

N°21
Octobre 2023

Méthode

Méthodologie de l'enquête Genre et Sécurité (Genese)

Auteurs : Danaé Odin-Steiner, Tiaray Razafindranovona, Sandra Zilloniz
Rédacteur en chef : Olivier Filatriau

Les auteurs remercient Hélène Guedj pour son travail sur le protocole de collecte de l'enquête et pour la constitution du dossier du label à l'attention du Conseil national de l'information statistique auquel elle a beaucoup contribué.

Table des matières

1	Présentation générale de l'enquête Genese : Contexte et objectif	5
2	Périmètre, questionnaire et protocole de collecte	5
2.1	Champ	5
2.2	Protocole de collecte	6
2.3	Questionnaire et principaux thèmes abordés	9
2.3.1	Questionnaire de phase 1	9
2.3.2	Questionnaire de phase 2 Eurostat : structure et thèmes abordés	10
3	Plan de sondage et échantillonnage	13
3.1	Base de sondage	13
3.2	Échantillonnage de la phase 1 (enquête filtre)	14
3.2.1	Objectifs et architecture générale	14
3.2.2	Taille d'échantillon	14
3.2.3	Stratification explicite et surreprésentation	15
3.2.4	Allocations cibles et calcul d'allocations par strates	16
3.2.5	Stratification implicite	17
3.2.6	Échantillonnage de la relance téléphonique	17
3.3	Échantillonnage de la phase 2 (enquête filtrée)	18
3.3.1	Objectifs et architecture générale	18
3.3.2	Taille d'échantillon	18
3.3.3	Stratification explicite et surreprésentation	18
4	Déroulement de la collecte	19
4.1	Enquête pilote	19
4.2	Enquête de phase 1 (enquête filtre)	20
4.3	Enquête de phase 2 (enquête filtrée)	22
5	Enquête filtre en face-à-face auprès des non répondants à la phase 1 de l'enquête	23
6	Traitements post-collecte de l'enquête de phase 1 (enquête filtre)	24
6.1	Identification et traitement des hors champ	24
6.2	Contrôles de cohérence et apurement	25
6.3	Correction de la non-réponse partielle	26
6.3.1	Principe général	27

6.3.2	Application et résultat	27
6.4	Traitement de la non-réponse totale	28
6.4.1	Principe général	28
6.4.2	Modélisation de la non-réponse et constitution des GRH	29
6.4.3	Résultats	30
6.5	Calage et calcul des pondérations	30
6.5.1	Principe général	30
6.5.2	Application et choix des marges	31
6.5.3	Résultats	31
6.6	Précision des indicateurs d'intérêt	32
6.6.1	Principe général	32
6.6.2	Application et résultats	32
7	Traitements post-collecte de l'enquête de phase 2 (enquête filtrée)	33
7.1	Identification et traitement des hors champ	34
7.2	Contrôles de cohérence et apurement	34
7.3	Traitement de la non-réponse totale	34
7.3.1	Modélisation de la non-réponse	35
7.3.2	Résultats	36
7.4	Calage et calcul des pondérations	36
7.4.1	Application et choix des marges	36
7.4.2	Résultats	37
7.5	Précision des indicateurs d'intérêt	38
A	Distributions départementales des poids de phase 1	40
B	Distributions des poids de phase 1 et phase 2	42

1 Présentation générale de l'enquête Genese : Contexte et objectif

L'enquête Genese (Genre et sécurité) est une enquête de victimation d'initiative européenne¹. Dans le cadre d'un appel à projets européen, le Service statistique ministériel de la sécurité intérieure (SSMSI) a obtenu en novembre 2019 un financement pour conduire en 2021, au niveau national, une enquête portant sur les violences liées au genre et basée sur le questionnaire Gender-Based Violence développé par Eurostat. L'enquête Genese a été conçue par le SSMSI pour répondre à deux ambitions majeures :

- répondre à la forte demande sociale d'informations sur les violences sexistes et sexuelles engendrée par l'affaire Weinstein fin 2017 et la vague #MeToo qui a suivi en recueillant de l'information plus complète que celle disponible trop partiellement dans l'enquête « Cadre de vie et sécurité » (CVS) ou dans les données administratives sur cette thématique mais également plus récente que les données de l'enquête Virage menée en 2015 par l'Ined ;
- réaliser une expérimentation grandeur réelle d'une enquête multimode pour mesurer les victimations et les perceptions en matière de sécurité en vue de la refonte de l'enquête « Cadre de vie et sécurité » à l'horizon 2022. L'enquête « Cadre de vie et sécurité » a été conduite chaque année de 2007 à 2021 par l'Insee en co-maîtrise d'ouvrage avec l'Observatoire national de la délinquance et des réponses pénales (ONDRP, jusqu'à sa disparition en 2020) et le SSMSI² (créé en 2014). Les enseignements de l'enquête Genese ont permis d'aboutir à la conception de la première édition de l'enquête « Vécu et ressenti en matière de sécurité » (VRS) réalisée par le SSMSI au deuxième trimestre 2022³.

L'enquête Genese a reçu le label d'intérêt général et de qualité statistique délivré par le Conseil national de l'information statistique (Cnis) et le caractère obligatoire⁴.

2 Périmètre, questionnaire et protocole de collecte

2.1 Champ

Le champ de l'enquête Genese est la population des individus âgés de 18 à 74 ans au 1^{er} janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire⁵. Ce périmètre s'applique aux deux phases de l'enquête Genese : l'enquête filtre et l'enquête filtrée.

1. <https://www.interieur.gouv.fr/Interstats/L-enquete-GENESE>

2. <https://www.interieur.gouv.fr/Interstats/L-enquete-Cadre-de-vie-et-securite-CVS>

3. <https://www.interieur.gouv.fr/Interstats/L-enquete-Vecu-et-ressenti-en-matiere-de-securite-VRS>

4. https://www.cnis.fr/wp-content/uploads/2021/02/AC_2021_SSMSI_Genese.pdf

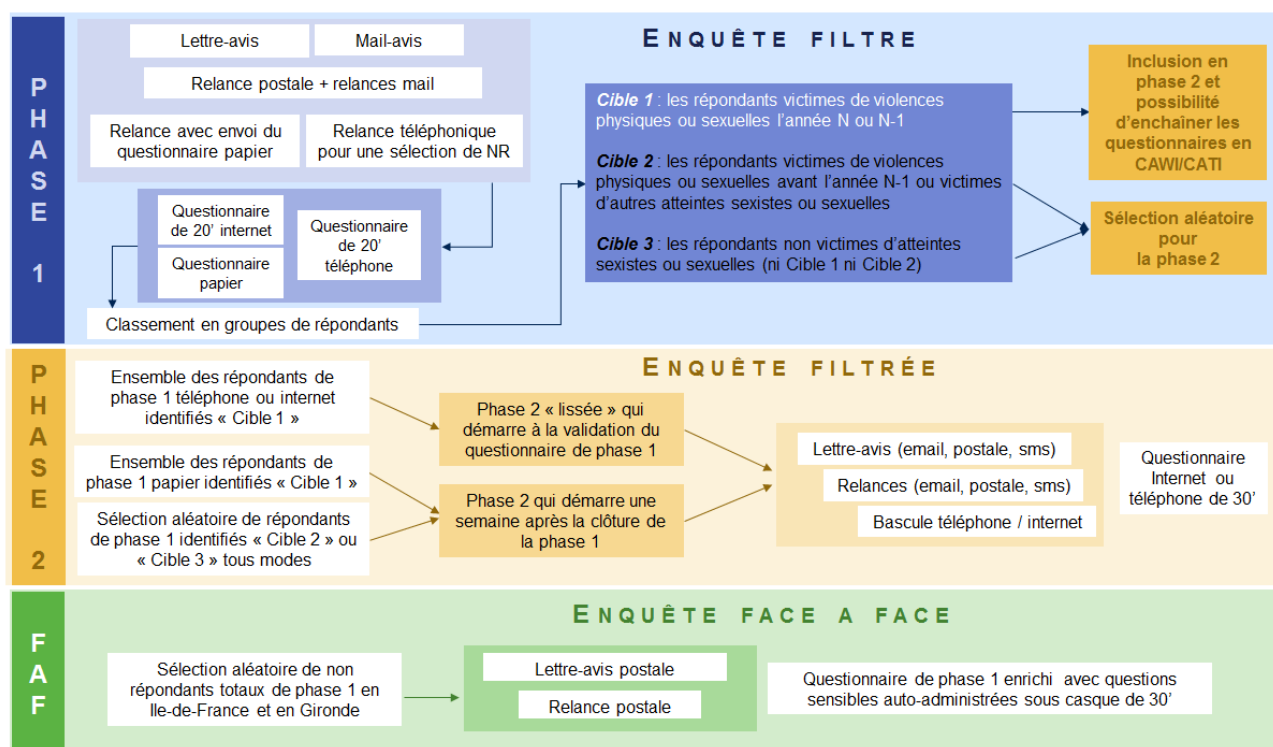
5. À l'exclusion des logements associés à une adresse de communauté, ainsi que quelques logements atypiques (camping, résidence hôtelière, bateau) et les domiciliations. Pour la plupart, ces logements exclus concernent les foyers fiscaux non redevables de la taxe d'habitation.

2.2 Protocole de collecte

Un protocole multimode en deux phases sur large échantillon pour permettre des estimations plus précises et des zooms sur des thématiques ou des populations particulières

En juillet 2019, le SSMSI a mis en place un groupe de travail méthodologique associant des experts des enquêtes ménages multimodes, la division Sondages ainsi que l'équipe de conception de CVS de l'Insee. L'objectif du nouveau protocole était de répondre à un coût soutenable de manière innovante et fiable aux besoins émergents de données localisées et plus précises pour permettre des suivis conjoncturels, tout en assurant une description détaillée des victimes et des atteintes subies. L'objectif d'estimation infranationale et de suivi conjoncturel de certains indicateurs requiert un échantillon important d'individus. L'objectif de description détaillée des victimations requiert un échantillon important de victimes. Pour atteindre ces deux objectifs, le SSMSI a ainsi conçu, en partenariat avec les experts de l'Insee, un protocole multimode en deux phases, en s'inspirant de l'enquête de victimation du Cespip menée en 1986 ou plus récemment de l'enquête VQS-Care de la Drees qui s'intéresse aux personnes dépendantes parmi les 60 ans ou plus (figure 1).

FIGURE 1 – Articulation générale du dispositif d'enquête Genese



La première phase consiste en une enquête multimode (internet, papier, téléphone) généraliste basée sur des questions de recensement des victimations, de perceptions et d'opinion en matière de sécurité

auprès d'un large échantillon. Cette première phase permet de produire les principaux indicateurs de victimation à l'échelle nationale et infranationale. Elle a interrogé un échantillon de 169 060 personnes âgées de 18 à 74 ans au 1er janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire, dans lequel les femmes étaient surreprésentées (70 %-30 %). Ces personnes ont été sélectionnées par l'Insee dans le répertoire des logements géré par l'institut à partir de sources administratives (Fidéli). La collecte a commencé par internet puis prévoyait 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. À l'issue de la phase 1, les répondants sont classés en 3 groupes en fonction de leurs réponses sur des faits d'atteintes à la personne. Le paragraphe 3.3.3 décrit la construction de ces groupes dits *cibles*.

La seconde phase consiste en une enquête de suivi permettant d'approfondir une thématique auprès d'un échantillon plus réduit de répondants de phase 1 surreprésentant les individus particulièrement concernés par la thématique. Lors de l'enquête Genese, elle était basée sur un questionnaire conçu par Eurostat permettant le recueil d'informations très détaillées sur les violences sexistes et sexuelles subies tout au long de la vie (« questionnaire Eurostat »). Elle a été proposée par internet et téléphone à un échantillon de 15 000 répondants de phase 1 dans lequel les victimes de violences physiques et sexuelles ont été surreprésentées. Les répondants de phase 1 identifiés cible 1 (les victimes de violences physiques ou sexuelles récentes) sont tous inclus en phase 2. Ceux ayant répondu par internet ou téléphone ont la possibilité d'enchaîner les deux questionnaires. Les répondants de phase 1 identifiés cible 2 ou cible 3 font l'objet d'une sélection aléatoire à la fin de la phase 1 surreprésentant les cible 2 (voir la partie 3.3.3 pour le détail de la stratification explicite et de la surreprésentation).

Enfin, au terme de la phase 1, un volet méthodologique en face-à-face a été conduit du 1er juin au 24 juillet 2021 auprès de plus de 3 000 personnes : les 1 000 non répondants totaux de Gironde et une sélection aléatoire de 2 000 non répondants totaux en Île-de-France.

Le protocole de la phase 1 combine 3 modes de collecte : internet, papier et téléphone. La relance par téléphone est proposée après environ six semaines à une sélection aléatoire de non répondants dont le nombre est fixé *a priori*. Ce choix permet de maîtriser les coûts de l'enquête en contrôlant le nombre d'entretiens téléphoniques à réaliser. Sur le plan méthodologique, il offre un cadre quasi-expérimental permettant la comparaison des réponses des enquêtés éligibles sélectionnés pour la relance CATI avec les enquêtés éligibles non sélectionnés dans une optique d'évaluation des effets de mode. Le questionnaire papier est proposé aux non répondants à différents stades de la collecte en fonction des coordonnées de contact disponibles pour les joindre afin d'offrir un mode supplémentaire aux enquêtés tout en réduisant le recours au papier qui engendre de la non-réponse partielle et des traitements

post-collecte beaucoup plus conséquents que les autres modes.

