

Victimations déclarées et effets de mode : enseignements de l'expérimentation panel multimode de l'enquête Cadre de Vie et Sécurité

Document de travail

N° M2023-04 – Octobre 2023



M 2023/04

**Victimations déclarées et effets de mode :
enseignements de l'expérimentation panel
multimode de l'enquête
Cadre de Vie et Sécurité**

**Laura CASTELL
Marie CLERC
Delphine CROZE
Stéphane LEGLEYE
Amandine NOUGARET**

Insee

Octobre 2023

Remerciements : Les auteurs remercient Patrick Sillard pour son investissement et son aide précieuse, ainsi que Valérie Albouy, Hélène Chaput, Christel Colin, Sylvie Lagarde, Corinne Prost et Julie Solard pour leurs relectures attentives et leurs encouragements.

Direction de la méthodologie et de la coordination statistique et internationale
Département des Méthodes Statistiques -Timbre L001 -
88 Avenue Verdier - CS 70058 - 92541 Montrouge Cedex - France -
Tél. : 33 (1) 87 69 55 00 - E-mail : DG75-L001@insee.fr - Site Web Insee : <http://www.insee.fr>

*Ces documents de travail ne reflètent pas la position de l'Insee et n'engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their author's views.*

Table des matières

Introduction.....	2
1. Les enquêtes de victimation et le multimode : pourquoi l'expérimentation CVS panel ?.....	3
1.1. L'enquête Cadre de vie et sécurité.....	3
1.2. Le multimode dans les enquêtes de victimation.....	3
1.3. Genèse de l'expérimentation panel.....	5
1.4. Protocole de l'expérimentation panel multimode.....	6
2. L'impact du protocole sur la participation.....	9
2.1. Bilan de la collecte.....	9
Taux de collecte et taux de réussite.....	9
Réception des questionnaires, effet des relances.....	10
Non-réponse partielle.....	13
Vérification de l'identité de l'individu panel (kish).....	14
2.2. Représentativité des répondants.....	15
Les indicateurs utilisés pour la sélection sur observables.....	16
La sélection sur observables dans CVS panel selon le protocole.....	17
Redressement du biais de sélection sur observables.....	19
Un biais de sélection inobservable ?.....	21
3. Les effets de mesure sur la victimation.....	23
3.1. Analyse brute des résultats.....	23
3.2. Travaux d'apurement.....	24
Traitement des doubles comptes.....	24
Traitement des libellés pour étudier le bon classement des victimations.....	28
L'impact de l'apurement sur les taux de victimation bruts.....	30
3.3. Comparaison avec les taux de victimations de CVS 2019.....	31
3.4. Méthodes d'estimation des effets de mode.....	32
Modèles d'estimation fondés sur le score de propension.....	33
Variables de contrôle de la sélection dans CVS panel.....	34
Estimation du score de propension.....	34
Méthode d'estimation fondée sur l'estimateur LATE.....	36
3.5. Résultats.....	38
Effets du mode de collecte sur les victimations.....	38
Analyses complémentaires.....	42
3.6. Correction des effets de mesure.....	44
Conclusion.....	48
Bibliographie.....	50
Annexes.....	51

Introduction

Les protocoles multimodes dans les enquêtes auprès des ménages permettent de bénéficier des avantages de différents modes de collecte. En offrant aux enquêtés plusieurs modalités de collecte, ils visent à augmenter les taux de réponse et à capter des populations différentes, tout en maîtrisant les coûts de collecte. Internet permet ainsi aux enquêtés peu disponibles et ayant une appétence pour les outils de communication numérique de répondre plus facilement alors que les modes intermédiés par enquêteur permettent d'établir un lien avec des enquêtés plus réfractaires aux outils informatiques ou plus confiants face à un contact humain direct.

Cependant, le fait que les répondants aux différents modes de collecte et le contexte d'interrogation diffèrent selon les modes de collecte pose des difficultés particulières de comparabilité des résultats. De fait, ces différences exposent à des biais de sélection et de mesure entre les différents modes de collecte, ce qui peut affecter la fiabilité des estimations. Ces biais peuvent être d'ampleur variable selon la thématique de l'enquête, sa complexité ou encore le contexte de la collecte. Suivant la quantité d'information auxiliaire disponible et son lien avec la thématique d'enquête, les estimations des biais seront elles-mêmes plus ou moins simples à réaliser.

Les enquêtes de victimation peuvent être particulièrement sujettes à ce type de biais, du fait de la sensibilité de la thématique et de l'impact du mode de collecte à la fois sur le fait de répondre et sur la façon de répondre. L'enquête de victimation annuelle Cadre de vie et sécurité (CVS), traditionnellement conduite en face-à-face, comportait déjà une composante multimode puisque les questions les plus sensibles sont posées sous casque, sans intermédiation d'un enquêteur. Plus largement, l'enquête CVS avait déjà fait l'objet d'une réplique en 2013 sous la forme de l'interrogation sur internet d'un échantillon indépendant. En 2019, les répondants de l'exercice 2018 ont été réinterrogés sur deux modes : internet ou téléphone, dans des protocoles séquentiels attribués aléatoirement. Cette expérimentation présente des caractéristiques rares permettant de mobiliser divers types d'informations auxiliaires susceptibles d'aider à produire une estimation fiable des effets de sélection et de mesure : elle est longitudinale et bénéficie du recueil des victimations personnelles ainsi que du sentiment d'insécurité antérieurs ; elle peut être enrichie de données externes sur l'exposition déclarées à des violences *via* le nombre de dépôts de plaintes pour divers crimes et délits au niveau de l'Iris (découpage infra-communal).

Ce document de travail a pour objet de présenter les principaux résultats issus de l'expérimentation CVS panel sur l'impact d'une interrogation par Internet sur la qualité des estimations en matière de victimation. Dans un premier temps, nous montrons que le téléphone est un mode de collecte complémentaire essentiel par rapport à un protocole exclusivement internet, pour disposer d'une meilleure représentativité des répondants. Dans un second temps, nous montrons qu'il persiste des effets de mesure plus ou moins importants selon les victimations. Enfin, nous proposons des résultats d'estimation des taux de victimation, fondés sur l'enquête et corrigés des effets de mesure estimés.

Mots-clés : Enquête multimode, cadre de vie et sécurité, information auxiliaire, biais de mesure.

Classification JEL : C21, C26, C83, C93

1. Les enquêtes de victimation et le multimode : pourquoi l'expérimentation CVS panel ?

1.1. L'enquête Cadre de vie et sécurité

L'enquête de victimation Cadre de vie et sécurité (CVS) vise à mieux connaître les faits de délinquance dont les ménages et leurs membres ont pu être victimes durant les deux années précédant l'enquête. En interrogeant directement un échantillon de la population, on mesure la prévalence (proportion de victimes) et l'incidence (nombre d'actes subis) de certaines atteintes contre les personnes physiques ou leurs biens. On peut également en étudier les variations annuelles, et alors effectuer des comparaisons de ces tendances avec celles des données administratives sur les faits déclarés à la police ou à la gendarmerie (notamment celles qui sont issues de l'enregistrement des crimes et délits par la police et la gendarmerie mais aussi celles issues des décisions de justice). De plus, le dispositif permet d'analyser les caractéristiques socio-démographiques des victimes, et d'aborder les effets de la victimation subie (dommages matériels et psychologiques, conséquences dans la vie quotidienne des victimes). L'enquête CVS permet aussi d'estimer la part des atteintes qui ne sont pas reportées aux autorités et de connaître les caractéristiques des victimes qui les ont subies. L'enquête porte également sur l'opinion des personnes en matière de sécurité, dans leur cadre de vie notamment, et permet aussi d'estimer leur sentiment d'insécurité.

L'enquête comporte des modules thématiques aperiodiques visant à intégrer des questions sur des atteintes non ou partiellement couvertes dans les modules permanents. Par exemple, trois nouvelles sections ont été introduites en 2018, la première portant sur les arnaques, la deuxième sur la corruption et la troisième sur les atteintes à caractère discriminatoire.

Après avoir fait l'objet de modules introduits en 2005 et 2006 dans le dispositif d'enquêtes permanentes des Conditions de vie (EPCV), dispositif qui intégrait depuis 1996 quelques questions sur les victimations¹, l'enquête CVS devient indépendante et annuelle en 2007. À l'exception du millésime 2020 qui n'a pu être collecté à cause de l'épidémie de Covid-19, l'enquête a été réalisée jusqu'en 2021 par l'Insee, en co-maîtrise d'ouvrage avec l'Observatoire national de la délinquance et des réponses pénales (ONDRP) et le service statistique ministériel de la sécurité intérieure du ministère de l'Intérieur (SSMSI), principalement en France métropolitaine². À partir de 2022, le SSMSI a pris en charge la collecte annuelle d'une nouvelle enquête de victimation.

1.2. Le multimode dans les enquêtes de victimation

Depuis sa mise en place en 2007, l'enquête CVS est une enquête annuelle, en coupe et multimode. Collectée majoritairement en face à face, une partie du questionnaire est réalisée en auto-administré sous casque. Cette partie du questionnaire concerne les victimations les plus sensibles (violences sexuelles et violences intra-familiales). De fait, on s'attend pour de telles victimations à une sous-déclaration dans une situation intermédiée par un enquêteur, sous l'effet du phénomène dit de désirabilité sociale. C'est ce qu'on appelle un effet de mode : l'estimation des victimations est différente selon le mode de collecte par lequel les enquêtés sont interrogés (voir encadré 1).

1 Les cambriolages, les vols de voitures et vols à la roulotte, les agressions dont les vols avec violence et les vols simples.

2 Des extensions en territoire d'Outre-mer ont eu lieu ponctuellement : La Réunion (2011), Guadeloupe, Martinique, Guyane (2015), Mayotte (2020) et en Nouvelle-Calédonie (2021 – collecte non assurée par des enquêteurs Insee).

Encadré 1 : Les effets de mode

Un effet de mode se définit comme l'effet du mode de collecte sur l'estimation du phénomène qu'on étudie. On parle le plus souvent d'effet de mode comme la différence d'estimation entre deux modes de collecte différents, pour des enquêtes multimodales. En théorie, les effets de mode peuvent également exister pour des enquêtes monomodes, si on suppose qu'il existe une « vraie » valeur du phénomène qu'on cherche à étudier (pour des questions factuelles principalement).

L'effet de mode se décompose en deux effets (Klausch, 2014) :

- un effet de mesure : il s'agit de l'effet du mode sur la réponse elle-même pour un même individu. La littérature identifie deux principaux mécanismes expliquant un biais de mesure : la désirabilité sociale et le *satisficing*. L'effet de désirabilité sociale correspond à la formulation de réponses consciemment ou inconsciemment insincères permettant de donner une image de soi en accord avec les attendus normatifs de son environnement. Cet effet est moins important en auto-administré qu'avec l'intermédiation d'un enquêteur. Le biais qui peut exister entre différents modes de collecte lié à cet effet est plus important pour des questions d'opinion ou des questions sensibles. L'effet de *satisficing* correspond au fait d'adapter ses efforts à donner la réponse la plus juste en fonction du niveau d'aspiration attendu. Cet effet conduit à une moindre justesse dans la réponse en auto-administré du fait de moindres efforts. Cet arbitrage peut se manifester de plusieurs manières : non-réponse partielle, valeurs refuges (NSP, refus, valeurs modales), effet de liste (réponse aux premières ou dernières modalités), etc. Comme nous le verrons dans ce document, il peut y avoir aussi des effets de mesure sur des variables plus factuelles, lorsqu'elles nécessitent une compréhension fine des concepts utilisés.
- un effet de sélection : il s'agit de l'effet de l'auto-sélection des répondants. Cet effet de sélection se caractérise par une composition de l'échantillon de répondants différente : l'estimation est alors différente car les personnes sur lesquelles elle est réalisée sont différentes. Cet effet de sélection peut différer selon le mode de collecte. En effet, imposer un mode de collecte, ou laisser le choix du mode de collecte, aux unités statistiques peut conditionner leur participation à l'enquête suivant leurs capacités pratiques (par exemple équipement internet ou téléphone), leurs compétences (par exemple maîtrise des outils, ressources cognitives, littératie) et leurs préférences pour un mode plutôt qu'un autre (par exemple sentiment de confidentialité et d'anonymat en lien avec la thématique du questionnaire).

Les enquêtes de victimation peuvent être sujettes aux effets de mode pour plusieurs raisons. Tout d'abord, la déclaration de victimation et l'opinion en matière de sécurité sont des sujets sensibles, potentiellement soumis à des effets de mesure. On peut s'attendre à ce que joue un phénomène de retenue ou de pudeur à déclarer des victimations ou des opinions négatives lorsque l'enquête est intermédiée par un enquêteur. Inversement, l'enquête CVS est construite autour d'un questionnement complexe, avec des concepts juridiques précis, qui peut s'avérer source d'erreurs en auto-administré : les victimations doivent être renseignées de la plus grave à la moins grave au sens juridique, en excluant les faits précédemment déclarés. Enfin, il peut également exister un biais de sélection lié à la thématique de l'enquête : des personnes victimes pourraient être plus réticentes, alors que d'autres pourraient être plus enclines, à répondre à une enquête dont la thématique affichée porte sur la sécurité et les victimations, et ce de façon éventuellement différenciée selon le mode de collecte proposée.

Les enquêtes de victimation ont déjà fait l'objet d'expérimentations de changement de protocole de collecte. Ainsi, en 2013, l'enquête Vols, violences et sécurité (VVS) a été menée par l'Insee en parallèle de l'enquête CVS pour mesurer les différences d'estimations liées à une collecte entièrement auto-administrée. Cette expérimentation a été menée auprès de 40 000 ménages *via* les modes de collecte internet et papier. Le taux de réponse était de 32 %, contre 63 % pour l'enquête CVS 2013. Les principaux résultats de cette expéri-

mentation sont des taux de victimations significativement plus élevés en auto-administré, y compris après divers calages. Ces résultats laissent supposer l'existence d'un biais de sélection inobservable, les personnes les plus concernées participant davantage à l'enquête, les autres ne répondant qu'après sollicitation et argumentation par l'enquêteur (Razafindranovona, 2016).

1.3. Genèse de l'expérimentation panel

À l'issue des différents débats consacrés depuis 2015 au dispositif CVS³, le Conseil national de l'information statistique (Cnis) a préconisé d'explorer :

- d'une part, les possibilités d'évolution du dispositif de collecte en dissociant les besoins d'informations annuelles (évolution de la prévalence et de l'incidence) de ceux plus « structurels » (description fine des victimations permise par l'empilement de plusieurs millésimes) ;
- et d'autre part, le recours à d'autres supports de collecte (internet), malgré les premiers enseignements tirés de l'expérimentation internet « Vols, violences et sécurité » (VVS) menée en 2013, qui avait révélé de forts biais induits par la collecte internet.

Ces recommandations débouchent sur deux expérimentations. La première consiste en l'insertion d'une quinzaine de questions de victimation dans le panel SRCV en 2018 et en 2019. La seconde repose sur la mise en place d'un dispositif expérimental visant à réinterroger en 2019 les ménages répondants de l'échantillon 2018 de CVS en utilisant d'autres modes de collecte que le face-à-face classiquement utilisé. C'est le bilan de cette seconde expérimentation, appelée par la suite « expérimentation CVS panel » qui fait l'objet du présent document.

L'objectif principal de l'expérimentation CVS panel multimode est de comparer différents protocoles pour voir lequel assure le meilleur taux de collecte. Le second objectif est de mesurer les effets de mode. Enfin, le troisième objectif est de mesurer les corrélations temporelles et étudier la pertinence d'un dispositif « panel » pour améliorer la précision des indicateurs conjoncturels.

Plus précisément, concernant les deux 1^{ers} objectifs, l'expérimentation vise à :

- mesurer l'acceptabilité d'une réinterrogation par le biais d'un questionnaire internet / téléphone portant sur la victimation et le sentiment d'insécurité. L'attrition sur une enquête sur ce thème était alors une inconnue, les variables d'intérêt pouvant être directement corrélées avec le fait de répondre ou non à l'enquête, comme l'enquête VVS 2013 l'a montré dans le cadre d'une collecte par questionnaire auto-administré ;
- mesurer le gain ou la perte de couverture et de représentativité due à l'introduction d'internet ;
- étudier les effets de mode (internet/téléphone) sur les mesures des victimations en dissociant effets de sélection et effets de mesure ;
- comparer la mesure de la prévalence de la victimation à partir des données de la collecte CVS 2019 classique et des données de l'expérimentation et expliquer si possible les écarts en fonction des points précédents ;
- mesurer l'impact des relances courriers et mail. Le protocole (décrit ci-dessous) prévoyait en effet qu'en l'absence de réponse plusieurs types de relances soient effectués à différents moments de la période de collecte.

Les indicateurs d'intérêt sont les mêmes que ceux analysés à partir de l'enquête CVS classique. Le questionnaire en réinterrogation a toutefois été allégé pour permettre l'interrogation par téléphone ou internet,

³ Se référer par exemple au compte-rendu de la séance du Cnis du 4 octobre 2016 (commission « services publics et services aux publics »)

donc sur une durée de questionnement plus courte. Ainsi, les modules du questionnaire qui ne se déclenchent que si l'enquêté a signalé une victimation ont été allégés pour ne garder que quelques informations sur la description de la victimation. Beaucoup d'éléments de contexte et ceux portant sur les conséquences ont été supprimés par rapport au questionnaire classique.

Plus précisément, l'expérimentation CVS collecte des données sur :

- les cambriolages ou vols sans effraction d'un logement ;
- les vols ou tentatives de vol de voiture, de deux roues ou de vélo ;
- les actes de vandalisme contre le logement ou la voiture ;
- les escroqueries bancaires ;
- les vols personnels (ou tentatives) avec violences ou menaces ;
- les vols personnels (ou tentatives) sans violences ou menaces ;
- les violences physiques ;
- les menaces ;
- les injures ou insultes ;
- le cadre de vie et sécurité (perception du risque de victimation, environnement, sentiment d'insécurité) ;
- les violences sexuelles hors ménages (pour les individus de 18-75 ans).

En 2018, l'Insee décide de ne plus réaliser l'enquête CVS dans sa configuration d'alors. Cette décision, qui conduira au transfert des enquêtes de victimation au SSMSI à compter de 2022, n'a pas remis en cause la conduite de l'expérimentation. Au contraire, en vue de la refonte du dispositif mais également pour avancer sur les expérimentations multimodes mises en place par l'Insee, il était important de maintenir cette expérimentation. En parallèle des travaux d'analyse sur l'expérimentation CVS panel multimode, le SSMSI a commencé à définir le contour du futur dispositif d'enquête sur les victimations aboutissant à la mise en place en 2022 de l'enquête Vécu et Ressenti en matière de sécurité (VRS)⁴. La mise en place d'un panel a vite été écartée amenant *de facto* à revoir les priorités en termes d'analyse des résultats de l'expérimentation panel. Ainsi, le 3^e objectif visant à mesurer les corrélations temporelles et étudier la pertinence d'un dispositif « panel » pour améliorer la précision des indicateurs conjoncturels, a été dépriorisé. Seules les corrélations ont été calculées ; les analyses de gain de précision sur les indicateurs conjoncturels n'ont pas été menées. Ces résultats ne sont pas détaillés dans ce document de travail.

1.4. Protocole de l'expérimentation panel multimode

Comme indiqué plus haut, le mode de collecte de l'expérimentation panel est différent de celui de l'enquête classique (voir annexe 1 pour une description du protocole de l'enquête CVS). Les individus panel ont été ré-interrogés par téléphone ou par internet. Principalement pour des raisons de coûts, l'expérimentation n'incluait pas d'échantillon interrogé en face-à-face. Le mode papier n'a pas été envisagé, le questionnaire CVS étant complexe (nombreux filtres).

4 En 2022, l'enquête VRS a donc remplacé les [enquêtes Cadre de vie et sécurité \(CVS\)](#). Entre 2019 et 2021, le SSMSI a été engagé dans le pilotage de cette refonte avec l'appui de l'Insee : concertation avec les administrations, les partenaires sociaux et les chercheurs ; conception des questionnaires ; élaboration du protocole de collecte ; construction du processus d'exploitation et de diffusion. Ces travaux se sont également appuyés sur les [enseignements de l'enquête GENre Et SEcurité \(Genese\)](#) réalisée par le SSMSI en 2021 sur financement européen et dont les résultats méthodologiques ont donné lieu à plusieurs communications/publications : [présentation aux journées de méthodologie statistique de 2022](#), publication dans le numéro 116 de la revue [Statéco](#).

Pour étudier l'effet des modes de collecte et surtout l'impact des différents types de relances (relances mails, téléphoniques, relances effectuées à différents stades de la collecte), la division « Conditions de vie des ménages », avec la division « Recueil et traitement de l'information » et le responsable du programme de développement des enquêtes ménages en multimode, ont proposé de définir trois sous-échantillons en réinterrogation. Ces trois sous-échantillons ont des volumes identiques pour simplifier l'exploitation. Ils sont ainsi constitués chacun de 4 374 répondants 2018.

- Le premier sous-échantillon a été interrogé par téléphone pendant les huit premières semaines. Après cette période, les non-répondants à l'enquête par téléphone ont été invités à répondre sur internet. Les deux modes étaient exclusifs l'un de l'autre.
- Le deuxième sous-échantillon a été d'abord invité à répondre au questionnaire sur internet (sur 4 semaines), puis par téléphone (sur 6 semaines). Les deux modes ont cohabité sur la deuxième partie de la collecte.
- Enfin, le troisième sous-échantillon a également été interrogé par internet (sur 6 semaines) puis par téléphone (4 semaines). Les deux modes ont cohabité sur la deuxième partie de la collecte. La différence entre ce protocole et le précédent résidait dans le temps plus long laissé aux enquêtés pour répondre par internet avant de basculer sur le mode téléphone.

Suite aux recommandations du comité du label, un protocole aléatoire de répartition des individus entre les différents sous-échantillons à probabilités différenciées selon les éléments de contact disponibles a été retenu (table 1).

Table 1 : probabilité d'appartenir à un des trois sous-échantillons selon le type de coordonnées disponibles

	Probabilité d'appartenir au			Part des individus panel ayant fourni ces coordonnées
	Sous-échantillon 1	Sous-échantillon 2	Sous-échantillon 3	
téléphone seul	0,56	0,24	0,20	43,4 %
mail seul	0,18	0,33	0,49	2,3 %
téléphone + mail	0,18	0,44	0,38	41,6 %
Aucune coordonnée	0,10	0,30	0,60	12,6 %

Lecture : 43,4 % des individus ayant répondu à l'enquête CVS 2018 ont fourni un numéro de téléphone en vue de l'enquête expérimentale de 2019. Ces individus ont une probabilité de 0,56 d'appartenir au sous-échantillon 1.
Source : Enquêtes CVS 2018 et expérimentale CVS panel 2019.

Les enquêtés sont tous contactés par une lettre-avis dont le contenu varie selon le sous-échantillon : elle les informe qu'un enquêteur les contactera bientôt par téléphone ou bien comporte l'adresse internet sécurisée permettant de répondre à l'enquête ainsi qu'un identifiant et un mot de passe de connexion. La lettre-avis est doublée, pour les enquêtés des sous-échantillon 2 et 3 ayant fourni un mail, d'un courrier électronique avis.

Un protocole de relance distinct a été établi pour les 3 sous-échantillons (figure 1).

