'enquête « Cadre de vie et sécurité » 2019 : victimation délinquance et sentiment d'insécurité, 2019.
- [25] Watson N., & Wooden M., « Identifying factors affecting longitudinal survey response », in Lynn P. (dir.), *Methodology of Longitudinal Surveys*, Chichester, John Wiley and Sons, 2009.
- [26] Warszawski J., Najos N., & Costemalle V. et al., « 4 % de la population a développé des anticorps contre le SARS-CoV-2 entre mai et novembre 2020 », *Etudes et résultats*, n°1202, 2021.
- [27] Yan, T., & Curtin, R., « The relation between unit nonresponse and item nonresponse: A response continuum perspective », *International Journal of Public Opinion Research*, vol.22, n°4, pp.535-551, 2010.

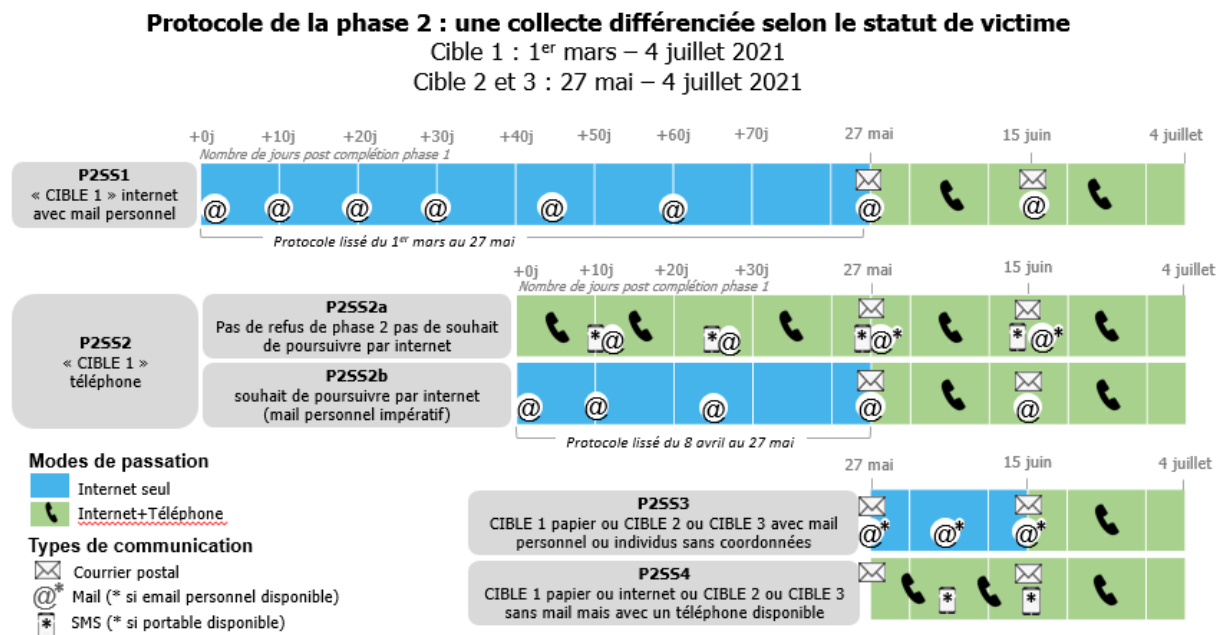
Annexes

Annexe 1 – Schéma de phase 1 de l'enquête



Source : SSMSI.

Annexe 2 – Schéma de phase 2 de l'enquête



Source : SSMSI.

Annexe 3 - Modélisation de la participation à l'enquête de phase 1 dans le P1SS2 (odds-ratio, modèles de régression logistiques)