La collecte multimode

Les deux phases de l'enquête Genese combinent plusieurs modes de passation : internet (CAWI), papier (uniquement pour le questionnaire de phase 1) et téléphone (CATI). Ce protocole s'inscrit dans les grands modèles promus par le programme multimode de l'Insee.

L'introduction de modes de collecte moins coûteux et plus modernes que le face-à-face constitue une piste de choix pour augmenter la taille d'échantillon et améliorer la précision des estimations. Néanmoins ces changements de protocole posent la question de la confidentialité et de la sécurité offertes aux enquêtés pour révéler des faits sensibles mais également de la continuité des séries historiques.

La passation auto-administrée en ligne pose également certaines limites concernant :

- le contrôle de l'identité de la personne qui répond ;
- la bonne compréhension du questionnaire ;
- l'équipement et l'aisance avec l'outil informatique (plus compliqués pour les enquêtés plus âgés ou les personnes de milieux sociaux plus modestes) ;
- les difficultés de lecture ;
- le risque de non-réponse partielle ou d'abandon avant la fin du questionnaire s'il est jugé trop long ou trop complexe.

Les avantages et inconvénients des différents modes sont désormais assez bien documentés. Les personnes qui vont répondre par internet seront *a priori* plus intéressées par le sujet, plus connectées et/ou plus à l'aise avec le support écrit. Les moyens de relance papier et/ou téléphone plus coûteux pourront être concentrés sur des populations plus difficiles à capter sur internet. D'un point de vue technique, la programmation combinée CAWI/CATI permet un questionnement long et complexe (grâce aux filtres programmés), la passation en plusieurs fois et la collecte de parodonnées (comme le nombre de tentatives d'appels avant d'obtenir un questionnaire validé, le nombre de connexions, la date et l'heure de chaque connexion, le temps mis pour répondre par partie du questionnaire...) utiles pour tirer des enseignements sur le déroulement de l'enquête et pour les enquêtes futures.

En dehors de ces généralités, dans le cas spécifique de l'enquête Genese, la taille d'échantillon envisagée pour parvenir à sur-échantillonner les victimes a exclu *de facto* le recours au face-à-face, mode trop coûteux. En outre, au-delà de la seule dimension budgétaire, l'ambition du projet était de proposer un protocole innovant à expérimenter pour alimenter la réflexion sur la refonte de l'enquête de victimation CVS. 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 TIC⁶ de l'Insee, car le questionnaire est court et simple et permet de réduire le recours à la relance téléphonique plus coûteuse. Le volet papier dans l'enquête TIC augmente significativement le taux de réponse global de l'enquête⁷. Enfin, les informations recueillies dans l'enquête Genese (en phase 2 principalement) sont particulièrement sensibles. Le téléphone comme l'utilisation d'internet sont des modes de collecte peu intrusifs et confidentiels qui permettent de limiter les biais de désirabilité sociale. Les grandes enquêtes statistiques sur la thématique des violences sensibles ont été réalisées en France par téléphone (enquête Enveff en 2000 et enquête Virage en 2015 réalisées par l'Ined⁸) ou de façon auto-administrée (passation sous casque) dans des protocoles face-à-face (enquête EVS⁹ de la Drees en 2005-2006 et les enquêtes annuelles CVS depuis 2007). Par ailleurs, les premières conclusions de l'expérimentation multimode (internet/téléphone) de l'enquête CVS conduite en 2019 auprès de répondants à l'enquête 2018 montrent que le volet téléphone améliore significativement la représentativité de l'échantillon final de répondants.

L'enchaînement entre les 2 phases : un protocole différencié selon le statut de victime

L'autre idée originale et innovante du protocole consiste à proposer un schéma de contact et de relance personnalisé pour l'enquête de suivi thématique (phase 2) : les profils les plus qualifiés pour l'enquête de suivi (cible 1) ayant une probabilité de sélection égale à 1 pour la phase 2 ont en effet la possibilité d'enchaîner les deux questionnaires en une seule interrogation (en CAWI/CATI). S'ils n'enchaînent pas les deux questionnaires, ces répondants sont sollicités et relancés à une cadence régulière et identique qui démarre le jour de complétion du questionnaire de phase 1. Ce protocole différencié permet à la fois de :

- proposer une durée de collecte plus longue et un calendrier de relance personnalisé favorisant la réponse et réduisant l'attrition des répondants de phase 1 cible 1 qui représentent la population d'intérêt pour la phase 2 ;
- conserver le contrôle sur la sélection des répondants cible 2 et cible 3 en fonction des résultats de la collecte de phase 1.

2.3 Questionnaire et principaux thèmes abordés

2.3.1 Questionnaire de phase 1

Le questionnaire de l'enquête filtre (phase 1) comprend schématiquement 5 blocs (figure 2) :

- la « phase contact » permet de vérifier l'identité (nom, prénom, date de naissance) de la personne

6. TIC : enquête sur les technologies de l'information et de la communication

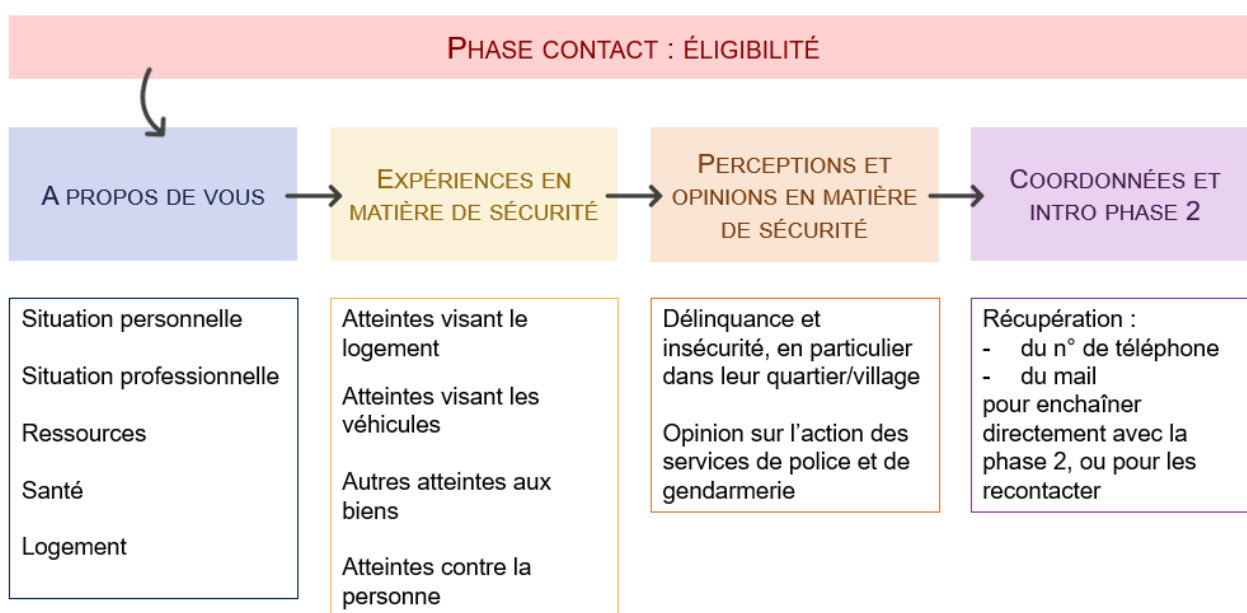
7. 45 % des répondants ont rempli le questionnaire papier en 2019. Celui-ci était envoyé 4 semaines après la lettre-avis.

8. Enveff : enquête nationale sur les violences faites aux femmes ; Virage : enquête sur les violences et rapports de genre

9. EVS : enquête Événements de vie et santé.

- connectée au questionnaire afin de s'assurer qu'elle est la personne sélectionnée pour l'enquête ;
- le bloc « À propos de vous » recueille les descripteurs socio-démographiques ;
 - les blocs « Expériences en matière de sécurité » et « Perceptions et opinions en matière de sécurité » permettent de recenser un large champ d'atteintes détaillées (cambriolages, 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 autres harcèlements etc.) au cours de différentes périodes (en 2020 et avant 2020), les taux de plainte associés en 2020, de mesurer le sentiment d'insécurité, les phénomènes délinquants observés dans le quartier ainsi que l'opinion sur l'action des forces de sécurité (police et gendarmerie). Les questions posées dans ces deux blocs sont reprises de l'enquête CVS et ont été adaptées au mode auto-administré ;
 - le bloc « coordonnées et intro phase 2 » annonce l'enquête de suivi et prévoit le recueil des coordonnées personnelles du répondant afin de le recontacter pour la phase 2 de l'enquête.

FIGURE 2 – Enquête filtre : objectif et contenu



2.3.2 Questionnaire de phase 2 Eurostat : structure et thèmes abordés

Une conception et un déploiement à l'échelle européenne

Début 2017, le groupe de travail « délinquance et justice pénale » (working group on statistics on crime and criminal justice) met en place une Task Force spécifiquement dédiée au développement de l'enquête Gender Based Violence Survey pilotée par la Direction des statistiques sociales au sein

d'Eurostat. Cette Task Force est composée de représentants des Etats membres, de la Direction générale de la Justice, de la FRA (Agence des droits fondamentaux de l'Union européenne), d'EIGE (Institut européen pour l'égalité entre les hommes et les femmes) et d'experts indépendants.

Entre 2017 et 2018, Eurostat a subventionné plusieurs actions de différents pays visant à tester les versions successives du questionnaire. Des enquêtes dites pilote, avec des protocoles variés, ont été conduites dans 11 pays suite à un appel à projets publié par Eurostat en mars 2017 afin de tester la pertinence du questionnaire pour mesurer les phénomènes d'intérêt (comportements sexistes et sexuels au travail, violences au sein du couple actuel ou passé, violences sexuelles, etc. : prévalence, profil des victimes et des auteurs et contextes des faits). Ces expérimentations ont conduit à un allègement et à des reformulations du questionnaire de l'enquête, jugé trop long et complexe. Des recommandations pour limiter l'interrogation aux 18-74 ans, en raison des difficultés rencontrées pour interroger les personnes plus âgées, ont été formulées par plusieurs pays ayant réalisé des enquêtes pilotes.

Depuis 2020, 17 États membres, dont la France, ont mis en œuvre l'enquête avec le soutien financier d'Eurostat : Bulgarie, Danemark, Estonie, Grèce, Espagne, France, Croatie, Lettonie, Lituanie, Malte, Pays-Bas, Autriche, Pologne, Portugal, Slovaquie, Slovénie, Finlande. En outre, l'agence des droits fondamentaux de l'Union européenne (FRA) et l'institut européen pour l'égalité entre les hommes et les femmes (EIGE) mettront en œuvre l'enquête conformément aux lignes directrices d'Eurostat dans huit autres États membres : Tchéquie, Allemagne, Irlande, Chypre, Luxembourg, Hongrie, Roumanie et Suède. Par ailleurs, la Belgique fournira les données d'une enquête mise en œuvre sur fonds nationaux à partir du matériel de collecte d'Eurostat et enfin, l'Italie partagera les données basées sur son enquête nationale sur la violence envers les femmes. L'ensemble des données devrait être transmis à Eurostat à la fin de l'année 2023 (sauf l'Italie qui a informé Eurostat d'une livraison en 2024). Le mode de collecte, la taille de l'échantillon et la population cible (hommes et femmes ou femmes seulement) ont été laissés au libre choix des États membres.

Par ailleurs, une proposition de directive européenne sur l'obligation de la part des États membres de mener tous les cinq ans une enquête sur les violences liées au genre a été adoptée le 8 mars 2022. La méthodologie pour recueillir ces données sera harmonisée au niveau de la Commission (Eurostat).