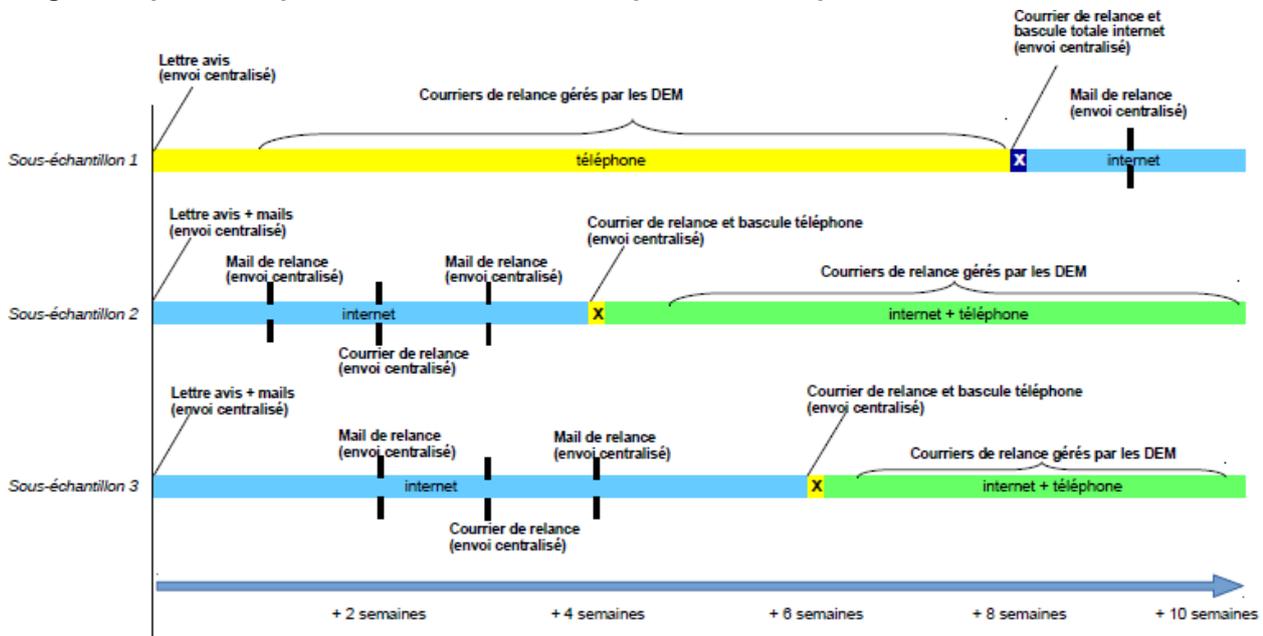
Pour le premier sous-échantillon, la relance se fait dans un premier temps par courrier postal. Cette relance est gérée par les divisions Enquêtes Ménages (DEM), chargées d'organiser la logistique de la collecte au sein des directions régionales de l'Insee, sur demande de l'enquêteur. En parallèle, l'enquêteur essaie toujours de joindre l'individu à partir du numéro que ce dernier avait fourni lors de la collecte 2018. En fin de période de collecte téléphone (8 semaines après le début de la collecte), une dernière relance courrier qui contient cette fois un identifiant et un mot de passe pour renseigner l'enquête *via* internet est effectuée. Ce

courrier est « doublé » d'un mail pour les personnes ayant communiqué une adresse électronique. La bascule téléphone vers internet est globale pour ce sous-échantillon, i.e. les enquêteurs n'essaient plus de joindre les enquêtés par téléphone⁵.

Pour le sous-échantillon 2, un mail contenant les identifiants de connexion est envoyé quelques jours après l'envoi de la lettre-avis papier. Les relances se font alternativement par mail et papier à une semaine d'intervalle. Une semaine après le début de la collecte, un premier mail de relance est envoyé. Deux semaines après le début de la collecte, un courrier de relance papier est adressé à l'enquêté. À trois semaines du début de la collecte, un nouveau mail de relance est adressé à l'enquêté. Enfin, si au bout d'un mois de collecte le questionnaire n'a pas été renseigné, l'enquêté est alors contacté par téléphone pour passer le questionnaire *via* ce mode. Il peut cependant encore répondre par internet jusqu'à la fin de l'enquête.

Enfin, le troisième sous-échantillon suit un protocole similaire au sous-échantillon 2 mais avec un calendrier différent. Les relances mails sont faites 15 jours et 30 jours après le début de la collecte. Les relances papier ont lieu au bout de trois et six semaines de collecte. À partir de la 7^e semaine de collecte, les enquêteurs entrent en jeu et cherchent à joindre les non-répondants pour leur faire passer l'enquête par téléphone. Les enquêtés peuvent toujours répondre par internet jusqu'à la fin de la collecte.

Figure 1 : protocole par sous-échantillon de l'expérimentation panel multimode CVS 2019



Pour les enquêtés répondant au questionnaire par internet, le choix d'un accès à la fois souple mais contrôlé a été fait. Quel que soit son sous-échantillon d'appartenance, l'enquêté répondant par internet peut, avant validation, interrompre et reprendre le questionnaire à sa convenance. Après validation, l'ensemble des questions, en particulier celles sur les violences sexuelles hors ménage, sont inaccessibles que ce soit pour l'enquêté ou une autre personne qui se connecterait avec les identifiants de l'enquêté. Lorsqu'une personne se reconnecte, elle a accès à une page de résultats statistiques disponibles sur le thème du cadre de vie et de la sécurité. Ce verrouillage a été mis en place pour garantir la protection de la personne répondante.

5 C'est ce qui est fait dans l'enquête TIC téléphone. Après 8 semaines de collecte, on peut considérer que si l'enquêteur n'a pas réussi à contacter par téléphone l'enquêté, il y a peu de chances qu'il y parvienne sur les dernières semaines (TIC recommande 20 essais de contacts), d'où le choix d'une bascule totale. Celle-ci permet en outre de simplifier la gestion des quotités enquêteurs : à partir de 8 semaines, ils ne travaillent plus que sur les sous-échantillons 2 et 3.

2. L'impact du protocole sur la participation

Le premier objectif de l'expérimentation consiste à évaluer quel serait le protocole multimode le plus adapté en matière de taux de réponse et de représentativité des répondants pour une telle enquête sur les victimations. Pour cela, on compare les répondants des trois sous-échantillons.

2.1. Bilan de la collecte

Taux de collecte et taux de réussite

Sur les 13 123 individus panel, 10 271 ont répondu à cette enquête expérimentale. 4 781 enquêtes ont été renseignées par internet alors que 5 490 l'ont été par téléphone. Le taux de collecte s'élève donc à 78 %. Pour mémoire, les taux de collecte de l'enquête CVS depuis 2014 sont compris entre 65 % (en 2019) et 71 % (en 2015).

Les taux de collecte sont meilleurs lorsqu'en 2018 les enquêtés ont fourni au moins un numéro de téléphone et une adresse courriel (87 %). Les taux de collecte sont très proches pour les enquêtés ayant donné soit un numéro de téléphone soit une adresse courriel en 2018 (respectivement 77 % et 75 %). À noter que 55 % des individus panel pour lesquels aucune coordonnée, ni mail ni téléphone, n'avait été récupérée ont finalement participé à l'enquête expérimentale.

Les taux de collecte sont légèrement différents entre les trois sous-échantillons : 78 % pour le sous-échantillon 1, 81 % pour le 2 et 76 % pour le 3 (table 2).

Ces différences de taux ne permettent toutefois pas de conclure sur une meilleure efficacité du sous-échantillon 2. L'attribution à un des trois sous-échantillons a en effet été faite aléatoirement mais à probabilités inégales suivant le type de coordonnées fournies. Ainsi le sous-échantillon 3 sur-représente les individus n'ayant donné aucune coordonnée.

Pour gommer ces différences de structure, un taux de collecte à structure constante est calculé à partir des taux de collecte observés sur 3 sous-échantillons par type de coordonnées récupérées en 2018 et de la structure par type de coordonnées soit observé sur le sous-échantillon 1 soit observé sur l'ensemble de l'échantillon de réinterrogation. Dans les deux cas, les taux de collecte observés sur les sous-échantillons avec le mode internet puis téléphone (sous-échantillon 2 et 3) sont légèrement supérieurs à ceux observés sur le sous-échantillon avec le mode téléphone puis internet (sous-échantillon 1).

Table 2 : taux de collecte par sous-échantillon et type de coordonnées récupérées en 2018

	Sous-échantillon 1	Sous-échantillon 2	Sous-échantillon 3	Ensemble
Ensemble	77,7 %	80,7 %	76,4 %	78,3 %
téléphone + courriel	86,3 %	86,6 %	87,1 %	86,7 %
téléphone uniquement	76,6 %	78,9 %	76,8 %	77,2 %
courriel uniquement	74,0 %	76,0 %	74,7 %	75,0 %
Aucune coordonnée	47,4 %	57,8 %	54,0 %	54,5 %
Taux de collecte avec structure de type de coordonnées du sous-échantillon 1	77,7 %	79,8 %	79,9 %	
Taux de collecte avec structure de type de coordonnées de l'échantillon total	76,9 %	79,4 %	78,2 %	

Source : Enquête expérimentale CVS panel 2019.

Étant donné le protocole retenu, la majorité des questionnaires téléphone sont issus du sous-échantillon 1 (3 098 questionnaires du sous-échantillon 1 contre 1 314 pour le sous-échantillon 2 et 1 078 pour le sous-échantillon 3). De manière analogue, les sous-échantillons 2 et 3 représentent 94 % des questionnaires internet. Pour le sous-échantillon 1, 91 % des questionnaires reçus ont été réalisés par téléphone. Pour les sous-échantillon 2 et 3, c'est le mode internet qui a permis de recueillir le plus de questionnaires. Plus précisément 63 % des questionnaires du sous-échantillon 2 ont été renseignés par internet et 68 % du sous-échantillon 3⁶.

Réception des questionnaires, effet des relances

Un des objectifs de l'expérimentation est de comparer l'efficacité des relances entre le calendrier défini pour le sous-échantillon 2 et celui défini pour le sous-échantillon 3.

Un courriel de relance a été envoyé aux individus du sous-échantillon 2 (pour lesquels on avait une adresse courriel) une semaine après le début de la collecte. 15 jours après le début de la collecte, c'est un courrier qui leur a été adressé. Un deuxième courriel de relance a été envoyé trois semaines après le début de la collecte. Enfin la lettre de bascule leur indiquant qu'un enquêteur les contacterait pour réaliser l'enquête a été envoyée 1 mois après le début de la collecte. Pour le sous-échantillon 3, le protocole de relance était le même (envoi d'un courriel puis une semaine après courrier puis une semaine après courriel) mais le premier courriel a été envoyé 2 semaines après le début de la collecte. Le courrier de bascule a également été envoyé deux semaines après le deuxième courriel de relance.

Les lettres ont mécaniquement un effet plus important⁷ (figure 2). Les courriels semblent plus efficaces sur le sous-échantillon 2 mais cela s'explique *a priori* par l'affectation à probabilités inégales des individus panel à un des trois sous-échantillons (sur-représentation des individus sans téléphone ni courriel dans le sous-échantillon 3).

En distinguant les individus qui avaient fourni une adresse courriel de ceux qui n'en avaient pas fourni, la figure 3 permet de comparer les sous-échantillons 2 et 3. La lettre-avis courrier était « doublée » d'un courriel annonçant le début de la collecte lorsque l'individu panel avait donné une adresse courriel en 2018. Cette double annonce incite l'enquêté à répondre plus rapidement : un quart des questionnaires internet des sous-échantillons 2 et 3 étaient rentrés après 5-6 jours de collecte pour les individus ayant fourni une adresse courriel en 2018 (contre 11-12 jours pour les individus n'ayant pas fourni d'adresse mail). L'effet de la première relance (par mail) est visible au jour 7 pour le sous-échantillon 2 et au jour 15 pour le sous-échantillon 3. Il est plus marqué pour le sous-échantillon 2 que pour le sous-échantillon 3. L'effet de la deuxième relance (par courrier) est visible au jour 16 pour le sous-échantillon 2 et au jour 22 pour le sous-échantillon 3. Cette fois-ci, l'effet est très légèrement plus marqué pour le sous-échantillon 3. L'effet de la troisième relance (par mail) est quant à lui très peu visible. En revanche, la lettre de bascule semble avoir eu un impact pour les individus n'ayant pas donné d'adresse mail (jour 30 pour le sous-échantillon 2 et jour 44 pour le sous-échantillon 3). On note en effet une légère hausse des répondants par internet pour ces individus. Ces personnes semblent donc plus enclines à répondre sur internet plutôt qu'au téléphone. En figure 4, la distinction détaillée des types de coordonnées connues confirme cette impression initiale : disposer d'un numéro de téléphone n'accroît que très marginalement la vitesse de collecte dans les phases internet des échantillons 2 et 3. Ainsi, se rendre volontairement plus joignable en fournissant son numéro de téléphone (en plus de l'adresse courriel) joue très marginalement lors de la réinterrogation *via* internet.

Au final, les relances sont efficaces et doubler courrier et courriel permet d'accélérer la collecte. Faire les relances le plus tôt possible pourrait permettre d'augmenter un peu le taux de collecte, sur une période donnée, si l'on se fie à la comparaison entre sous-échantillon 2 et sous-échantillon 3.

⁶ Le détail du nombre de questionnaires reçus suivant le mode et le sous-échantillon figure en annexe 2.

⁷ Une adresse postale était *a priori* disponible pour tous les individus panel mais une adresse courriel n'était disponible que pour 58 % du sous-échantillon 2 et 51 % du sous-échantillon 3.

Figure 2 : nombre de questionnaires internet par date de réception

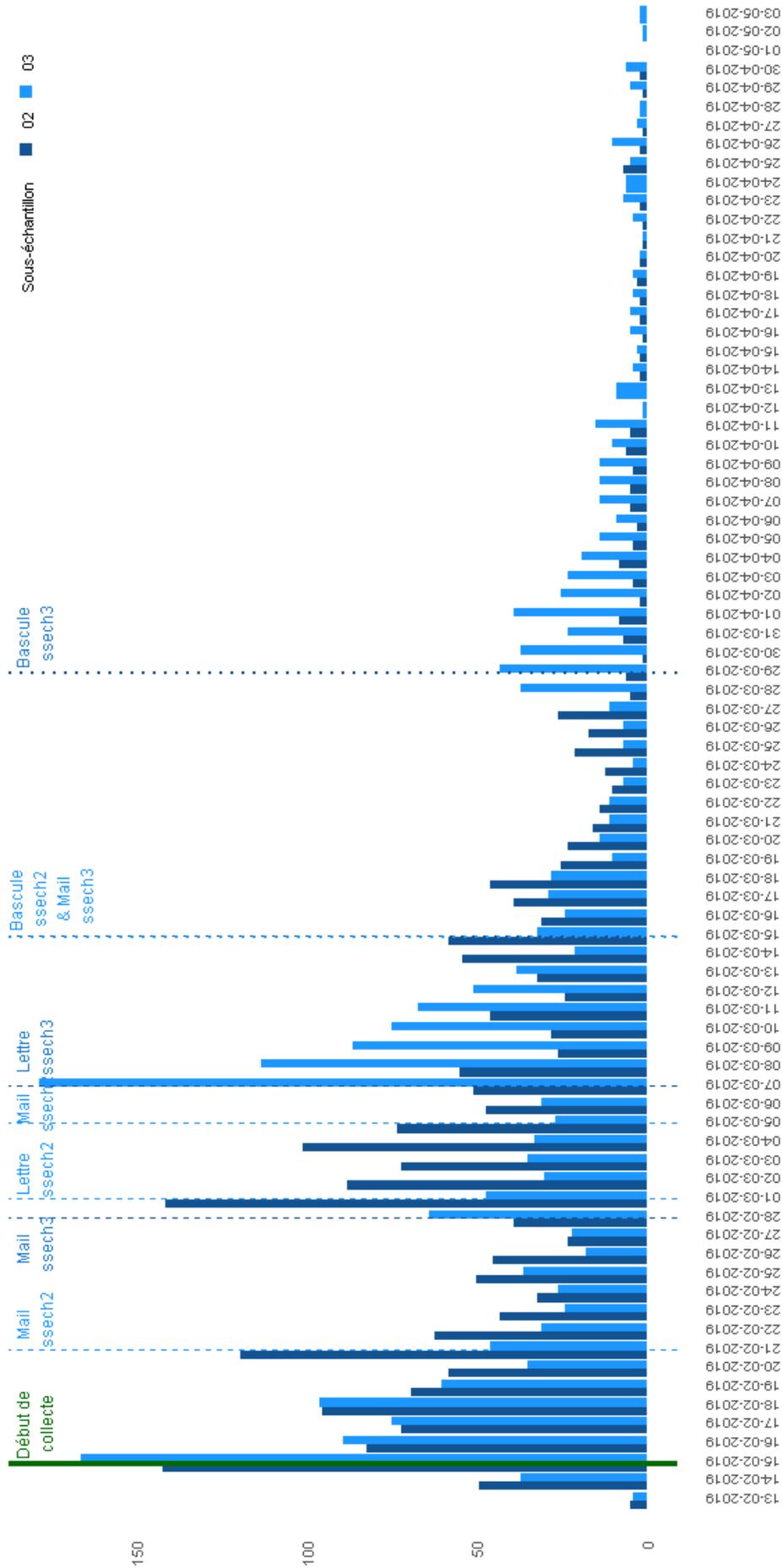
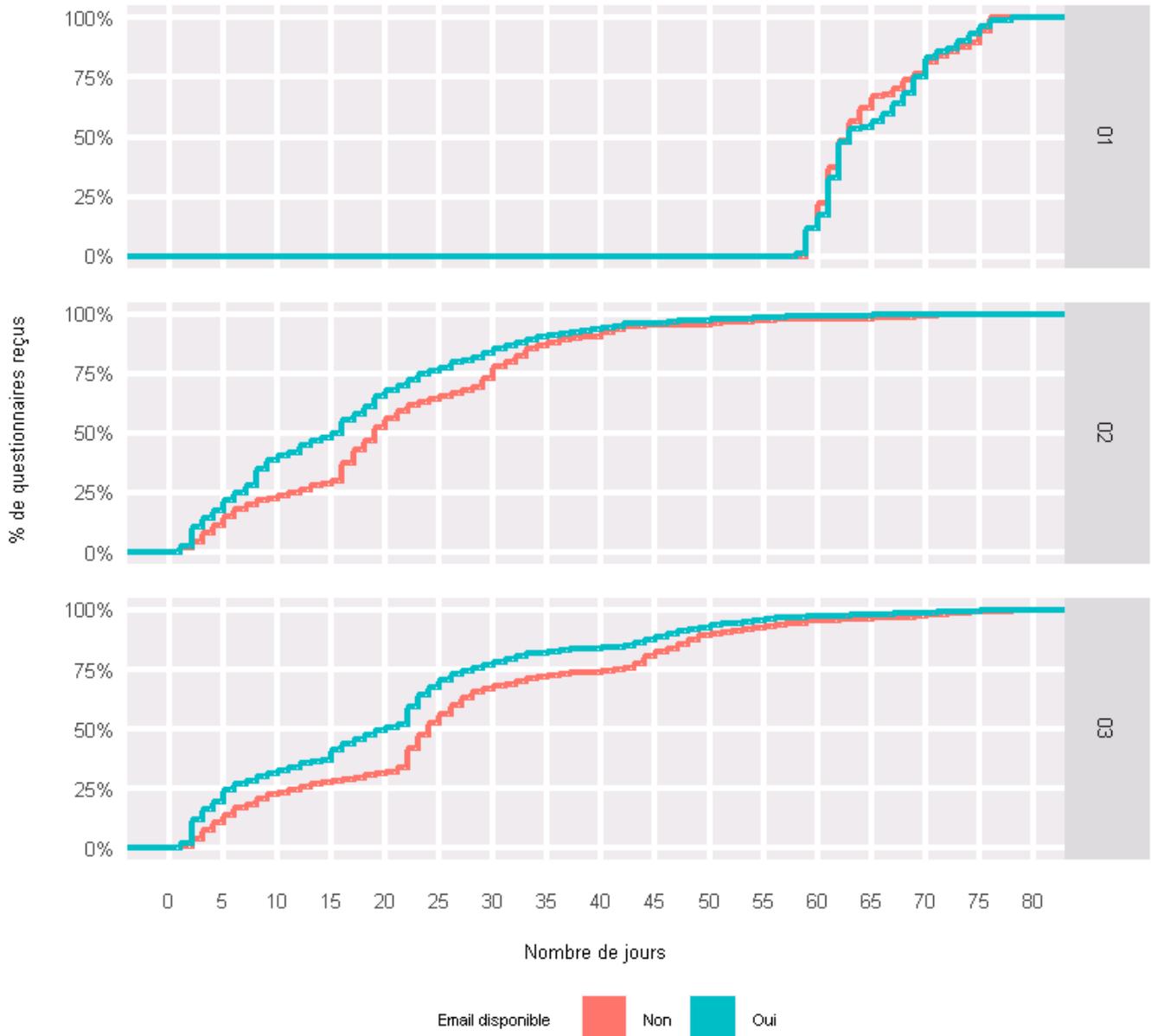
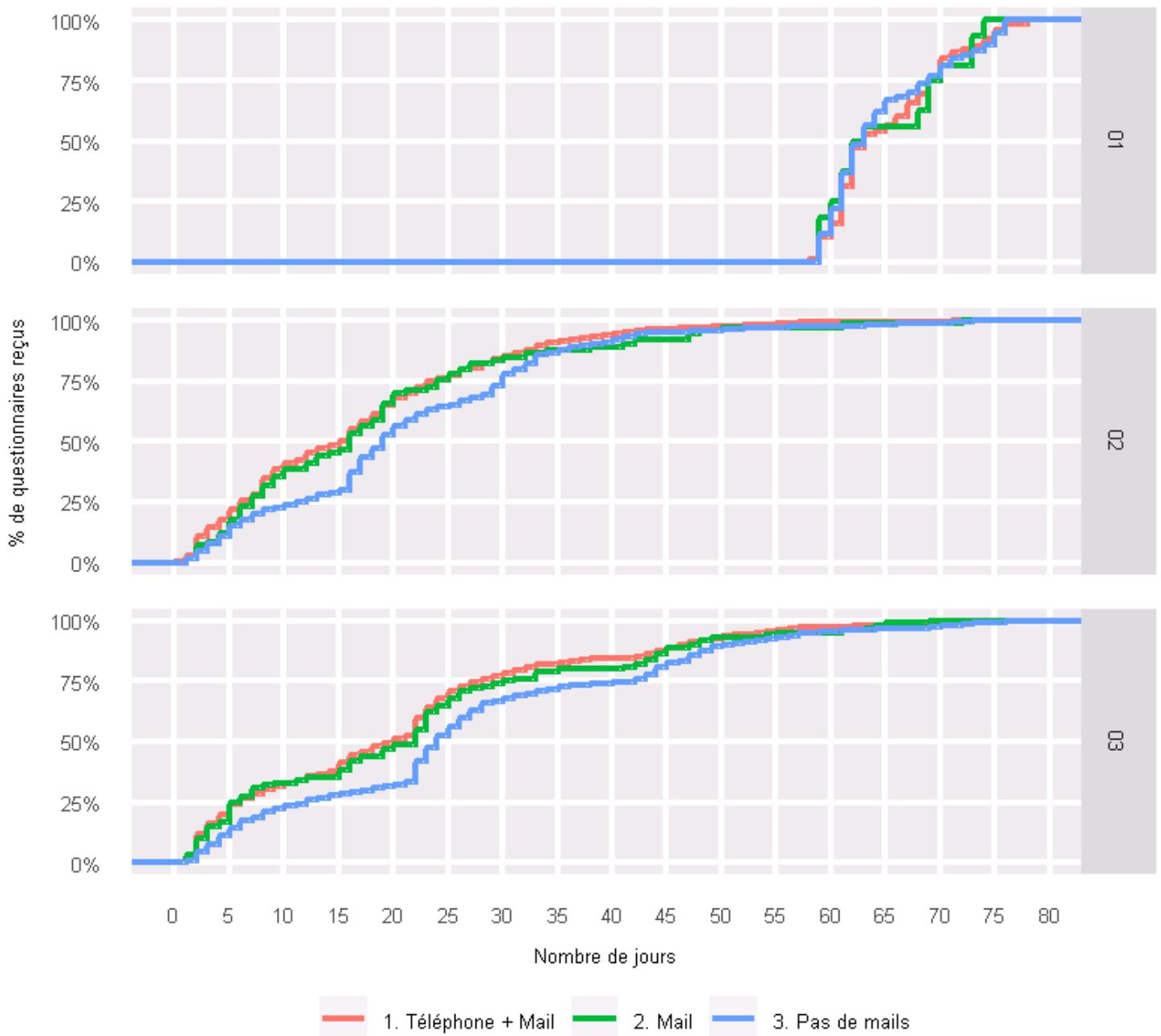


Figure 3 : fonction de répartition des questionnaires internet reçus par sous-échantillon, suivant le nombre de jours de collecte et la disponibilité d'une adresse e-mail



Source : Enquête expérimentale CVS panel 2019.

