
AGREGER LES ÉCHANTILLONS D'UNE ENQUETE MULTIMODE EN LIMITANT L'EFFET DE MESURE : UNE PROPOSITION D'IMPUTATION RAISONNABLE ET PRAGMATIQUE

Stéphane Legleye (), Tiaray Razafindranovona (**), Gaël de Peretti (*)*

() Insee, Division Recueil et traitement e l'information*

*(**) Insee, Division Emploi*

stephane.legleye@insee.fr

Mots-clés : Enquête multimode, agrégation, effet de mesure, effet de sélection, imputation, score de propension

Résumé

Lorsque l'effet de mesure est important dans une enquête multimode qui doit être comparée à une enquête traditionnelle en mode unique, les concepteurs de l'enquête doivent faire des choix. Soit pondérer à nouveau l'enquête afin de tenir compte de l'effet de sélection en équilibrant les deux échantillons spécifiques au mode, soit imputer les réponses recueillies dans l'un des échantillons. Nous proposons une méthode simple pour mettre en œuvre la deuxième stratégie. Nous utilisons des données réelles : deux enquêtes aléatoires indépendantes ont été menées en 2013 pour étudier la victimisation (avec une série de questions communes) : une enquête traditionnelle en face à face (n=14585) et une autre sur Internet/papier (n=12896) qui a donné des taux de victimisation beaucoup plus élevé. Nous étudions une situation fictive où les deux ensembles de données sont combinés sans aucun prétraitement. Nous identifions d'abord les questionnaires Internet appariés à un jumeau face-à-face sur la base des caractéristiques socio-démographiques. Ensuite, nous identifions les paires de jumeaux dans lesquelles certaines victimisations diffèrent et impute les valeurs d'Internet. Les résultats montrent une diminution importante du taux de victimisation dans l'ensemble des données combinées, qui sont plus proches de celles calculées dans le cadre de l'enquête traditionnelle en face à face. L'efficacité, les limites et les combinaisons possibles avec d'autres méthodes sont discutées.

Abstract

When measurement effect is important in a mix-mode survey that has to be compared to a traditional single-mode survey, survey designer have to choices. Either reweight the survey in order to adjust for selection effect by balancing the two mode-specific samples, or to impute the responses collected in of one of the samples. We propose a simple method to implement the second strategy. We use real data: two independent random surveys have been conducted in 2013 to study victimization (with a common set of questions): a traditional face-to-face (n=14585) and another on Internet/paper (n=12896) that yielded much higher victimization rates. We study a fictitious situation where the two datasets are combined without any pretreatment such as inverse propensity weighting. We first identify the Internet questionnaires that have an almost perfect face-to-face twin based on a propensity score matching. Then we identify the pairs of twins in which some victimizations differ and impute the Internet values. The results show an important diminution of the victimization rate in the combined data set, that are closer to the ones computed in the traditional face-to-face survey. Efficiency, limits and possible combination with other methods such as inverse propensity score weighting are discussed.

Article long

1. Introduction

Les enquêtes multimodes sont à la mode depuis quelques années. La première raison est la baisse du taux de participation des enquêtes traditionnelles en face-à-face ou au téléphone [1-3]. Certes, calculer un taux de participation de façon transparente et universelle est une tâche ardue malgré les recommandations diffusées par certains organismes comme l'AAPOR (American association for public opinion research) et qui deviennent des standards [4]. Certes, le lien entre le taux de réponse et la qualité de la mesure est généralement discutable, ce qui en fait un indicateur assez médiocre de la qualité global de la collecte [5-8]. Aux Etats-Unis, le taux de réponse moyen dans les enquêtes téléphoniques aléatoire avoisine les 9% en 2016, sans que les estimations d'indicateurs courants ne soient fortement affectées [3]. En France, dans certaines populations, par exemple les médecins [9-12], obtenir un taux de participation supérieur à 35% est une gageure : pourtant, les résultats ne sont que rarement mis en doute. Malgré tout, afficher un bon taux de réponse reste un objectif désirable et valorisé, et fournir un ou plusieurs modes de collecte adaptés peut y contribuer [13, 14].