Cinq grands modules

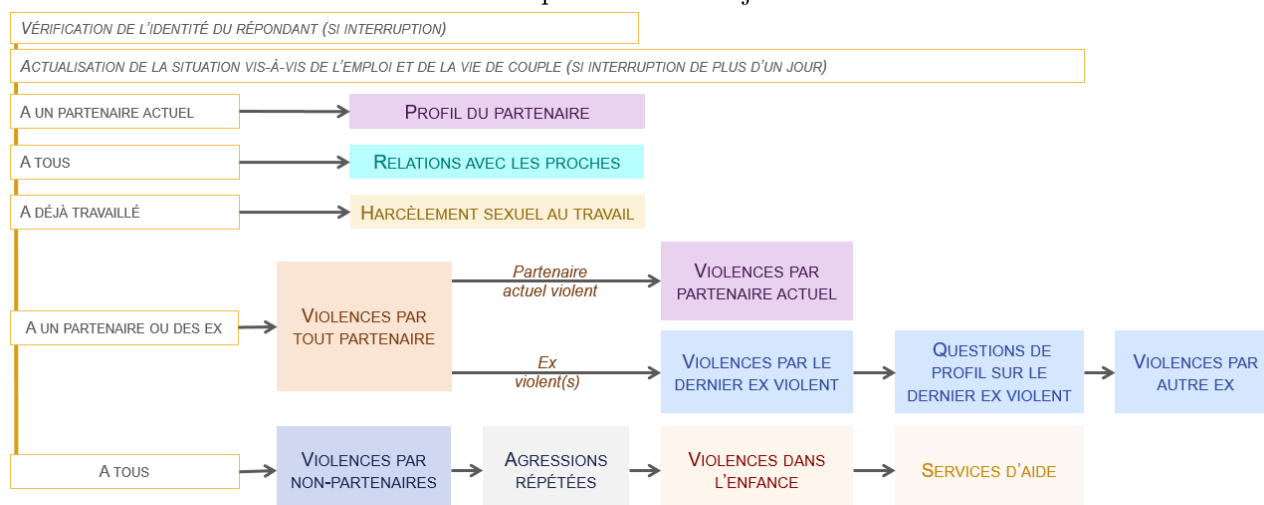
Le « questionnaire Eurostat » interroge des hommes et des femmes sur les violences sexistes et sexuelles subies au cours de la vie, sans jamais utiliser le terme de violence, ni de victime. Le questionnaire Gender-Based Violence comporte cinq grandes sections (figure 3) pour aider le répondant à se remé-

morer des épisodes spécifiques, dans l'ordre suivant :

- comportements sexistes ou sexuels au travail : ils font référence aux conduites verbales, non verbales ou physiques non désirées de nature sexuelle qui ont pu arriver au cours de l'ensemble de la vie professionnelle et aux auteurs avec lesquels la victime a ou a eu une relation professionnelle (comme un collègue, un patron, un client). Ce module permet de connaître les faits subis, les auteurs en cause, la fréquence des incidents et leur signalement par les victimes. Concernant l'emploi actuel ou le dernier emploi, des questions sur l'existence de ce type de comportements ainsi que sur la mise en place de procédures pour lutter contre ces agissements sur le lieu de travail (formation, sensibilisation, référents, etc.) sont également posées dans ce module ;
- violences par partenaire (actuel/ex) : ce module permet d'identifier les différents types de violences – psychologiques, y compris économique, menaces, violences physiques et sexuelles – exercées par un partenaire intime actuel (au moment de l'enquête) et/ou par tout ancien partenaire intime au cours de la vie. En plus du type de violence, ce module permet de dresser une description des faits, tels que leur fréquence, leur gravité, leurs impacts et conséquences (blessures, violence pendant la grossesse, enfants témoins/victimes, etc.) et les mesures prises par les victimes (signalement, recours aux services d'aide, etc.) ;
- violences par un non-partenaire depuis l'âge de 15 ans : elles comprennent les menaces, les violences physiques et sexuelles commises par tout autre personne qu'un partenaire actuel ou un ancien partenaire intime (inconnu, connaissance, ami, collègue, enseignant, voisin, membre de la famille, etc.) subies depuis l'âge de 15 ans. Les informations sont collectées sous forme d'épisodes pour chaque auteur ou groupe d'auteurs impliqués pour dater, renseigner la fréquence des faits, leur durée, leur gravité, leur impact et conséquences, leur signalement/recours aux services de santé ou d'aide ;
- agressions répétées : elles concernent les comportements ou les actes offensants ou menaçants répétés survenus au cours de la vie et commis par tout type d'auteur. Ce module permet d'identifier les situations d'agressions répétées vécues, les types d'auteurs impliqués, la durée des faits, leur fréquence, l'âge des premiers faits, de préciser la date du dernier fait et leur(s) auteur(s) et de connaître les mesures prises par les victimes (signalement, recours aux services d'aide, etc.) ;
- violences dans l'enfance avant l'âge de 15 ans : elles concernent les violences physiques exercées par les parents, les violences sexuelles commises par toute personne et le fait d'avoir été témoin de violences entre ses parents avant l'âge de 15 ans. Ce module collecte, pour toutes les formes de violences, la fréquence des faits et, pour les violences à caractère sexuel, le type d'auteur impliqué, l'âge au premier incident et leur signalement/recours aux services de santé ou d'aide/le

fait d'en parler à quelqu'un (famille, proches, école, etc.).

FIGURE 3 – Enquête filtrée : objectif et contenu



3 Plan de sondage et échantillonnage

3.1 Base de sondage

La base de sondage utilisée pour l'échantillonnage de l'enquête Genese est une table d'individus issue des fichiers fiscaux et plus précisément du millésime 2019 du dispositif Fidéli.

Cette base est construite à partir des données de la Direction générale des finances publiques recoupant 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.

La table utilisée comme base de sondage est composée de tous les individus présents dans le champ du dispositif Fidéli au 1er janvier 2019 et ayant 18 à 74 ans au 1er janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire, soit environ 46 millions d'individus. En particulier, 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.

L'utilisation de Fidéli permet de disposer de données quasi-exhaustives (la sous-couverture a été estimée à environ 2 % concernant les personnes en situation irrégulière et des cas très particuliers). Un intérêt majeur de cette base est de disposer de nombreuses informations de contact et d'une multitude

d'informations auxiliaires qui ont pu être mobilisées lors des traitements statistiques post-collecte. Une limite de cette base est qu'elle ne donne pas d'information sur la catégorie socio-professionnelle ou le diplôme des individus.

Par ailleurs, la table utilisée pour le tirage ne couvre pas complètement le champ de diffusion de l'enquête :

- il manque les individus qui vivent en France métropolitaine en 2021 et qui n'y vivaient pas au 1er janvier 2019 ;
- cette table n'intègre pas non plus les individus qui vivent en logement ordinaire en 2021 et qui vivaient dans une communauté d'après le dispositif Fidéli au 1er janvier 2019.

Ces défauts de couverture seront corrigés par calage sur marges.

Outre cette base principale utilisée pour le tirage de l'échantillon de phase 1 de l'enquête Genese, plusieurs autres bases de sondage - correspondant toutes à des sous-échantillons de l'échantillon de phase 1 - ont été mobilisées pour des tirages d'échantillons au cours de la collecte. Pour le dispositif Genese, ont ainsi été constituées :

- une base de non répondants éligibles à la relance téléphonique pour la phase 1 ;
- une base de répondants à la phase 1 pour le tirage des individus à enquêter en phase 2.

3.2 Échantillonnage de la phase 1 (enquête filtre)

3.2.1 Objectifs et architecture générale

Comme évoqué en 2.2, l'enquête Genese vise à assurer une certaine représentativité départementale ainsi qu'une représentativité nationale des quartiers prioritaires de la politique de la ville (QPV) et des quartiers de reconquête républicaine (QRR). Une autre dimension importante, compte tenu de la thématique de l'enquête est celle du genre. Ainsi, pour répondre à ces objectifs tout en intégrant les contraintes inhérentes au recours à une base de sondage dont l'utilisation n'est pas restreinte à cette seule enquête¹⁰, l'architecture générale du plan de sondage est, en accord avec la division Sondages de l'Insee, un tirage stratifié selon le département, le sexe et la résidence en QRR.

3.2.2 Taille d'échantillon

170 000 individus ont été tirés dans la base de sondage issue du millésime 2019 de Fidéli constituée d'environ 46 millions d'individus. Cette taille d'échantillon, contrainte notamment par des considérations budgétaires, correspond à un objectif de collecter environ 85 000 questionnaires avec un taux de

10. Pour assurer la disjonction avec les enquêtes futures tout en conservant une base de sondage équilibrée, il convient de limiter le nombre de strates et éviter les trop fortes surreprésentations afin de ne pas nécessiter le marquage d'un nombre d'individus trop important dans Nautile.

réponse anticipé global de 50 %. Cette taille cible de 85 000 questionnaires répond à la fois aux besoins en termes de diffusion sur les indicateurs clés de cette enquête de phase 1 mais aussi aux objectifs liés à l'enquête de seconde phase, de recueil détaillé d'informations sur les victimations subies. En particulier, l'objectif est de disposer, à l'issue de l'enquête de phase 1, d'un vivier suffisant de victimes récentes de violences physiques ou sexuelles à interroger pour la phase 2.

3.2.3 Stratification explicite et surreprésentation

Cette taille d'échantillon permet ainsi d'envisager la diffusion de résultats à un niveau géographique assez fin. Une stratification départementale a ainsi été mise en œuvre avec une surreprésentation des départements peu peuplés : le nombre minimal anticipé de répondants par département est fixé à 500 répondants. Ce nombre minimal de répondants par département permet d'estimer les paramètres d'intérêt avec une précision « acceptable » : approximativement (sous l'hypothèse simplificatrice de sondage aléatoire simple) 2,7 % lorsque le facteur de prévalence est de 10 % dans la population, 3,6 % pour une prévalence de 20 % et 4,5 % pour une prévalence de 50 %. Cette précision est suffisante pour réaliser des typologies de départements sur les différents paramètres d'intérêt en distinguant, par exemple, ceux qui se situent nettement au-dessus de la valeur moyenne nationale et ceux qui se situent nettement en dessous. Aussi, fixer un seuil minimal à 500 répondants limite le nombre de départements concernés par cette borne : une quarantaine de départements ont été surreprésentés.

Un autre objectif de diffusion est d'assurer une représentativité sur les QPV et les QRR avec un nombre minimal de 3 000 répondants pour chacun de ces zonages. L'idée est de disposer d'un nombre au moins équivalent de répondants pour ces zonages que ceux vivant en QPV dans les précédentes éditions des enquêtes CVS. Compte tenu du taux de sondage et de la population en QPV (3,2 millions d'individus dans Fidéli pour le périmètre de Genese), même avec un taux de réponse anticipé assez pessimiste, le seuil de 3 000 répondants en QPV est atteint : il n'a donc pas été nécessaire d'inclure cette dimension dans la stratification. En revanche, une surreprésentation des individus en QRR (860 000 individus dans Fidéli pour le périmètre de Genese) a dû être appliquée pour répondre aux objectifs de diffusion.

Enfin, la thématique de l'enquête impose assez naturellement de surreprésenter les femmes dans l'échantillon pour cibler au mieux les potentielles victimes à interroger en phase 2. L'objectif était de disposer au final d'environ un répondant de sexe masculin pour deux répondants de sexe féminin.

3.2.4 Allocations cibles et calcul d’allocations par strates

Les allocations par strates sont calculées pour respecter les objectifs cibles suivants : 1 000 répondants hommes en QRR, 2 000 répondants femmes en QRR, 27 000 répondants hommes hors QRR, 55 000 répondants femmes hors QRR avec au minimum 500 répondants par département.

Enfin, comme demandé par la division Sondages de l’Insee et afin de limiter les volumes en jeu pour rééquilibrer la base de sondage, seule la surreprésentation départementale est appliquée dans certains départements peu peuplés ou sans QRR, où le taux de sondage des femmes est déjà élevé compte tenu de cette surreprésentation.

L’hypothèse retenue pour estimer le nombre de répondants est un taux de réponse anticipé global de 50 % : un peu plus de 50 % pour les personnes résidant hors QRR, 35 % pour les personnes vivant en QRR. Comme il s’agit d’une première édition de l’enquête, nous ne disposons pas de références suffisamment robustes pour décliner plus finement ces taux de réponse. En effet, les enquêtes sur lesquelles on aurait pu s’appuyer se distinguent de Genese sur des aspects importants du point de vue de la participation à l’enquête : le protocole des enquêtes CVS est radicalement différent (face-à-face, tirage de logements dans Octopusse) et si l’enquête Epicov¹¹ s’appuie sur la même base de sondage, le mode de collecte n’est pas vraiment similaire (pas de recours au questionnaire papier) et de plus, sa thématique est assez éloignée de celle de Genese. Par ailleurs, les effectifs du pilote sont beaucoup trop limités pour fournir des estimations robustes de la participation à un niveau fin. Ainsi, les taux de réponse anticipés en QRR et hors QRR sur lesquels repose le calcul des allocations sont homogènes par sexe et département¹².

Les allocations par strates sont calculées, à partir de taux de réponse anticipés : on initialise à partir d’allocations proportionnelles, on intègre les facteurs de surreprésentations puis les différentes contraintes et on itère de manière à être suffisamment proche des objectifs cibles globaux tout en respectant les contraintes. Les allocations globales ainsi que les proportions comparées dans la base de sondage et dans l’échantillon sont présentées dans le tableau 1 suivant.

11. Enquête Epicov : Épidémiologie et Conditions de vie sous le Covid-19, réalisée par l’Inserm et la DREES.

12. Des simulations de calculs d’allocations qui reposeraient sur des taux déclinés par département (estimés à partir de ceux d’Epicov) indiquent que les principales différences porteraient sur les départements peu peuplés et donc surreprésentés, dont le taux de réponse anticipé pourrait être significativement moins élevé que la moyenne nationale : cela concernerait en particulier la Corse-du-Sud et la Haute-Corse.

TABLEAU 1 – Distribution de la base de sondage et allocation de tirage par strate

	Effectifs de la base de sondage	Allocations du tirage de phase 1	Base de sondage (%)	Échantillon (%)
Sexe				
Hommes	22 852 984	56 012	50,1	32,9
Femmes	22 730 528	113 988	49,9	67,1
QRR				
Personnes vivant hors QRR	44 724 277	161 426	98,1	95,0
Personnes vivant en QRR	859 235	8 574	1,9	5,0
Sexe et QRR				
Hommes vivant hors QRR	22 414 333	53 154	49,2	31,3
Hommes vivant en QRR	438 651	2 858	1,0	1,7
Femmes vivant hors QRR	22 309 944	5 716	48,9	63,7
Femmes vivant en QRR	420 584	108 272	0,9	3,4

3.2.5 Stratification implicite

Le tirage effectué est un tirage systématique stratifié. Ainsi, pour contrôler au mieux la représentativité selon différents facteurs, au-delà des variables de stratification, les strates de tirage sont triées, dans l'ordre, par aire urbaine, tranche d'unité urbaine, identifiant de la commune, identifiant QPV, décile de niveau de vie du logement, identifiant du logement et identifiant de l'individu.