Figure 4 : fonction de répartition des questionnaires internet reçus par sous-échantillon suivant le nombre de jours de collecte et les coordonnées disponibles



Source : Enquête expérimentale CVS panel 2019.

Non-réponse partielle

Si l'enquêté ne valide pas son questionnaire sur internet, celui-ci n'est pas réceptionné et l'enquêté est considéré comme non répondant. 119 enquêtes ont été initiées sur internet et n'ont pas été finalisées (soit moins de 2,5 % des répondants internet ou 0,9 % de l'échantillon panel).

Un quart des questionnaires internet non finalisés est abandonné dès les toutes premières questions. Le début de l'enquête rappelant le prénom de l'enquêté, on peut faire l'hypothèse que ce n'est pas l'individu Kish (c'est-à-dire l'individu qui a été tiré au hasard pour répondre au questionnaire de l'enquête parmi les

membres du ménage éligibles pour cette enquête⁸) qui s'est connecté : l'individu se rend alors compte que l'enquête ne s'adresse pas à lui. Un tiers des questionnaires est arrêté dans le bloc final de validation .constitué de trois questions et du bouton de validation :

- Est-ce que l'enquêté a répondu seul au questionnaire ?
- Y a-t-il d'autres faits dont il a été victime et qu'il aurait souhaité aborder ?
- Commentaire libre ;
- Bouton de validation.

La majorité des individus ayant stoppé le questionnaire dans ce bloc se sont arrêtés juste avant le bouton de validation. La formulation « Ce questionnaire est maintenant terminé... » les a probablement induits en erreur.

Il est intéressant de noter que seul un questionnaire internet a été interrompu au niveau des questions des violences sexuelles hors ménage (table 3). Cette section, pressentie comme sensible, ne semble donc pas être particulièrement cause de non-réponse partielle.

Table 3 : répartition des questionnaires internet non finalisés selon la partie du questionnaire où l'internaute s'est arrêté

Arrêt	Nombre de questionnaires	Pourcentage
Introduction	27	22,5 %
Tableau des habitants du logement	14	12 %
Questions de victimation « ménage »	18	15 %
Questions de victimation « individu »	21	17,5 %
Questions de victimation « sensibles »	1	1 %
Validation	38	32 %
Total	119	100 %

Champ : questionnaires internet non finalisés.

Source : Enquête expérimentale CVS panel 2019.

Parmi ces 119 enquêtes initiées sur internet, la moitié ont finalement donné lieu à une enquête téléphonique.

Concernant les questionnaires réalisés par téléphone, 34 questionnaires ont été commencés mais n'ont pas été finalisés. Cela représente moins de 1 % des questionnaires téléphoniques (le taux d'enquête partielle sur CVS classique s'élève à 1,7 %). Ces 34 cas couvrent des situations très différentes : abandon au tout début du questionnaire, abandon vers la fin du questionnaire, erreur dans la programmation du code résultat.

Vérification de l'identité de l'individu panel (kish)

Dans le cadre de cette expérimentation panel, il était primordial que ce soit l'individu kish, sélectionné pour répondre à l'enquête CVS 2018, qui réponde à nouveau en 2019. À la fin de l'enquête CVS 2018, il était demandé aux répondants d'accepter ou de refuser que leurs réponses à l'enquête 2018 soient utilisées pour raccourcir la passation de l'enquête expérimentale en 2019.

⁸ Pour l'enquête CVS, l'individu tiré au sort est celui dont la date d'anniversaire est la plus proche du 1^{er} janvier lors de l'année d'enquête.

Si le mode téléphone ne peut assurer dans la totalité des cas la bonne identité de la personne devant être enquêté, ce mode, *via* l'intermédiation d'un enquêteur, offre toutefois plus de garanties que le mode internet. Le prénom et le nom de l'enquêté était toutefois rappelé au début du questionnaire internet (figure 5).

Figure 5 : question de vérification de l'identité de l'individu panel sur internet

Ce sont vos réponses qui font nos statistiques. Pour cette enquête, il est important que ce soit la personne enquêtée en 2018 qui participe à ce second et dernier questionnaire.

Êtes-vous « **V1Prénom V1Nom** » ?

1. Oui
2. Non mais « **V1Prénom V1Nom** » habite avec vous et pourra répondre plus tard
3. Non « **V1Prénom V1Nom** » n'habite plus dans ce logement

Source : Extrait du questionnaire de l'enquête expérimentale CVS panel 2019.

L'analyse des données collectées a mis en lumière des changements de kish. Deux contrôles différents ont été réalisés suivant l'acceptation de la remontée des données antérieures par le kish en 2018.

Pour les individus qui avaient refusé en 2018 la remontée de données antérieures, la vérification a consisté en la comparaison des prénoms, du sexe et de la date de naissance de l'individu Kish, renseignées en 2018 et à nouveau en 2019. 45 « changements » de Kish ont été ainsi repérés. Ces cas ont été traités manuellement en comparant la composition du ménage sur les 2 années, ainsi que les dates de naissance. Dans un seul cas, il a été identifié qu'il s'agissait de la même personne. La plupart des autres cas sont des couples dont l'un des conjoints a complété l'enquête en 2018 et le second en 2019 : ils ont été recodés selon un nouveau code déchet. À noter qu'il s'agit à 80 % de questionnaires collectés par téléphone.

Pour les individus qui avaient accepté la remontée de données antérieures, la vérification sur les prénoms, sexe et date de naissance n'étaient pas possibles car il n'était pas demandé en 2019 de renseigner à nouveau ces informations. En revanche, dans le tableau des habitants du logement allégé, l'enquêté devait indiquer quels individus présents en 2018 ne faisaient plus partie du ménage en 2019. 77 « départs » de Kish ont été ainsi identifiés : ils ont également été recodés selon un nouveau code déchet. À noter que l'intégralité de ces questionnaires ont été collectés par internet.

Au total, 237 questionnaires sont éliminés lors de l'apurement, soit seulement 2,3 % des fiches-adresses (FA) collectées lors de cette expérimentation. Au final, ce sont donc 10 034 questionnaires qui sont conservés pour l'exploitation. Les taux de collecte calculés après apurement sont inférieurs de 1,4 point (sous-échantillon 1) ou 2 points (sous-échantillon 2 ou 3) à ceux observés avant apurement (table 4).

Table 4 : taux de collecte par sous-échantillon avant et après apurement

	Taux de collecte avant apurement	Taux de collecte après apurement
Sous-échantillon 1	77,7 %	76,3 %
Sous-échantillon 2	80,7 %	78,7 %
Sous-échantillon 3	76,4 %	74,4 %
Ensemble	78,3 %	76,5 %

Source : Enquête expérimentale CVS panel 2019.

2.2. Représentativité des répondants

L'effet de mode peut se traduire par un biais de sélection, lié à une auto-sélection différente des répondants selon les modes de collecte sur lesquels ils sont interrogés. Nous allons voir dans cette partie quel est l'impact des différents protocoles sur la représentativité des répondants, en prenant en compte tout d'abord la sélection sur observables puis celle, potentielle, dépendant de facteurs non observables.

Les indicateurs utilisés pour la sélection sur observables

Dans un premier temps, on va estimer l'effet de sélection observable. Pour cela, on va étudier la représentativité des différents sous-échantillons à partir de deux indicateurs.

Les différences standardisées permettent de quantifier les différences, pour une variable donnée, entre deux échantillons a et b sans tenir compte de la taille de ces échantillons, contrairement aux statistiques usuelles. En général, on considère une différence standardisée comme importante si elle dépasse 10 %.

On note m_e la moyenne, S_e l'écart-type et p_e la proportion pour les échantillons $e \in \{a, b\}$. Dans le cas de variables continues, la formule est la suivante :

$$d = \frac{100 * (m_a - m_b)}{\sqrt{\frac{S_a^2 + S_b^2}{2}}}$$

Dans le cas de variables catégorielles, on fait la moyenne en valeur absolue des différences standardisées de chaque modalité⁹. La formule pour une variable binaire est la suivante :

$$d_i = \frac{100 * (p_a - p_b)}{\sqrt{\frac{p_a(1-p_a) + p_b(1-p_b)}{2}}}$$

Le R indicateur est une mesure multivariée de similitude entre répondants et population totale. Il s'agit d'une mesure de la dispersion des propensions à répondre. L'indicateur est compris entre 0 et 1. Plus il est proche de 1, plus l'échantillon des répondants est représentatif de la population totale. Le R indicateur se calcule de la façon suivante : $R = 1 - 2S(\theta)$, avec θ la propension individuelle estimée à répondre et $S(\theta)$ l'écart-type de la propension à répondre.

Les variables prises en compte dans cette analyse de la représentativité sont issues de l'enquête CVS 2018, habituellement utilisées dans les redressements de la non-réponse :

- Le type de logement (individuel/collectif) ;
- La taille d'unité urbaine du logement (classée en 4 catégories) ;
- La composition du ménage (personne seule/famille monoparentale/couple sans enfant/couple avec enfant/autre) ;
- Le sexe de l'individu ;
- L'âge de l'individu ;
- Le plus haut niveau de diplôme de l'individu (supérieur/bac ou bac+2/CAP-BEPC/sans diplôme ou non déclaré) ;
- La nationalité de l'individu (française ou étrangère) ;
- Le statut d'activité de l'individu (actif occupé/chômeur/inactif) ;
- La catégorie socio-professionnelle de l'individu (classée en 7 catégories).

On prend également en compte le type de coordonnées fournies en 2018 (téléphone+mail / téléphone / mail / aucun des deux) et le fait d'avoir accepté la remontée de données antérieures.

⁹ Pour les variables catégorielles (type de ménage, taille d'unité urbaine, activité et catégorie socioprofessionnelle), la différence standardisée présentée correspond à la moyenne en valeur absolue des différences standardisées de chaque modalité. Contrairement aux autres variables, il n'y a donc pas de sens qui puisse être interprété.

La sélection sur observables dans CVS panel selon le protocole

La structure des répondants à l'enquête CVS panel est assez proche de celle de l'échantillon, qui correspond aux répondants de CVS 2018 (table 5). Derrière ce résultat, il y a cependant de fortes disparités entre les répondants internet et les répondants téléphone, en lien avec les différents protocoles et les comportements de réponse en fonction du mode. Ainsi, les répondants internet sont nettement plus diplômés que les répondants téléphone (58 % ont un diplôme de niveau baccalauréat ou supérieur, contre 42 % des répondants téléphone), plus jeunes (50 ans en moyenne, contre 56 ans), plus souvent en couple avec des enfants (un tiers d'entre eux). Ils sont également plus souvent actifs occupés (56 %, contre 42 % des répondants téléphone) et cadres ou de profession intermédiaire (45 %, contre 32%).

Le taux de collecte étant très élevé, il semble que les phases téléphoniques ont permis de compléter les phases internet et que l'ensemble des répondants est assez similaire à l'échantillon. Comme point de comparaison, on peut également mettre en parallèle les répondants au panel avec l'échantillon initial de l'enquête CVS 2018, représentatif de la population en 2018. Pour cela, on utilise les informations issues de la base de sondage, qui correspondent aux variables de calage et de correction de la non-réponse¹⁰. Les répondants à l'enquête CVS panel sont nettement plus différents de l'échantillon initial de CVS 2018. Les principales différences observées proviennent donc de la non-réponse à CVS 2018. Par exemple, les ménages résidant dans un logement individuel représentent 57 % de l'échantillon initial de l'enquête CVS 2018, alors qu'ils représentent 64 % des répondants à CVS 2018 (échantillon de l'enquête CVS panel) et 67 % des répondants à CVS panel.

Table 5 : structure des répondants à CVS panel (en %)

Variable	Modalité	Échantillon CVS Panel	Répondants CVS Panel	Répondants téléphone CVS Panel	Répondants internet CVS Panel
Type de logement	Individuel	64	67	66	67
	Collectif ou autres	36	33	34	33
Type de ménage	Personne seule	34	32	37	26
	Famille monoparentale	9	8	9	8
	Couple sans enfant	30	31	30	32
	Couple avec enfant(s)	26	27	22	32
	Autres	2	2	2	2
Tranche d'unité urbaine de la commune	0 – Commune rurale	23	25	25	24
	1 à 5 – UU de 2 000 à moins de 100 000 habitants	32	32	34	30
	6 et 7 – UU de 100 000 à moins de 2 000 000 habitants	32	33	33	33
	8 – Agglomération de Paris	12	11	9	13
Sexe	Homme	45	45	43	47
	Femme	55	55	57	53
Age moyen	3 ^e quintile	53	53	56	50
Diplôme	Diplôme du supérieur	17	18	14	22
	Bac ou bac+2	30	32	28	36
	CAP ou BEPC	31	31	33	30
	Sans diplôme ou non déclaré	22	19	26	12
Nationalité	Etrangère	5	4	5	4
	Française	95	96	95	96
Activité	Actif occupé	47	48	42	56
	Chômeur	6	6	6	5
	Inactif	47	46	52	39
Catégorie socio-professionnelle	Agriculteurs	2	2	3	1
	Commerçant ou chef d'entreprise	6	6	6	5
	Profession libérale ou cadre	15	16	12	20
	Profession intermédiaire	21	22	20	25
	Employé	26	27	29	23
	Ouvrier	19	17	20	15
	Autres	11	10	10	11
Nombres d'individus		13123	10034	5379	4655

Source : Enquête expérimentale CVS panel 2019.

¹⁰ La taille d'unité urbaine, le nombre de pièces du logement, le nombre de personnes habitant dans le logement, le statut d'occupation, l'âge et sexe croisés de la personne de référence (PR), la situation d'activité de la PR, le diplôme de la PR, la nationalité de la PR.

Ces différences entre les répondants à l'enquête CVS panel par mode et l'échantillon total de CVS panel peuvent cacher à la fois des différences liées à la structure des échantillons (ceux-ci sont tirés aléatoirement mais avec des probabilités inégales) et des différences de comportement de réponse selon les modes de collecte. Pour différencier ces effets, on s'intéresse ici à la comparaison des répondants et des échantillons pour chacun des sous-échantillons, ce qui permet d'étudier le seul impact des comportements de réponse par mode.

Les répondants au sous-échantillon 1, ayant principalement répondu par téléphone, sont plus diplômés, plus souvent actifs occupés et de nationalité française que l'ensemble des personnes appartenant initialement au sous-échantillon 1 (table 6) même si ces particularités sont limitées puisque le R-indicateur s'élève à 0,80. L'ensemble des répondants au sous-échantillon 1 est relativement similaire aux seuls répondants par téléphone. Proposer internet en deuxième intention ne permet pas de rattraper des profils très particuliers et complémentaires aux répondants téléphone. Ainsi le R-indicateur des seuls répondants téléphone est le même que celui de l'ensemble des répondants (0,80). Quand on regarde plus précisément au niveau des variables, on remarque qu'internet permet cependant de récupérer des personnes vivant en agglomération parisienne : alors que seulement 9 % des répondants téléphone vivent en agglomération parisienne, c'est le cas de 24 % des répondants internet (la proportion dans l'échantillon étant de 12 %). Le mode internet permet également d'avoir plus de répondants hommes et de rajeunir l'ensemble des répondants. Cependant, les répondants internet sont encore plus diplômés et plus souvent actifs occupés, des profils qui répondent déjà davantage au téléphone.

Les répondants au sous-échantillon 2 sont plus diplômés et vivent plus souvent en maison que l'ensemble des personnes échantillonnées pour répondre au sous-échantillon 2. Cependant, les répondants téléphone en deuxième intention contribuent à augmenter nettement la représentativité de l'ensemble des répondants. De fait, les répondants internet sont très particuliers : très souvent plus diplômés, beaucoup plus jeunes, très souvent actifs occupés, plus souvent cadres ou de profession intermédiaire, plus souvent en couple, etc. L'ensemble des différences standardisées diminue quand on ajoute les répondants téléphone et le R-indicateur montre que globalement l'ensemble des répondants est nettement plus représentatif que les seuls répondants par internet. Ces résultats montrent que le fait de proposer le téléphone comme mode complémentaire apparaît très efficace pour améliorer la représentativité.

On trouve des résultats similaires pour le sous-échantillon 3, qui diffère du sous-échantillon 2 par une part un peu moins importante de répondants téléphone, du fait d'une bascule plus tardive. Les répondants du sous-échantillon 2 sont finalement plus proches de la structure de leur échantillon que ne le sont les répondants du sous-échantillon 3 : le R-indicateur est plus important dans le cas du sous-échantillon 2 que du sous-échantillon 3, alors que les répondants internet sont très similaires. Laisser plus de temps au mode internet ne fait gagner qu'une cinquantaine de répondants supplémentaires alors que laisser plus de temps au mode téléphone fait gagner près de 200 répondants. Ces résultats montrent que proposer le téléphone comme mode alternatif et laisser du temps pour que les personnes puissent répondre sous ce mode semble donc efficace en termes de représentativité et d'effectifs.

Les résultats de l'expérimentation CVS panel montrent qu'il est important de compléter une collecte internet par un mode de collecte intermédiaire, ici le téléphone. De fait, les répondants internet sont spécifiques et non représentatifs de l'ensemble de la population. Il est également important de laisser suffisamment de temps à la collecte téléphone alors qu'une période de collecte longue par internet apporte peu pour récolter des répondants avec des profils différents.

Par ailleurs, pour la suite des exploitations, et notamment la comparaison à CVS 2019, il est intéressant de noter que la représentativité des répondants à l'expérimentation CVS panel est très proche de celle des répondants à CVS 2018.

Table 6 : Indicateurs de représentativité pour chaque sous-échantillon¹¹

	Par rapport au sous-échantillon 1		Par rapport au sous-échantillon 2		Par rapport au sous-échantillon 3	
	Répondants 1 ^{er} mode (téléphone)	Total répondants	Répondants 1 ^{er} mode (internet)	Total répondants	Répondants 1 ^{er} mode (internet)	Total répondants
Effectif	3048	3336	2156	3441	2211	3257
Différences standardisées (a)						
Type de logement (<i>maison</i>)	3,9	3,4	7,1	5,8	7,6	5,3
Tranche d'unité urbaine (<i>4 cat.</i>)	3,8	2,5	2,1	1,2	1,3	2,5
Type de ménage (<i>5 cat.</i>)	1,7	1,9	8,0	2,4	8,0	3,8
Sexe (<i>femme</i>)	1,4	0,5	-2,3	0,4	-5,1	-0,7
Âge (<i>continu</i>)	1,3	-1,7	-13,6	0,4	-13,8	-2,4
Diplôme (<i>continu</i>)	4,7	6,5	23,7	6,5	22,2	7,3
Nationalité (<i>française</i>)	5,4	5,6	6,7	2,6	6,6	3,8
Activité (<i>3 cat.</i>)	1,5	3,3	11,9	2,8	10,4	2,9
Catégorie socioprofessionnelle (<i>7 cat.</i>)	2,4	2,3	7,4	2,2	7,0	2,3
R indicateur (variables a)	<i>0,80</i>	<i>0,80</i>	<i>0,66</i>	<i>0,80</i>	<i>0,66</i>	<i>0,77</i>
Différences standardisées (b)						
Type de coordonnées fournies (<i>4 cat.</i>)	5,8	4,6	15,7	5,3	14,9	8,1
Remontée de données antérieures acceptée	12,6	9,7	6,1	8,9	6,7	12,2
R indicateur (variables a+b)	<i>0,72</i>	<i>0,74</i>	<i>0,62</i>	<i>0,74</i>	<i>0,62</i>	<i>0,67</i>

Lecture : les répondants téléphone du sous-échantillon 1 résident plus souvent en maison (avec une différence standardisée de 3,9) que l'ensemble du sous-échantillon 1. Le R indicateur tenant compte de l'ensemble des variables de la liste est de 0,80 dans le sous-échantillon 1.

Source : Enquête expérimentale CVS panel 2019.

Redressement du biais de sélection sur observables

Pour corriger la sélection sur observables, une pondération représentative de la population générale en 2019 est calculée. Elle part des poids corrigés de la non-réponse de CVS 2018 (voir encadré 2), puis une correction de la non-réponse à l'enquête panel est ensuite modélisée. Pour ce calcul, un travail sur les données sociodémographiques intervenant dans le calcul de pondération a été nécessaire (ce travail est exposé en annexe 3). Par ailleurs, le modèle de non-réponse a été adapté par rapport à celui utilisé pour l'enquête CVS 2019 en prenant en compte certaines caractéristiques connues grâce à l'édition CVS 2018 (le fait d'avoir accepté que les données de l'enquête 2018 soient réutilisées pour raccourcir la réinterrogation, le fait d'avoir déclaré en 2018 se sentir en insécurité à son domicile ou dans son quartier, le fait d'avoir fourni des coordonnées téléphoniques ou une adresse mail pour participer à la réinterrogation). Les étapes suivantes (construction des groupes de réponses homogènes, calage) sont strictement identiques.

Deux poids ont ainsi été calculés : un poids ménage (pour l'étude des taux de prévalence au niveau ménage) et un poids individu-kish (pour l'étude des variables au niveau individu).

À noter que cette pondération ne distingue pas les répondants selon leur mode d'interrogation. Elle permet uniquement de corriger l'auto-sélection de l'ensemble des répondants à l'enquête par rapport à l'échantillon initial.

¹¹ Les R-indicateurs des répondants par sous-échantillon et de l'ensemble des répondants ne sont pas comparables. De fait, les R-indicateurs pour chaque sous-échantillon reposent sur la comparaison entre les répondants d'un sous-échantillon à l'ensemble de l'échantillon, alors qu'une partie ne peut pas être répondant au sous-échantillon en question. Cet indicateur est donc difficilement interprétable mais il permet de comparer les différents sous-échantillons entre eux.

Encadré 2 : la pondération dans CVS

Pour l'exploitation de l'enquête CVS, trois poids distincts sont calculés : un poids associé aux ménages répondants (quel que soit le comportement de réponse de l'individu kish du ménage), à chaque individu kish-répondant et à chaque individu-kish répondant au questionnaire sous casque.

Les redressements suivent plusieurs étapes :

- d'abord, la non-réponse totale des ménages est corrigée par repondération suivant la méthode des groupes de réponse homogène (GRH). Les GRH sont constitués suivant une méthode des scores :
 - La probabilité de répondre à l'enquête est d'abord estimée pour chaque ménage de l'échantillon (hors ménages hors-champs, i.e. identifiés par l'enquêteur lors de la collecte comme inoccupés, détruits ou occupés à titre de résidence secondaire) à l'aide d'un modèle de régression logistique expliquant le fait de répondre par les caractéristiques des ménages disponibles dans la base de sondage. En 2019, les caractéristiques suivantes ont été retenues pour modéliser le fait de répondre : région administrative, indicatrice région de gestion = « île-de-France », âge de la personne de référence, type d'habitat (individuel/collectif), période d'achèvement du logement, statut d'occupation du logement, nombre de pièces du logement, appartenance à un organisme HLM, nombre de voitures du ménage, décile de revenu médian communal et le quintile de taux de chômage communal.
 - Ensuite, les GRH sont constitués par la méthode des quantiles¹², i.e. en regroupant dans un même GRH les ménages, répondants et non-répondants, dont la probabilité de réponse estimée est considérée identique. Les quantiles sont définis suivant la procédure proposée par D.Haziza et J.-F. Beaumont¹³. Le poids corrigeant de la non-réponse de niveau ménage, pour une observation d'un quantile donné, est défini comme l'inverse de la proportion de répondants observée dans le quantile.
- Ensuite, un calage sur marges est appliqué aux poids corrigés de la non-réponse des ménages.
- Les poids des individus répondants sont ensuite calculés en partant des poids corrigés de la non-réponse des ménages.
 - Les poids de sondage des individus-kish interrogés dans les ménages répondants sont calculés en multipliant les poids corrigés de la non-réponse des ménages par le nombre de membres du ménage dans le champ de l'enquête.
 - Puis la non-réponse spécifique des individus-kish est corrigée avec la même méthode que la non-réponse totale des ménages. Les variables intervenant dans le modèle de non-réponse sont issues de l'enquête CVS. En 2019, les caractéristiques suivantes de l'individu-kish ont été retenues : croisement sexe-âge, croisement statut d'activité-catégorie socio-professionnelle, nationalité, diplôme, type et nombre de personnes du ménage.
 - Enfin, les poids corrigés de la non-réponse des individus-kish sont ajustés par calage sur marges.