La deuxième raison du recours au multimode est l'amélioration de la qualité des données recueillies : il peut être préférable d'utiliser un mode auto-administré et donc d'écarter l'enquêteur pour recueillir certaines informations, notamment des informations personnelles et sensibles. Internet apparaît comme une ressource de choix dans ce cas.

La dernière motivation est évidemment la contrainte budgétaire. Les instituts de statistique et les instituts de recherche sont soumis à de fortes pressions pour basculer tout ou partie de leurs enquêtes ou questionnaires sur Internet, comme l'ont fait les instituts de sondage privés depuis quelques années.

Les deux premières raisons de recourir au multimode trouvent place dans le cadre général de l'erreur totale d'enquête (Total survey error). Le TSE définit la liste de toutes les sources d'erreurs du processus d'enquête [15, 16] que les protocoles multimodes invitent à reconsidérer avec un regard attentif et renouvelé. Les principales sont le biais de non-réponse l'erreur de mesure. L'erreur de mesure se présente souvent à l'esprit dès qu'il est question de multimode et de très nombreuses méthodes ont été développées pour tenter de l'estimer, et de l'éliminer. Toutefois, nous verrons que décider de corriger est difficile.

Nous présenterons ici un prolongement de travaux antérieurs déjà publiés sur l'enquête de victimation Cadre de vie et sécurité (CVS) et sa réplique sur Internet, Vols, violence et sécurité (VVS) menées toutes deux en 2013 [17]. CVS est une enquête annuelle conduite en face à face par l'Insee ; le taux de participation s'élève à 63 % en 2013. Le taux de participation à VVS s'élève à 32 %. Les deux enquêtes ont été réalisées en parallèle : il s'agit de deux échantillons indépendants. Les plans de sondages des deux enquêtes sont différents mais visent à représenter la même population cible, celle des résidents métropolitains francophones âgés de 14 ans et plus en ménages ordinaires.

Nous nous plaçons dans le cas fictif où ces deux enquêtes auraient été les deux branches d'une enquête multimode : les deux échantillons de répondants doivent donc être agrégés avant l'exploitation des résultats. Cela est rendu possible par le tirage indépendant des deux échantillons. Le mode de référence de CVS étant le face-à-face, notre objectif est de proposer une méthode simple permettant de limiter les biais de mesure dans les agrégations de données d'enquête multimode.

2. Approche générale

2.1. Définition des effets de mesure et de sélection

Plaçons-nous dans le cas simple de la comparaison de deux échantillons indépendants A et B définis par le mode de collecte $M=A$ et $M=B$, où l'on cherche à mesurer un jeu de variables cibles Y, et où l'on dispose de variables auxiliaires X liées aux Y et d'un jeu de variables sociodémographiques SD. A est le mode de référence.

L'effet de mode se définit par $EM= BS + BM$. Autrement dit, comme la somme du biais de sélection et du biais de mesure. Le premier se définit comme la contribution de la différence des compositions des échantillons A et B sur la mesure de Y ; le second comme la contribution de la différence des modes sur la

mesure de Y qui n'est pas expliquée par la différence de composition ou, d'une façon plus restrictive, par la différence de mesure induite par le mode pour les individus ayant le même profil socio-démographique SD-X. La première définition fait référence aux méthodes de décomposition économétriques type Oaxaca-Blinder [18] ; elle envisage le problème en termes de variables. La seconde fait plus spécifiquement référence à un problème causal qui envisage davantage le problème en termes de profils des personnes [19].

C'est le biais de mesure que l'on a généralement en tête lorsqu'on parle d'effet de mode. Il se définit conceptuellement comme l'effet causal du mode sur la mesure de Y pour chaque individu : autrement dit comme l'effet de répondre sur le mode B plutôt que A pour chaque répondant. Comme chaque répondant n'a répondu qu'à un seul questionnaire sur un seul mode, cet effet est généralement inestimable (sauf éventuellement dans des plans d'expérience en cross-over où chaque individu est son propre témoin). En revanche on peut l'estimer en moyenne en s'appuyant sur les échantillons de répondants similaires en SD-X sur les deux modes : $BM = E[Y(B) - Y(A) | SD-X]$. Cette estimation causale est celle du modèle contrefactuel de Rubin [20]. La définition doit alors être adaptée au niveau « échantillon » : l'effet causal du mode de collecte sur la réponse à une question, corrigé de l'effet de composition.