3.2.6 Échantillonnage de la relance téléphonique

40 jours après le début de la collecte, une sélection est effectuée parmi les individus éligibles à la relance téléphonique : il s'agit des individus n'ayant pas répondu au questionnaire en ligne et pour lesquels on dispose d'une coordonnée téléphonique pour les contacter. Il est ainsi prévu de sélectionner 25 000 individus parmi ces éligibles, correspondant à 7 500 entretiens téléphoniques budgétés¹³. Une stratification implicite a été mise en œuvre, le tirage systématique reposant sur un tri selon, dans l'ordre : le département, l'aire urbaine, la tranche d'unité urbaine, l'identifiant de la commune, l'identifiant QPV, l'identifiant QRR, le décile de niveau de vie du logement, l'identifiant du logement et l'identifiant de l'individu.

13. Estimation réalisée à partir d'un taux de participation anticipé, révisé à partir des résultats du pilote.

3.3 Échantillonnage de la phase 2 (enquête filtrée)

3.3.1 Objectifs et architecture générale

L'objectif de la phase 2 (enquête filtrée) de l'enquête est de recueillir de l'information détaillée sur les victimations subies par les personnes, en particulier sur les atteintes sexistes et sexuelles. Comme pour l'enquête filtre, le périmètre de l'enquête de phase 2 est celui des individus ayant 18 à 74 ans au 1er janvier 2021, vivant en France métropolitaine et résidant dans un logement ordinaire. Ainsi, la base de sondage de l'échantillon de cette enquête est l'ensemble des répondants à l'enquête filtre dont les réponses sont mobilisées pour cibler au mieux certaines formes de victimation et les surreprésenter dans l'échantillon. La représentativité de cette enquête de phase 2 se limite à la dimension nationale (métropole).

3.3.2 Taille d'échantillon

La taille de l'échantillon de l'enquête de phase 2 est de 15 000 individus. Cette taille d'échantillon répond au besoin de disposer d'un nombre global de répondants à peu près équivalent à celui des enquêtes CVS tout en surreprésentant certaines formes de victimation ciblées par la thématique de l'enquête Genese.

3.3.3 Stratification explicite et surreprésentation

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 leur vie ;
- cible 3 : répondants n'appartenant ni au groupe 1 ni au groupe 2.

Les individus du groupe cible 1 ont été systématiquement sélectionnés¹⁴ et ceux du groupe cible 2 ont été surreprésentés. En raison de la sélection systématique des individus du groupe cible 1, les allocations globales pour ces différents groupes ont été fonction des remontées de l'enquête de phase 1.

Les travaux sur les victimations recensées dans les différents millésimes des enquêtes CVS ainsi que l'expérience du pilote permettaient d'envisager des allocations prévisionnelles : 2 000 individus du groupe cible 1, 10 000 individus du groupe cible 2 et 3 000 individus du groupe cible 3. Au sein du groupe cible 2, certaines sélections systématiques ont été envisagées comme la sélection systématique,

14. Une limite est néanmoins fixée pour cette sélection systématique. Si le nombre d'individus du groupe cible 1 avait dépassé 5 000 (ce qui était peu probable d'après les résultats du pilote et les enquêtes historiques CVS), ils n'auraient pas été systématiquement sélectionnés mais auraient fait l'objet d'une forte surreprésentation.

opérée lors du pilote, des victimes de conjoints, des victimes de viols et des victimes récentes (2021 ou 2020). Une stratification implicite a été mise en œuvre, le tirage systématique reposant sur un tri selon, dans l'ordre : le sexe, le département, l'aire urbaine, la tranche d'unité urbaine, l'identifiant de la commune, l'identifiant QPV, l'identifiant QRR, le décile de niveau de vie du logement, l'identifiant du logement et l'identifiant de l'individu.

4 Déroulement de la collecte

4.1 Enquête pilote

L'enquête Genese a été précédée d'une enquête pilote afin de tester le dispositif mis en place et les questionnaires. Elle a été réalisée entre septembre et octobre 2020 sur 8 semaines. L'enquête pilote avait pour but d'évaluer :

- l'approche des répondants et leur attitude face au thème de l'étude ;
- la compréhension des questions et leur formulation ;
- l'apport des relances téléphoniques, postales, SMS ou par courrier électronique ;
- les flux programmés pour assurer les relances lissées ;
- la durée des différents questionnaires selon le mode utilisé ;
- la qualité des données recueillies.

Un échantillon de 2 000 individus âgés de 18 à 74 ans a été interrogé lors de l'enquête pilote, elle aussi articulée en deux phases. Au total, 786 questionnaires complets de phase 1 ont été collectés, soit un taux de participation global de 39 %. La phase 2 a permis de collecter 94 questionnaires complets, soit un taux de collecte de 63 %.

L'objectif de recueillir suffisamment de personnes du groupe cible 1 a été atteint, ce qui a permis notamment de bien représenter les groupes cible 1 et cible 2 dans l'échantillon de phase 2. Le pilote a permis d'évaluer la bonne compréhension des questionnaires via les remarques des enquêtés, l'analyse des champs ouverts renseignés sur les questions proposant des listes, la non-réponse et les retours des enquêteurs CATI. Il a également permis de réaliser divers ajustements du protocole pour l'enquête en population générale, dont les principaux sont :

- l'augmentation du nombre et du rythme de relances mail pour augmenter la réponse par internet ;
- le décalage de quelques jours entre les envois de mails et les envois postaux pour maximiser l'effet incitatif des mails grâce au caractère officiel des lettres qui ont préalablement été envoyées ;
- l'adaptation du calendrier d'envoi du questionnaire papier selon les coordonnées disponibles pour favoriser la réponse par internet ou téléphone (lors du pilote le questionnaire papier avait

- été envoyé à tous les non répondants avant la relance CATI) ;
- enfin, la hausse de la taille d'échantillon de la relance CATI (25 000 au lieu de 15 000 prévu) pour compenser la faible participation sur ce mode (50 % attendu *versus* 14 % observé lors du pilote) afin d'utiliser le budget alloué à ce mode.

4.2 Enquête de phase 1 (enquête filtre)

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 RNVP) et enrichissement des coordonnées téléphoniques. 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.

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 groupe P1SS1 des personnes sans mail ni téléphone et celui P1SS2 des personnes avec mail ou téléphone.

L'ensemble des individus sélectionnés sont tous destinataires d'une lettre-avis nominative, qui spécifie clairement l'interdit du proxy (une autre personne que la cible ne peut pas répondre à sa place) 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, son prénom et son année de naissance pour accéder au questionnaire. Les courriers et mails informent l'enquêté de la mise à disposition d'une hotline disponible tout au long de l'enquête par téléphone sur un numéro vert gratuit ou par mail pour répondre à leurs questions, les informer sur l'enquête et prendre rendez-vous pour les enquêtés concernés par la relance téléphonique. La hotline était composée de huit personnes préalablement formées au cours d'une session d'une demi-journée assurée conjointement par Ipsos et le SSMSI. L'adresse du site dédié à l'enquête qui sert de plateforme de connexion pour la réponse au questionnaire en ligne est également communiquée dans tous les courriers et mails envoyés. Véritable outil d'information et d'animation de la collecte, le site internet confère une dimension institutionnelle à l'enquête et contribue à légitimer l'opération pour les enquêtés.

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.

La relance téléphonique a été proposée à un nombre fixé à l'avance de 25 000 non répondants à J40 afin d'améliorer la participation de populations moins connectées tout en restant dans l'enveloppe

budgétaire de la subvention européenne. Pour interroger ces personnes, 68 enquêteurs ont été formés à distance avec le réseau « enquêteurs à domicile » d'Ipsos. Quatre sessions de formations d'une journée et demie ont été organisées. Une partie théorique et une partie pratique (élaboration de scénarios pour les exercices) étaient prévues. Également, une psychologue a réalisé une intervention afin d'aider les enquêteurs à garder de la distance et adopter le bon comportement face à des victimes de violences et à des personnes qui leur demanderaient de l'aide. 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. Concernant les mails-avis, 6 % n'ont pas pu être distribués. Sur les 25 000 individus sélectionnés pour la relance téléphonique, 51 % n'ont pas pu être joints par téléphone. Parmi eux, 20 % ne présentaient que des faux numéros. Au terme de la collecte, on estime à environ 3 % la part d'individus qui n'ont pu être contactés par aucun des moyens mis en place.

À l'issue des 11 semaines de terrain, sur les 169 060 individus de l'échantillon, près de 109 000 individus (64 %) ont répondu à la première phase. Environ 2 % des enquêtés ont commencé le questionnaire sans le compléter entièrement et 33 % sont non répondants. Près de 86 000 individus ont répondu par internet, ce qui représente 79 % des répondants. Les autres modes sont minoritaires : 17 500 ont répondu par papier (16 %) et 5 500 (5 %) par téléphone. La participation globale varie de 43 % parmi le groupe des individus sans mail ni téléphone (P1SS1) à 68 % parmi les individus éligibles à l'envoi de mail (P1SS2). Dans ce dernier groupe, la participation internet atteint 85 % des répondants. Dans le P1SS1, l'envoi du questionnaire papier contribue de moitié à la participation totale observée dans ce groupe.

L'évaluation du protocole de l'enquête Genese et l'effet des relances a fait l'objet d'une communication aux Journées de méthodologie statistiques de l'Insee en 2022¹⁵. Des analyses multivariées conduites ont montré que l'envoi de mails a eu 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 a contribué 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 multimodes concurrentiels différés qui permettent de

15. Guedj, H., Quiviger, A., Zilloniz, S., & Razafindranovona, T. (2022). Un protocole innovant, multimode en deux phases, pour mesurer la victimation : résultats et principaux enseignements de la collecte de l'enquête genre et securite (Genese). XIV^{ème} Journées de méthodologie statistique de l'Insee http://www.jms-insee.fr/2022/S03_1_ACTE_GUEDJ_JMS2022.pdf

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, il a également été montré 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 socio-démographiques, en particulier les marqueurs de situations sociales défavorisées (habitants en QPV, déciles de niveau de vie plus faibles, habitants en logement social).

4.3 Enquête de phase 2 (enquête filtrée)

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. 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.

L'efficacité du protocole tient alors 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 déménagés ; 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).

La phase 2 de l'enquête a concerné 15 000 répondants à la phase 1 : environ 2 000 individus du groupe cible 1, 10 000 individus du groupe cible 2 et 3 000 individus du groupe cible 3. Les personnes cible 1 ont été systématiquement sélectionnées, les personnes cible 2 ont été surreprésentées (elles représentent 67 % des sélectionnées pour la phase 2 alors que leur part 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). La phase 2 a permis de collecter plus de 10 000 questionnaires complets : près de 8 000 par internet et environ 2 000 par téléphone, soit un taux de collecte de 68 %. L'objectif de participation en phase 2 initialement fixé a été dépassé grâce à une bonne participation par internet (77 % des répondants) qui compense une participation par téléphone

en-deçà des attentes (23 % des répondants). La part de questionnaires incomplets représente 4 % de l'ensemble des personnes interrogées.

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.

5 Enquête filtre en face-à-face auprès des non répondants à la phase 1 de l'enquête

L'enquête a porté sur 3 114 personnes sélectionnées pour être interrogées à leur domicile avec un enquêteur sur le questionnaire de l'enquête filtre (comme dans CVS, la partie concernant les atteintes à la personne était auto-administrée avec une passation sous casque-audio). Un réseau de 40 enquêteurs a été déployé pour se rendre au domicile de tous les non répondants à l'enquête filtre de Gironde et d'une sélection de non répondants d'Île de France. Au préalable, 3 sessions d'une journée, ont été organisées pour former ces enquêteurs.

Les enquêtés étaient destinataires d'une lettre-avis annonçant le passage de l'enquêteur. La prise de rendez-vous était possible sur le site internet de l'enquête. Un SMS de relance ou un mail de relance était envoyé une semaine avant la fin de la collecte selon les coordonnées disponibles pour les enquêtés.

Sur les 8 semaines de terrain, les enquêteurs ont réalisé 1 049 interviews complètes, soit un taux de participation de 34 %. En moyenne, 2,6 contacts ont été nécessaires pour réaliser l'entretien. Les principaux motifs de la non-participation ont été : l'impossibilité de réaliser l'entretien pendant toute la durée du terrain (personne absente du domicile pendant la période de l'enquête, handicap, etc. ; 17 %) et le refus de participer (15 %).

En moyenne, la durée d'un entretien a été d'environ 27 minutes.

Le taux de participation varie selon le territoire : de 16 % dans le Val-D'Oise à 44 % dans les Yvelines. En Gironde, où l'ensemble des non répondants étaient contactés, le taux de participation a été de 38 %.