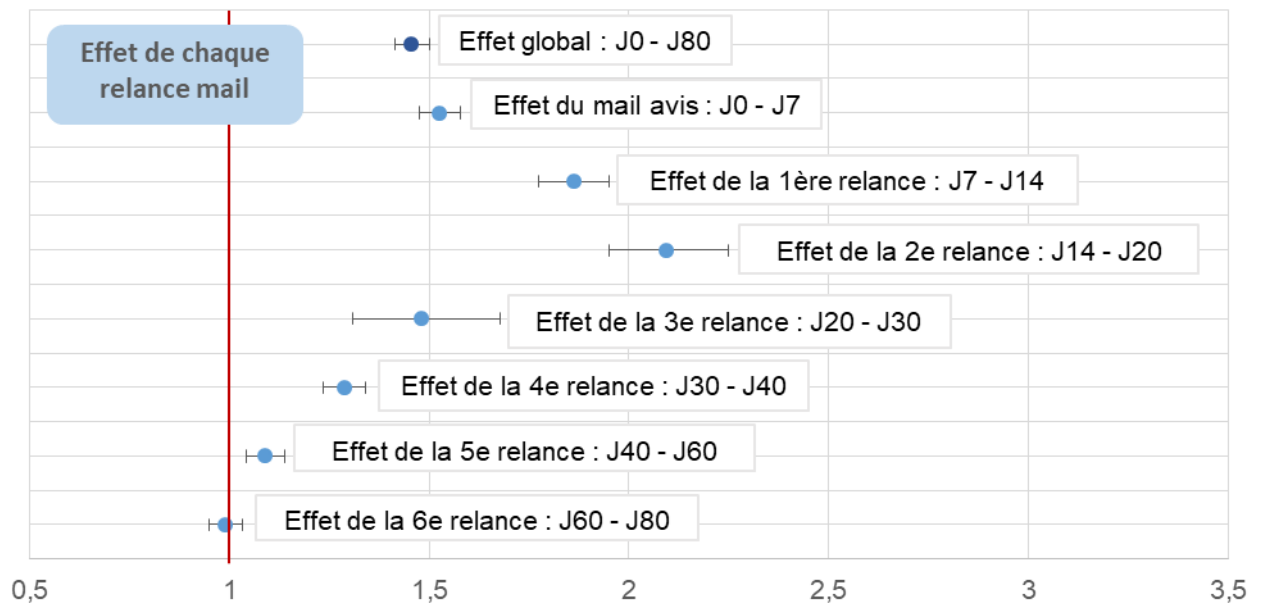
| Variables | Modalités | ODDS-RATIO | | | | |
|-----------------------------------|---|-----------------|-----------------|------------------|------------------|------------------|
| | | Modèle J0 - J80 | Modèle J0 - J20 | Modèle J20 - J40 | Modèle J40 - J60 | Modèle J60 - J80 |
| | | N = 142 101 | N = 142 101 | N = 91 055 | N = 73 905 | N = 56 635 |
| Sexe | Femmes | 0,793 *** | 0,873 *** | 0,894 *** | 0,843 *** | 0,829 *** |
| | Hommes | Réf. | Réf. | Réf. | Réf. | Réf. |
| Âge | 18 - 29 ans | 0,877 *** | 0,937 *** | 0,991 | 1,000 | 0,797 *** |
| | 30 - 39 ans | 0,788 *** | 0,838 *** | 0,837 *** | 0,917 *** | 0,833 *** |
| | 40 - 49 ans | Réf. | Réf. | Réf. | Réf. | Réf. |
| | 50 - 59 ans | 1,030 | 0,958 ** | 1,054 * | 1,015 | 1,073 ** |
| | 60 - 69 ans | 1,029 | 0,881 *** | 0,911 *** | 0,982 | 1,339 *** |
| Type de ménage | Personnes seules | 0,748 *** | 0,795 *** | 0,746 *** | 0,798 *** | 0,887 *** |
| | Couples sans enfant | 0,979 | 1,052 *** | 0,964 | 0,991 | 0,964 |
| | Couples avec enfant | Réf. | Réf. | Réf. | Réf. | Réf. |
| | Familles monoparentales | 0,731 *** | 0,758 *** | 0,856 *** | 0,761 *** | 0,862 *** |
| | Autres types de ménage | 0,648 *** | 0,707 *** | 0,770 *** | 0,748 *** | 0,753 *** |
| Statut dans le foyer fiscal | Référent fiscal | Réf. | Réf. | Réf. | Réf. | Réf. |
| | Conjoint ou concubin | 0,927 *** | 0,884 *** | 1,042 | 0,964 | 0,959 |
| | Enfant à charge | 1,116 *** | 1,151 *** | 1,038 | 1,097 ** | 1,007 |
| | Autres personnes à charge | 0,738 *** | 0,795 *** | 0,829 *** | 0,773 *** | 0,773 *** |
| Décile de niveau de vie du ménage | Non renseigné | 0,550 *** | 0,716 *** | 0,698 *** | 0,682 *** | 0,552 *** |
| | D1 | 0,593 *** | 0,701 *** | 0,690 *** | 0,702 *** | 0,656 *** |
| | D2 | 0,685 *** | 0,724 *** | 0,747 *** | 0,753 *** | 0,791 *** |
| | D3 | 0,770 *** | 0,829 *** | 0,821 *** | 0,840 *** | 0,827 *** |
| | D4 | 0,884 *** | 0,907 *** | 0,897 *** | 0,962 | 0,906 |
| | D5 | Réf. | Réf. | Réf. | Réf. | Réf. |
| | D6 | 1,053 * | 1,113 *** | 1,009 | 1,073 * | 0,933 ** |
| | D7 | 1,221 *** | 1,247 *** | 1,139 *** | 1,129 *** | 1,035 |
| | D8 | 1,411 *** | 1,399 *** | 1,232 *** | 1,246 *** | 1,118 *** |
| | D9 | 1,515 *** | 1,566 *** | 1,251 *** | 1,326 *** | 1,061 |
| | D10 | 1,489 *** | 1,637 *** | 1,261 *** | 1,247 *** | 1,018 |
| Taille d'unité urbaine | Communes rurales | 0,985 | 0,865 *** | 0,916 *** | 1,022 | 1,114 *** |
| | 2 000 - moins de 20 000 hab. | 0,965 * | 0,902 *** | 0,922 *** | 1,015 | 1,041 |
| | 20 000 - moins de 200 000 hab. | 1,000 | 0,956 ** | 0,944 ** | 1,038 | 1,053 * |
| | 200 000 hab. ou plus Ubité urbaine de Paris | Réf. | Réf. | Réf. | Réf. | Réf. |
| Quartier prioritaire de la ville | Non | Réf. | Réf. | Réf. | Réf. | Réf. |
| | Oui | 0,726 *** | 0,688 *** | 0,810 *** | 0,871 *** | 0,839 *** |
| Présence d'un numéro de téléphone | Non | Réf. | Réf. | Réf. | Réf. | Réf. |
| | Oui | 1,407 *** | 1,134 *** | 1,128 *** | 1,490 *** | 1,671 *** |
| Envoi de mail | Non | Réf. | Réf. | Réf. | Réf. | Réf. |
| | Oui | 1,457 *** | 1,859 *** | 1,309 *** | 1,089 *** | 0,991 |
| Relance téléphonique | Non | | | | 3,013 *** | 1,550 *** |
| | Oui | | | | Réf. | Réf. |