Si les échantillons A et B étaient de grande taille et représentatifs de la population cible, si l'attribution du mode aux répondants était strictement aléatoire et si aucune non-réponse totale n'était observée dans les deux modes (le taux de participation à l'enquête sur le mode A étant de 100% comme celui observé sur le mode B), alors $E(Y(A) - Y(B))$ nous donnerait l'effet de mesure. En effet, il n'y aurait aucune sélection des observations sur un mode plutôt que l'autre, les caractéristiques moyennes des répondants dans les deux échantillons seraient identiques et aucun effet de composition serait à l'œuvre.

Dans les cas réels, il y a plusieurs problèmes. D'abord, les échantillons sont de taille souvent réduites, ce qui peut amener à des erreurs d'échantillonnage. Ensuite, A et B souffrent de non-réponse totale, ce qui induit des biais de composition. Si la non-réponse totale ne dépend que de SD-X, et le mode est indépendant de Y, ce qui est le cas des mécanismes d'assignation réguliers [20] alors on peut utiliser plusieurs méthodes équivalentes.

2.2. Estimation du biais de mesure

Parce que le biais de mesure est fondé sur une différence individuelle, on ne peut l'estimer que pour les individus comparables dans les deux modes A et B, ce qui suppose l'existence d'un sous-ensemble non-vides d'individus ayant des profils SD-X similaires, ce que l'on appelle le support commun. L'idée est de se rapprocher au maximum de l'expérimentation randomisée, soit d'une situation hypothétique où l'attribution du mode de collecte est indépendante de Y (idéalement faite avant toute mesure de Y) et des caractéristiques SD-X.

On utilise souvent un score de propension pour comparer globalement les échantillons A et B : on modélise la probabilité de répondre en B plutôt qu'en A conditionnellement à des variables SD et X choisies, c'est-à-dire en lien avec Y et le mode. La propriété équilibrante du score de propension fait que conditionnellement au score l'attribution de M est aléatoire [20] : les deux échantillons A et B présentent alors des distributions SD-X équivalentes conditionnellement au score.

Pour mener l'estimation il faut se ramener au support commun. Cela peut se faire par matching direct ou par matching sur le score de propension ou bien encore par restriction d'un sous-ensemble d'individus dans la plage commune du score de propension [20, 21]. L'estimation ne porte donc quasiment jamais sur la totalité des échantillons A et B et donc *a fortiori* pas sur la population cible. Plus on est exigeant dans le choix des variables SD-X, plus le support commun est constitué d'individus similaires, mais plus il est réduit : on gagne en validité interne ce que l'on perd en validité externe c'est-à-dire en généralisation.

Les méthodes d'estimation sur le support commun sont très nombreuses : pondérations, régressions, régressions pondérées, régressions stratifiées sur le score de propension, matching et régressions etc. [20, 22]. Il existe aussi des méthodes par imputation [23-25].

On trouve parfois des pondérations directes d'un échantillon sur l'autre, par calage ou post-stratification [26] : l'idée est ici encore de neutraliser les biais de composition en SD-X dans les échantillons A et B. Toutefois, cela est souvent insuffisant relativement aux méthodes traditionnelles : il faut en effet en théorie prendre en considération de nombreux croisements des variables SD-X et cela peut conduire à des poids très variables, ce qui peut nuire à la stabilité des estimations. En fait, tout est affaire de précision dans le choix du support commun et de généralisation.

Notez que le biais de sélection peut lui-même être affecté d'un biais de mesure : ce n'est pas parce que le biais de mesure n'est estimé que sur le support commun que les valeurs Y des individus hors support commun ne sont pas différentes suivant le mode, par définition. Le mettre en évidence, par exemple relativement à une variable donnée (disons l'âge) implique des analyses particulières et peut-être d'autres jeux de données.

Au final, corriger le biais de composition par pondération, par exemple, n'est pas une correction de l'effet de mesure, mais seulement un préalable à son estimation : s'il persiste une différence de mesure en Y, l'écart ne peut plus être corrigé ainsi. Conceptuellement, l'effet de mesure est un problème d'imputation et non de poids.