Les poids des individus-kish répondant au questionnaire auto-administré sous casque sont déterminés de manière analogue aux poids des individus-kish répondants. Les variables intervenant dans le modèle de non-réponse sont également issues de l'enquête CVS. En 2019, les caractéristiques suivantes de l'individu-kish ont été retenues : croisement sexe-âge, statut d'activité, nationalité, diplôme, type de ménage et le fait d'avoir été victime de menaces ou d'injures en 2017 ou 2018.

12 Voir par exemple J.L.Eltinge, I.S.Yansaneh, « Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey », 1997, vol.23 n°1, 37-45.

13 Voir D. Haziza, J.-F. Beaumont, « On the construction of imputation classes in surveys », International Statistical Review, 2007, vol.75, 25-43.

L'impact du redressement sur l'estimation des victimations n'est pas très important dans CVS panel (table 7). Une grande partie des victimations ont une prévalence estimée semblable avec ou sans pondération. D'autres victimations, comme les vols de voiture, de deux-roues ou de vélo, le vandalisme de la résidence principale, les vols sans violence et les injures, ont une prévalence légèrement plus élevée après redressement du biais de sélection sur observables. Une seule victimation, le vandalisme de la voiture, a une prévalence estimée inférieure après pondération. Cependant, ces écarts restent minimes et probablement pas significatifs. Dans CVS 2019, le demi intervalle de confiance à 95 % pour le taux de cambriolages de la résidence principale par exemple est de 0,21 (voir table 17).

Table 7 : prévalence des victimations à partir des données brutes et pondérées de l'expérimentation CVS panel

	CVS Panel	CVS Panel
	d. brutes	d. pondérées
<i>Victimation ménage</i>		
Cambriolage (ou tentative) RP	1,72	1,71
Vol sans effraction RP	1,00	1,04
Vol avec ou sans effraction d'un autre logement	0,64	0,68
Vol (ou tentative) de voiture	0,87	1,03
Vol (ou tentative) d'objets dans ou sur la voiture	2,04	2,06
Vol (ou tentative) de deux roues à moteur	0,16	0,25
Vol (ou tentative) de vélo	1,30	1,52
Vandalisme de la RP	2,97	3,14
Vandalisme de la voiture	5,66	5,37
<i>Victimation individu</i>		
Vol avec violence	0,35	0,37
Vol sans violence	1,37	1,49
Violences physiques hors ménage	0,64	0,68
Menaces	3,27	3,30
Injures	8,36	8,70

Source : Enquête expérimentale CVS panel 2019.

Note : RP est l'abréviation pour résidence principale. La première colonne présente les prévalences brutes observées parmi les répondants à CVS panel ; la seconde présente les prévalences pondérées représentant la population générale en 2019.

Un biais de sélection inobservable ?

Les analyses précédentes permettent d'étudier et de corriger la déformation de la structure des répondants par rapport à l'échantillon initial en fonction de caractéristiques socio-démographiques observées à la fois pour les répondants et les non-répondants. Cependant, l'auto-sélection des répondants selon les modes de collecte peut également dépendre de facteurs non observés.

C'est notamment le cas de l'intérêt des enquêtés pour la thématique de l'enquête, qui est un levier important de la motivation à participer ou non à l'enquête (Groves, Singer *et alii*, 2000). Ainsi, avoir été soi-même victime de violence ou de dégradation, s'estimer en situation d'insécurité ou être préoccupé par la délinquance peut inciter les enquêtés à participer à l'enquête. À l'inverse, les victimes de violences peuvent ne pas avoir envie de se remémorer ces moments pénibles, ce qui les désincite alors à participer à l'enquête. Ce biais d'intéressement (ou de désintéressement selon le cas) peut exister dans toutes les enquêtes. Cependant, dans la plupart des enquêtes, on considère que ce biais inobservable peut être entièrement contrôlé par des caractéristiques observables, comme l'âge, le diplôme, le revenu, etc. Si l'intermédiation par l'enquêteur permet de convaincre les individus qui ne se sentent pas concernés de répondre, c'est moins le cas en auto-ad-

ministré. L'hypothèse selon laquelle le biais d'intéressement est bien corrélé aux caractéristiques observables est alors plus forte.

Dans l'expérimentation Vols, violence et sécurité (VVS) menée en 2013 par internet et papier en parallèle de l'enquête CVS réalisée en face-à-face, l'importance de l'effet de mode résiduel, après contrôle des caractéristiques observables, semble s'expliquer par ce biais d'intéressement inobservable. Ce biais, lié à la thématique de l'enquête, dépend également du taux de réponse et du protocole de collecte. Or l'expérimentation VVS diffère sur ces deux points de l'expérimentation CVS panel : elle interroge dès la première interrogation les enquêtés par des modes de collecte uniquement auto-administrés (internet et papier) ; le taux de réponse de 32 % est relativement faible.

Dans VVS, on observe un biais d'intéressement (les personnes les plus concernées participent davantage à l'enquête). Dans le cadre d'un panel, ce biais peut suivre la logique inverse (désintéressement). Comme le montre le taux de collecte selon la déclaration de victimations lors de l'enquête CVS 2018, les répondants 2018 ayant déclaré avoir été victime en 2017 de violences physiques ou sexuelles ont été moins nombreux à répondre à l'enquête CVS panel un an plus tard (table 8). À l'inverse, les répondants 2018 ayant déclaré avoir été victime en 2017 d'une autre agression que celles évoquées dans l'enquête CVS 2018 ont été plus nombreux à répondre à l'enquête CVS panel.

Dans la suite, l'utilisation des méthodes d'estimation sur variables observables (régression doublement robuste et appariement) qui incluent les variables de victimation de CVS 2018 permet vraisemblablement de contrôler de ce biais de sélection. La méthode LATE utilisée ensuite est robuste même s'il y a un biais de sélection, si les répondants aux différents sous-échantillons sont similaires, hypothèse crédible au vu des taux de participation et des R-indicateurs très proches, pour chacun de ces différents sous-échantillons.

Table 8 : taux de réponse à CVS panel 2019 selon la déclaration de victimations N-1 lors de CVS 2018

		a déclaré avoir été victime en 2017 de ce fait	n'a pas déclaré avoir été victime en 2017 de ce fait
Victimations ménage			
vm123	Au moins une atteinte ménage sur le logement	75,0%	76,5%
vm4567	Au moins une atteinte ménage sur les véhicules	77,3%	76,4%
vm89	Au moins une dégradation (logement ou voiture)	74,9%	76,6%
Victimations individu			
vi1	Vol avec violence	72,2%	76,5%
vi2	Vol sans violence	74,7%	76,5%
vi3	Violences physiques hors ménage	70,1%	76,5%
vi4	Menaces	74,6%	76,5%
vi5	Injures	78,1%	76,3%
AUTRAG	Autres agressions	84,8%	76,3%
Victimations individu – partie casque			
tvps	au moins une violence physique ou sexuelle	71,2%	79,7%
tvp	au moins une violence physique (hors ou intra ménage)	71,0%	79,6%
tvs	au moins une violence sexuelle (hors ou intra ménage)	65,1%	79,6%
vshm	violence sexuelle hors ménage	70,8%	79,5%
vsm	violence sexuelle intra-ménage	52,2%	79,5%
vpm	violence physique intra-ménage	63,8%	79,6%
Ensemble		76,9 %	76,4 %

Lecture : 84,8 % des personnes ayant déclaré un autre type d'agression subie en 2017 lors de l'enquête CVS 2018 ont répondu à l'enquête expérimentale. 76,3 % des personnes n'ayant pas déclaré un autre type d'agression lors de l'enquête CVS 2018 ont répondu à l'enquête expérimentale.

Note : taux de participation bruts (pondération unitaire).

Source : Enquêtes CVS 2018 et expérimentale CVS panel 2019.

3. Les effets de mesure sur la victimation

Le deuxième objectif de l'expérimentation consiste à évaluer l'existence d'un biais (ou effet) de mesure lié aux modes de collecte utilisés. L'effet de mesure correspond à l'effet du mode de collecte sur la réponse elle-même pour un même individu. Ainsi, l'écart dans les réponses observées entre deux modes, une fois neutralisé des différences de composition, révèle l'existence d'un biais de mesure. Ce biais de mesure peut être lié à un effet de désirabilité sociale qui, dans le cas des victimations, peut conduire les enquêtés à moins déclarer de victimations face à un enquêteur. Il peut également être lié à un effet de *satisficing*, lequel peut se manifester par une sous-déclaration des victimations ou un mauvais remplissage en auto-administré, d'autant plus favorisé que le questionnaire est complexe.

3.1. Analyse brute des résultats

Dans un premier temps, on observe que les taux de victimation des répondants par internet sont systématiquement supérieurs à ceux des répondants par téléphone (table 9). Les différences entre les deux sous-populations sont variables suivant les victimations :

- différence très faible pour les victimations telles que les vols de deux roues ou les violences physiques hors ménage ;
- différence plus marquée pour les victimations telles que les vols avec ou sans effraction d'un autre logement que la résidence principale, le vandalisme de la résidence principale ou de la voiture ou les vols avec violence.

Table 9 : effectifs et taux bruts de victimation CVS panel selon le mode

	Taux sur l'ensemble des répondants CVS Panel	Effectifs des victimes CVS Panel (tout mode)	Taux sur les répondants téléphone CVS Panel	Taux sur les répondants internet CVS Panel
<i>Victimation ménage</i>				
Cambriolage (ou tentative) RP	1,7	173	1,3	2,2
Vol sans effraction RP	1,0	100	0,8	1,2
Vol avec ou sans effraction d'un autre logement	0,6	64	0,3	1,0
Vol (ou tentative) de voiture	0,9	87	0,6	1,2
Vol (ou tentative) d'objets dans ou sur la voiture	2,0	205	1,5	2,6
Vol (ou tentative) de deux roues à moteur	0,2	16	0,2	0,2
Vol (ou tentative) de vélo	1,3	130	0,9	1,7
Vandalisme de la RP	3,0	298	1,6	4,6
Vandalisme de la voiture	5,7	568	3,1	8,6
<i>Victimation individu</i>				
Vol avec violence	0,4	35	0,2	0,6
Vol sans violence	1,4	137	1,0	1,8
Violences physiques hors ménage	0,6	64	0,6	0,7
Menaces	3,3	328	2,4	4,2
Injures	8,4	839	6,8	10,2
Effectif	10034	2096	5379	4655

Source : Enquête expérimentale CVS panel 2019.

Note : calculs de taux bruts, pondérations unitaires.

3.2. Travaux d'apurement

Traitement des doubles comptes

Le recueil des faits dans l'enquête CVS intègre une échelle de gravité reprenant les actes de violence tels que hiérarchisés dans le Code pénal. Ils sont différenciés selon les atteintes aux personnes ou les atteintes aux biens puis déclinés par niveau de gravité. Pour chaque fait, chaque nouveau type de fait proposé exclut les précédents. Ce procédé permet d'éviter les doubles comptes : c'est un principe essentiel de l'enquête auquel sont formés les enquêteurs qui aident les enquêtés à attribuer correctement le fait qu'ils ont subi à la victimation correspondante. En auto-administré, les enquêtés sont *de facto* moins sensibles au classement correct des victimations et peuvent recenser à différents niveaux un même fait si celui-ci recouvre différentes victimations. Les doubles comptes concernent donc majoritairement les questionnaires collectés sur internet ; l'analyse conduite ci-après établit que, selon le type de victimation, entre 63 % et 100 % des doubles comptes concernent des questionnaires réalisés sur internet.

Grâce à plusieurs variables décrivant le dernier fait subi (date de l'évènement, circonstances dans lesquelles se sont déroulées l'évènement), il est en effet possible d'identifier les doubles comptes en comparant les victimations deux à deux (voir encadré 3).

Encadré 3 : Repérage des doubles comptes

L'étape préalable à la correction des doubles comptes est le repérage de ces potentiels doubles comptes. Deux indicatrices ont été créées pour isoler les potentiels cas : une indicatrice double-compte « sûr » et une indicatrice double compte « douteux ». Le tableau suivant explicite les variables retenues pour la création de ces indicatrices pour chaque binôme de victimation.

Binôme de victimations	Doubles comptes sûrs	Doubles comptes douteux
Cambriolage et vol sans effraction	- même mois - même estimation de la valeur des objets volés - même auteur	- même mois - même estimation de la valeur des objets volés ou même auteur
Cambriolage et dégradation logement	- même mois - même estimation de la valeur des objets volés - même auteur	- même mois - même estimation de la valeur des objets volés ou même auteur
Vol sans effraction et dégradation logement	- même mois - même estimation de la valeur des objets volés - même auteur	- même mois - même estimation de la valeur des objets volés ou même auteur
Cambriolage et dégradation voiture	- même mois - même estimation de la valeur des objets volés - même auteur	- même mois - même estimation de la valeur des objets volés ou même auteur
Vol de voiture et vol d'objet dans la voiture	- même mois - même lieu de vol (garage/parking/rue...) - même moment de la journée (jour/nuit/NSP)	- même mois - même lieu de vol (garage/parking/rue...) ou même moment de la journée (jour/nuit/NSP)
Vol de voiture et dégradation voiture	- même mois - même lieu de vol ou de dégradation (garage/parking/rue...) - même moment de la journée (jour/nuit/NSP)	- même mois - même lieu de vol ou de dégradation (garage/parking/rue...) ou même moment de la journée (jour/nuit/NSP)

Vol d'objet dans la voiture et dégradation voiture	<ul style="list-style-type: none"> - même mois - même lieu de vol ou de dégradation (garage/parking/rue...); - même moment de la journée (jour/nuit/NSP) 	<ul style="list-style-type: none"> - même mois - même lieu de vol ou de dégradation (garage/parking/rue...) ou même moment de la journée (jour/nuit/NSP)
Vol avec violence et violence physique	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de vol ou de violence (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre) 	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de vol ou de violence (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)
Vol avec violence et menace	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de vol ou de menace (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre) 	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de vol ou de menace (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)
Vols avec violences et vols sans violence	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de vol (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre) 	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de vol (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)
Violence physique et menace	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de violence ou de menace (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre) 	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de violence ou de menace (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)
Violence physique et injures	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de violence ou d'injure (do- 	<ul style="list-style-type: none"> - même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de violence ou d'injure (domicile/logement de qqun d'autre/immeuble du Kish/lieu de tra-

	micile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)	vail ou d'études/transport en commun/commerce/autre)
Menaces et injures	- même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) - même moment de la semaine (jour de semaine/week-end/NSP) - même moment de la journée (jour/nuit/NSP) - même lieu de menace ou d'injure (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)	- même mois - même zone géographique (France métropolitaine/Outre-mer/étranger) ou même moment de la semaine (jour de semaine/week-end/NSP) ou même moment de la journée (jour/nuit/NSP) ou même lieu de menace ou d'injure (domicile/logement de qqun d'autre/immeuble du Kish/lieu de travail ou d'études/transport en commun/commerce/autre)

Une fois repérés les potentiels doubles comptes à partir de ces deux indicatrices, sur l'ensemble des fiches-adresses (FA) collectées dans le cadre de l'enquête, ceux-ci ont été étudiés un à un en comparant les libellés de victimation. Si certains libellés sont identiques ou suffisamment explicites (« déjà indiqué », « c'est le même fait »), les descriptions indiquées dans les libellés des deux faits ne sont pas toujours suffisamment précises pour être sûr qu'il s'agit effectivement de doubles comptes. En cas de doute, aucune correction n'est apportée.

La table 10 présente les doubles comptes identifiés comme sûrs (dates et circonstances identiques) et douteux (dates identiques et au moins une variable de circonstances différentes), sur l'ensemble des FA collectées. Les binômes « Menaces et injures », « Vol de voitures – Vol d'objet dans/sur voiture » et « Vol d'objet dans/sur voiture – Vandalisme voiture » sont les plus sujets à double déclaration des enquêtés. Au total, 252 cas de doubles comptes considérés comme sûrs ont été identifiés et 72 cas de doubles comptes « douteux ».

Table 10 : nombre de doubles comptes (sûrs et douteux)

	Nombre de cas avec les 2 variables de victimation renseignées	... dont le même mois	Double compte sûr	Double compte douteux
Cambriolage RP – Vol sans effraction RP	20	14	11	3
Cambriolage RP – Vandalisme RP	27	16	0	10
Vol sans effraction RP – Vandalisme RP	12	2	0	1
Cambriolage RP – Vandalisme voiture	17	1	0	1
Vol de voitures – Vol d'objet dans/sur voiture	52	48	43	5
Vol de voitures – Vandalisme voiture	30	20	17	3
Vol d'objet dans/sur voiture – Vandalisme voiture	60	35	33	2
Vol personnel avec violence – Violences physiques	2	2	2	0
Vol personnel avec violence – Vol personnel sans violence	11	8	6	2

Vol personnel avec violence – Menaces	8	6	0	6
Violences physiques – Menaces	31	19	13	6
Violences physiques – Injures	32	18	15	3
Menaces – Injures	204	142	112	30

Source : Enquête expérimentale CVS panel 2019.

Tous les doubles comptes sûrs et douteux ont été étudiés manuellement en comparant les libellés des deux victimations déclarées. Au final, ce sont 260 indicatrices de victimation qui ont été corrigées (table 11).

Table 11 : nombre de corrections suite au traitement des doubles comptes et part de ces corrections issues des réponses par internet

Victimation	Nombre de cas corrigés	Dont part internet
Cambriolage RP	2	100 %
Vol sans effraction RP	10	90 %
Cambriolage ou vol sans effraction autre lieu	2	100 %
Vol de voitures	2	100 %
Vol d'objet dans/sur la voiture	47	83 %
Vol de deux roues à moteur	0	-
Vol de vélos	0	-
Vandalisme RP	10	80 %
Vandalisme voiture	35	94 %
Vol personnel avec violence	2	100 %
Vol personnel sans violence	6	100 %
Violences physiques	1	100 %
Menaces	16	63 %
Injures	127	74 %

Source : Enquête expérimentale CVS panel 2019.

Ainsi, 2,6 % des FA de l'expérimentation CVS panel nécessitent une correction suite à l'étude des doubles-comptes (table 12). Si le même exercice était appliqué à CVS classique sur le millésime 2019, ce sont seulement 27 corrections qui seraient apportées soit des corrections sur 0,2 % des FA répondantes.

Table 12 : comparaison du nombre de doubles comptes entre CVS panel et CVS 2019

	CVS Panel	CVS panel répondants internet	CVS panel répondants téléphone	CVS 2019 classique
Nombre total de doublons « sûrs »	252	197	55	93
Nombre total de doublons « douteux »	72	58	14	62
Nombre total de correction	260	207	53	27
Nombre total de FA	10 034	4 655	5 379	12 397
% des corrections dans le total des FA	2,6 %	4,4 %	1,0 %	0,2 %

Source : Enquête expérimentale CVS panel 2019 et enquête CVS 2019 (enquête annuelle CVS « classique » collectée en 2019).

Traitement des libellés pour étudier le bon classement des victimations

Outre le principe d'exclusion des faits déjà comptabilisés avant, les enquêteurs sont également rodés à l'exclusion de certains cas qui ne doivent pas être comptabilisés dans l'enquête CVS : exclusion des biens professionnels, faits concernant une tierce personne, etc. De fait, les victimations recensées dans l'enquête correspondent à des délits précis, pour pouvoir être comparées aux faits enregistrés par la police et la gendarmerie. Pour chaque victimation déclarée, l'enquêté est amené à la décrire dans un champ libre. L'étude de ces libellés de victimations de l'enquête expérimentale met en lumière la déclaration de faits ne rentrant pas dans le champ de l'enquête CVS ou un mauvais classement des faits décrits.

En cas de doute ou de libellés particulièrement imprécis, aucun recodage n'a été effectué. En effet, le classement selon le réel type de faits peut être particulièrement compliqué :

- par exemple entre la tentative de vol de voiture et le vandalisme, il peut être difficile pour l'enquêté de connaître l'objectif initial de l'auteur ;
- par exemple entre une menace et une injure, un même propos pourra être interprété différemment suivant les enquêtés.

De plus, certains champs de la victimation ne sont pas couverts par l'enquête CVS comme le harcèlement ou les violences psychologiques.

La table 13 présente pour chaque type de victimation le nombre de libellés renseignés, le nombre de libellés jugés douteux au regard de la description de la victimation qui en est faite et le nombre de libellés amenant à une correction car ne correspondant pas de façon certaine à la définition précise de la victimation.

Table 13 : nombre de corrections suite à l'analyse des libellés et part de ces corrections issues des réponses par internet

Victimation	Nombre de libellés renseignés	Nombre de libellés jugés douteux	Nombre de libellés finalement repris	Dont part internet
Cambriolage RP	173	11	7	71 %
Vol sans effraction RP	100	7	1	100 %
Cambriolage ou vol sans effraction autre lieu	64	12	8	88 %
Vol de voitures	87	18	16	94 %
Vol d'objet dans/sur la voiture	205	19	1	100 %
Vol de deux roues à moteur	16	1	0	
Vol de vélos	130	9	8	100 %
Vandalisme RP	298	81	76	83 %
Vandalisme voiture	568	9	4	100 %
Vol personnel avec violence	35	11	10	80 %
Vol personnel sans violence	137	12	9	67 %
Violences physiques	64	7	6	67 %
Menaces	328	41	28	64 %
Injures	839	53	24	71 %
Total	3044	291	198	79 %

Source : Enquête expérimentale CVS panel 2019.

Lecture : 173 déclarations de victimation « cambriolage résidence principale » ont été enregistrées ; 11 sont jugées douteuses à l'issue de l'analyse et 7 ont donné lieu à une requalification. 71 % des 7 déclarations modifiées sont collectées par internet.

Sur 291 libellés jugés douteux en première lecture, 198 cas ont amené à une correction d'au moins une indicatrice de victimation.

La table 14 présente pour chaque type de victimation la requalification des victimations mal classées à partir des libellés renseignés par les enquêtés. Plus de la moitié des corrections correspondent à des faits qui ne devraient pas être pris en compte dans le champ de l'enquête CVS (105 hors champ). Le vandalisme de la résidence principale explique à elle seule les deux tiers des hors champ. Pour cette victimation, beaucoup d'enquêtés ont décrit des faits qui ne concernent pas directement leur logement (appartements de voisins, boîtes aux lettres de voisins, parties collectives de l'immeuble).

Dans les victimations individuelles, plusieurs reclassements entre les « Menaces » et les « Injures » ont été effectués.

Table 14 : requalification des mauvais classements suite à l'analyse des libellés

Victimation	Nombre de libellés finalement repris	Requalification en :
Cambriolage RP	7	Vol sans effraction RP : 2 Vandalisme logement : 2 Hors champ CVS : 3
Vol sans effraction RP	1	Cambriolage RP : 1
Cambriolage ou vol sans effraction autre lieu	8	Hors champ CVS : 7 Vol personnel : 1
Vol de voitures	16	Vol d'objet dans/sur la voiture : 7 Vandalisme voiture : 8 Hors champ CVS : 1
Vol d'objet dans/sur la voiture	1	Vandalisme voiture : 1
Vol de vélos	8	Hors champ CVS : 8
Vandalisme RP	76	Hors champ CVS : 69 Cambriolage RP : 5 Vandalisme voiture : 2
Vandalisme voiture	4	Vol d'objet dans/sur la voiture : 4
Vol personnel avec violence	10	Hors champ CVS : 7 Menaces : 1 Injures : 1 Vol de voitures : 1
Vol personnel sans violence	9	Vol personnel avec violence : 5 Hors champ CVS : 4
Violences physiques	6	Menaces : 3 Injures : 1 Violence sexuelle hors ménage : 1 Hors champ CVS : 1
Menaces	28	Injures : 25 Vol de voitures : 1 Hors champ CVS : 2
Injures	24	Menaces : 15 Violences physiques : 6 Hors champ CVS : 3

Source : Enquête expérimentale CVS panel 2019.

L'impact de l'apurement sur les taux de victimation bruts

Les tables 15 et 16 synthétisent les impacts des redressements, exposés aux deux paragraphes précédents, sur les taux de victimation selon le mode. Comme attendu, ce sont surtout les taux de victimation du mode internet qui sont revus à la baisse.