Légende: Réf. : situation de référence ; *** significatif à 1 % ; ** significatif à 5 % ; * significatif à 10 %.

Note : modèles conduits dans le P1SS2 uniquement.

Source : enquête Genese, SSMSI.

Annexe 4 – Effet des relances par mail sur la probabilité de répondre au cours des différentes périodes du terrain de phase 1 dans le P1SS2



Note : 8 modèles de régression logistique ont été mis en œuvre afin d'évaluer l'effet de chaque envoi de mails dans le P1SS2. Le premier modèle « J0 – J80 » (N=142 101) mesure la probabilité de répondre entre J0 et la fin du terrain parmi les individus du P1SS2 ayant reçu la lettre-avis ; le deuxième modèle « J0-J7 » (N=142 101) mesure la probabilité de répondre entre J0 et l'envoi de la 1ère relance par mail à J7 parmi les individus du P1SS2 ayant reçu la lettre-avis ; le troisième modèle « J7-J14 » (N=113 366) mesure la probabilité de répondre entre J7 et l'envoi de la 2e relance par mail à J14 parmi les individus du P1SS2 ayant reçu la lettre-avis ; le quatrième modèle « J14-J20 » (N=100 457) mesure la probabilité de répondre entre J14 et l'envoi de la 3e relance par mail à J20 parmi les individus du P1SS2 ayant reçu la lettre-avis ; le cinquième modèle « J20-J30 » (N=91 055) mesure la probabilité de répondre entre J20 et J30 parmi les individus du P1SS2 ayant reçu la première lettre de relance ; le sixième modèle « J30-J40 » (N=89 493) mesure la probabilité de répondre entre J30 et J40 parmi les individus du P1SS2 ayant reçu la première lettre de relance ; le septième modèle « J40-J60 » (N=73 905) mesure la probabilité de répondre entre J40 et J60 parmi les individus du P1SS2 ayant reçu la deuxième lettre de relance ; le huitième modèle « J60-J80 » (N=55 635) mesure la probabilité de répondre entre J60 et J80 parmi les individus du P1SS2 ayant reçu la troisième lettre de relance. Les variables de contrôle introduites dans le modèle sont les informations disponibles dans l'échantillon (sexe, âge en 5 tranches, type de ménage en 5 catégories, statut fiscal en 4 catégories, décile de niveau de vie, TUU en 5 tranches, indicatrice de résidence en QPV et présence d'un numéro de téléphone). Les données représentées sur le graphique correspondent aux odds ratios et leurs intervalles de confiance à 95 % des indicatrices « envoi de mail » obtenus dans les 8 différents modèles.