2.3. *Note sur la sélection des variables SD-X à prendre en compte*

Pour modéliser un score de propension lié à un mode de collecte, il suffit en théorie de retenir toutes les variables liées à la fois au mode de collecte et aux variables Y [20]. Dans le doute, il faut privilégier les variables liées aux Y et non au mode de collecte, afin de ne pas intégrer de variable instrumentale biaisant les estimations ultérieures [27]. La première difficulté, inhérente à toute enquête multithématique, est la profusion potentielle des variables Y : il faut s'attacher à les étudier toutes, mais il est impossible de les traiter toutes aussi efficacement en une étape, si elles sont diversement associées à des variables SD-X.

Une difficulté spécifique dans les enquêtes multimodes est le fait que toute variable tirée d'une réponse au questionnaire de l'enquête est potentiellement affectée d'un effet de mesure. Il faut donc se restreindre à celles qui sont apparemment les plus stables [28]... mais lesquelles ? Dans le cas de CVS, toutes les variables d'opinion, qui sont les plus liées aux victimations, sont susceptibles d'un tel effet. Si certaines semblent nécessaires malgré un fort biais apparent, il est possible de les traiter en amont, mais cela nécessite des hypothèses particulières liées à l'effet de mode... Les variables socio-démographiques, portant sur des caractéristiques stables dans le temps et que le répondant a l'habitude de communiquer à des administrations, semblent assez indépendantes à la fois des questions d'opinions liées, des Y et du mode. Ce sont celles qui se rapprochent le plus des variables prétraitement chères à Rubin [29].

2.4. *Application à CVS-VVS agrégé*

Dans notre étude, CVS et VVS seront considérées comme une seule enquête, un seul échantillon. Cela ne modifie en rien l'approche consistant à identifier le support commun. Toutefois, nous nous en tiendrons aux variables socio-démographiques.

3. Méthode

3.1. *Variables sociodémographiques et de calage (SD)*

Les variables socio-démographiques utilisées pour le calage des échantillons sont les suivantes : Sexe, âge (6 tranches, de 14 ans à 70 ans et plus), taille d'unité urbaine (4 modalités), diplôme (5 modalités), le type de ménage (4 modalités), la CSP (9 modalités dont une pour les non-codés), le type de logement (2 modalités) et le statut d'occupation du logement (2 modalités).

3.2. Variables auxiliaires (X)

Les variables auxiliaires potentiellement utiles pour notre analyse sont au nombre de 14 : elles évaluent les opinions des répondants relativement à certains problèmes de société, la présence et l'action de la police, etc.

3.3. Variables cibles (Y)

Les variables cibles mesurant la victimation sont binaires, au nombre de 6 ; elles décrivent des faits survenus au cours des 24 mois précédant l'enquête : vol sans violence, vol avec violence, violence physique, cambriolage, vol de véhicule, menace.

3.4. Procédure et analyses statistiques

Notre objectif global sera d'agréger les deux échantillons en tentant de minimiser les biais de mesure dus au recours à deux modes de collecte. Pour ce faire, nous procéderons en plusieurs temps. Avant toute chose, nous allons évaluer la représentativité des échantillons séparés et agrégés relativement à la population cible telle que décrite par le recensement de la population au moyen des variables de calage. La distribution des variables SD dans le recensement et dans les échantillons de répondants CVS, VVS et CVS-VVS est présentée Tableau 1. L'évaluation de la représentativité se fait par le calcul de distances standardisées et d'indicateurs R [30].

La différence standardisée d est définie suivant [31]. Ainsi, pour une variable continue,

$$d = 100 \times \frac{\bar{X}_i - \bar{X}_j}{0.5 \times (\sqrt{S_i^2} + \sqrt{S_j^2})}$$

Pour la modalité i d'une variable catégorielle :

$$d = 100 \times \frac{X_i - X_j}{0.5 \times (\sqrt{X_i} + \sqrt{X_j})}$$

Pour fournir une mesure synthétique de l'équilibrage d'une variable catégorielle, nous proposons également de calculer la moyenne des valeurs absolues des d de toutes les modalités.