Sur les données recueillies par internet, le traitement des doubles-comptes permet de diminuer la prévalence des injures (-1,1 point), des vols d'objets dans/sur la voiture (-0,8 point), du vandalisme sur la voiture (-0,4 point) et dans une moindre mesure des vols sans effraction de la résidence principale, du vandalisme de la résidence principale, et des menaces (-0,2 point) mais également des vols de voiture, vol personnel avec ou sans violence (-0,1 point). L'analyse des libellés a abouti à la diminution des taux de prévalence du vandalisme de la résidence principale (-1,3 point), les vols de voitures (-0,3 point), des cambriolages ou vol sans effraction dans un autre lieu, des menaces et des injures (-0,2 point). Elle amène également à l'augmentation du taux de prévalence pour les vols d'objets dans/sur la voiture (+0,2 point) et le vandalisme sur la voiture (+0,1 point). Après les deux correctifs, le taux de prévalence de 11 types de victimation sur 14 a été modifié.

Sur les données recueillies par téléphone, après les deux traitements (doubles-comptes et libellés), seules les victimations suivantes sont corrigées, toutes à la baisse, et dans une moindre importance : vols d'objets dans/sur la voiture (-0,1 point), vandalisme de la résidence principale (-0,3 point), violences physiques (-0,1 point), menaces (-0,1 point) et enfin injures (-0,3 point).

Table 15 : taux de victimation issus des réponses par le mode **internet après traitements des doubles comptes et des libellés**

Victimation	Taux de victimation bruts	Taux de victimation après traitement des doubles comptes	Taux de victimation après traitement des doubles comptes et des libellés
Victimation ménage			
Cambriolage RP	2,2	2,2	2,2
Vol sans effraction RP	1,2	1,0	1,0
Cambriolage ou vol sans effraction autre lieu	1,0	1,0	0,8
Vol de voitures	1,2	1,1	0,8
Vol d'objet dans/sur la voiture	2,6	1,8	2,0
Vol de deux roues à moteur	0,2	0,2	0,2
Vol de vélos	1,7	1,7	1,6
Vandalisme RP	4,6	4,4	3,1
Vandalisme voiture	8,6	8,2	8,3
Victimation individu			
Vol personnel avec violence	0,6	0,5	0,4
Vol personnel sans violence	1,8	1,7	1,6
Violences physiques	0,7	0,7	0,7
Menaces	4,2	4,0	3,8
Injures	10,2	9,1	8,9

Source : Enquête expérimentale CVS panel 2019. Taux calculés avec des pondérations unitaires.

Table 16 : taux de victimation du mode **téléphone après traitements des doubles comptes et des libellés**

Victimation	Taux de victimation bruts	Taux de victimation après traitement des doubles comptes	Taux de victimation après traitement des doubles comptes et des libellés
<i>Victimation ménage</i>			
Cambriolage RP	1,3	1,3	1,3
Vol sans effraction RP	0,8	0,8	0,8
Cambriolage ou vol sans effraction autre lieu	0,3	0,3	0,3
Vol de voitures	0,6	0,6	0,6
Vol d'objet dans/sur la voiture	1,5	1,4	1,4
Vol de deux roues à moteur	0,2	0,2	0,2
Vol de vélos	0,9	0,9	0,9
Vandalisme RP	1,6	1,5	1,3
Vandalisme voiture	3,1	3,1	3,1
<i>Victimation individu</i>			
Vol personnel avec violence	0,2	0,2	0,2
Vol personnel sans violence	1,0	1,0	1,0
Violences physiques	0,6	0,6	0,5
Menaces	2,4	2,4	2,3
Injures	6,8	6,5	6,5

Source : Enquête expérimentale CVS panel 2019. Taux calculés avec des pondérations unitaires.

3.3. Comparaison avec les taux de victimations de CVS 2019

La comparaison des taux de prévalence entre l'enquête CVS 2019 et l'enquête panel expérimentale sur données non pondérées montre que seuls 4 types de victimations affichent une différence de moins de 0,05 en valeur absolue sur 14 victimations étudiées (table 17). La pondération corrigeant la non-réponse pour rendre les répondants à CVS panel représentatifs de l'ensemble de la population ne permet pas de resserrer les écarts observés entre les taux de victimation calculés sur les deux enquêtes. Au contraire, les écarts en valeur absolue sont plus importants sur résultats pondérés. Seuls les taux de prévalence du vandalisme de la voiture et des vols avec violence sont plus proches entre les deux enquêtes après pondération.

Seuls trois taux de prévalence calculés sur les données de l'expérimentation se trouvent dans l'intervalle de confiance du taux de prévalence calculé sur les données de l'enquête CVS 2019 : cambriolage de la résidence principale, vol sans effraction de la résidence principale, vol avec violence.

Sur les victimations individuelles, les prévalences calculées à partir des données de l'expérimentation sont systématiquement plus basses que celles calculées à partir des données de l'enquête classique. Cela est variable pour les victimations au niveau ménage.

Table 17 : comparaison des taux de victimation entre CVS panel et CVS 2019

	CVS Panel	CVS Panel	CVS Panel		CVS 2019	CVS 2019	Intervalle de confiance CVS 2019	
	brut	brut corrigé des DC et libellés	pondéré		brut	pondéré	Borne inférieure	Borne supérieure
<i>Victimation ménage</i>								
Cambriolage (ou tentative) RP	1,72	1,67	1,65		1,69	1,68	1,47%	1,89%
Vol sans effraction RP	1,00	0,90	0,94		0,90	0,86	0,71%	1,01%
Vol avec ou sans effraction d'un autre logement	0,64	0,54	0,58		0,29	0,27	0,19%	0,35%
Vol (ou tentative) de voiture	0,87	0,72	0,90		0,66	0,68	0,55%	0,81%
Vol (ou tentative) d'objets dans ou sur la voiture	2,04	1,67	1,65		1,93	1,97	1,75%	2,19%
Vol (ou tentative) de deux roues à moteur	0,16	0,16	0,25		0,14	0,15	0,09%	0,21%
Vol (ou tentative) de vélo	1,30	1,22	1,42		1,01	1,09	0,92%	1,26%
Vandalisme de la RP	2,97	2,11	2,28		1,97	1,99	1,76%	2,22%
Vandalisme de la voiture	5,66	5,49	5,18		3,62	3,70	3,40%	4,00%
<i>Victimation individu</i>								
Vol avec violence	0,35	0,29	0,31		0,31	0,32	0,23%	0,41%
Vol sans violence	1,37	1,24	1,38		1,59	1,85	1,60%	2,10%
Violences physiques hors ménage	0,64	0,63	0,65		1,19	1,36	1,15%	1,56%
Menaces	3,27	2,98	3,05		3,24	3,45	3,10%	3,80%
Injures	8,36	7,60	7,98		8,53	9,32	8,74%	9,90%

Source : Enquête expérimentale CVS panel 2019 ; enquête CVS 2019.

Note : Pondération par les poids issus du plan de sondage et de la correction de non-réponse.

3.4. Méthodes d'estimation des effets de mode

En théorie, le biais de mesure est défini au niveau individuel ; c'est l'effet du mode de collecte sur la réponse apportée à une question. Cependant, on observe en général la réponse d'un même individu sur un seul mode de collecte. La réponse sur l'autre mode n'étant pas observée, l'estimation de l'effet de mesure va consister à estimer un contrefactuel de cette réponse non observée. Pour cela, la stratégie consiste à rendre comparables les répondants à chacun des modes de collecte pour que la différence résiduelle éventuellement observée entre des répondants comparables ne s'explique que par la différence de mode et non par des effets de composition des deux ensembles de répondants comparés.

Le cadre théorique pour cette estimation est celui du modèle causal de Rubin (Imbens et Rubin, 2015). Ce modèle repose sur deux hypothèses :

- Le mécanisme de réponse est probabiliste, c'est-à-dire que tous les répondants ont une probabilité non nulle de répondre à chacun des modes. De fait, si ce n'est pas le cas, il n'existe pas de contrefactuel possible pour ces individus, et l'estimation de l'effet de mesure ne pourra pas être faite sur ces répondants. Plus généralement, il est utile de disposer du support commun (i.e. une plage de variation de caractéristiques — ou un score sur ces caractéristiques — pour laquelle il existe des répondants comparables sur les différents modes de collecte) le plus large possible. En dehors de ce support commun, il n'existe pas de contrefactuel comparable sur l'autre mode de collecte.
- Le processus de sélection propre à chacun des modes de collecte est indépendant des variables d'intérêt, conditionnellement aux caractéristiques observables. Cette hypothèse d'indépendance conditionnelle implique que les différences entre les répondants des deux modes sont entièrement contrôlées par les variables prises en compte pour rendre les répondants comparables. En d'autres termes, il n'existe pas de caractéristique inobservable qui explique à la fois la participation à un mode et les variables d'intérêt. Si ce n'est pas le cas, ces méthodes permettent d'estimer un effet de mode contrôlé des caractéristiques observables mais sans pouvoir distinguer l'effet de mesure qu'on cherche à estimer d'un effet de sélection inobservable.

Ces deux hypothèses sont des conditions suffisantes pour la convergence des estimateurs de l'erreur de mesure étudiés ici et elles sont nécessaires pour tous sauf pour l'estimateur LATE. En effet, pour le LATE, la seconde hypothèse ci-dessus est remplacée par une hypothèse de similarité des répondants entre les différents sous-échantillons (voir Angrist *et al.*, 1996). Dans ce cas, l'estimateur LATE estime sans biais l'erreur

de mesure sur la sous-population des répondants qui changeraient de mode si leur affectation à l'un des sous-échantillons changeait également (les « compliers »-répondants au sens de Angrist *et al.*, 1996). Ainsi, sous cette hypothèse de similarité, l'estimateur LATE est robuste à un effet de sélection résiduel de la participation à l'enquête qui serait par exemple lié à une variable inobservable également corrélée aux variables d'intérêt, contrairement aux autres estimateurs étudiés ici.

Modèles d'estimation fondés sur le score de propension

Dans la plupart des cas, la comparabilité des répondants est réalisée en alignant les caractéristiques observables des deux groupes comparés (observé/contrefactuel). Plusieurs méthodes existent pour estimer cet effet de mode contrôlé des caractéristiques observables. Parmi elles, une méthode consiste à régresser les variables d'intérêt sur le mode de collecte et un certain nombre de variables de contrôle, pour contrôler de ces différences observables. Cette régression est facile à mettre en œuvre mais elle repose sur des hypothèses de linéarité des effets. D'autres méthodes reposent sur le concept de score de propension (Rosenbaum et Rubin, 1983) défini comme la probabilité de répondre sur un mode plutôt qu'un autre. Ce score permet de réduire le nombre de dimensions à comparer entre les répondants à chaque mode. À partir de ce score, on peut réaliser une repondération par l'inverse du score de propension. Si le score est bien modélisé, la repondération permet de rendre comparables les répondants aux différents modes de collecte. La différence pondérée observée est alors une estimation de l'erreur de mesure d'un mode par rapport à l'autre.

Ici, nous présentons les résultats d'une méthode qui consiste à combiner les deux précédentes en réalisant une régression pondérée (Quantin, 2018). Cette méthode est doublement robuste dans le sens où le résultat sera sans biais dès lors que la modélisation du score ou de la régression est correctement spécifiée. En revanche, l'utilisation de la pondération peut conduire à des estimations instables ou imprécises si la variance des poids issus de la repondération par l'inverse du score de propension est élevée. Pour éviter cela, les poids issus de la repondération par l'inverse du score de propension sont tronqués aux premier et dernier percentiles. Ils sont également normalisés à la taille de l'échantillon de répondants pour éviter des effets artificiellement significatifs.

À partir du score de propension, on peut également réaliser un appariement pour trouver le contrefactuel le plus proche parmi les répondants à chacun des modes de collecte. Cette méthode a l'avantage de ne pas faire d'hypothèses paramétriques sur le lien entre le mode de collecte et les variables d'intérêt. De même que pour la méthode précédente, elle est relativement aisée à mettre en œuvre. Cependant, il existe de nombreuses façons d'apparier en fonction du nombre de voisins à apparier à chaque observation, de la distance autorisée pour l'appariement ou encore du fait de faire un tirage avec ou sans remise. Le choix de ces paramètres dépend d'un arbitrage biais-variance : avec un appariement strict avec un voisin très proche, le risque de biais est faible mais la variance peut être élevée et cela peut conduire à exclure de nombreuses observations de la comparaison ; à l'inverse un appariement plus lâche conduit à un risque de biais plus élevé mais une variance plus faible. Ici, les résultats présentés sont issus d'un appariement auprès de 2 voisins, avec une distance bornée à 0,15 point de score (caliper) pour ne pas retenir des voisins trop éloignés et un tirage avec remise.

Ces différentes méthodes conduisent à des résultats relativement similaires si elles prennent en compte les mêmes caractéristiques observables de comparaison. Cependant, leurs hypothèses diffèrent ; il est donc intéressant d'en tester plusieurs pour tester la robustesse des résultats. Toutes ces méthodes sont applicables quel que soit le protocole à partir du moment où on dispose de deux groupes ayant répondu sur deux modes de collecte différents. Cependant, les résultats obtenus seront sans biais sous les deux conditions indiquées précédemment. C'est la vérification de ces deux hypothèses qui permet d'appliquer ces méthodes sans risque de biais sur les résultats.

Variables de contrôle de la sélection dans CVS panel

Au-delà de la méthode utilisée, c'est le choix des variables de contrôle qui est primordiale pour vérifier l'hypothèse d'indépendance conditionnelle. Tous les facteurs de confusion, c'est-à-dire les variables corrélées à la fois à la participation et aux variables d'intérêt, doivent être pris en compte, pour s'assurer que toutes les différences entre répondants téléphone et répondants internet sont purgées, hormis le mode de réponse. L'expérimentation CVS panel constitue une occasion rare de mesurer proprement les effets de mesure sur les victimations puisqu'on dispose d'un très grand nombre d'informations, grâce aux informations issues de la première interrogation réalisée en 2018 auprès de l'ensemble de l'échantillon de CVS panel sur un même mode de collecte. Nous nous situons donc dans une situation très favorable pour évaluer l'existence d'effets de mesure, avec peu de risques qu'ils soient biaisés par des différences de composition inobservables. Il apparaît dans ce cas raisonnable de penser que l'hypothèse d'indépendance conditionnelle est vérifiée.

Les variables prises en compte dans le conditionnement sont les suivantes :

- (a) **Les variables socio-démographiques issues de la base de sondage**, habituellement disponibles dans les enquêtes et utilisées classiquement dans les modèles de correction de la non-réponse : l'âge, le sexe, le diplôme, la situation d'activité, la catégorie socio-professionnelle ; le type de logement, le type de ménage ; le niveau de vie du ménage ; la taille d'unité urbaine ; la Zeat ; le fait de vivre dans un quartier prioritaire de la ville.
- (b) **Les variables de contexte, issues de données administratives** : ici il s'agit des faits enregistrés à l'Iris pour les cambriolages, les vols sans violence, les violences physiques, les vols de véhicule, les vols dans ou sur la voiture et le vandalisme. Ces variables permettent de prendre en compte d'autres informations qui peuvent être corrélées à la thématique de l'enquête, tout en étant disponibles pour l'ensemble de l'échantillon.
- (c) **Les variables antérieures issues de l'enquête CVS 2018**, dont on dispose rarement dans les enquêtes non panélistées : l'ensemble des victimations subies en 2017 déclarées en 2018 ; le type de coordonnées et le fait d'avoir accepté la remontée de données antérieures ; les équipements du logement en lien avec la sécurité (alarme, caméra...) ; les opinions sur la justice et la police ; le sentiment d'insécurité dans le quartier et dans son logement ; le fait de déclarer son quartier comme sûr, de renoncer à sortir ; la déclaration de problèmes de délinquance dans le quartier et de cambriolages dans le voisinage. Ces variables sont directement liées à l'intérêt (retardé d'un an cependant) que peuvent avoir les enquêtés pour la thématique de l'enquête, intérêt qui est souvent inobservé et capté uniquement par des caractéristiques issues de la base de sondage. Par ailleurs, elles sont recueillies de la même manière pour l'ensemble de l'échantillon, et donc non affectées d'un biais de mesure. La connaissance de ces variables permet de contrôler d'un éventuel biais d'intéressement (ou de désintéressement) qu'on a observé en fonction des victimations déclarées dans CVS 2018 (voir paragraphe 2.2).

Estimation du score de propension

Pour vérifier l'hypothèse de mécanisme de réponse probabiliste, il est utile de vérifier qu'il existe des répondants comparables dans les deux modes de collecte. Pour cela, on étudie le score de propension : d'après la propriété équilibrante du score, les individus ayant un même score de propension ont en moyenne des caractéristiques similaires concernant celles qui sont prises en compte dans le score. On va donc vérifier que pour chaque score de propension, il existe des répondants ayant répondu par téléphone et d'autres par internet.

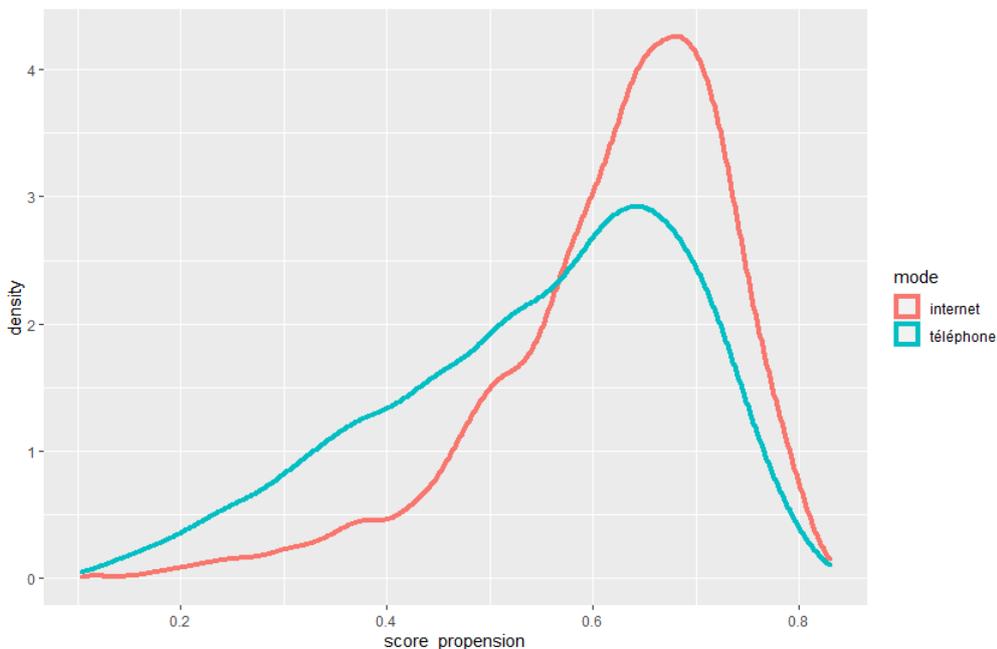
Les résultats présentés ici visent à comparer les répondants internet des sous-échantillons 2 et 3 et les répondants téléphone du sous-échantillon 1. Ces répondants correspondent à des répondants de première intention (i.e. ayant répondu spontanément à l'enquête par le mode de réponse qui leur a été proposé en pre-

mier). Ils sont donc *a priori* davantage comparables que les répondants internet et les répondants téléphone d'un même sous-échantillon. De fait, si on compare les répondants d'un même sous-échantillon, on compare des répondants de première intention avec des répondants relancés plusieurs fois avant d'être interrogés sur l'autre mode de collecte. Le risque que ces répondants soient différents, notamment du point de vue de caractéristiques inobservables, et que le support commun soit restreint est donc *a priori* plus grand. Des analyses ont cependant également été réalisées en comparant les répondants internet et les répondants téléphone des sous-échantillons 2 et 3 qui correspondent aux sous-échantillons web puis téléphone. Elles sont présentées dans la suite.

Le score de propension correspond à la probabilité de répondre par internet plutôt que par téléphone, contrôlée par un certain nombre de variables. La figure 6 présente la distribution de ce score pour les répondants internet et pour les répondants téléphone tenant compte des variables socio-démographiques ((a) de la liste ci-dessus). On peut voir que la distribution est relativement différente : les répondants internet ont des caractéristiques spécifiques qui leur confèrent une forte probabilité de répondre par ce mode plutôt que par téléphone ; les répondants téléphone ont une probabilité de répondre par internet plus faible et surtout plus étalée. Cependant, le support commun – c'est-à-dire la plage de variation du score commune aux deux échantillons de répondants – est très large. De fait, si on considère la méthode du min/max¹⁴, cela conduit à supprimer seulement 7 observations ; aucune observation n'a une probabilité de répondre par internet plutôt que par téléphone inférieure à 0,1 ou supérieure à 0,9.

Dans la suite, nous utiliserons le score de propension fondé sur les seules variables socio-démographiques (liste (a) ci-dessus) comme pondération dans les régressions pondérées. Cela permet de réaliser la régression sur le maximum d'observations (plus le score tient compte de variables, moins le support commun est large) tout en tenant compte des autres types de variables en les ajoutant comme variables de contrôle dans la régression. Pour les méthodes d'appariement sur le score de propension, trois modèles seront présentés en fonction des variables prises en compte dans le score : les variables de la liste (a) pour le modèle 1 ; les variables des listes (a) et (b) pour le modèle 2 et les variables des listes (a), (b) et (c) pour le modèle 3.

Figure 6 : distribution du score de propension à répondre par internet plutôt que par téléphone, à partir des variables socio-démographiques ((a) de la liste précédente)



14 Cette méthode consiste à supprimer les observations au-delà des bornes pour lesquelles on observe à la fois des répondants internet et des répondants téléphone. Elle repose sur une définition assez large du support commun, donc on peut parfois préférer supprimer les observations dont la probabilité de réponse est extrême (dans le premier ou le dernier décile).

Note : la courbe rouge représente la distribution du score de propension, modélisant la propension à répondre par internet plutôt que par téléphone, pour les répondants internet ; la courbe bleue représente la distribution du même score pour les répondants téléphone.

Source : Enquête expérimentale CVS panel 2019.

Méthode d'estimation fondée sur l'estimateur LATE

Dans les méthodes précédentes, l'identification de l'erreur de mesure se fonde sur l'hypothèse d'indépendance conditionnelle selon laquelle il y a indépendance entre l'affectation au mode (ou du choix du mode) et la variable d'intérêt (ici une variable de victimation), conditionnellement aux variables utilisées pour rapprocher les répondants téléphone et les répondants internet. Cette hypothèse est vraisemblablement raisonnable dans le cas présent, *a fortiori* dans la mesure où l'on dispose de variables caractérisant les réponses apportées par les enquêtés à CVS 2018, c'est-à-dire une enquête de même nature, collectée l'année précédente en face-à-face. Pour autant, il est intéressant de confronter cette méthode avec une méthode alternative ne reposant pas sur la même hypothèse d'indépendance conditionnelle.

Compte tenu de la configuration du protocole de CVS panel, il est possible d'utiliser, comme source d'identification de l'erreur de mesure, l'affectation aléatoire aux différents sous-échantillons de l'enquête. En effet, il apparaît que la répartition des répondants en termes de modes diffère nettement d'un sous-échantillon à l'autre (voir annexe 2). Or l'affectation des individus aux sous-échantillons est aléatoire.