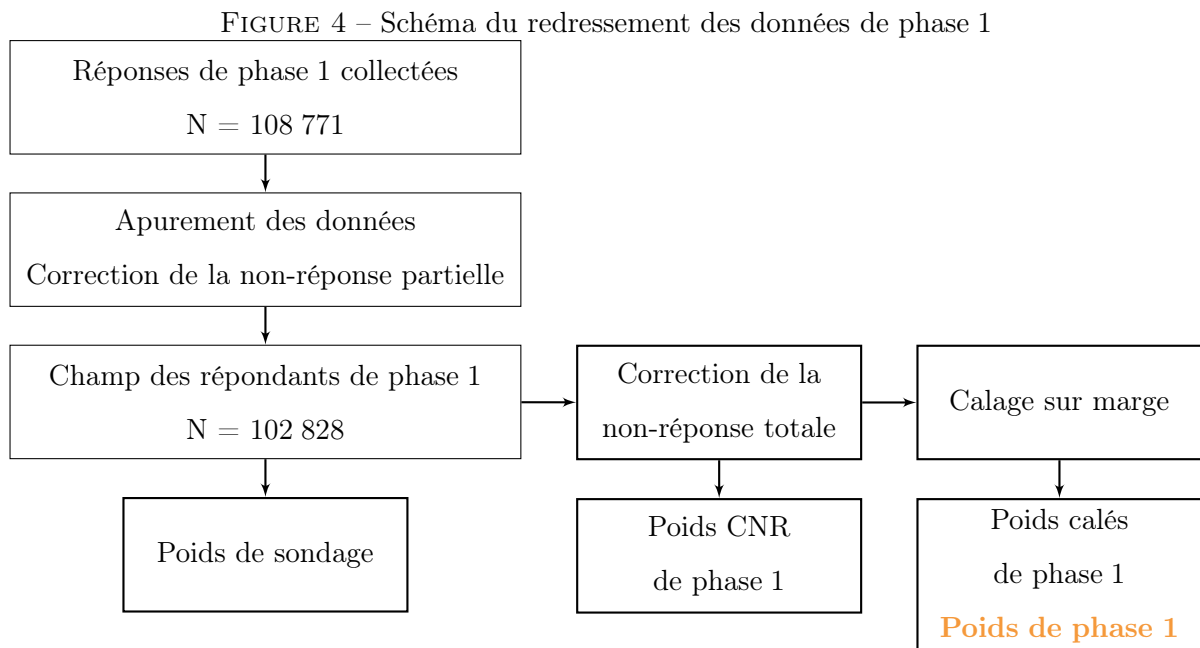
Cette enquête en face-à-face auprès des non répondants a permis de collecter les réponses de certains profils qui avaient un peu moins bien répondu dans le cadre du dispositif principal de l'enquête multimode internet, téléphone, papier : les habitants des quartiers de reconquête républicaine (QRR) et de quartiers prioritaires de la politique de la ville (QPV) et les jeunes de 18 à 29 ans. Ainsi, les habitants d'un QRR représentent 9 % des répondants contre 5 % de l'échantillon et 4 % des répondants internet, téléphone ou papier. Les habitants d'un QPV représentent 17 % des répondants contre 8 % de l'échantillon et 6 % des répondants internet, téléphone ou papier. Enfin, les 18-29 ans représentent

20 % des répondants contre 19 % de l'échantillon et 16 % des répondants internet, téléphone, papier. On observe par ailleurs, que la répartition par cible est proche de celle obtenue dans le dispositif principal par CAWI ou CATI (4 % de cible 1 contre 2 % en CAWI, 3 % CATI et 1 % en papier, 43 % de cible 2 contre respectivement 41 % en CAWI, 43 % en CATI et, enfin, 20 % en papier et 53 % de cible 3 contre 57 % en CAWI, 55 % en CATI et 79 % en papier).

6 Traitements post-collecte de l'enquête de phase 1 (enquête filtre)

Une fois la collecte terminée, 108 771 réponses sont obtenues tous modes confondus. Un processus d'apurement des données a été nécessaire afin d'identifier les répondants hors champ, et d'exclure les réponses trop partielles pour être exploitables (en particulier pour les répondants papier).

On présente ensuite les trois étapes de redressement des données de phase 1 réalisées : imputation de la non-réponse partielle, imputation de non-réponse totale et calage sur marge suivant l'ordre présenté dans la figure 4 et qui permettent d'obtenir les poids de phase 1.



6.1 Identification et traitement des hors champ

Afin de nous assurer que l'enquêté est bien la personne sélectionnée, une vérification de l'identité est effectuée en début de questionnaire en CAWI et en CATI. Le nom et le prénom de l'individu, lorsqu'ils sont présents dans la base Fidéli 2019, sont affichés et l'enquêté doit valider son identité. Sa date de naissance est redemandée et est comparée à celle de la base de sondage. Dans le cas minoritaire où la personne qui s'est connectée n'est pas la personne tirée, il lui est demandé de bien vouloir transmettre le lien du questionnaire à remplir à l'individu concerné. Cette vérification est faite *a posteriori* pour

les répondants papier qui sont classés en non répondants si les informations ne coïncident pas.

Une fois cette vérification faite, nous nous assurons que l'unité est bien dans le champ de l'enquête. En effet, les individus ne vivant pas en France métropolitaine dans un logement ordinaire (communauté, camping, résidence hôtelière, bateau etc.) ou ne pouvant pas être interrogés pour raisons de santé ou de handicap (y compris les personnes décédées) sont considérés comme « hors champ de l'enquête », c'est-à-dire qu'ils ne font pas partie du champ d'intérêt de l'enquête. La base de tirage Fidéli ne permet pas d'écarter l'intégralité de ces individus, notamment en cas de changement de situation depuis la dernière actualisation de la base en 2019. Environ 1 500 enquêtés sont exclus du champ de l'enquête lors de cet apurement.

6.2 Contrôles de cohérence et apurement

Les données individuelles collectées sont ensuite vérifiées afin de nous assurer que leur qualité et de l'améliorer le cas échéant. Plusieurs éléments sont traités :

- les cas de doublons : certains enquêtés ont rempli à la fois le questionnaire par internet et par papier. La version internet est conservée si elle est complète et la version papier est conservée sinon ;
- la taille d'unité urbaine (TUU) est actualisée : si l'enquêté a déclaré vivre à la même adresse que celle du fichier Fidéli 2019, il conserve cette adresse, et donc la même TUU. Si l'enquêté est un répondant internet ou téléphone et qu'il a déclaré une nouvelle adresse, sa TUU est corrigée en fonction de cette nouvelle adresse. Dans les autres cas, l'adresse et la TUU utilisées sont celles issues de de Fidéli 2020 quand elles sont renseignées et de Fidéli 2019 sinon ;
- le type de ménage du répondant est reconstitué à partir de deux questions de l'enquête socle : le nombre de personnes qui vivent dans le logement et sa composition.

Les variables identifiées comme variables de calage potentielles ont également conduit à la suppression de certaines observations. Les variables de calage doivent en effet être connues pour l'ensemble des unités répondantes, il a fallu écarter les non répondants à ces questions.

- les individus pour lesquels le sexe n'est pas renseigné dans la base Fidéli 2019 et qui n'ont pas répondu à cette question du questionnaire sont écartés de la base finale des répondants ;
- les individus dont le diplôme n'est pas renseigné sont écartés ;
- les individus dont la situation face à l'emploi n'est pas renseignée sont écartés¹⁶.

Enfin, pour aboutir au champ final des répondants de phase 1, un minimum de complétude du questionnaire a été imposé.

16. Du fait de la formulation de la question dans le questionnaire Genese, cette variable n'a finalement pas pu être utilisée en tant que variable de calage. Les modalités de réponse ne permettaient en effet pas de reconstituer les catégories d'emploi au sens du BIT utilisées par l'Insee.

Le questionnaire socle comporte cinq tableaux de recensement des atteintes :

1. les atteintes visant le logement, composées de quatre atteintes différentes ;
2. les atteintes visant les véhicules, composées de cinq atteintes différentes ;
3. les atteintes aux biens, composées de cinq atteintes différentes ;
4. les atteintes graves à la personne, composées de quatre atteintes différentes ;
5. les autres atteintes à la personne, composées de quatre atteintes différentes.

On ne considère comme répondants que les individus qui ont complété au moins un de ces blocs d'atteintes ¹⁷.

Pour les répondants papiers, on considère un bloc d'atteintes complet lorsqu'au moins une réponse est donnée pour chacune des atteintes de ce bloc. Pour les répondants internet et téléphone, on considère un bloc d'atteintes complet lorsqu'au moins une réponse, autre que « ne sait pas » / « ne veut pas », est donnée pour chacune des atteintes de ce bloc.

Les 102 828 individus qui remplissent l'ensemble de ces conditions sont alors considérés comme les répondants à l'enquête socle.

6.3 Correction de la non-réponse partielle

Dans le but de disposer d'information sur l'ensemble des répondants dans le cadre de la repondération des données, il a été décidé de corriger la non-réponse partielle de la variable de type de ménage plutôt que de supprimer des répondants sur ce seul motif.

La composition du ménage du répondant, très corrélée avec le sujet de l'enquête Genese, sera en effet utilisée à plusieurs reprises dans le redressement aval des données. Cette variable n'est pas issue directement du questionnaire de phase 1 mais est reconstituée à partir de deux autres variables du questionnaire : le nombre de personnes dans le logement et sa composition. La variable ainsi agrégée prend les 5 modalités suivantes :

- personne seule ;
- couple sans enfant ;
- couple avec enfant ;
- famille monoparentale ;
- autre.

17. On considère le bloc d'atteintes visant les véhicules complet si l'individu déclare ne pas posséder de véhicule ou n'en avoir jamais possédé. Dans ces deux cas de figure, le détail des atteintes visant les véhicules n'est alors pas demandé à l'enquêté.

Du fait de la non-réponse partielle, il n'a pas été possible de reconstruire cette variable pour 4,7 % des répondants à l'enquête. Afin de la compléter, une imputation par hotdeck aléatoire a été mise en œuvre.

Seule la variable de type de ménage a fait l'objet d'une imputation de la non-réponse. Il a été décidé pour des raisons de lourdeur du processus de ne pas imputer la non-réponse aux questions de victimation et de perception.

6.3.1 Principe général

L'imputation par hotdeck consiste à choisir aléatoirement un répondant, le donneur, dont la réponse est utilisée pour imputer la valeur manquante d'un non répondant, le receveur. Afin d'améliorer la qualité de l'imputation, celle-ci se fait au sein d'une classe d'imputation. Les classes d'imputation sont constituées par le croisement des modalités d'un jeu de variables, corrélées avec la variable à imputer mais pas avec le fait d'y répondre.

En pratique, on parcourt une classe d'imputation jusqu'à rencontrer une unité dont la valeur est manquante. On attribue la valeur d'un donneur tiré aléatoirement au sein de la classe à ce receveur puis on reprend le parcours de la classe jusqu'à ce que tous les receveurs aient une valeur imputée.

Il faut veiller à ce que les tailles des classes d'imputation soient suffisamment importantes afin de ne pas attribuer les valeurs d'un trop petit nombre de donneurs à un trop grand nombre de receveurs, ce qui conduirait à la déformation de la distribution.

6.3.2 Application et résultat

L'imputation, via hotdeck aléatoire, du type de ménage des répondants est réalisée au sein de classes d'imputation, elles-mêmes constituées par le croisement des variables suivantes :

- le type de ménage fiscal issu de la base Fidéli ;
- le sexe ;
- la tranche d'âge en 5 positions ;
- le quintile de niveau de vie du ménage.

Les déciles de niveau de vie, qui conduisaient à des classes d'imputation plus fines mais à effectifs trop faibles, ont été regroupés en quintiles et permettent alors l'imputation de l'ensemble des valeurs manquantes. La distribution de la variable après imputation des valeurs manquantes, présentée en tableau 2, ne s'écarte pas trop de celle avant traitement.

TABLEAU 2 – Distribution du type de ménage déclaré avant et après imputation des valeurs manquantes

Type de ménage	Avant imputation	Après imputation
1- Personne seule	16,5 %	17,7 %
2- Couple sans enfant	30,6 %	30,2 %
3- Couple avec enfant	38,4 %	37,5 %
4- Famille monoparentale	9,6 %	9,6 %
5- Autre	4,9 %	4,9 %

6.4 Traitement de la non-réponse totale

Avec un total de 102 828 questionnaires exploitables après apurement, le taux de réponse à l'enquête socle est de 60,8 %. La non-réponse introduit un biais dans les estimations, qui tient au fait que les caractéristiques des non répondants peuvent différer de celles des répondants. Afin de réduire le biais induit par la non-réponse et d'améliorer la précision des estimateurs produits, la non-réponse a été corrigée par repondération.

6.4.1 Principe général

Le principe de la correction de la non-réponse par repondération est d'augmenter les poids initiaux des répondants pour compenser le biais introduit par les non répondants. Dans cette méthode, on suppose qu'il est possible de découper l'échantillon en parties disjointes, appelées groupes de réponse homogène (GRH), telles qu'à l'intérieur de ces groupes, toutes les unités de l'échantillon aient des comportements de réponse indépendants et aient la même probabilité de réponse.

Il existe plusieurs méthodes de repondération des unités répondantes par groupes de réponse homogène mais seule la méthode des scores, mobilisée pour les redressements de l'enquête Genese, sera présentée ici¹⁸.

La méthode des scores permet de construire des groupes de repondération en deux étapes :

1. Estimation de la probabilité de réponse individuelle par régression logistique à partir de variables auxiliaires disponibles pour toutes les unités dans l'échantillon, qu'elles soient répondantes ou non répondantes. Les variables auxiliaires choisies doivent expliquer le choix d'être répondant à l'enquête pour être pertinentes. La probabilité estimée de réponse ainsi obtenue servira de critère d'homogénéité des GRH ;

18. Les explications qui suivent reprennent très largement l'article suivant : Deroyon, T. (2017). La correction de la non-réponse par repondération. *Document de travail, Insee*. <https://www.insee.fr/fr/statistiques/fichier/2838097/8-correction-de-la-non-reponse-par-reponderation.pdf>

2. Formation des GRH. Là encore, plusieurs méthodes de constitution des groupes coexistent, mais seule la méthode d’Haziza et Beaumont mobilisée dans le cadre de l’enquête Genese est présentée.