La distance d permet ainsi d'évaluer l'équilibrage (identité de distribution) de deux échantillons (répondants à CVS et recensement de la population, par exemple).

Les indicateurs R pour un échantillon sont calculés à partir de la modélisation de l'appartenance à l'échantillon plutôt qu'à l'échantillon annuel du recensement de 2013 (pondéré) via une régression logistique (le poids de sondage est utilisé pour l'échantillon, les données du recensement étant également pondérées par leur poids au sein des 5 vagues) relativement aux 8 variables de calage socio-démographiques. Le premier indicateur (R^1) prend en compte les effets propres des 8 variables SD, tandis que le second (R^2) prend en compte en plus toutes les interactions bivariées entre ces dernières, soit 44 effets.

Dans un deuxième temps, nous allons identifier les liens existant entre les variables socio-démographiques et auxiliaires et les variables cibles dans l'agrégation CVS-VVS.

Dans un troisième temps, nous allons repérer les individus interrogés sur Internet qui portent le plus l'effet de mesure. Pour ce faire, nous allons appairer les individus interrogés sur Internet et en Face-à-face à partir de leurs caractéristiques socio-démographiques et des variables auxiliaires au sein de l'échantillon agrégé calé sur le recensement. Parmi les appariés, nous identifierons les individus interrogés sur Internet qui répondent différemment de leur homologue interrogé en face-à-face : ce sont eux qui portent l'essentiel de l'effet de mesure. Nous imputerons leurs réponses. Enfin, nous comparerons les estimations de victimations dans CVS (qui est la référence), VVS (qui est une expérimentation), et dans l'échantillon apparié CVS-VVS initial et imputé. Cette comparaison de proportions est doublée d'une comparaison d'effets causaux de la réponse sur Internet plutôt qu'en face-à-face via une régression (pondérée par le poids calé de l'agrégation CVS-VVS) contrôlant le décile de score de propension, les 8 variables SD et les quatre variables auxiliaires les plus stables associées aux variables Y (méthode de régression [20]).

4. Résultats

4.1. Représentativité

Les marges de la population cible et des différents échantillons sont données Tableau 1. Elles montrent que CVS et VVS peinent à interroger des jeunes, que les hommes manquent dans CVS plus que dans VVS, que les tailles d'UU des répondants collent davantage aux marges du RRP dans VVS que CVS, que les types de ménages ou les PCS sont aussi mal représentées dans VVS que CVS, mais que la distribution des types de logements est de meilleure qualité dans VVS que CVS. La synthèse de ces résultats, exprimée en différence standardisée de niveau variable montre que globalement, VVS est un peu plus proche du RRP que CVS, et que l'agrégation occupe une position encore plus avantageuse (Tableau 2).

Si on modélise la représentativité par une mesure multivariée, le R-indicateur, on trouve d'abord les valeurs suivantes (indicateurs simples) : $R^1(\text{CVS})=0.81$, $R^1(\text{VVS})=0.82$, $R^1(\text{CVS-VVS})=0.78$. Sur ces mesures classiques, VVS apparaît ainsi aussi représentative que CVS. En revanche, avec la mesure intégrant les interactions bivariées, VVS et l'agrégation CVS-VVS apparaissent nettement inférieures à CVS : $R^2(\text{CVS})=0.76$; $R^2(\text{VVS})=0.63$ et $R^2(\text{agrégation})=0.69$. Sur ce critère plus contraignant, CVS est donc de qualité supérieure à VVS¹.

4.2. Constitution des échantillons analysés

L'échantillon de référence historique sera CVS calé ; l'échantillon qui y sera comparé est CVS-VVS calé, avec les mêmes paramètres de calage. A titre de comparaison, on pondérera également VVS avec le même calage. Le calage commun dont il est question est le calage classique de CVS : calage simple sur les marges présentées Tableau 1. Les statistiques de poids initiales de CVS calé seul figurent dans le tableau 3. Avant tout traitement, les pourcentages de victimations dans CVS, VVS et dans l'agrégation calées sont présentés Tableau 3 (partie de gauche) : CVS calé présente des prévalences nettement inférieures à celles recueillies dans VVS calé. En moyenne, le taux de victimation y est de 3,12% vs 5,37%. Dans l'agrégation calée avant correction, les prévalences sont à un niveau intermédiaire, mais toujours nettement supérieur à ce qui est mesuré dans CVS calé (4,07% en moyenne). L'analyse de l'effet causal initial est présentée Tableau 8 (partie de gauche). Il est souvent important et toujours significatif.