Établissons plus précisément les hypothèses. On s'intéresse à une variable y (ici une victimation) et on cherche à connaître sa dépendance à une variable de mode M qui vaut 0 pour le mode de référence et 1 pour le mode alternatif¹⁵. On dispose d'un instrument du mode, variable notée D , rôle joué ici par l'indicatrice d'appartenance à un sous-échantillon donné¹⁶, noté 1. On peut poser les relations suivantes :

$$\begin{cases} y = a + \Delta M + \varepsilon \\ M = c + f(D) + \nu \end{cases}$$

où ε et ν sont deux aléas et f une fonction inconnue mais déterministe de D . Dans cette expression, le coefficient Δ s'interprète comme l'erreur de mesure du mode 1 par rapport au mode 0. Cependant, la variable de mode M résultant d'un choix, il y a tout lieu de penser qu'elle se forme en même temps que la variable d'intérêt y , c'est-à-dire que¹⁷ $E(\varepsilon|M) \neq 0$ dans l'expression précédente. Par conséquent, si par exemple on estime une quantité de la forme $E(y|M=1) - E(y|M=0)$, qui correspondrait à calculer une moyenne de la variable d'intérêt pour les répondants téléphone d'une part, et les répondants internet d'autre part, cette quantité ne sera vraisemblablement pas égale à Δ puisqu'il n'est donc pas exclu que $E(\varepsilon|M=1) - E(\varepsilon|M=0) \neq 0$. C'est là où la variable D , c'est-à-dire la variable d'affectation au sous-échantillon 1 peut jouer un rôle intéressant, puisque cette variable est par construction indépendante de toute autre variable considérée ici (i.e. sa répartition ne dépend d'aucune autre variable examinée ici), l'affectation aux sous-échantillons résultant d'une sélection aléatoire. En conditionnant par D plutôt que par M , on a :

$$\begin{cases} E(y|D) = a + \Delta E(M|D) + E(\varepsilon|D) \\ E(M|D) = c + f(D) + E(\nu|D) \end{cases}$$

Puis comme D est un instrument du mode, c'est-à-dire qu'elle est indépendante, d'une part des aléas ε et ν et, d'autre part qu'elle explique le mode (i.e. la fonction f n'est pas constante), les espé-

¹⁵ Ce faisant, on cherche donc à identifier l'« erreur de mesure » du mode 1 par rapport au mode 0.

¹⁶ On distingue ici deux sous-échantillons 0 et 1 pour le propos théorique. On précisera ensuite à quoi correspondent ces sous-échantillons par rapport à ceux du protocole d'enquête.

¹⁷ E désigne l'espérance mathématique.

rances conditionnelles des aléas sont nulles et il vient, par substitution des deuxièmes équations dans les premières que :

$$\Delta = \frac{E(y|D=1) - E(y|D=0)}{E(M|D=1) - E(M|D=0)}$$

Grâce à cette propriété, on peut en déduire un estimateur connu sous le nom d'estimateur de Wald ou estimateur LATE¹⁸ (Angrist and Pische, 2009), chacune des composantes du ratio précédent pouvant être estimée séparément grâce à des estimateurs de moyennes portant sur la population des répondants dans chacun des sous-échantillons de l'enquête. En effet, le protocole de CVS Panel est suffisamment homogène entre les différents sous-échantillons (les mêmes modes de collecte, Internet et téléphone, sont proposés) pour justifier que si une personne ayant répondu à l'enquête en étant affectée à un des sous-échantillons, avait été affectée à un autre sous-échantillon, elle aurait également répondu (pas nécessairement sous le même mode). Moyennant quoi, les populations répondantes aux trois sous-échantillons sont les mêmes, sous réserve de corriger les différences de probabilités d'inclusion non-uniformes de chaque individu dans les différents sous-échantillons (cf. paragraphe 1.4 et en particulier la table 1). Pour cela, on procède par pondération en fonction de l'inverse de la probabilité d'inclusion. Plus précisément, on note ω_i^s l'inverse de la probabilité d'inclusion de l'individu i dans le sous-échantillon $s \in \{0,1\}$ considéré. Compte tenu de ce qui précède, on peut donc estimer Δ par :

$$\hat{\Delta} = \frac{\frac{\sum_{i \in ECH1} \omega_i^1 y_i}{\sum_{i \in ECH1} \omega_i^1} - \frac{\sum_{i \in ECH0} \omega_i^0 y_i}{\sum_{i \in ECH0} \omega_i^0}}{\frac{\sum_{i \in ECH1} \omega_i^1 M_i}{\sum_{i \in ECH1} \omega_i^1} - \frac{\sum_{i \in ECH0} \omega_i^0 M_i}{\sum_{i \in ECH0} \omega_i^0}}$$

Le mode de référence M_i est le téléphone ($M_i=0$) et le mode alternatif est l'internet ($M_i=1$). Le poids ω_i^s est l'inverse des probabilités d'inclusion figurant dans la table 1 s'agissant des individus du sous-échantillon 1 (échantillon « traité » dans le vocabulaire de l'économétrie de l'évaluation). Le sous-échantillon 0 auquel il est fait référence dans la formule précédente (échantillon « témoin ») est soit le sous-échantillon 2, soit le sous-échantillon 3, soit la réunion des sous-échantillons 2 et 3. Dans les deux premiers cas, les probabilités correspondantes se lisent dans la table 1. Pour le troisième cas (réunion des sous-échantillons 2 et 3), la probabilité concernée est la somme des probabilités des sous-échantillons 2 et 3.

Le facteur clé de l'identification de $\hat{\Delta}$ est le « take-up », c'est-à-dire la différence, apparaissant au dénominateur de $\hat{\Delta}$, entre les probabilités moyennes de répondre à l'enquête, sous le mode alternatif, des répondants des lots traité et témoin. La table 18 donne ces valeurs (en valeur absolue) pour les trois sous-échantillons de l'enquête. Ici, le « take-up » est assez grand, supérieur à 1/2, entre le sous-échantillon 1 (témoin) et les autres sous-échantillons, ce qui est de nature à permettre un niveau d'identification très satisfaisant (i.e. une bonne précision de l'estimateur $\hat{\Delta}$).

Table 18 : probabilités moyennes de répondre par internet selon les sous-échantillons considérés (%)

sous-échantillon	non-pondéré	pondéré
1	8,6	10,5
2	62,7	59,1

18 pour *Local average treatment effect*

3	67,9	63,8
2-3	65,2	61,3

Lecture : dans le sous-échantillon 1, 8,6 % des répondants répondent par internet. La probabilité moyenne (pondérée pour tenir compte de l'affection à probabilité inégale aux différents sous-échantillons) de répondre par internet dans ce sous-échantillon est de 10,5 %.

Source : Enquête expérimentale CVS panel 2019.

3. 5. Résultats

Effets du mode de collecte sur les victimations

La table 19 présente les résultats issus des régressions des variables de victimation sur le mode de collecte et un certain nombre de variables de contrôle. L'effet présenté correspond à l'odd-ratio associé au fait de répondre par internet plutôt que par téléphone. Pour évaluer l'effet des différentes variables de contrôle, nous réalisons plusieurs modèles. Le premier modèle correspond à l'effet brut du mode sur les victimations ; il s'agit d'une régression non pondérée et sans aucune variable de contrôle. On observe que le fait de répondre par internet plutôt que par téléphone a toujours un effet positif et significatif sur la déclaration de victimation, avec des effets qui peuvent aller du simple au triple pour certaines victimations.

Trois régressions pondérées par l'inverse du score de propension¹⁹ sont ensuite présentées. Le modèle 1 utilise comme variables de contrôle dans la régression les variables socio-démographiques classiques²⁰. La prise en compte de ces variables socio-démographiques modifie relativement peu la significativité et l'ampleur des effets de mode bruts. L'ajout des variables de contexte sur les faits enregistrés à l'Iris comme variables de contrôle dans la régression ne modifie quasiment pas les effets de mode observés (modèle 2). L'ajout des variables issues de l'enquête CVS 2018 modifie également peu les résultats (modèle 3). La prise en compte des variables antérieures joue surtout sur les victimations de niveau individu, pour lesquelles seul l'effet de mode sur les menaces reste significatif.

La table 20 présente les résultats issus de la méthode d'estimation par appariement sur score de propension. Les modèles 1, 2 et 3 se distinguent par les variables de contrôles prises en compte pour calculer le score de propension, de façon analogue aux modèles présentés pour la régression pondérée.

Quelle que soit la méthode utilisée (régression pondérée ou appariement), les effets du mode de collecte sur les prévalences de victimation sont relativement similaires entre les modèles. Les différences sont incluses dans l'intervalle de confiance à 95 %, révélant que les résultats des différents modèles sont sensiblement les mêmes. Ce résultat témoigne d'une robustesse certaine des constats quant aux erreurs de mesure : cette stabilité des effets, y compris après le contrôle des informations de CVS 2018, conduit à assimiler les effets estimés aux seuls effets de mesure.

Le tableau 21 donne les estimations des erreurs de mesure avec l'estimateur LATE (Local Average Treatment Effect). On observe que les valeurs estimées de l'erreur de mesure diffèrent peu de celles présentées à la table 20 et obtenues par des méthodes d'appariement. Cela confirme donc indirectement que l'hypothèse d'indépendance de la sélection au mode, conditionnellement aux variables utilisées dans l'appariement répondants téléphone-répondants internet, est globalement vérifiée dans le cas présent. Cependant, les écarts-types, calculés par Bootstrap, sont un peu plus élevés avec ce nouvel estimateur, sans doute plus réalistes de ce point de vue.

19 Pour ces trois régressions, le score de propension utilisé pour cette pondération tient compte des seules variables socio-démographiques (variables de la liste (a)).

20 À noter que les variables étant déjà prises en compte dans le score et donc dans la pondération, les résultats du modèle 1 sont similaires à ceux d'une régression pondérée sans contrôles.

Au final, on observe des effets très forts de déclaration plus importante du fait de répondre par internet plutôt que par téléphone sur les victimations de niveau ménage. Quelle que soit la méthode d'estimation, le fait de répondre par internet double les chances de déclarer une victimation de niveau ménage par rapport à une réponse par téléphone. C'est principalement le cas pour les dégradations, que ce soit pour la résidence principale ou les véhicules, et les vols d'objets dans ou sur la voiture. Par exemple, le fait de répondre par internet plutôt que par téléphone conduit à augmenter le taux de victimes de vandalisme entre 5,2 et 6,7 points de pourcentage, soit un taux 2 à 3 fois plus important par internet que par téléphone²¹.

L'effet est également très fort pour les cambriolages ou vols sans effraction dans un autre lieu bien qu'il repose sur de faibles effectifs, le taux de prévalence de cette victimation étant faible. Dans une moindre mesure, on observe un effet positif sur les vols sans effraction dans la résidence principale et les vols de véhicules (voiture, 2 roues ou vélo). Bien que les effets soient importants, leur significativité est cependant variable selon les méthodes d'estimation en lien avec de plus faibles effectifs, ce qui ne permet pas de conclure à une erreur de mesure nette.

En ce qui concerne les victimations de niveau individu, les effets de mode sont moindres et s'observent principalement sur les menaces. Si ces effets sont certes de moindre ampleur, ils ne sont pas négligeables pour autant puisqu'ils conduisent à une augmentation de 30 à 50 % environ. Pour les vols et dans une moindre mesure les injures, des effets sont observés mais ils ne sont plus significatifs lorsque les variables issues de CVS 2018 sont prises en compte. Ce résultat suggère l'absence d'erreur de mesure pour ces victimations, ce qui est corroboré par les résultats issus de l'estimateur LATE.

On peut toutefois noter, à propos de l'estimateur LATE, que ce dernier a tendance à produire des estimations d'erreur de mesure plus marquée que celles des estimateurs par appariements, mêmes si ces différentes estimations restent compatibles au vu de leurs intervalles de confiance (comparer le tableau 20-colonnes « effet de mode » et le tableau 21-colonne sous-échantillon 2-3, par exemple). Ce phénomène est peut-être à mettre en relation avec l'existence possible d'une sélection endogène à l'enquête (voir discussion du paragraphe « Un biais de sélection inobservable »), puisque les conditions de convergence de ces deux groupes d'estimateurs diffèrent si un tel phénomène survient, les estimateurs par appariement étant fondés sur l'hypothèse d'absence de sélection endogène, contrairement aux estimateurs LATE.

Table 19 : effet du mode internet par rapport au mode téléphone estimé par régression pondérée

	Effec- tif	Aucun contrôle	1. Var socio- démogra- phiques	2. 1+faits à l'IRIS	3. 2+informa- tions CVS 2018
Victimations ménage	922	2,07 ***	1,93 ***	1,94 ***	2,09 ***
Au moins une atteinte ménage sur le logement	224	1,67 ***	1,76 ***	1,78 ***	1,98 ***
Cambriolage RP	126	1,83 ***	1,89 ***	1,87 ***	2,13 ***
Vol sans effraction RP	68	1,28	1,49	1,52 *	1,90 **
Cambriolage ou vol sans effraction autre lieu	46	2,89 ***	2,79 ***	2,87 ***	2,63 **
Au moins une atteinte ménage sur les véhicules	263	1,69 ***	1,74 ***	1,76 ***	1,96 ***
Vol de voiture	50	1,63	1,88 **	1,92 **	1,69
Vol d'objets dans/sur la voiture	122	1,82 ***	1,85 ***	1,88 ***	2,72 ***

21 Par rapport à une prévalence moyenne au téléphone de 4,0 %, un effet de 5,3 points (obtenu par appariement) conduit à une prévalence 2,3 fois supérieure à la prévalence par téléphone. Avec la régression pondérée, cet effet de répondre par internet plutôt que par téléphone est de 2,7. La différence entre les deux méthodes n'est pas significative au regard des intervalles de confiance des deux estimateurs : [1,93;2,72] pour l'estimateur par appariement et [2,18;3,44] pour l'estimateur par régression doublement robuste.

Vol de 2 roues	11								
Vol de vélo	96	2,00	***	2,09	***	2,08	***	1,78	**
Au moins une dégradation (logement ou voiture)	568	2,76	***	2,49	***	2,50	***	2,74	***
Vandalisme RP	169	2,29	***	2,25	***	2,25	***	2,58	***
Vandalisme voiture	448	3,06	***	2,64	***	2,66	***	2,77	***
Victimations individu	825	1,33	***	1,15	*	1,15	*	1,09	
Au moins un vol (avec ou sans violence)	110	2,07	***	1,64	**	1,68	***	1,37	
Violences physiques hors ménage	47	1,36		1,12		1,15		1,15	
Menaces	234	1,60	***	1,31	**	1,31	**	1,43	**
Injures	589	1,31	***	1,19	**	1,20	**	1,16	

Note : les effets sont présentés sous forme de odd-ratio. La significativité des effets est indiquée par les * : * effet significatif au seuil de 10 %; ** au seuil de 5 %; *** au seuil de 1 %. Les intervalles de confiance de l'effet du mode dans le modèle 3 sont les suivants : [1,75 ; 2,48] pour les victimations ménages, [1,42 ; 2,75] pour les atteintes sur le logement, [1,47 ; 2,61] pour les atteintes sur les véhicules, [2,18 ; 3,44] pour les dégradations, [0,91 ; 1,29] pour les victimations individu.

Lecture : le fait de répondre par internet plutôt que par téléphone multiplie les chances de déclarer une victimation ménage par 2,07.

Source : Enquête expérimentale CVS panel 2019.

Table 20 : effet du mode internet par rapport au mode téléphone estimé par appariement

	prévalences non pondérées		effet de mode		
	répondants téléphone	répondants internet	Modèle 1	Modèle 2	Modèle 3
Victimations ménage	8,1	15,4	6,3 *** (0,83)	6,6 *** (0,83)	6,1 *** (0,99)
Au moins une atteinte ménage sur le logement	2,2	3,6	1,3 *** (0,44)	1,3 *** (0,43)	1,4 ** (0,54)
Cambriolage RP	1,1	2,1	0,7 ** (0,33)	0,8 ** (0,32)	1,0 ** (0,41)
Vol sans effraction RP	0,8	1,0	ns	ns	ns
Cambriolage ou vol sans effraction autre lieu	0,3	0,8	0,5 ** (0,21)	0,5 ** (0,20)	0,4 * (0,25)
Au moins une atteinte ménage sur les véhicules	2,6	4,2	1,3 *** (0,46)	1,4 *** (0,47)	1,5 *** (0,58)
Vol de voiture	0,5	0,8	ns	ns	ns
Vol d'objets dans/sur la voiture	1,1	2,0	0,8 *** (0,32)	1,0 *** (0,32)	1,1 *** (0,40)
Vol de 2 roues	0,1	0,2	ns	ns	ns
Vol de vélo	0,8	1,6	0,6 ** (0,28)	0,6 * (0,29)	ns
Au moins une dégradation (logement ou voiture)	4,0	10,2	5,7 *** (0,66)	5,9 *** (0,67)	5,3 *** (0,81)
Vandalisme RP	1,3	3,0	1,7 *** (0,38)	1,7 *** (0,38)	1,7 *** (0,47)
Vandalisme voiture	2,9	8,3	4,7 ***	5,0 ***	4,3 ***

			(0,59)	(0,59)	(0,72)
Victimations individu	9,5	12,2	ns	ns	ns
Au moins un vol (avec ou sans violence)	0,9	1,9	0,7 ** (0,31)	0,8 *** (0,30)	ns
Violences physiques hors ménage	0,7	1,6	0,7 ** (0,28)	0,7 *** (0,27)	0,6 * (0,34)
Menaces	2,4	3,7	1,2 *** (0,44)	0,7 * (0,44)	1,2 ** (0,53)
Injures	6,9	8,7	ns	ns	ns

Note : les effets sont présentés en points de pourcentage, les écarts-types sont présentés entre parenthèses. La significativité des effets est indiquée par les * : * effet significatif au seuil de 10 %; ** au seuil de 5 %; *** au seuil de 1 %; ns non significatif.

Lecture : le fait de répondre par internet plutôt que par téléphone, à caractéristiques données (variables socio-démographiques, faits à l'Iris et informations de CVS 2018), augmente de 6,1 points de pourcentage la prévalence des victimations ménages.

Source : Enquête expérimentale CVS panel 2019.

Table 21 : Erreur de mesure moyenne du mode internet par rapport au mode téléphone estimé par l'estimateur de Wald (LATE) sur la population des répondants

	Echantillon témoin		
	sous-échantillon 2	sous-échantillon 3	sous-échantillon 2-3
Victimations ménage	8,8*** (1,7)	6,9*** (1,6)	7,8*** (1,5)
Au moins une atteinte ménage sur le logement	1,3 (0,9)	1,7 (0,9)	1,5* (0,8)
Cambrionage RP	0,2 (0,7)	0,9 (0,7)	0,5 (0,6)
Vol sans effraction RP	0,4 (0,4)	0,7 (0,5)	0,6 (0,4)
Cambrionage ou vol sans effraction autre lieu	1,0*** (0,37)	0,6 ** (0,29)	0,8*** (0,27)
Au moins une atteinte ménage sur les véhicules	2,4** (1,02)	1,9 ** (0,95)	2,1 ** (0,86)
Vol de voiture	0,2 (0,52)	0,1 (0,46)	0,1 (0,43)
Vol d'objets dans/sur la voiture	1,8 *** (0,66)	1,5 *** (0,62)	1,7 *** (0,52)
Vol de 2 roues	-0,1 (0,17)	0,1 (0,20)	0,0 (0,16)
Vol de vélo	0,8 (0,59)	0,7 (0,57)	0,7 (0,49)
Au moins une dégradation (logement ou voiture)	6,7*** (1,34)	5,2*** (1,26)	5,9*** (1,13)
Vandalisme RP	1,9** (0,79)	1,0 (0,69)	1,5** (0,64)
Vandalisme voiture	5,3*** (1,23)	4,7*** (1,14)	4,9*** (0,96)
Victimations individu	-0,3	-1,5	-0,8

	(1,74)	(1,60)	(1,50)
Au moins un vol (avec ou sans violence)	0,9 (0,71)	0,8 (0,70)	0,8 (0,59)
Violences physiques hors ménage	0,0 (0,42)	0,1 (0,39)	0,1 (0,37)
Menaces	1,4 (0,92)	0,2 (0,85)	0,8 (0,77)
Injures	-0,4 (1,50)	1,2 (1,42)	0,8 (1,25)

Note : les effets sont présentés en points de pourcentage. La significativité des effets est indiquée par les * : * effet significatif au seuil de 10 %; ** au seuil de 5 %; *** au seuil de 1 %; ns non significatif. Ecart-types estimés par Bootstrap.

Lecture : l'effet de mesure lié à une réponse par internet plutôt que téléphone est estimé à 8,8 points de pourcentage sur les victimations ménage, en utilisant le sous-échantillon 2 comme sous-échantillon témoin .

Source : Enquête expérimentale CVS panel 2019.

Analyses complémentaires

Les résultats présentés jusqu'ici portent sur les victimations corrigées après le travail d'apurement des doubles-comptes et de l'analyse des libellés présenté précédemment. La table 22 présente ces effets si ce travail d'apurement n'avait pas été réalisé. Les résultats présentés correspondent à ceux du modèle 3, contrôlés de l'ensemble des variables de confusion (variables socio-démographiques, faits à l'Iris et informations collectées dans CVS 2018), obtenus par régression pondérée. On peut voir que l'apurement présenté précédemment s'avère particulièrement utile pour corriger une partie des effets de mesure observés. La correction des doubles comptes diminue principalement l'effet pour les vols d'objets dans ou sur la voiture, et dans une moindre mesure les vols sans effraction dans la résidence principale et les cambriolages ou vol sans effraction dans un autre lieu. La correction supplémentaire à partir des libellés de victimations diminue principalement l'effet du mode sur la déclaration de vol de voiture et le vandalisme de la résidence principale. Au total, ces corrections apportées lors d'un apurement manuel en amont conduisent à diminuer nettement l'effet du mode de collecte sur le vandalisme de la résidence principale et les vols de voiture. Malgré ces corrections, ces effets restent importants pour un certain nombre de victimations, notamment pour l'ensemble des victimations ménages. Tout d'abord, il peut rester des problèmes de doubles-comptes ou d'incompréhension de la définition précise des victimations demandées ; l'apurement n'a pu être fait qu'en cas d'informations complètes et sans ambiguïté. Par ailleurs, un phénomène de désirabilité sociale peut expliquer également des déclarations plus importantes de victimations par internet plutôt qu'au téléphone avec un enquêteur.

Table 22 : effet du mode sur les victimations avant et après apurement, estimé par régression pondérée

	Victimations avant correction		Victimations après correction des doubles comptes		Victimations après correction des doubles comptes et des libellés	
Victimations ménage	2,27	***	2,27	***	2,09	***
Au moins une atteinte ménage sur le logement	2,10	***	2,09	***	1,98	***
Cambriolage RP	2,29	***	2,26	***	2,13	***
Vol sans effraction RP	2,24	***	1,93	**	1,90	**
Cambriolage ou vol sans effraction autre lieu	2,93	***	2,74	***	2,63	**

Au moins une atteinte ménage sur les véhicules	2,08	***	2,07	***	1,96	***
Vol de voiture	2,45	***	2,41	***	1,69	
Vol d'objets dans/sur la voiture	3,07	***	2,51	***	2,72	***
Vol de 2 roues						
Vol de vélo	1,88	***	1,88	***	1,78	**
Au moins une dégradation (logement ou voiture)	3,02	***	2,94	***	2,74	***
Vandalisme RP	3,18	***	3,08	***	2,58	***
Vandalisme voiture	2,83	***	2,75	***	2,77	***
Victimations individu	1,08		1,08		1,09	
Au moins un vol (avec ou sans violence)	1,44	*	1,44	*	1,37	
Violences physiques hors ménage	1,12		1,11		1,15	
Menaces	1,48	**	1,45	**	1,43	**
Injures	1,24	**	1,19	*	1,16	

Note : les effets sont présentés sous forme de odd-ratio. La significativité des effets est indiquée par les * : * effet significatif au seuil de 10 %; ** au seuil de 5 %; *** au seuil de 1 %. Calcul fondé sur le modèle de régression 3 : le calcul présenté ici « Victimations après correction des doubles comptes et des libellés » correspond au modèle 3 de la table 18.

Source : Enquête expérimentale CVS panel 2019.

Les différentes méthodes d'estimation présentées permettent de tester la robustesse des résultats. Celle-ci peut également être testée en comparant les répondants au protocole séquentiel web puis téléphone, soit les répondants internet et les répondants téléphone des sous-échantillons 2 et 3. La table 23 présente les résultats du modèle 3 (contrôlé de l'ensemble des variables socio-démographiques, des faits à l'Iris et des informations issues de CVS 2018) par régression pondérée. La comparaison de ces répondants est plus sujette à un risque de sélection inobservable, conjuguée à l'effet de mesure, du fait qu'on ne compare plus des répondants de première intention dans les deux modes de collecte. On s'attend alors à avoir des effets de mode plus importants que lorsqu'on compare les répondants internet des sous-échantillons 2 et 3 aux répondants téléphone du sous-échantillon 1. C'est le cas pour les victimations de niveau individu, et notamment les injures pour lesquelles l'effet de mode est significatif contrairement aux résultats présentés plus haut. En revanche, on observe le phénomène inverse pour les victimations de niveau ménage pour lesquelles les effets de mode sont moindres et sont parfois non significatifs, comme pour les atteintes aux véhicules ou les cambriolages de la résidence principale. Certains résultats semblent donc relativement sensibles, du fait des faibles effectifs et des faibles taux de prévalence. Cependant, les principaux résultats portant sur le vandalisme et les cambriolages et vols sans effraction dans un autre lieu persistent et sont robustes aux différentes méthodes d'estimation.