Les GRH sont construits en appliquant un algorithme des centres mobiles, la distance entre unités étant définie comme le carré de la différence entre leurs probabilités de réponse estimées. Le nombre de GRH est déterminé en l’augmentant progressivement et en s’arrêtant au nombre le plus faible de GRH rendant compte d’une partie suffisante de la dispersion des probabilités de réponse estimées. Plus précisément :

- on construit d’abord deux GRH ;
- on estime ensuite la régression linéaire des probabilités de répondre estimées sur les indicatrices d’appartenance aux GRH ;
- si le coefficient de détermination du modèle est supérieur à un seuil fixé *a priori*, par exemple de 95 % ou 99 %, alors le modèle rend compte de 95 % ou 99 % de la dispersion des probabilités de réponse estimées. On s’arrête donc à deux GRH. À l’inverse, si le R^2 du modèle est inférieur au seuil, on recommence le processus avec trois GRH ;
- on augmente le nombre de GRH jusqu’à obtenir des GRH rendant compte d’une part de la dispersion des probabilités de réponse estimées supérieure au seuil fixé *a priori*.

Une fois le découpage en GRH obtenus, on applique un coefficient de correction de la non-réponse aux poids individuels des répondants à l’enquête. Ce coefficient s’obtient en divisant la somme des poids de sondage du groupe par la somme des poids de sondage des unités répondantes du groupe. On multiplie alors ce coefficient correctif aux poids de sondage des répondants du GRH pour obtenir les poids corrigés de la non-réponse (poids CNR).

6.4.2 Modélisation de la non-réponse et constitution des GRH

Les variables à la fois corrélées avec la participation à l’enquête ainsi qu’avec certaines variables d’intérêt sont sélectionnées dans la modélisation de la non-réponse via régression logistique. Le modèle retenu, avec la meilleure aire sous la courbe ROC et le meilleur pseudo R^2 de McFadden, mobilise les variables suivantes :

- le croisement du sexe et de la tranche d’âge ;
- le type de ménage fiscal issu de Fidéli ;
- le décile de niveau de vie du ménage ;
- une indicatrice d’allocation chômage ;
- une indicatrice de résidence en logement social ;
- une indicatrice de résidence en appartement ;

- la source d’origine du numéro téléphone ;
- une indicatrice disponibilité d’une adresse e-mail ;
- la taille d’unité urbaine de résidence ;
- le département de résidence ;
- une indicatrice de résidence en quartiers prioritaires de la politique de la ville ;
- une indicatrice de résidence en quartier de reconquête républicaine.

À partir des probabilités de réponse estimées ainsi obtenues, on construit les GRH selon la méthode d’Haziza et Beaumont, en choisissant un coefficient de détermination du modèle à 99 %.

6.4.3 Résultats

On aboutit à la constitution de 13 groupes de réponse homogène dont le plus petit comporte 3 685 unités (tableau 3).

TABLEAU 3 – Groupes de réponse homogène par taille

Groupe	GRH 1	GRH 7	GRH 11	GRH 9	GRH 3	GRH 2	GRH 6	GRH 4	GRH 8	GRH 13	GRH 10	GRH 12	GRH 5
Taille	3 685	5 306	7 267	9 473	10 681	12 145	13 196	13 608	15 522	17 201	19 528	20 385	21 063

La distribution des poids corrigés de la non-réponse est présentée dans le tableau 4 ci-dessous.

TABLEAU 4 – Distribution des poids corrigés de la non-réponse de phase 1

Min	10 %	25 %	50 %	75 %	90 %	Max
117,1	199,5	292	350,4	606,4	770,5	2 762,4

Enfin, le rapport du poids calé maximal sur le poids calé minimal est de 23,6.

6.5 Calage et calcul des pondérations

Une fois la non-réponse corrigée, un calage sur marge a été effectué à partir du package Icarus¹⁹ créé au plus proche de la macro SAS CALMAR et disponible sur CRAN.

6.5.1 Principe général

Le calage sur marge permet de redresser un échantillon, par repondération des individus, en utilisant une information auxiliaire disponible sur un certain nombre de variables, appelées variables de calage. Les pondérations produites par ces méthodes assurent le calage de l’échantillon sur des totaux

19. Rebecq, Antoine (2017). Icarus : an R package for calibration in survey sampling. R package version 0.3.0. <https://github.com/haroine/icarus>

de variables quantitatives connus sur la population, et sur des effectifs de modalités de variables catégorielles connus sur la population. Elles permettent également d'améliorer la précision des estimations des totaux des variables d'intérêt corrélées aux variables de calage. Le calage sur marges permet également de présenter des estimations cohérentes entre les différentes sources statistiques.

Par ailleurs, ces variables de calage ne doivent pas comporter de valeurs manquantes, les totaux ou effectifs de ces variables doivent donc être connus sur l'ensemble de la population étudiée.

6.5.2 Application et choix des marges

La méthode logit a été retenue pour ce calage. Elle permet d'éviter les poids négatifs ainsi que de définir des bornes inférieure et supérieure au rapport $\frac{\text{poids calés}}{\text{poids initiaux}}$ afin d'éviter une déformation des poids extrêmes. Les poids initiaux qu'on utilise ici sont les poids corrigés de la non-réponse obtenus à l'étape précédente.

Les variables suivantes, disponibles pour l'ensemble des répondants et corrélées avec la participation à l'enquête, ont été mobilisées pour le calage :

- la variable croisée du sexe, de l'âge en 5 tranches et du département de résidence ;
- le type de ménage imputé ;
- la taille d'unité urbaine ;
- une indicatrice de résidence en quartiers prioritaires de la politique de la ville ;
- le diplôme en 4 modalités ;
- une indicatrice de nationalité française.

Il convient enfin de choisir des bornes de calage suffisamment serrées pour éviter les déformations des poids extrêmes sans toutefois chercher les bornes convergentes les plus resserrées afin d'éviter l'accumulation de poids aux bornes. Les bornes finalement retenues pour ce calage sont [0.5-2]. La courbe de densité du rapport $\frac{\text{poids calés}}{\text{poids initiaux}}$ a bien une allure de loi normale centrée autour de 1.

6.5.3 Résultats

À l'issue de ces étapes de redressement successives, il est possible de réaliser des analyses au niveau national et départemental sur les données de phase 1. Le tableau 5 ci-dessous synthétise la distribution des poids de phase 1 aux différentes étapes des redressements.

TABLEAU 5 – Distribution des poids de phase 1 après redressements

	Poids de sondage	Poids CNR de phase 1	Poids calés de phase 1
Moyenne	269,3	442,8	430,2
Minimum	94,8	117,1	68,4
10 %	112,6	199,5	187,0
25 %	224,8	292,0	259,0
50 %	225,0	350,4	341,4
75 %	365,5	606,4	540,7
90 %	491,2	770,5	764,1
Maximum	491,8	2 762,4	4 495,5
Maximum / minimum	5,2	23,6	65,7

Les distributions départementales détaillées des poids de phase 1 sont présentées en annexe [A](#)

6.6 Précision des indicateurs d'intérêt

6.6.1 Principe général

Les résultats présentés dans cette section sont, comme tout résultat de sondage et par opposition à un relevé exhaustif dans la population, assortis d'une certaine précision. Quand le sondage est aléatoire, comme c'est le cas de l'enquête Genese, la notion d'intervalle de confiance permet de donner une idée de la précision de l'estimation. Lorsqu'un intervalle de confiance à 95 % est fourni pour une grandeur, cela signifie que cet intervalle a 95 % de chances de contenir la valeur qu'aurait donnée une interrogation exhaustive.

La précision des indicateurs de phase 1 est calculée à l'aide du package R Gustave²⁰ disponible sur CRAN. Le package fournit une fonction d'estimation de la variance adaptée aux enquêtes avec une méthodologie de redressement usuelle. La fonction permet de prendre en compte facilement, dans le cadre d'un plan de sondage stratifié, la correction de la non-réponse par groupe de réponse homogène et le calage sur marges.

6.6.2 Application et résultats

L'enquête Genese s'inscrit dans le cadre usuel de la fonction d'estimation de la variance pré-codée du package Gustave (plan de sondage stratifié, repondération par GRH puis calage sur marge). Elle est

20. Gustave : a User-oriented Statistical Toolkit for Analytical Variance Estimation.
<https://github.com/InseeFr/gustave>

donc directement mobilisable dans les calculs de précision des indicateurs de phase 1, présentés dans le tableau 6 ci-dessous.

TABLEAU 6 – Précisions nationales des indicateurs de phase 1 en 2020

	Effectif	Effectif interrogé	Estimation	Intervalle de confiance à 95 %	Précision à 95 %
Cambriolages et tentatives	1 743	102 828	757 579	[714 564 - 800 594]	± 43 015
Vols sans effraction dans le logement	859	102 828	394 173	[361 894 - 426 452]	± 32 279
Actes de vandalisme contre le logement	1 833	102 828	849 084	[801 608 - 896 560]	± 47 476
Vols de voiture ¹	90	40 183	38 774	[29 523 - 48 026]	± 9 252
Vols d'objets dans ou sur la voiture ¹	1 760	40 188	790 640	[746 432 - 834 848]	± 44 208
Vols de vélos ²	81	29 362	37 300	[27 464 - 47 137]	± 9 836
Autres atteintes contre le véhicule ³	1 619	41 372	705 084	[663 330 - 746 839]	± 41 754
Vols et tentatives avec violence ou menaces	499	102 716	264 654	[236 403 - 292 906]	± 28 252
Vols et tentatives sans violence ou menaces	1 906	102 716	840 222	[792 450 - 887 994]	± 47 772
Escroqueries bancaires	7 417	102 716	3 066 520	[2 984 732 - 3 148 308]	± 81 788
Rapports sexuels forcés et tentatives	302	102 673	111 579	[95 955 - 127 203]	± 15 624
Sentiment d'insécurité au domicile	14 438	102 025	6 164 325	[6 055 522 - 6 273 128]	± 108 803
Sentiment d'insécurité dans le quartier	22 049	101 506	9 443 230	[9 313 587 - 9 572 873]	± 129 643

¹ Parmi les répondants qui déclarent une voiture dans le ménage

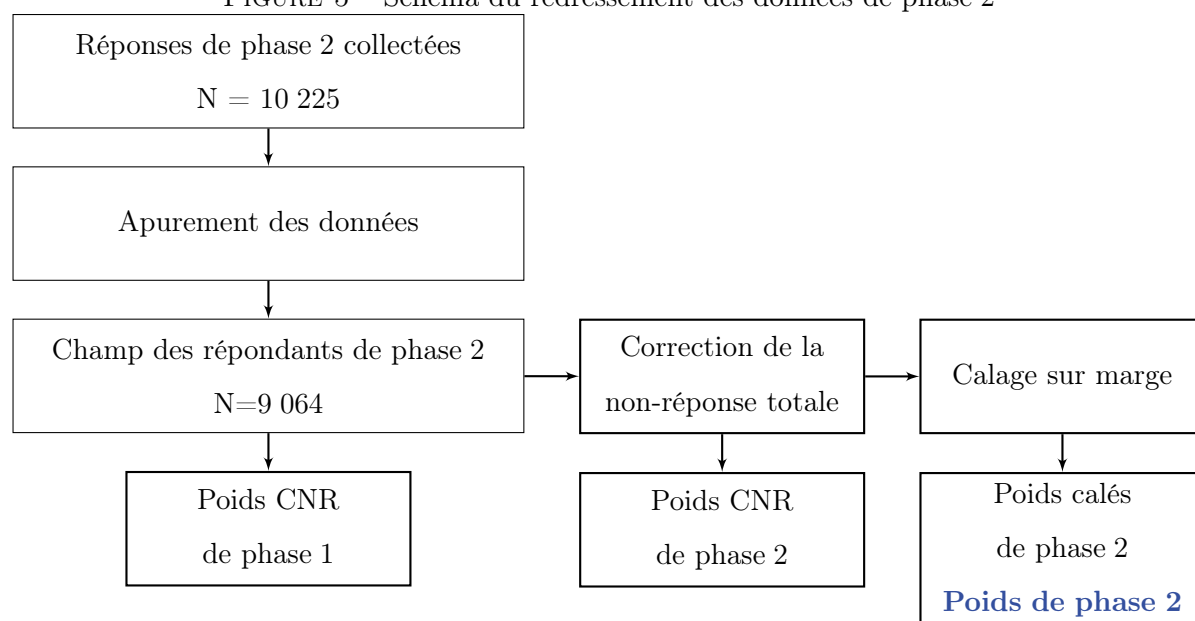
² Parmi les répondants qui déclarent un vélo dans le ménage

³ Parmi les répondants qui déclarent un véhicule dans le ménage

7 Traitements post-collecte de l'enquête de phase 2 (enquête filtrée)

De la même manière qu'en phase 1, les données collectées lors de la phase 2 de l'enquête doivent être apurées et redressées avant d'être exploitées. On commence par exclure les répondants hors champ et les réponses trop partielles puis on redresse les données par correction de la non-réponse totale et calage sur marge pour aboutir aux poids de phase 2. Ces poids finaux sont obtenus à partir des poids corrigés de la non-réponse de phase 1 (et non les poids calés de phase 1) comme décrit dans la figure 5 ci-dessous.