4.3. Sélection des variables liées aux variables cibles

Toutes ou presque les variables auxiliaires sont liées aux variables cibles ainsi qu'à l'échantillon (CVS ou bien VVS), à partir d'un critère d'une différence standardisée supérieure à 10. Le sexe n'apparaît cependant lié qu'aux violences physiques : c'est la variable socio-démographique la moins liée aux victimations dans l'échantillon CVS-VVS. Nous ne présentons qu'un aperçu des associations pour les plus stables Tableau 4. Ce sont donc de bonnes candidates pour étudier l'effet de mesure dans VVS relativement à CVS. Toutefois, comme explicité plus haut, par prudence, nous les écarterons provisoirement de l'analyse.

4.4. Score de propension

Pour identifier notre support commun nous modélisons l'appartenance à VVS plutôt qu'à CVS parmi les répondants, en pondérant par le poids de sondage. Le modèle de score de propension se restreint donc aux variables SD et intègre toutes les interactions d'ordre 2 entre les variables SD. L'histogramme de la distribution est fourni Figure 1. Le support commun exclut 0 et 1 seules 65 observations en sont exclues

¹ 665 valeurs manquantes de PCS ont été recodées à l'aide d'un hotdeck aléatoire (42 dans CVS et 556 dans VVS).

dans CVS et 52 dans VVS². La procédure d'appariement se fait par sélection du meilleur voisin et égalité stricte des huit variables socio-démographiques.

4.5. *Imputation*

Cet appariement assemble 7253 paires de parfaits jumeaux socio-démographiques, représentant 56% des observations de VVS. L'équilibrage est parfait sur les huit variables socio-démographiques ($d=0$) et il est également relativement bon sur les quatre variables d'opinions relatives à la police (*policer* : $d=8.5$), sa présence (*polpresr* : $d=8.9$), le sentiment d'insécurité dans le quartier (*insecqrqb* : $d=18.1$) et au domicile (*insecurrb* : $d=15.2$). Au sein de ces paires, 31.5% présentent au moins une différence de variable Y, 9% seulement en présentent au moins 2 (Tableau 5).

Au sein de ces 7253 paires de jumeaux SD CVS-VVS, deux choix d'imputations s'offrent à nous, si l'on considère que CVS est l'enquête de référence : soit imputer directement les variables Y de VVS dès qu'elles diffèrent de celles du jumeau CVS ; soit sélectionner approximativement la proportion d'observations qui, après imputation stochastique, permettra de restituer, au sein des observations VVS, la proportion observée dans CVS pour chaque Y. La première méthode permet de rendre totalement identiques les paires SD CVS-VVS : ce seront de parfaits jumeaux SD-Y. Elle est cependant maximaliste car elle impose une égalité parfaite ; or ce qui est visé est plutôt l'égalité au sein de l'ensemble des paires. Pour être totalement légitime, elle nécessiterait un modèle très fort liant les caractéristiques socio-démographiques des individus à leurs victimations et la certitude que le lien mesuré dans CVS est totalement correct. Par ailleurs, elle ne permet pas de calculer une variance d'imputation, car elle est totalement déterministe.

La seconde méthode est plus parcimonieuse et repose sur l'idée suivante. Le modèle stochastique d'imputation utilisé devrait, approximativement, imputer les valeurs manquantes à la moyenne de ce qui est observé dans CVS (contrainte 1), tandis qu'au sein des paires SD, la proportion de $Y=1$ parmi observations VVS est calculée à partir de la proportion de $Y=1$ au sein des vrais jumeaux SD-Y et des jumeaux SD divergeant en Y (contrainte 2). La première proportion sera laissée inchangée par notre procédure, nous n'agissons que sur la seconde en mettant à manquer la proportion idoine compte tenu de ces deux contraintes. Le calcul s'effectue variable Y par variable Y. La sélection des observations à imputer se fait aléatoirement suivant une loi uniforme. Le second intérêt de cette méthode est qu'elle devrait offrir une meilleure adéquation des liens entre variables socio-démographiques et variables d'opinions et victimation.