Source : enquête Genese, SSMSI.

Annexe 5 - Différences standardisées selon le sous-échantillon et l'avancée de la collecte

| | P1SS1 | | P1SS2a | | P1SS2b | |
|--------------------------------|-------|-------|--------|-------|--------|-------|
| | J20 | J80 | J60 | J80 | J60 | J80 |
| Sexe | | | | | | |
| Femme | 15,1 | 11,8 | 7,4 | 5,5 | 7,8 | 9,2 |
| Homme | -15,1 | -11,8 | -7,4 | -5,5 | -7,8 | -9,2 |
| Age | | | | | | |
| 18-29 ans | 6,1 | -21,4 | -4,6 | -4,9 | -5,4 | -16,3 |
| 30-39 ans | 2,5 | -7,0 | -2,9 | -2,3 | -12,8 | -16,3 |
| 40-49 ans | 7,4 | 2,4 | 2,0 | 1,3 | 0,4 | -0,3 |
| 50-59 ans | 13,3 | 8,6 | 3,1 | 1,7 | 6,4 | 6,6 |
| 60-74 ans | -25,1 | 13,6 | 2,2 | 3,8 | 9,9 | 22,0 |
| Type de commune | | | | | | |
| Rurale | 15,3 | 14,7 | 8,1 | 7,1 | 11,1 | 11,7 |
| Urbaine | -15,3 | -14,7 | -8,1 | -7,1 | -11,1 | -11,7 |
| Résidence en QPV | | | | | | |
| QPV | -24,3 | -12,6 | -13,2 | -11,7 | -18,7 | -14,2 |
| Hors QPV | 24,3 | 12,6 | 13,2 | 11,7 | 18,7 | 14,2 |
| Type de ménage (fiscal) | | | | | | |
| Personne seule | -46,3 | -26,2 | -12,5 | -6,2 | -14,0 | -8,2 |
| Couple sans enfant | 1,9 | 13,1 | 8,3 | 5,6 | 11,7 | 12,7 |
| Couple avec enfants | 44,6 | 17,7 | 12,4 | 7,4 | 13,6 | 6,3 |
| Famille monoparentale | -0,3 | -2,6 | -9,6 | -5,5 | -10,4 | -9,6 |
| Autre type de ménage | -7,9 | -5,0 | -7,5 | -7,1 | -12,8 | -9,3 |
| Décile de niveau de vie | | | | | | |
| d01 | -10,4 | -10,6 | -13,9 | -12,4 | -19,7 | -15,3 |
| d02 | 12,3 | -1,8 | 11,0 | -7,4 | -13,3 | -7,4 |
| d03 | -4,4 | 4,1 | -4,2 | -2,6 | -10,8 | -5,0 |
| d04 | 3,2 | 4,9 | -0,5 | 1,0 | -2,3 | -0,5 |
| d05 | 5,7 | 9,5 | 2,4 | 1,8 | 0,4 | 3,7 |
| d06 | 6,1 | 8,9 | 6,5 | 4,5 | 3,2 | 3,0 |
| d07 | 11,2 | 10,0 | 4,4 | 5,0 | 8,5 | 6,6 |
| d08 | 12,0 | 11,1 | 8,5 | 5,8 | 12,2 | 9,3 |
| d09 | 22,0 | 12,2 | 9,1 | 6,6 | 16,1 | 8,9 |
| d10 | 27,7 | 13,4 | 6,7 | 5,5 | 13,8 | 7,1 |
| NA | -38,4 | -44,4 | -12,9 | -11,6 | -16,5 | -17,0 |

Source : enquête Genese, SSMSI.