FIGURE 5 – Schéma du redressement des données de phase 2



7.1 Identification et traitement des hors champ

Parmi les 15 000 personnes sollicitées pour participer à la phase 2 de l'enquête, 10 225 ont répondu au questionnaire (par internet ou par téléphone).

De nombreuses variables nécessaires au calage étant issues de questions posées lors de la phase 1 de l'enquête et pour des raisons de calculs des poids, seuls les répondants conservés en phase 1 sont également conservés en phase 2.

7.2 Contrôles de cohérence et apurement

L'aval de la phase 2 de l'enquête a été construit en suivant les instructions du guide méthodologique d'Eurostat²¹ et en vérifiant les données au moyen des contrôles de cohérence prévus.

7.3 Traitement de la non-réponse totale

Les 9 064 réponses exploitables après apurement correspondent à un taux de participation de 60,4 %.

En procédant de la même manière qu'en phase 1 (voir 6.4), la non-réponse de phase 2 a été corrigée en constituant des groupes de réponse homogène par la méthode d'Haziza et Beaumont. Afin de tenir compte des différentes probabilités d'inclusion dans l'échantillon de phase 2, les poids obtenus suite à la repondération par GRH sont ensuite multipliés par l'inverse de la probabilité de sélection à la phase 2.

21. <https://ec.europa.eu/eurostat/fr/web/products-manuals-and-guidelines/-/ks-gq-21-009>

7.3.1 Modélisation de la non-réponse

Plusieurs variables mobilisées pour modéliser la non-réponse de phase 1 ont été de nouveau incluses dans la modélisation de la non-réponse de phase 2.

On retrouve les variables suivantes :

- le croisement du sexe et de la tranche d'âge ;
- le type de ménage imputé ;
- le décile de niveau de vie du ménage ;
- une indicatrice d'allocation chômage ;
- une indicatrice de résidence en logement social ;
- une indicatrice de résidence en appartement ;
- la taille d'unité urbaine ;
- le département de résidence ;
- une indicatrice de résidence en quartiers prioritaires de la politique de la ville ;
- une indicatrice de résidence en quartier de reconquête républicaine.

En tirant profit du protocole en deux phases de l'enquête, il a été possible d'ajouter à ce modèle certaines variables socio-démographiques recueillies dans le questionnaire socle. Les variables suivantes, disponibles pour tous les répondants de phase 1 et corrélées à la participation à l'enquête de phase 2 ont ainsi servies à la modélisation de la non-réponse :

- le diplôme ;
- une indicatrice de nationalité française ;
- une indicatrice d'appartenance à la cible 1 des répondants.

De la même manière, certaines variables d'intérêt collectées dans l'enquête de phase 1, corrélées à la participation à la phase 2 ainsi qu'au sujet de l'enquête thématique, ont également été ajoutées dans le modèle :

- une indicatrice de sentiment d'insécurité dans le quartier ;
- une indicatrice de sentiment d'insécurité au domicile ;
- une variable de comptage du nombre de types différents d'atteintes à la personne subis par période puis agrégés sur toutes les périodes. Cette variable permet de mieux prendre en compte la multivictimation au cours de la vie.

Par exemple, subir un attouchement au niveau du sexe en 2021 rajoute 1 à la variable de comptage ; l'avoir subi en 2020 et en 2021 rajoute 2 à cette variable.

Les probabilités de réponse estimées sont obtenues par régression logistique et permettent de construire des GRH selon la méthode d'Haziza et Beaumont. Le coefficient de détermination du mo-

dèle est cette fois fixé à 95 %. Un seuil à 99 % conduisait à plus de GRH de plus petites taille qui les rendaient moins interprétables.

En plus d’appliquer le coefficient de correction de la non-réponse à chaque unité répondante de phase 2 (comme détaillé en 6.4.1), les poids CNR sont multipliés par l’inverse de la probabilité de sélection à la phase 2 pour tenir compte de l’échantillonnage réalisé entre les deux phases de l’enquête.

7.3.2 Résultats

La modélisation de la non-réponse et les paramètres retenus, aboutissent à la constitution de 7 GRH dont le plus petit comporte 619 unités (tableau 7).

TABLEAU 7 – Groupes de réponse homogène de phase 2 par taille

Groupe	GRH 7	GRH 1	GRH 2	GRH 5	GRH 4	GRH 6	GRH 3
Taille	619	1 342	1 565	2 135	2 778	2 819	3 208

La distribution des poids corrigés de la non-réponse est présentée dans le tableau 8 ci-dessous.

TABLEAU 8 – Distribution des poids corrigés de la non-réponse de phase 2

Min	10 %	25 %	50 %	75 %	90 %	Max
147,0	406,4	638,3	2 468,4	5 710,1	11 987,2	100 345,7

Enfin, le rapport du poids calé maximal sur le poids calé minimal est de 682,4.

7.4 Calage et calcul des pondérations

En procédant de la même manière qu’en phase 1 (voir 6.5), on procède au calage sur marge des données des répondants de l’enquête thématique à l’aide du package R Icarus.

7.4.1 Application et choix des marges

Tout comme en phase 1, la méthode logit est utilisée pour le calage de phase 2. On réutilise la plupart des marges que l’on avait mobilisées pour la phase 1 :

- le type de ménage imputé ;
- la taille d’unité urbaine ;
- une indicatrice de résidence en quartiers prioritaires de la politique de la ville ;
- le diplôme en 4 modalités ;
- une indicatrice de nationalité française.

On retire en revanche la variable croisée du sexe, de l'âge en 5 tranches et du département de résidence, au profit d'une variable avec nettement moins de modalités, la variable croisée du sexe et de l'âge en 5 tranches.

On ajoute également en tant que marges certaines variables issues du questionnaire socle particulièrement liées au sujet de l'enquête thématique. Elles sont obtenues en calculant leurs effectifs pondérés sur la population de phase 1. Ces variables correspondent aux atteintes à la personnes subies en 2020 :

- agression physique ;
- viol ;
- tentative de viol ;
- attouchement au niveau du sexe ;
- autre attouchement ;
- harcèlement sexuel ;
- harcèlement moral.

Les bornes retenues pour ce calage sont [0.5-2]. Elles permettent d'obtenir une courbe de densité du rapport $\frac{\text{poids calés}}{\text{poids initiaux}}$ ayant une allure de loi normale centrée autour de 1.

7.4.2 Résultats

La suppression de la marge sexe, tranche d'âge et département s'est imposée du fait de la taille de l'échantillon, certaines modalités de cette marge étaient de taille trop faibles voire nulles pour être conservée, notamment dans les plus petits départements ou sur les tranches d'âge moins répondantes. Du fait de cette suppression, les données issues de l'enquête de phase 2 ne sont plus représentatives au niveau départemental mais seulement au niveau national. Le tableau 9 ci-dessous synthétise la distribution des poids de phase 2 aux différentes étapes des redressements.

TABLEAU 9 – Distribution des poids de phase 2 aux différentes étapes de redressement

	Poids calés de phase 1	Poids CNR de phase 2	Poids calés de phase 2
Moyenne	430,2	4 758,7	4 880,4
Minimum	68,4	147,0	85,6
10 %	187,0	406,4	388,6
25 %	259,0	638,3	641,8
50 %	341,4	2 468,4	2 360,9
75 %	540,7	5 710,2	5 462,8
90 %	764,1	11 987,2	12 633,1
Maximum	4 495,5	100 345,7	116 564,2
Maximum / minimum	65,7	682,4	1 361,0

Enfin, la distribution des poids de phase 1 et de phase 2 aux différentes étapes de redressement est présentée en annexe B.

7.5 Précision des indicateurs d'intérêt

Les calculs de précision des indicateurs de l'enquête thématique sont plus complexes à effectuer que ceux de l'enquête socle du fait du plan de sondage en deux phases. L'application directe de la fonction pré-codée d'estimation de la variance du package Gustave n'est plus possible et la création d'une fonction tenant compte de la méthodologie de l'enquête est en théorie nécessaire.

Par soucis de simplification et de gain de temps, il a été choisi de se rapprocher du cadre théorique d'utilisation de la fonction d'estimation de la variance du package Gustave plutôt que de créer une fonction qui permette de prendre en compte les spécificités du protocole de l'enquête avec exactitude. Les poids de phase 1 ont été multipliés par la probabilité de la sélection en phase 2 afin de prendre en compte le tirage en deux phases propre à l'enquête Genese.

TABLEAU 10 – Précisions nationales des indicateurs de phase 2 en 2020

	Effectif	Estimation	Intervalle de confiance à 95 %	Précision à 95 %
Violences dans l'enfance avant l'âge de 15 ans				
Climat de violence entre les parents	1 873	5 488 815	[5 098 527 - 5 879 104]	± 390 288
Humiliations entre les parents	1 200	3 140 528	[2 832 886 - 3 448 170]	± 307 642
Violences physiques entre les parents	1 391	4 253 463	[3 904 960 - 4 601 966]	± 348 503
Violences psychologiques	1 688	3 825 901	[3 535 904 - 4 115 899]	± 289 998
Humiliations répétées des parents ou harcèlement intrafamilial	1 071	2 399 154	[2 151 794 - 2 646 515]	± 247 360
Harcèlement hors intrafamilial	772	1 669 694	[1 483 468 - 1 855 921]	± 186 226
Violences physiques exercées par les parents	1 738	5 348 354	[4 978 265 - 5 718 442]	± 370 088
Violences sexuelles	1 808	3 372 746	[3 089 567 - 3 655 925]	± 283 179
Violences sexuelles intrafamiliales	852	1 470 637	[1 273 627 - 1 667 648]	± 197 010
Violences sexuelles hors intrafamilial	956	1 902 108	[1 683 420 - 2 120 797]	± 218 688
Violences commises par partenaire				
Victimes au moins une fois depuis l'âge de 15 ans				
Violences psychologiques	3 845	10 175 885	[9 672 640 - 10 679 129]	± 503 244
Contrôle ou emprise	2 850	7 452 910	[7 007 864 - 7 897 957]	± 445 046
Harcèlement moral ou dénigrement	2 924	6 274 584	[5 887 213 - 6 661 956]	± 387 372
Intimidations ou menaces	2 638	5 477 427	[5 105 151 - 5 849 702]	± 372 276
Violences physiques ou sexuelles	2 650	4 814 554	[4 470 799 - 5 158 308]	± 343 754
Violences physiques exclusivement	938	2 641 545	[2 377 659 - 2 905 430]	± 263 886
Violences sexuelles exclusivement	510	785 508	[637 899 - 933 117]	± 147 609
Violences physiques et sexuelles	1 202	1 387 501	[1 179 356 - 1 595 646]	± 208 145
Victimes au moins une fois au cours des 5 dernières années				
Violences psychologiques	1 726	4 396 265	[4 030 115 - 4 762 416]	± 366 150
Violences physiques ou sexuelles	925	1 387 328	[1 189 769 - 1 584 887]	± 197 559
Victimes au moins une fois au cours des 12 derniers mois				
Violences psychologiques	759	1 794 217	[1 567 877 - 2 020 558]	± 226 340
Violences physiques ou sexuelles	319	414 032	[309 072 - 518 992]	± 104 960
Violences commises par un non partenaire depuis l'âge de 15 ans				
Victimes au moins une fois depuis l'âge de 15 ans				
Violences physiques	2 564	7 855 805	[7 460 216 - 8 251 394]	± 395 589
dont violences exclusivement physiques	1 318	5 679 045	[5 329 757 - 6 028 333]	± 349 288
Violences sexuelles	2 527	4 535 268	[4 241 432 - 4 829 105]	± 293 836
dont violences exclusivement sexuelles	1 281	2 358 508	[2 125 370 - 2 591 646]	± 233 138
Violences physiques et sexuelles	1 246	2 176 760	[1 965 577 - 2 387 944]	± 211 184
Victimes au moins une fois au cours des 5 dernières années				
Violences physiques	535	1 469 354	[1 292 256 - 1 646 452]	± 177 098
dont violences exclusivement physiques	379	1 229 609	[1 062 255 - 1 396 963]	± 167 354
Violences sexuelles	500	811 736	[660 881 - 962 591]	± 150 855
dont violences exclusivement sexuelles	344	571 991	[430 663 - 713 318]	± 141 328
Violences physiques et sexuelles	156	239 745	[177 941 - 301 549]	± 61 804
Victimes au moins une fois au cours des 12 derniers mois				
Violences physiques	191	389 754	[283 888 - 495 621]	± 105 866
dont violences exclusivement physiques	159	364 426	[261 504 - 467 349]	± 102 922
Violences sexuelles	128	189 397	[97 680 - 281 114]	± 91 717
dont violences exclusivement sexuelles	96	164 069	[75 556 - 252 583]	± 88 514