Le détail des variables concernées figure Tableau 6. La solution stochastique est nettement plus économe que la solution déterministe en nombre d'individus à imputer (8,0% vs 17,7% des répondants de VVS) ou de variables à imputer (1,7% vs 4,0% des valeurs des 6 variables Y).

L'imputation déterministe ne pose aucun problème d'implémentation. La stochastique se fait par la proc MI, en imposant les contraintes suivantes : le modèle d'imputation provient des observations CVS appariées SD ($n=7253$) ; les variables explicatives sont, pour tous les Y, le score de propension, ainsi que les quatre variables auxiliaires X retenues précédemment (on utilise des rangs calculés dans CVS pour chacune).

4.6. *Estimations post imputation*

Les 5 imputations stochastiques calculées sont étudiées conjointement par la suite grâce à la proc MIANALYZE, qui permet de calculer une variance d'imputation et d'en tenir compte, alors que les calculs pour l'imputation déterministe ne sont opérés qu'une fois. La variance d'imputation stochastique négligeable (<0.5%). Nous ne reproduisons pas les estimations post-imputations au sein des paires, car les

² Si l'on modélise le score sans les interactions bivariées, la superposition des histogrammes est excellente, confirmant le faible écart des indicateurs R1 mesurant la représentativité de CVS et VVS relativement au recensement : aucune observation n'est exclue du support commun dans CVS et seules 2 le sont dans VVS.

résultats sont naturellement très bons : l'imputation déterministe comme l'imputation stochastique éliminent presque totalement la différence de prévalence pour chacune des victimations, et aucun t-test n'est significatif. Notez la persistance de petites différences, même pour l'imputation déterministe, dès lors que l'on prend les poids en compte (les différences sont nulles en l'absence de pondération).

Le résultat des imputations sur l'ensemble CVS-VVS agrégé est fourni Tableau 7. L'imputation stochastique fait aussi bien que la déterministe en réduction de l'écart de proportion de chaque Y par rapport aux valeurs avant imputation : en moyenne, la réduction du taux de victimation (pour les 6 variables étudiées ici) est de 11% (contre 12% pour l'imputation déterministe) par rapport à la situation pré-imputation. Relativement à CVS seul, considéré comme la référence, cela correspond à une augmentation de 11% au lieu de 25% initialement. Le succès varie suivant les victimations, celui des vols avec violence (VAVN) étant l'un des plus faibles. Néanmoins, les valeurs imputées (et les individus) ne sont pas les mêmes.

Il est toutefois attendu que l'imputation stochastique reflète mieux la structure des liens entre variables socio-démographiques (et variables d'opinions) et les victimations que l'imputation déterministe. Cela devrait se voir dans la seconde estimation de l'effet de mesure. On peut en effet affiner l'appréciation de l'efficacité des imputations en considérant l'effet causal de l'interrogation sur Internet plutôt qu'en face-à-face au sein de l'échantillon agrégé initial et final (imputé), pour chaque méthode d'imputation. Cette analyse est censée ôter l'essentiel du biais introduit par la différence de structure des échantillons CVS et VVS dans l'ensemble agrégé [20]. Elle est opérée ici sur l'ensemble des observations de l'agrégation, car les observations hors support commun sont très peu nombreuses (cf. supra) : le biais éventuel est extrêmement faible. Les variables prises en compte pour l'ajustement sont les 8 variables socio-démographiques de calage et les rangs des 4 variables d'opinion (calculés séparément dans CVS et VVS). L'analyse est faite avant imputation puis post-imputation (Tableau 8) ; les calculs pour l'imputation déterministe sont faits une seule fois (pas de variance d'imputation, ce qui est fautif). Les régressions opérées sur les 5 imputations stochastiques sont analysées conjointement par la proc MIANALYZE : la variance ajoutée est ici encore tout à fait négligeable (<0,5%).