Table 23 : effet du mode sur les victimations selon les répondants pris en compte dans la comparaison, estimé par régression pondérée

	Comparaison des répondants internet des sous-échantillons 2 et 3 avec	
	les répondants téléphone du sous-échantillon 1	les répondants téléphone des sous-échantillons 2 et 3
Victimations ménage	2,09	1,81
	***	***

Au moins une atteinte ménage sur le logement	1,98	***	1,51	***
Cambriolage RP	2,13	***	1,21	
Vol sans effraction RP	1,90	**	1,69	*
Cambriolage ou vol sans effraction autre lieu	2,63	**	4,75	***
Au moins une atteinte ménage sur les véhicules	1,92	***	1,20	
Vol de voiture	1,69		1,09	
Vol d'objets dans/sur la voiture	2,72	***	1,35	
Vol de 2 roues				
Vol de vélo	1,78	**	1,32	
Au moins une dégradation (logement ou voiture)	2,74	***	2,67	***
Vandalisme RP	2,58	***	2,73	***
Vandalisme voiture	2,77	***	2,85	***
Victimations individu	1,09		1,33	**
Au moins un vol (avec ou sans violence)	1,37		1,23	
Violences physiques hors ménage	1,15		1,08	
Menaces	1,43	**	1,53	***
Injures	1,16		1,49	***

Note : les effets sont présentés sous forme de odd-ratio. La significativité des effets est indiquée par les * : * effet significatif au seuil de 10 %; ** au seuil de 5 %; *** au seuil de 1 %.

Source : Enquête expérimentale CVS panel 2019.

3.6. Correction des effets de mesure

Une fois évalué l'effet de mesure sur les différentes victimations, il est possible de le corriger pour obtenir des estimations exemptes d'un effet de mesure à partir de l'expérimentation CVS panel. L'effet de mesure s'assimile à une réponse partielle erronée par rapport à ce qu'on cherche à estimer. Une méthode de correction relativement naturelle est alors l'imputation : on considère la réponse déclarée par l'enquêté sur un mode de collecte donné comme erronée et on va lui substituer une valeur imputée, corrigée de l'effet de mode.

On considère ici que le mode de collecte de référence est le mode historique de l'enquête, soit le face-à-face, et on assimile la collecte par téléphone (autre mode intermédiaire) comme proxy de cette réponse en face à face. Ainsi, on va imputer les réponses des répondants internet uniquement, sans modifier les réponses des répondants téléphone.

La méthode consiste à identifier un support d'imputation, c'est-à-dire à définir les individus pour lesquels on dispose d'un contrefactuel. Pour cela, on réalise un appariement²² sur modèle logistique à partir de toutes les variables de confusion utilisées dans le document, qu'il s'agisse des variables socio-démographiques, des variables de contexte et des variables antérieures issues de l'enquête CVS 2018. Cet appariement conduit à appairer 6 550 individus, soit les deux tiers des répondants à l'enquête.

Parmi ces paires de répondants internet et de répondants téléphone contrefactuels, on identifie alors les observations à imputer, parmi les répondants internet. Plusieurs méthodes sont envisageables :

²² L'appariement est réalisé sans remise, avec un caliper de 0,20 (borne maximale de la distance entre 2 voisins calculée à partir de score de propension).

- Imputer la réponse du contrefactuel aux répondants internet. Cela conduit à modifier plus d'un quart des répondants internet appartenant au support d'imputation, et quasiment tous les individus ayant déclaré une victimation. Cette méthode n'est pas opérante dans le cas d'indicateurs et de prévalences faibles comme les victimations.
- Imputation aléatoire des victimations déclarées sur internet pour obtenir la même prévalence que pour les répondants téléphone. Cette prévalence est calculée sur le seul support d'imputation et tous les facteurs de confusion égaux par ailleurs. Comme la prévalence par téléphone est toujours inférieure à celle par internet, l'imputation consiste uniquement à modifier les valeurs pour les répondants internet ayant déclaré des victimations. Cette méthode conduit à modifier 9 % des répondants internet appartenant au support d'imputation, soit 297 individus et 3 % de l'ensemble des répondants à l'enquête. Parmi eux, 251 individus sont modifiés pour 1 victimation, 35 pour 2 victimations et 11 pour plus de 3 victimations. C'est cette méthode qui est réalisée ici et présentée par la suite (méthode (1) table 24).
- Imputation pour obtenir la même prévalence que par téléphone mais en privilégiant l'imputation de certains répondants, par exemple selon leur poids limiter le nombre d'imputations en corrigeant les individus avec le plus de poids ou encore selon la distance à leur contrefactuel pour limiter les imputations aux individus les plus proches de leur contrefactuel.

Il est aussi possible de réaliser une correction directement sur les agrégats, sans modifier les réponses individuelles. Pour cela, on calcule des estimateurs de prévalence corrigés des erreurs de mesure en modifiant la contribution des observations affectées par l'erreur de mesure (ici, on considère, de manière conventionnelle, que cela concerne les observations collectées par internet) avec l'estimation moyenne de l'erreur de mesure (méthode (2) table 24). Techniquement, on estime une moyenne μ sur une population générale à l'aide d'un estimateur de Hajek portant sur des observations y_i sur un ensemble de répondants $i \in \mathcal{R}$, chaque répondant ayant une probabilité π_i d'être dans \mathcal{R} . Cet estimateur a pour expression :

$$\hat{\mu} = \left(\sum_{i \in \mathcal{R}} \pi_i^{-1} \right)^{-1} \sum_{i \in \mathcal{R}} \pi_i^{-1} y_i$$

Si un sous-ensemble $\mathcal{R}^e \subset \mathcal{R}$ d'observations est affecté d'une erreur moyenne Δ , alors, étant donnée la moyenne $\hat{\mu}$ calculée avant correction, une estimation de la moyenne corrigée s'écrit simplement en :

$$\hat{\mu}^c = \underbrace{\left(\sum_{i \in \mathcal{R}} \pi_i^{-1} \right)^{-1} \sum_{i \in \mathcal{R}} \pi_i^{-1} y_i}_{\hat{\mu}} - \underbrace{\Delta \left(\sum_{i \in \mathcal{R}} \pi_i^{-1} \right)^{-1} \sum_{i \in \mathcal{R}} \pi_i^{-1} \mathbf{1}(i \in \mathcal{R}^e)}_{\text{correction d'erreur}}$$

Il est possible d'estimer Δ grâce à l'estimateur $\hat{\Delta}$ proposé au paragraphe 3.4 (estimateur LATE). La variance de l'estimateur $\hat{\mu}^c$ peut être obtenue par Bootstrap sur les répondants.

Pour illustrer l'impact d'une correction de l'effet de mesure, on présente ici les résultats issus de la méthode d'imputation aléatoire, ainsi que les résultats de l'estimateur $\hat{\mu}^c$ précédent²³. La table 24 présente les prévalences avant et après correction. À noter que seules les victimations pour lesquelles un effet de mesure significatif a été évalué par la méthode de régression pondérée sont corrigées ici.

On peut voir que les prévalences des victimations diminuent après correction de l'effet de mesure. C'est notamment le cas du vandalisme de la voiture qui passe d'une prévalence de 5,18 % à 3,65 % avec l'estimateur par imputation et 2,94 avec l'estimateur $\hat{\mu}^c$. Dans les deux cas, la prévalence corrigée est beaucoup plus proche de celle estimée à partir de CVS 2019 que ne l'est la prévalence non corrigée. En outre, un test

23 Dans l'expression de $\hat{\mu}^c$, l'ensemble \mathcal{R}^e des observations affectées d'une erreur de mesure est celui des répondants internet. Le lot « traité » est le sous-échantillon 1, tandis que le lot « témoin » est la réunion des sous-échantillons 2 et 3.

statistique d'égalité avec les estimations de CVS 2019 ne rejette pas l'hypothèse nulle (égalité), tandis que cette hypothèse est rejetée avec les moyennes avant correction. On observe un résultat analogue avec les cambriolages ou vols sans effraction dans un autre lieu ou les vols de vélos. Ainsi, la compatibilité des estimations CVS panel avec celles de CVS 2019 s'améliore nettement avec la correction d'erreur de mesure proposée, lorsque l'erreur de mesure estimée est significativement non nulle. Les estimations par sous-échantillons après correction sont ainsi similaires alors qu'elles étaient différentes avant correction (annexe 5).

Cependant, pour certaines victimations, la correction de l'effet de mesure conduit à s'en éloigner légèrement : c'est le cas du cambriolage de la résidence principale, du vol d'objets dans ou sur la voiture et des menaces, même si les estimations restent compatibles au sens d'un test d'égalité. La seule anomalie résiduelle concerne les vols d'objets dans la voiture : l'estimation CVS panel avant correction est compatible avec l'estimation de CVS 2019 ; en revanche, un test d'égalité de l'estimation après correction et de l'estimation CVS 2019 est rejeté. Cela montre que les estimations sur d'aussi faibles prévalences peut s'avérer fragile. Ce peut aussi être le signe que CVS 2019 et CVS panel n'estiment pas rigoureusement la même grandeur, différentes causes exogènes sont possibles, comme un décalage dans les dates de collecte, ou un biais dans la réponse téléphonique (prise comme mode de référence dans CVS panel) par rapport à la réponse en face-à-face.

Cet exercice de correction est réalisé à titre d'illustration ; il montre que la non prise en compte des effets de mesure peut avoir un impact non négligeable sur les résultats. Cependant, ces méthodes comportent des limites. Tout d'abord, la méthode 1 d'imputation est ici relativement frustrée, elle consiste à faire des imputations aléatoires pour certains individus sans tenir compte de leurs caractéristiques. Par ailleurs, elle consiste à modifier la réponse des enquêtés, ce qui est un choix fort, d'autant plus lorsqu'il s'agit de variables sensibles comme les victimations. Par ailleurs, si l'imputation est trop frustrée, elle peut conduire à modifier certaines corrélations entre variables. La méthode 2 de correction sur les agrégats permet de ne corriger que la variable à laquelle on s'intéresse sans modifier les données individuelles. Cependant, cela implique de mener cette correction à chaque fois qu'on s'intéresse à un agrégat ; la correction ne peut pas être faite a priori et diffusée pour une exploitation directe.

Comme noté en commentaire des estimateurs des erreurs de mesure (§ 3.5), l'estimateur LATE (associé ici à la méthode de correction 2) conduit à des estimations de victimation plus faibles que les estimateurs par appariement (associés ici à la méthode de correction 1). Les estimateurs de la méthode 2 sortent, dans 4 cas sur les 5 où une erreur de mesure significative a été détectée (tableau 21), de l'intervalle de confiance des estimations fondées sur CVS 2019 (cf. table 24). Comme évoqué à propos de l'estimation de l'erreur de mesure, il est possible qu'un effet résiduel de sélection endogène dans l'enquête CVS Panel puisse engendrer ce phénomène, la correction appliquée ici ne corrigeant qu'un effet de mesure. Des travaux complémentaires pourront éventuellement examiner cette question.

Table 24 : Prévalences pondérées avant et après correction de l'effet de mesure

	CVS panel			CVS 2019		
	Avant correction	Après correction (1)	Après correction (2)	Moyenne	Borne inférieure	Borne supérieure
Victimations ménage	11,86	9,47	8,36	10,70	10,19	11,21
Cambriolage RP	1,65	1,37	1,41	1,68	1,47	1,89
Vol sans effraction RP	0,94	0,82	0,66	0,86	0,71	1,01
Cambriolage ou vol sans effraction autre lieu	0,58	0,48	0,21	0,27	0,19	0,35
Vol de voiture	0,90	-	-	0,68	0,55	0,81

Vol d'objets dans/sur la voiture	1,65	1,38	0,89	1,97	1,75	2,19
Vol de 2 roues	0,25	-	-	0,15	0,09	0,21
Vol de vélo	1,42	1,19	1,11	1,09	0,92	1,26
Vandalisme RP	2,28	1,75	1,62	1,99	1,76	2,22
Vandalisme voiture	5,18	3,65	2,94	3,70	3,40	4,00
Victimations individu	11,12	10,93	11,61	13,71	12,46	14,96
Au moins un vol (avec ou sans violence)	1,68	-	-	2,14	1,88	2,40
Violences physiques hors ménage	0,65	-	-	1,36	1,15	1,56
Menaces	3,05	2,52	2,68	3,45	3,10	3,80
Injures	7,98	-	-	9,32	8,74	9,90

Lecture : (1) estimation de l'erreur de mesure par régression pondérée et correction par imputation aléatoire ; (2) estimation de l'erreur moyenne de mesure par l'estimateur LATE et correction correspondante des contributions des répondants internet aux moyennes estimées ($\hat{\mu}^c$). Seules les victimations pour lesquelles une erreur de mesure est significative dans la régression pondérée sont corrigées.

Source : Enquête expérimentale CVS panel 2019 ; enquête CVS 2019.

Note : prévalences pondérées pour représenter la population générale en 2019, Intervalle de confiance à 95 % pour CVS 2019.

Conclusion

Concernant l'impact du protocole multimode mis en place sur les taux de réponse et la représentativité des répondants, l'expérimentation CVS panel montre que le recours à une phase internet plus longue ou plus courte ne joue que très peu sur les taux de collecte. En revanche, il apparaît nécessaire de maintenir une phase téléphonique, en complément de la réponse par internet, suffisamment longue, pour assurer un taux de réponse et une représentativité optimaux.

Concernant l'impact du mode de collecte sur les estimations, l'expérimentation CVS panel constitue une expérience rare d'évaluation des biais de mesure tout en s'assurant de contrôler au maximum des effets de sélection, grâce aux nombreuses informations dont on dispose, notamment celles relatives à l'enquête CVS 2018 directement liées à la thématique de l'enquête. Les estimations montrent néanmoins que les variables sociodémographiques, dans ce cas précis, semblent suffire pour contrôler de la sélection ; l'ajout des variables antérieures changent peu les résultats.

On établit dans ce document l'existence d'effets de mesure importants, le fait de répondre par internet plutôt que par téléphone conduisant à déclarer davantage de victimations. Ainsi, le fait de répondre par internet plutôt que par téléphone multiplie par 2 le taux de victimation au niveau ménage. C'est notamment le cas du vandalisme, de la résidence principale ou de la voiture, ainsi que dans une moindre mesure des vols d'objets dans et sur la voiture et des cambriolages et vols sans effraction dans un autre lieu que la résidence principale.

Les modes auto-administrés sont réputés générer des biais de mesure comparativement aux modes intermédiés, en particulier pour les variables relatives à des questions subjectives ayant trait à des opinions ou des attitudes par exemple (Shork and Riillo 2021), ou à des questions sur des thématiques sensibles, des comportements discriminés ou socialement dévalorisés. Le questionnaire étudié ici comporte des victimations ayant trait à des événements ou faits *a priori* objectifs (vol, vandalisme, violence physique) mais aussi des faits *a priori* plus subjectifs (insultes, menaces). Ce document de travail montre que des biais existent, même pour des questions *a priori* peu exposées à de tels biais. Cependant, il peut exister une part de subjectif et surtout des difficultés à bien comprendre et cerner le concept attendu dans les questions (de victimation), ce qui peut se révéler d'autant plus vrai que les questions concernées sont complexes. Ainsi, le travail d'apurement réalisé à l'aval de la collecte montre qu'il existe des erreurs de déclarations, liées à une incompréhension du questionnaire (principe de double-compte) et des concepts, dont le risque est forcément accru lorsque la collecte n'est pas accompagnée d'un enquêteur formé spécifiquement à ces sujets. Ces résultats prouvent que les questionnaires auto-administrés doivent être véritablement pensés pour réduire la charge cognitive reposant sur les enquêtés (connaissances sur la thématique, complexité des concepts, respect des consignes de remplissage), le travail d'apurement des questionnaires individuels, reporté à l'aval, ne pouvant être que partiel et imprécis.

Les effets de mode observés sont intrinsèquement liés à la thématique de l'enquête, à son protocole et au questionnaire associé, ces différents éléments pouvant jouer sur la façon de répondre en fonction du mode de collecte. Si les résultats de ce document de travail donnent des éléments sur la sensibilité des questions de victimation au mode de collecte dans l'enquête CVS, ils méritent d'être retestés dès lors que le protocole et le questionnaire sont modifiés, comme c'est le cas de la nouvelle enquête de victimation, Vécu et ressenti en matière de sécurité (VRS), réalisée à partir de 2022 par le SSMSI. Les résultats issus de l'expérimentation CVS panel présentés dans ce document ont été pris en compte pour adapter le questionnaire sur le recensement des victimations dans l'enquête VRS afin de limiter au maximum des effets de mesure. Ainsi, les questions similaires sont posées sur un même écran pour limiter les erreurs de doubles-comptes puisque l'enquêté peut voir sur quelles victimations il va être interrogé ; des consignes supplémentaires ont été données sous la forme d'info-bulles ou de notices explicatives et des formulations de question ont été revues pour expliciter la définition et le périmètre de certaines victimations qui posent des difficultés,

comme le vandalisme par exemple. Après ce re-travail sur le questionnaire et avec un protocole différent, d'autres études ont montré des effets de mode moindres sur les victimations (Rosenzweig *et alii*, 2022).

Bibliographie

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444-455.
- Angrist J. D. and Pischke J. S. (2009), *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Castell L. et Sillard P. (2021), Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman, Document de travail INSEE M2021/02.
- Castell L., Favre-Martinoz C., Paliot N., Sillard P. (2023), Redressement des la première vague de l'enquête EpiCov : un exemple de correction des effets de sélection dans les enquêtes multimodes. Document de travail INSEE M2023/02.
- De Leeuw, E.D. (2018), Mixed-Mode: Past, Present, and Future. *Survey Research Methods*, 2018. 12(2): p. 75-89.
- Groves, R. M., E. Singer and A. Corning (2000), Leverage-Saliency Theory of Survey Participation: Description and an Illustration, *Public Opinion Quarterly* 64(3): p299-308.
- Klausch, T. (2014), Informed Design of Mixed-Mode Surveys: Evaluating mode effects on measurement and selection error, *Utrecht University Repository*.
- Imbens, G. W. and D. B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge, MA, Cambridge University Press.
- Quantin S. (2018), *Estimation avec le score de propension sous R*, Document de travail Insee n°M2018/01.
- Razafindranovona T. (2016), *Exploitation de l'enquête expérimentale Vols, Violences et sécurité*, Document de travail Insee n°M2016/03.
- Rosembaum et Rubin (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika* 70(1): p. 41-55.
- Rosenzweig A., Razafindranovona T, Guedj H, Quiviger A et Zilloniz S. (2022), *Effets de mode relatifs aux victimations et aux perceptions en matière de sécurité : estimations à partir de l'enquête Genre et Sécurité (GENESE)*, Journées de méthodologie statistique, Insee.
- Schork, J. and C.F.A. Riillo (2021), Survey Mode Effects on Objective and Subjective Questions: Evidence from the Labour Force Survey, *Journal of Official Statistics* 37(1): p. 213-237.

Annexes

Annexe 1 – description du protocole de l'enquête Cadre de vie et sécurité

L'enquête Cadre de Vie et Sécurité est une enquête en population générale auprès d'un échantillon de logements de France métropolitaine et, selon les années, dans un ou plusieurs territoires ou collectivités d'Outre-Mer.

Période de collecte

Jusqu'en 2019, l'enquête métropolitaine se déroulait de février à fin avril ou début mai de chaque année. En 2021, l'enquête a eu lieu d'avril à juin.

Mode de collecte

La passation du questionnaire est réalisée pour sa plus grande partie en face-à-face, par un enquêteur du réseau Insee équipé d'un ordinateur (Capi pour Computer Assisted Personal Interviewing). Pour les questions portant sur les violences sexuelles et les violences au sein du ménage, il est demandé au répondant de saisir lui-même ses réponses sur l'ordinateur au moyen d'un protocole sous casque (Casi pour Computer Assisted Self Interviewing), qui permet aux personnes interrogées de répondre de façon autonome et dans la confidentialité par rapport à toute tierce personne, y compris l'enquêteur ou les autres personnes du ménage. Après verrouillage du questionnaire par l'enquêté, les réponses à ces questions ne sont plus accessibles à l'enquêteur, ce qui garantit une complète confidentialité sur cette partie du questionnaire.

Les questionnaires

Le questionnaire « ménage » s'adresse à un adulte du ménage-unité de vie, de préférence la personne de référence ou un membre du groupe de référence.

Le questionnaire « individuel » en face à face s'adresse à une personne par ménage-unité de vie présente dans le logement, choisie parmi les personnes de 14 ans et plus ayant leur date anniversaire la plus proche du 1er janvier de l'année d'enquête. Le proxy n'est pas autorisé ; une seconde visite est possible.

Le questionnaire « individuel auto-administré » s'adresse à la personne retenue pour le questionnaire « individuel en face-à-face » sous condition qu'elle soit âgée de 18 ans au moins au jour de l'enquête et qu'elle ait 75 ans ou moins au 1^{er} janvier.

Durée d'enquête et contraintes pour l'enquêté

L'enquête dure 50 minutes en moyenne dont 8 minutes pour le questionnaire auto-administré (20 minutes maximum).

Le temps de réponse varie selon les événements vécus par l'enquêté(e) : en l'absence de victimations subies, la passation du questionnaire dure 40 minutes en moyenne. En cas de victimations multiples, elle peut dépasser une heure.

Un répondant sur quatre déclare au moins une victimation « ménage ». Deux sur dix déclarent au moins une victimation « individuelle » (hors violences « sensibles »). Six sur cent déclarent avoir été victimes d'au moins une violence « sensible ».

L'impact sur la personne enquêtée peut être important, en particulier pour les violences sexuelles. En effet, certaines questions posées portent sur des événements dont l'évocation peut s'avérer délicate (recensement et description des violences sexuelles, des violences dans le ménage, désignation de l'auteur). Le protocole retenu (questionnaire informatisé auto-administré) permet à l'enquêté de s'isoler et de répondre en toute confidentialité. Un dispositif d'accompagnement est prévu notamment avec la remise d'une fiche comportant des numéros de téléphone d'associations d'aides aux victimes. Par ailleurs un numéro d'association est transmis sous casque aux enquêtés, qu'ils aient ou pas déclaré des violences.