A Distributions départementales des poids de phase 1

	total des poids			moyenne	Ecart-type		quantile des poids calés							rapport poids calés sur poids tirage						
Niveau	tirage	CNR	calés	Poids calés	Poids calés	poids calés/ poids tirage	min	0.10	0.25	0.5	0.75	0.9	max	min	0.10	0.25	0.5	0.75	0.9	max
National	27 758 279,9	45 531 261,9	44 236 135,0	430,2	277,3	0,6	68,4	187,0	259,0	341,4	540,7	764,1	4 495,5	0,6	1,1	1,2	1,4	1,8	2,3	9,3
Départements																				
01	282 720,0	443 617,7	449 239,1	483,1	278,2	0,5	164,5	252,3	291,7	364,6	605,0	864,8	1 943,2	0,7	1,1	1,2	1,4	1,7	2,2	4,0
02	212 863,0	371 524,2	358 069,9	525,0	282,8	0,6	178,3	280,7	330,4	449,6	615,9	868,9	2 365,5	0,8	1,1	1,3	1,5	2,0	2,6	4,8
03	143 474,3	228 014,7	225 994,3	369,3	197,8	0,6	126,1	197,8	230,8	294,5	465,6	627,0	1 716,9	0,7	1,1	1,2	1,5	1,8	2,3	4,8
04	66 163,7	114 692,4	112 177,7	196,5	91,4	0,8	84,3	123,2	138,0	167,8	223,0	319,6	881,5	0,7	1,1	1,2	1,5	1,9	2,7	7,4
05	62 142,6	97 330,6	90 228,8	145,8	47,8	0,5	69,6	97,8	112,7	134,3	164,3	208,6	387,2	0,7	1,0	1,1	1,3	1,7	2,1	3,7
06	470 199,3	809 216,1	720 341,9	454,2	260,0	0,6	94,4	230,0	277,2	366,3	564,6	785,4	2 680,6	0,7	1,0	1,2	1,4	1,8	2,3	5,5
07	137 062,2	227 474,1	220 019,9	366,7	205,6	0,6	115,0	201,7	229,0	298,1	448,9	602,1	1 805,7	0,7	1,1	1,3	1,5	1,8	2,3	4,9
08	107 914,3	187 389,4	183 222,0	320,9	194,3	0,6	102,9	148,6	176,3	257,6	407,2	569,8	1 365,2	0,7	1,1	1,2	1,5	2,0	2,5	4,5
09	67 720,0	102 551,8	102 203,6	158,5	55,8	0,5	75,8	107,7	122,2	144,4	179,5	224,4	481,3	0,7	1,0	1,2	1,4	1,7	2,1	4,4
10	121 250,9	206 859,7	204 106,1	351,9	199,6	0,7	115,8	179,1	209,6	286,0	426,6	607,3	1 582,5	0,8	1,1	1,2	1,5	2,0	2,4	5,3
11	147 874,5	249 654,9	254 685,8	436,1	255,2	0,6	145,7	224,2	261,3	339,4	552,9	768,8	2 029,9	0,8	1,2	1,3	1,5	1,9	2,5	5,0
12	118 880,6	186 196,4	184 787,9	292,8	174,4	0,5	98,2	157,7	182,3	226,3	387,3	488,7	2 740,0	0,7	1,1	1,3	1,4	1,7	2,1	9,1
13	820 426,9	1 466 416,3	1 359 815,5	475,6	297,9	0,8	81,0	238,7	293,1	383,8	570,0	811,0	4 163,8	0,7	1,1	1,2	1,5	2,0	2,6	9,1
14	311 239,5	488 541,6	469 594,6	463,1	275,2	0,5	164,5	255,8	295,8	362,3	554,6	778,7	3 964,8	0,7	1,0	1,2	1,4	1,7	2,1	8,1
15	59 586,4	97 103,5	95 642,1	159,4	61,8	0,6	69,8	109,1	123,5	143,8	179,2	228,9	817,5	0,7	1,1	1,2	1,4	1,8	2,3	8,0
16	146 915,9	240 276,5	242 279,6	401,1	220,5	0,6	103,9	208,5	245,6	324,4	509,3	666,1	1 602,6	0,7	1,1	1,3	1,5	1,9	2,5	5,6
17	290 475,1	454 078,9	438 186,8	459,8	261,3	0,5	161,4	246,0	288,4	359,9	574,7	799,3	2 565,4	0,7	1,0	1,2	1,4	1,7	2,1	5,2
18	125 361,6	205 749,5	202 782,3	334,1	196,8	0,6	103,7	172,7	199,5	253,9	434,3	591,4	1 842,9	0,7	1,1	1,2	1,5	1,8	2,3	6,3
19	102 231,7	163 709,6	159 878,0	258,3	131,3	0,6	86,4	148,4	168,2	216,4	317,8	401,6	1 326,2	0,7	1,1	1,3	1,5	1,8	2,2	5,6
21	241 278,8	376 418,6	351 989,0	454,8	249,7	0,6	161,9	251,8	286,6	380,2	564,6	726,0	3 084,6	0,7	1,0	1,2	1,3	1,7	2,0	6,3
22	252 229,9	411 923,9	405 735,9	500,9	279,9	0,6	177,0	286,6	330,2	401,9	581,2	829,9	2 817,8	0,7	1,1	1,3	1,5	1,8	2,3	6,1
23	46 989,2	77 675,2	78 871,4	176,4	60,4	0,6	75,0	121,0	136,7	163,9	198,9	241,3	508,2	0,7	1,1	1,3	1,5	1,9	2,3	4,7
24	170 311,0	280 476,5	278 383,7	459,4	308,7	0,6	149,6	232,4	270,9	356,8	574,7	815,0	3 647,5	0,7	1,1	1,2	1,5	1,8	2,3	8,0
25	238 484,9	367 076,4	360 545,9	442,4	258,4	0,6	83,4	221,8	274,6	358,5	537,4	781,1	2 681,9	0,7	1,0	1,1	1,4	1,8	2,2	6,8
26	219 272,7	360 586,7	355 216,9	488,6	305,9	0,6	159,5	229,2	276,3	384,3	622,7	864,1	2 865,3	0,7	1,0	1,2	1,4	1,8	2,4	5,8
27	239 377,1	407 747,5	405 684,1	514,2	335,1	0,7	166,7	253,1	299,2	399,8	635,5	886,1	3 858,0	0,7	1,1	1,3	1,5	1,9	2,3	7,9
28	187 918,0	293 610,2	284 647,3	449,7	233,4	0,5	107,0	237,0	288,6	367,7	562,2	751,5	1 653,2	0,7	1,1	1,2	1,4	1,8	2,3	4,7
29	424 300,3	631 364,3	624 132,5	454,9	227,7	0,5	168,9	254,5	296,1	363,0	560,1	779,5	2 160,1	0,7	1,0	1,2	1,4	1,6	2,0	6,2
2A	47 264,5	105 497,8	110 537,1	239,3	118,3	1,1	82,7	133,9	162,1	205,6	282,0	382,1	974,2	0,8	1,3	1,6	2,0	2,7	3,8	9,7
2B	54 631,0	111 357,9	127 566,5	268,0	129,7	1,1	82,5	141,7	179,0	240,7	328,5	420,3	1 053,8	0,7	1,2	1,6	2,1	2,8	3,7	9,0
30	300 604,3	520 080,2	515 128,1	510,0	336,5	0,7	89,6	253,0	305,8	390,7	609,5	888,3	3 987,3	0,7	1,1	1,3	1,5	1,9	2,5	8,1
31	614 864,4	990 803,2	984 452,4	479,8	291,1	0,6	76,1	251,4	290,0	371,6	600,7	830,3	2 849,4	0,7	1,1	1,2	1,4	1,8	2,3	6,4
32	80 572,7	130 116,8	132 963,6	217,6	84,0	0,6	96,8	137,4	164,5	201,3	243,8	317,1	1 188,7	0,7	1,0	1,2	1,5	1,8	2,4	8,7
33	734 763,4	1 164 986,0	1 145 011,7	459,7	265,7	0,6	84,7	240,6	289,8	361,8	577,6	803,8	3 802,6	0,7	1,1	1,2	1,4	1,7	2,2	7,7
34	480 933,1	822 978,1	798 529,9	455,5	299,5	0,7	77,2	199,8	269,0	355,1	570,0	827,2	3 681,9	0,7	1,1	1,2	1,5	1,9	2,4	8,3
35	494 374,3	743 390,8	733 336,3	438,9	240,1	0,5	89,8	235,1	279,0	348,9	570,1	740,6	2 167,6	0,7	1,1	1,2	1,4	1,7	2,0	6,0
36	85 440,2	145 315,3	146 034,0	251,3	146,0	0,7	99,3	132,6	153,0	195,0	298,9	462,1	1 061,4	0,8	1,1	1,3	1,5	2,0	2,5	8,3
37	264 732,2	436 336,4	414 971,8	462,1	271,9	0,7	92,5	241,7	285,9	357,2	576,9	804,7	2 494,2	0,7	1,1	1,2	1,4	1,8	2,2	8,5
38	559 429,3	873 969,1	854 434,5	426,8	261,3	0,6	72,4	197,3	256,0	341,4	548,1	751,8	2 519,5	0,7	1,0	1,2	1,4	1,8	2,2	8,6
39	113 269,4	181 091,2	176 971,4	282,7	165,0	0,6	111,1	155,6	178,2	225,0	339,2	467,3	1 717,8	0,8	1,1	1,2	1,5	1,7	2,2	8,8
40	186 957,9	287 604,5	281 245,2	438,8	257,5	0,5	144,2	229,7	265,3	342,9	513,1	780,8	1 619,5	0,7	1,0	1,2	1,4	1,7	2,1	6,8
41	134 949,8	225 071,1	218 275,0	365,0	214,0	0,6	123,8	179,4	213,6	286,0	466,7	636,4	1 467,6	0,7	1,1	1,2	1,4	1,8	2,2	6,6
42	317 143,9	519 956,5	520 294,0	477,3	292,5	0,7	71,0	235,8	286,5	359,7	605,4	867,6	2 269,2	0,7	1,1	1,2	1,5	1,9	2,4	8,1
43	97 219,6	155 554,1	153 721,4	250,4	131,9	0,5	100,9	134,5	153,4	201,9	318,6	415,2	897,4	0,8	1,1	1,2	1,4	1,8	2,2	4,8
44	662 301,6	1 007 890,9	997 137,1	443,7	241,4	0,5	96,5	245,8	278,1	343,0	564,4	774,7	2 595,6	0,7	1,1	1,2	1,4	1,6	2,1	8,8
45	280 914,1	460 526,3	451 472,0	489,7	298,9	0,6	161,3	252,3	292,9	377,4	615,2	880,3	4 124,1	0,7	1,1	1,2	1,4	1,8	2,3	8,4
46	70 994,5	118 455,2	117 790,7	198,6	74,6	0,6	100,5	132,0	150,2	177,6	224,9	284,0	822,5	0,9	1,1	1,3	1,5	1,9	2,4	6,6
47	135 592,2	223 629,2	221 776,1	367,2	226,3	0,6	128,4	189,0	219,3	300,6	448,7	619,0	2 826,0	0,8	1,1	1,2	1,5	1,9	2,3	7,8
48	30 250,2	49 616,4	51 944,5	174,9	56,6	0,5	68,4	118,2	137,8	161,5	197,1	252,6	390,6	0,7	1,2	1,4	1,6	1,9	2,4	3,8
49	369 963,7	566 148,0	543 010,3	452,9	239,3	0,5	163,7	252,3	288,5	358,6	566,1	756,0	1 958,9	0,7	1,1	1,2	1,4	1,6	2,0	5,8
50	217 074,6	341 900,5	341 000,1	467,5	299,4	0,6	169,7	240,5	278,8	373,5	587,1	804,9	3 667,2	0,7	1,0	1,2	1,4	1,7	2,1	7,5
51	240 051,2	387 716,5	376 033,9	476,6	275,4	0,6	160,1	242,8	281,5	386,4	584,8	809,8	2 144,3	0,7	1,0	1,2	1,4	1,8	2,3	8,6
52	72 684,5	119 927,7	116 598,6	193,4	77,9	0,6	83,3	123,3	146,8	180,5	217,8	276,0	1 010,6	0,7	1,0	1,2	1,5	1,8	2,3	8,1
53	134 612,9	201 184,1	200 231,4	311,0	167,3	0,4	105,0	165,7	192,3	239,1	401,1	530,4	1 107,9	0,7	1,0	1,2	1,4	1,6	2,0	3,6
54	308 611,9	510 374,7	497 863,1	491,5	303,5	0,6	166,5	242,1	280,5	397,8	614,0	866,0	2 859,7	0,7	1,0	1,2	1,4	1,9	2,3	5,8
55	78 174,2	125 066,9	122 662,9	201,1	73,5	0,6	95,4	135,2	152,7	178,1	228,0	303,9	615,5	0,8	1,1	1,2	1,4	1,8	2,3	4,7
56	320 704,3	535 438																		

B Distributions des poids de phase 1 et phase 2

	Poids de sondage	Poids CNR de phase 1	Poids calés de phase 1	Poids CNR de phase 2	Poids calés de phase 2
Moyenne	269,3	442,8	430,2	4 758,7	4 880,4
Minimum	94,8	117,1	68,4	147,0	85,6
10 %	112,6	199,5	187,0	406,4	388,6
25 %	224,8	292,0	259,0	638,3	641,8
50 %	225,0	350,4	341,4	2 468,4	2 360,9
75 %	365,5	606,4	540,7	5 710,2	5 462,8
90 %	491,2	770,5	764,1	11 987,2	12 633,1
Maximum	491,8	2 762,4	4 495,5	100 345,7	116 564,2
Maximum / minimum	5,2	23,6	65,7	682,4	1 361,0