Annexe 6 – Modélisation de la participation à l'enquête de suivi (phase 2) (odds-ratio, modèle de régression logistique)

| | Odds-ratio |
|--|-------------|
| Sexe | |
| Femme | 1,03 |
| <i>Homme</i> | <i>Réf.</i> |
| Âge | |
| 18 à 29 ans | 0,81*** |
| 30 à 39 ans | 0,77*** |
| <i>40 à 49 ans</i> | <i>Réf.</i> |
| 50 à 59 ans | 1,18*** |
| 60 à 74 ans | 1,41*** |
| Type de ménage (fiscal) | |
| Personne seule | 0,9* |
| Couple sans enfant | 1,09 |
| <i>Couple avec enfant</i> | <i>Réf.</i> |
| Famille monoparentale | 0,82*** |
| Autre type de ménage | 0,86* |
| Décile de niveau de vie du ménage | |
| NA | 0,77** |
| 1er décile | 0,89 |
| 2e décile | 1,04 |
| 3e décile | 1,07 |
| 4e décile | 1,06 |
| <i>5e décile</i> | <i>Réf.</i> |
| 6e décile | 1,19** |
| 7e décile | 1,26*** |
| 8e décile | 1,24** |
| 9e décile | 1,20** |
| 10e décile | 1,16* |
| Plus haut niveau de diplôme | |
| Aucun diplôme | 0,55*** |
| Inférieur au bac | 0,63*** |
| Bac | 0,79*** |
| Niveau bac+2 | 0,86** |
| <i>Niveau supérieur à bac+2</i> | <i>Réf.</i> |
| Nationalité | |
| <i>Française</i> | <i>Réf.</i> |
| Etrangère | 0,72*** |
| Type de logement | |
| <i>Maison</i> | <i>Réf.</i> |
| Appartement | 0,92* |
| Autre type de logement | 1,31 |
| NR | 0,98 |

| | |
|--|---------|
| Habite en Ile-de-France | |
| Oui | 0,93 |
| Non | Réf. |
| Type de territoire | |
| Rural | 1,06 |
| Urbain | Réf. |
| Quartier prioritaire de la ville | |
| Oui | 0,88 |
| Non | Réf. |
| Quartier de reconquête républicaine | |
| Oui | 0,99 |
| Non | Réf. |
| Etat de santé perçu | |
| Bon, très bon | Réf. |
| Moyen, mauvais | 0,97 |
| Très mauvais | 0,69** |
| NR | 1,03 |
| A laissé un numéro de téléphone personnel en fin de phase 1 | |
| Oui | 2,16*** |
| Non | Réf. |
| A laissé un mail personnel en fin de phase 1 | |
| Oui | 1,36*** |
| Non | Réf. |
| Est joignable par un autre numéro de téléphone (Fidéli ou enrichissement) | |
| Oui | 1,71*** |
| Non | Réf. |
| Cible | |
| Cible 1 | 2,48*** |
| Cible 2 | 1,03 |
| Cible 3 | Réf. |
| Mode de réponse en phase 1 | |
| Papier | 0,64*** |
| Internet | Réf. |
| Téléphone | 0,98 |
| Période de réponse en phase 1 | |
| Avant J20 | 1,79*** |
| Entre J20-J40 | Réf. |
| Entre J41-J60 | 0,69*** |
| Après J60 | 0,59*** |
| Temps de remplissage du questionnaire de phase 1 | |
| Moins de 20 min | 0,87*** |
| De 20 à 30 min | Réf. |
| De 30 à 40 min | 1,03 |
| Plus de 40 min | 0,99 |

Légende: Réf. : situation de référence ; *** significatif à 1 % ; ** significatif à 5 % ; * significatif à 10 %.
Source : enquête Genese, SSMSI.