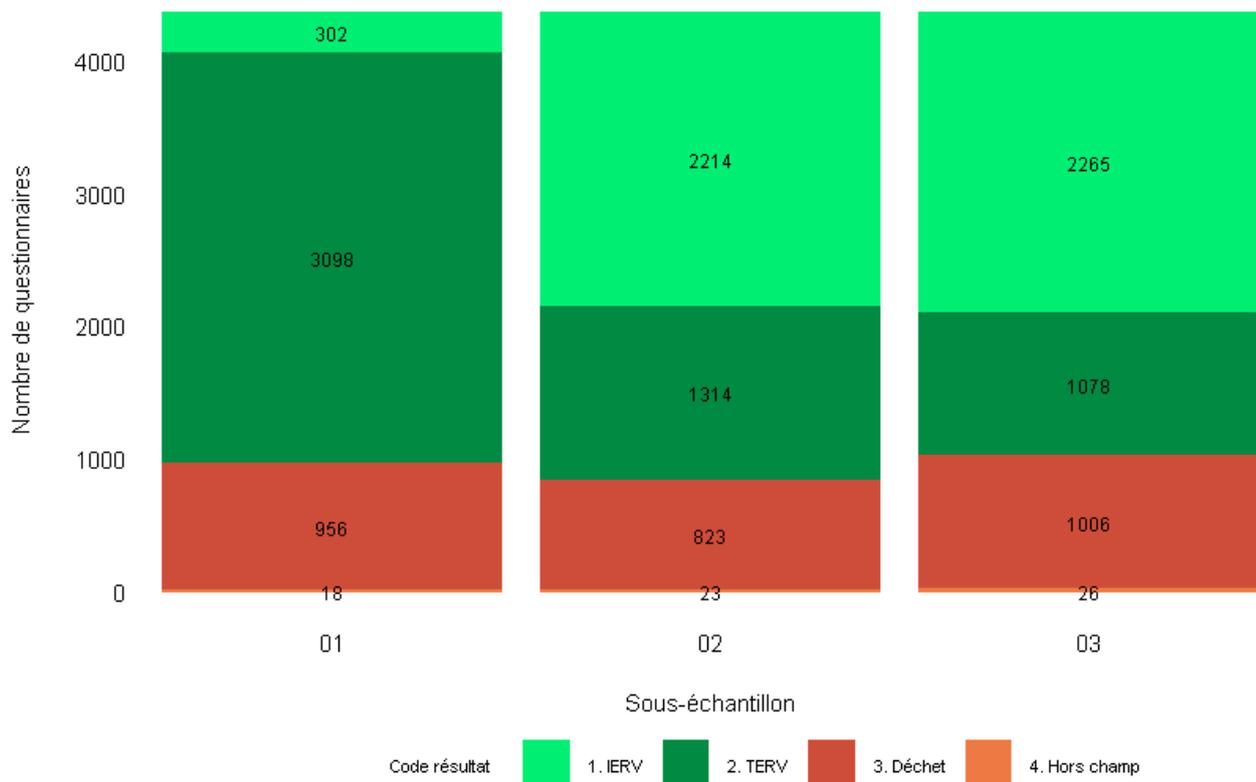
Échantillon

L'échantillon de l'enquête se compose généralement de deux échantillons : un échantillon principal couvrant le territoire métropolitain et un sur-échantillon sur les zones sensibles. Jusqu'en 2015, les logements en zones urbaines sensibles (ZUS) étaient en effet surreprésentés. L'entrée en vigueur des nouveaux « quartiers prioritaires » (QP) au 1^{er} janvier 2015 a entraîné l'abandon du sur-échantillon ZUS à partir de 2016 et la mise en place d'un sur-échantillon « quartiers prioritaires » pour l'enquête de 2016. Les quartiers prioritaires étant définis à partir des données carroyées de l'Insee, le sur-échantillon est tiré dans les fichiers fiscaux.

Le plan de sondage de l'enquête CVS est à deux degrés : tirage des logements uniquement dans les zones d'action enquêteurs de l'échantillon maître qui sont les unités primaires.

Annexe 2 – Détails des résultats par sous-échantillon

Figure : nombre de questionnaires par sous-échantillon



Lecture : IERV = questionnaire internet validé, TERV = questionnaire téléphone validé, Déchet = questionnaire non renseigné mais l'individu est dans le champ de l'enquête, Hors champ = questionnaire non renseigné car l'individu n'est plus dans le champ de l'enquête (décès notamment).

Annexe 3 – Traitement des variables socio-démographiques nécessaires à la pondération et/ou à l'exploitation

Pour limiter le temps de passation de l'enquête expérimentale, le Tronc Commun des Ménages (TCM), partie préliminaire à la grande majorité des enquêtes ménages permettant la description des habitants du logement, n'était pas posé et le nombre de questions posées sur les caractéristiques socio-démographiques ou le logement était limité. En effet, lors de la conception du questionnaire en réinterrogation, il a été décidé de n'actualiser les informations 2018 que sur un nombre très limité de variables, en l'occurrence la commune de résidence et le type de logement (appartement ou maison) en cas de déménagement et la composition du ménage.

La grande majorité des informations socio-démographiques sont donc celles de 2018 sauf pour :

- l'âge ;
- la nouvelle commune de résidence et le type de logement (appartement ou maison) en cas de déménagement ;
- le nombre de personnes dans le ménage ;
- le type de ménage.

Ce choix de conception implique toutefois d'avoir recours à des imputations dans plusieurs situations :

- en cas de déménagement avec une imprécision ou une absence de précision sur la nouvelle commune de résidence ;
- en cas de départ de la personne de référence du ménage ;
- en cas d'arrivée de nouvelles personnes au sein du ménage dont le lien de parenté n'est ni descendant direct ou ascendant direct.

Traitements des variables de département et de commune

En cas de déménagement de l'individu kish, il était indispensable de récupérer la nouvelle commune de résidence. Cette variable permet notamment l'appariement avec des variables de calage (quintile taux de chômage communal, décile de niveau de vie communal).

La table commune utilisée en sortie CAPI donne le nom de communes mais sans précision du code commune. Ainsi si plusieurs communes de départements différents ont le même nom, il est impossible de savoir à quel département la commune appartient. Pour 30 FA, il a donc fallu choisir parmi les différents codes postaux possibles au vu du nom de commune. Les déménagements étant souvent des déménagements de proximité, il a été décidé de retenir :

- en premier la commune de 2018 (si le nom était identique),
- sinon le code postal appartenant au même département que la commune de 2018,
- sinon le code postal appartenant à la même région que la commune de 2018,
- sinon, en dernier ressort, le code postal de la commune ayant le plus d'habitant.

En cas de refus de l'enquêté de fournir le nom de la nouvelle commune de résidence, on attribue par défaut la commune de 2018. Une seule FA a dû faire l'objet d'une telle imputation.

Traitements des variables relatives à la composition du ménage

Dans l'enquête expérimentale, il était demandé à l'enquêté d'actualiser le tableau des habitants du logement (en cas d'acceptation de remontée des données antérieures) ou de recenser les membres de son ménage (en cas de refus de remontée des données antérieures). À partir de la composition 2019 et des informations de 2018, il fallait reconstituer les variables nombre d'enfants (nenfants) et nombre de couples (ncouples) dans le ménage qui permettent le calcul de la variable type de ménage.

Plus précisément, ces variables ont été calculées de la manière suivante :

- En cas d'acceptation de la remontée des données antérieures
 - le nombre d'individus dans le ménage est identique en 2018 et 2019 : les valeurs de nenfants et de ncouples sont donc celles de 2018 ;
 - le nombre d'individus a changé entre les deux années
 - le kish est désormais seul : les deux variables nenfants et ncouples sont égales à 0.
 - le ménage du kish est composé d'au moins deux personnes, on analyse manuellement les départs et les arrivées en 2019 de chaque FA. Pour les départs, on récupère les variables enfant et couple de 2018 pour connaître le statut 2018 des individus partis. Pour les arrivées, une variable du questionnaire 2019 précise les liens avec l'individu kish.
- En cas de refus de la remontée des données antérieures
 - le nombre d'individus dans le ménage est identique en 2018 et 2019 : on affecte les valeurs de nenfants et de ncouples de 2018 ;
 - le nombre d'individus a changé entre les deux années
 - le kish est désormais seul : les deux variables nenfants et ncouples sont égales à 0.
 - le ménage du kish est composé d'au moins deux personnes, on analyse les variables de lien du questionnaire 2019 (COENTR_NOI et LIENTR_NOI) en fonction du nombre de personnes présentes en 2019 pour déterminer les liens des différents membres du ménage et en déduire le nombre d'enfants et le nombre de couples dans le ménage
 - Exemple 1 : si le ménage est composé de deux personnes et que le kish est en couple avec la deuxième personne alors ncouples=1 et nenfants=0 ;
 - Exemple 2 : si le ménage est composé de quatre personnes, le kish est en couple avec un des individus et les deux autres membres du ménage sont les enfants du kish alors ncouples=1 et nenfants=2 ;

À l'issue du codage des variables ncouples et nenfants, la variable type de ménage (typmen) est créée comme habituellement dans le TCM.

Affectation de la personne de référence en cas de départ en 2019 de celle-ci

En cas de départ, de la personne de référence du ménage, une nouvelle personne de référence a donc dû être attribuée. Le choix de la nouvelle personne de référence est effectué en fonction des personnes présentes en 2019 et des informations disponibles en 2018. On regarde toute d'abord si le groupe de référence était composé de plusieurs personnes en 2018 et si les autres membres de ce groupe sont encore présents. Sinon on retient l'adulte actif. S'il y a plusieurs actifs, le plus âgé est sélectionné. Enfin, en l'absence d'actifs, on retient le retraité et finalement les autres inactifs en dernier ressort. Étant donné la stabilité des ménages et la part importante de ménages composés d'une seule personne, l'imputation d'une nouvelle personne de référence n'a été nécessaire que pour 69 ménages.

Annexe 4 – Caractéristiques des répondants les plus sensibles au mode de collecte

Pour aller plus loin, on peut regarder si les effets de mesure évalués dans la partie 3.5 sont spécifiques à certaines sous-populations. Pour cela, on utilise la méthode de la G-computation. Cette méthode consiste à évaluer l'effet de mesure en calculant la différence entre deux contrefactuels. Pour estimer ces contrefactuels, on réalise deux régressions des variables d'intérêt sur l'ensemble des variables de contrôle, d'une part pour les répondants internet, et d'autre part pour les répondants téléphone. On utilise alors comme contrefactuel la valeur prédite à partir des coefficients du modèle pour tous les individus, d'une part s'ils avaient répondu sur internet et d'autre part s'ils avaient répondu par téléphone. L'avantage de cette méthode est de pouvoir étudier les caractéristiques des individus pour lesquels la différence entre les deux contrefactuels est la plus importante sans avoir d'*a priori* sur les caractéristiques à étudier.

Ici, on étudie toujours la même population – les répondants internet des sous-échantillons 2 et 3 et les répondants téléphone du sous-échantillon 1 – en utilisant dans les régressions l'ensemble des facteurs de confusion. On regarde ensuite les caractéristiques des 25 % d'individus dont la différence de contrefactuels est la plus importante. Au vu des résultats moyens, on se concentre ici sur les victimations suivantes : les cambriolages ou vols sans effraction, les dégradations et les menaces pour les victimations individu.

Pour les cambriolages ou vols sans effraction, on observe très peu de différences entre les 25 % d'individus avec l'effet de mesure le plus important et l'ensemble des répondants. Les seules différences notables portent sur les variables de CVS 2018 : les individus pour lesquels l'effet de mesure est le plus important se sont davantage déclarés victimes et en situation d'insécurité dans leur quartier dans CVS 2018. Ainsi, les individus avec l'effet de mesure le plus important sont un quart à être victime d'une atteinte de niveau ménage dans CVS 2018, contre 12 % pour l'ensemble des répondants.

Pour les dégradations, en plus de ces caractéristiques, les effets de mesure sont plus déterminés au plan socio-démographique : ils sont plus importants pour les hommes, les moins de 60 ans, les actifs occupés plutôt que les inactifs, les plus diplômés, les personnes vivant en QPV et en appartement. Par exemple, 11 % des individus avec l'effet de mesure le plus important vivent en QPV et 69 % ont un diplôme supérieur ou égal au baccalauréat, contre respectivement 5 % et 52 % pour l'ensemble des répondants.

Annexe 5 – Étude des estimations de victimation par sous-échantillon avant et après correction de l'erreur de mesure LATE

Il est intéressant d'examiner les prévalences estimées par sous-échantillon, après correction de la non-réponse et de les comparer avec les prévalences estimées dans les mêmes conditions, avant correction. En effet, comme les sous-échantillons sont, après repondération, représentatifs de la population CVS 2019, ces prévalences estimées par sous-échantillon se rapportent à la même population. Sous l'hypothèse d'absence d'erreur de mesure et d'un processus de sélection des répondants indépendant des variables d'intérêt, conditionnellement aux observables entrant dans le modèle de non-réponse (Castell et Sillard 2021 ; Castell et al. 2023), ces estimateurs estiment sans biais la même grandeur. Les pondérations utilisées sont celles calculées pour rendre les répondants à l'enquête CVS panel représentatifs de la population générale de l'enquête CVS 2019, puis pour chaque sous-échantillon, multipliées par la probabilité d'inclusion qui découle de la répartition de l'échantillon des répondants CVS 2018 au sein des trois sous-échantillons (cf. table 1 au paragraphe 1.4). Ce faisant, la population répondante à chaque sous-échantillon est représentative de la population générale de l'enquête CVS 2019.

Tout d'abord, dans ce contexte, des prévalences significativement différentes entre les différents sous-échantillons témoignent d'une anomalie, c'est-à-dire de la violation d'une des hypothèses sous-jacentes à la construction d'estimateurs sans biais. Puis, en cas d'anomalie, si la correction proposée est faite convenablement, l'hypothèse d'égalité des estimateurs corrigés sur chaque sous-échantillon doit être retenue lors d'un test d'égalité.

La table ci-après donne les valeurs estimées et les écarts-types avant et après correction pour les variables de victimation présentées au tableau 24. Pour les quatre victimations élémentaires sur lesquelles une erreur de mesure a été détectée, les moyennes avant correction diffèrent significativement entre l'échantillon 1 d'une part, et les échantillons 2 et 3 d'autre part. Après correction, les moyennes par échantillon ne diffèrent plus significativement.

Table : Prévalences pondérées avant et après correction de l'effet de mesure (points de %)

	Avant correction			Après correction		
	ECH 1	ECH 2	ECH 3	ECH 1	ECH 2	ECH 3
Victimation ménage	9,10	13,69	12,92	8,20	9,22	8,10
	(0,49)	(0,59)	(0,58)	(0,40)	(0,88)	(0,94)
Cambriolage (ou tentative) RP	1,51	1,75	2,28	1,45	1,44	1,94
	(0,20)	(0,21)	(0,25)	(0,16)	(0,39)	(0,42)
Vol sans effraction RP	0,67	0,91	1,19	0,59	0,54	0,80
	(0,16)	(0,15)	(0,19)	(0,12)	(0,24)	(0,26)
Vol avec ou sans effraction d'un autre logement	0,19	0,72	0,62	0,09	0,26	0,12
	(0,09)	(0,14)	(0,14)	(0,09)	(0,15)	(0,16)
Vol (ou tentative) de voiture	0,75	1,19	0,86	0,74	1,12	0,79
	(0,13)	(0,15)	(0,15)	(0,15)	(0,27)	(0,29)
Vol (ou tentative) d'objets dans ou sur la voiture	1,05	1,90	2,12	0,86	0,94	1,09
	(0,19)	(0,23)	(0,24)	(0,15)	(0,32)	(0,34)
Vol (ou tentative) de deux roues à moteur	0,12	0,16	0,32	0,12	0,14	0,30
	(0,07)	(0,06)	(0,08)	(0,09)	(0,12)	(0,12)
Vol (ou tentative) de vélo	0,95	1,65	1,67	0,87	1,25	1,23
	(0,15)	(0,21)	(0,20)	(0,16)	(0,29)	(0,31)
Vandalisme de la RP	1,69	2,74	2,37	1,52	1,90	1,47
	(0,22)	(0,27)	(0,26)	(0,19)	(0,40)	(0,43)
Vandalisme de la voiture	3,78	5,99	5,66	3,20	3,15	2,59
	(0,30)	(0,41)	(0,42)	(0,26)	(0,56)	(0,60)
Victimation individu	11,66	11,33	10,37	11,76	11,84	10,90
	(0,51)	(0,53)	(0,54)	(0,45)	(1,00)	(1,05)

Vol avec ou sans violence	1,50 (0,18)	2,03 (0,21)	1,64 (0,23)	1,39 (0,19)	1,50 (0,41)	1,08 (0,43)
Violences physiques hors ménage	0,59 (0,13)	0,52 (0,13)	0,64 (0,15)	0,58 (0,10)	0,47 (0,24)	0,58 (0,25)
Menaces	2,37 (0,28)	3,54 (0,33)	2,63 (0,29)	2,28 (0,25)	3,05 (0,52)	2,12 (0,55)
Injures	8,39 (0,46)	8,06 (0,47)	7,47 (0,46)	8,50 (0,39)	8,57 (0,88)	8,01 (0,92)

Lecture : En fond rosé, les victimations pour lesquelles l'erreur de mesure LATE est significativement non nulle (cf. tableau 21).

Note : prévalences pondérées estimées en population générale CVS 2019 ; écarts-types entre parenthèses.

Source : Enquête expérimentale CVS panel 2019 ; enquête CVS 2019.

Série des Documents de Travail « Méthodologie Statistique »

- 9601** : Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population.
G. DECAUDIN, J.-C. LABAT
- 9602** : Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises.
N. CARON, P. RAVALET, O. SAUTORY
- 9603** : La procédure **FREQ** de **SAS** - Tests d'indépendance et mesures d'association dans un tableau de contingence.
J. CONFAIS, Y. GRELET, M. LE GUEN
- 9604** : Les principales techniques de correction de la non-réponse et les modèles associés.
N. CARON
- 9605** : L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration.
P. RAVALET
- 9606** : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**).
S. LOLLIVIER, M. MARPSAT, D. VERGER
- 9607** : Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes.
N. CARON, D. LE BLANC
- 9701** : Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?
J.-C. DEVILLE
- 9702** : Modèles univariés et modèles de durée sur données individuelles.
S. LOLLIVIER
- 9703** : Comparaison de deux estimateurs par le ratio stratifiés et application aux enquêtes auprès des entreprises.
N. CARON, J.-C. DEVILLE
- 9704** : La faisabilité d'une enquête auprès des ménages.
1. au mois d'août.
2. à un rythme hebdomadaire
C. LAGARENNE, C. THIESSET
- 9705** : Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine.
P. GIRARD
- 9801** : Les logiciels de désaisonnalisation **TRAMO** & **SEATS** : philosophie, principes et mise en œuvre sous **SAS**.
K. ATTAL-TOUBERT, D. LADIRAY
- 9802** : Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation.
J.-C. DEVILLE
- 9803** : Pour essayer d'en finir avec l'individu Kish.
J.-C. DEVILLE
- 9804** : Une nouvelle (encore une !) méthode de tirage à probabilités inégales.
J.-C. DEVILLE
- 9805** : Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish.
J.-C. DEVILLE
- 9806** : Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel **POULPE**.
N. CARON, J.-C. DEVILLE, O. SAUTORY
- 9807** : Estimation de données régionales à l'aide de techniques d'analyse multidimensionnelle.
K. ATTAL-TOUBERT, O. SAUTORY
- 9808** : Matrices de mobilité et calcul de la précision associée.
N. CARON, C. CHAMBAZ
- 9809** : Échantillonnage et stratification : une étude empirique des gains de précision.
J. LE GUENNEC
- 9810** : Le Kish : les problèmes de réalisation du tirage et de son extrapolation.
C. BERTHIER, N. CARON, B. NEROS
- 9901** : Perte de précision liée au tirage d'un ou plusieurs individus Kish.
N. CARON
- 9902** : Estimation de variance en présence de données imputées : un exemple à partir de l'enquête Panel Européen.
N. CARON
- 0001** : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**) (version actualisée).
S. LOLLIVIER, M. MARPSAT, D. VERGER
- 0002** : Modèles structurels et variables explicatives endogènes.
J.-M. ROBIN
- 0003** : L'enquête 1997-1998 sur le devenir des personnes sorties du RMI - Une présentation de son déroulement.
D. ENEAU, D. GUILLEMOT
- 0004** : Plus d'amis, plus proches ? Essai de comparaison de deux enquêtes peu comparables.
O. GODECHOT
- 0005** : Estimation dans les enquêtes répétées : application à l'Enquête Emploi en Continu.
N. CARON, P. RAVALET
- 0006** : Non-parametric approach to the cost-of-living index.
F. MAGNIEN, J. POUGNARD
- 0101** : Diverses macros **SAS** : Analyse exploratoire des données, Analyse des séries temporelles.
D. LADIRAY
- 0102** : Économétrie linéaire des panels : une introduction.
T. MAGNAC
- 0201** : Application des méthodes de calages à l'enquête EAE-Commerce.
N. CARON
- C 0201** : Comportement face au risque et à l'avenir et accumulation patrimoniale - Bilan d'une expérimentation.
L. ARRONDEL, A. MASSON, D. VERGER
- C 0202** : Enquête Méthodologique Information et Vie Quotidienne - Tome 1 : bilan du test 1, novembre 2002.
J.-A. VALLET, G. BONNET, J.-C. EMIN, J. LEVASSEUR, T. ROCHER, P. VRIGNAUD, X. D'HAULTFOEUILLE, F. MURAT, D. VERGER, P. ZAMORA
- 0203** : General principles for data editing in business surveys and how to optimise it.
P. RIVIERE
- 0301** : Les modèles logit polytomiques non ordonnés : théories et applications.
C. AFSA ESSAFI
- 0401** : Enquête sur le patrimoine des ménages - Synthèse des entretiens monographiques.
V. COHEN, C. DEMMER
- 0402** : La macro **SAS** **CUBE** d'échantillonnage équilibré
S. ROUSSEAU, F. TARDIEU
- 0501** : Correction de la non-réponse et calage de l'enquêtes Santé 2002
N. CARON, S. ROUSSEAU

0502 : Correction de la non-réponse par ré pondération et par imputation
N. CARON

0503 : Introduction à la pratique des indices statistiques - notes de cours
J-P BERTHIER

0601 : La difficile mesure des pratiques dans le domaine du sport et de la culture - bilan d'une opération méthodologique
C. LANDRE, D. VERGER

0801 : Rapport du groupe de réflexion sur la qualité des enquêtes auprès des ménages
D. VERGER

M2013/01 : La régression quantile en pratique
P. GIVORD, X. D'HAULTFOEUILLE

M2014/01 : La microsimulation dynamique : principes généraux et exemples en langage R
D. BLANCHET

M2015/01 : la collecte multimode et le paradigme de l'erreur d'enquête totale
T. RAZAFINDROVONA

M2015/02 : Les méthodes de Pseudo-Panel
M. GUILLERM

M2015/03 : Les méthodes d'estimation de la précision pour les enquêtes ménages de l'Insee tirées dans Octopusse
E. GROS

K. MOUSSALAM

M2016/01 : Le modèle Logit Théorie et application.
C. AFSA

M2016/02 : Les méthodes d'estimation de la précision de l'Enquête Emploi en Continu
E. GROS K.MOUSSALAM

M2016/03 : Exploitation de l'enquête expérimentale Vols, violence et sécurité.
T. RAZAFINDROVONA

M2016/04 : Savoir compter, savoir coder. Bonnes pratiques du statisticien en programmation.
E. L'HOURL R. LE SAOUT B. ROUPPERT

M2016/05 : Les modèles multiniveaux
P. GIVORD M. GUILLERM

M2016/06 : Econométrie spatiale : une introduction pratique
P. GIVORD R. LE SAOUT

M2016/07 : La gestion de la confidentialité pour les données individuelles
M. BERGEAT

M2016/08 : Exploitation de l'enquête expérimentale Logement internet-papier
T. RAZAFINDROVONA

M2017/01 : Exploitation de l'enquête expérimentale Qualité de vie au travail
T. RAZAFINDROVONA

M2018/01 : Estimation avec le score de propension sous R
S. QUANTIN

M2018/02 : Modèles semi-paramétriques de survie en temps continu sous R
S. QUANTIN

M2019/01 : Les méthodes de décomposition appliquées à l'analyse des inégalités
B. BOUTCHENIK E. COUDIN S. MAILLARD

M2020/01 : L'économétrie en grande dimension
J. L'HOURL

M2021/01 : R Tools for JDemetra+ - Seasonal adjustment made easier
A. SMYK A. TCHANG

M2021/02 : Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman
L. CASTELL P. SILLARD

M2021/03 : Conception de questionnaires auto-administrés
H. KOUMARIANOS A. SCHREIBER

M2022/01 : Introduction à la géomatique pour le

statisticien : quelques concepts et outils innovants de gestion, traitement et diffusion de l'information spatiale
F. SEMECURBE E. COUDIN

M2022/02 : Le zonage en unités urbaines 2020
V. COSTEMALLE S. OUJIA C. GUILLO A. CHAUVET

M2023/01 : Les réseaux de neurones appliqués à la statistique publique : méthodes et cas d'usages
D. BABET Q. DELTOUR T. FARIA S. HIMPENS

M2023/02 : Redressements de la première vague de l'enquête epicov : un exemple de correction des effets de sélection dans les enquêtes multimodes
L. CASTELL C. FAVRE-MARTINOZ N. PALIOD P. SILLARD

M2023/03 : Appariements de données individuelles : concepts, méthodes, conseils
L. MALHERBE

M2023/04 : Victimations déclarées et effets de mode : enseignements de l'expérimentation panel multimode de l'enquête cadre de vie et sécurité
LAURA CASTELL MARIE CLERC DELPHINE CROZE STÉPHANE LEGLEYE AMANDINE NOUGARET