Les résultats sont synthétisés tableau 8 (les résultats sous forme de moyenne ont été assimilés à des pourcentages). Initialement, l'effet est très fort : par exemple, pour les vols avec violence, on a 1,53% de déclarations en plus sur Internet relativement au face-à-face, toutes choses égales par ailleurs. Cet effet Internet est toujours significatif. Après imputation déterministe, les effets sont tous réduits, mais restent significatifs. Après imputation stochastique, les résultats sont bien meilleurs : la réduction est massive et aucun effet n'est significatif. L'imputation stochastique neutralise mieux l'effet de mesure, comme nous pouvions le supposer en vertu de notre modèle d'imputation.

5. Discussion

Dans le cas hypothétique traité, CVS et VVS sont considérées comme deux branches de la même enquête dont le design impose aléatoirement à chacune des personnes sélectionnées un mode unique (Internet pour VVS et le face-à-face pour CVS). Nous montrons d'abord que la représentativité de l'échantillon VVS est, en première approche aussi bonne que celle de CVS, mais qu'une analyse plus poussée y révèle des faiblesses, tandis que la représentativité de l'échantillon multimode est intermédiaire.

Des analyses montrent que la réponse sur Internet plutôt qu'en face-à-face est associée à un fort effet de mesure pour les 6 variables de victimations étudiées ici (voir aussi : [17]). Nous nous attachons à circonscrire cet effet par imputation d'observations choisies au sein de VVS. Ces individus sont des jumeaux-SD parfaits, sélectionnés par appariement sur le score de propension et contrainte d'égalité sur les 8 variables socio-démographiques de calage. Les imputations sont faites soit déterministiquement au sein des paires, soit de façon stochastique (ce qui conduit à des imputations moins nombreuses : 8.0% vs 17.7% d'individus de VVS concernés ou 1.7% vs 4.0% des valeurs des 6 variables cibles Y).

Notre méthode permet la neutralisation totale de l'effet de mesure sur les jumeaux socio-démographiques ainsi qu'une réduction de la contribution de cet effet dans l'estimation des prévalences de victimations au sein de l'échantillon agrégé CVS-VVS. De plus, elle réduit (pour l'imputation déterministe) ou neutralise totalement (pour l'imputation stochastique) l'effet de mesure tel qu'estimé par régression (méthode de régression) au sein de l'intégralité de l'agrégation CVS-VVS. La variance additionnelle de l'imputation stochastique est négligeable.

Nous recommandons donc d'envisager une méthode stochastique pour des raisons de parcimonie et d'efficacité.

Les points que l'on peut discuter sont nombreux :

1. La méthode doit être appliquée uniquement si on juge l'effet de mesure important ou néfaste pour la diffusion des données (rupture de séries d'enquêtes, ou démonstration d'un effet fort de satisficing, par exemple). Elle est donc postérieure à une estimation d'effet causal de mesure (que nous avons produit ici Tableau 8). Le jugement d'opportunité doit se fonder sur une bonne connaissance des effets de mode pour la thématique concernée. Dans le cas des enquêtes sur des sujets sensibles, il est probable que la référence en termes de mesure soit autoadministrée et non intermédiée : Internet pourrait être supérieur. Si c'est bien le cas, faire en sorte qu'une partie de l'échantillon réponde mieux accroît la qualité globale du recueil, même si effets de mesure subsistent au sein de l'échantillon multimode. Coût et représentativité sont également à prendre en compte.
2. Comme pour toutes les études d'enquêtes multimodes, la méthode est contrainte par le choix de variables stables relativement au mode (ici les 8 variables socio-démographiques de calage) : utiliser d'autres variables associées aux Y et à la réponse nécessiterait un traitement, au minimum un recodage afin que les distributions soient proches dans les échantillons CVS et VVS calés, comme nous l'avons fait pour les variables d'opinion relatives à l'insécurité (le calcul de variables de rang type ridit [32, 33] est aussi possible). Nous avons préféré utiliser ces variables auxiliaires en complément d'ajustement pour les estimations d'effets causaux multivariés, mais le choix peut être discuté.
3. Précisons que l'on pourrait envisager une méthode radicale consistant à imputer toutes les valeurs déclarées par les répondants VVS de façon à ce qu'elles soient similaires à celles déclarées dans CVS. Toutefois, cela reviendrait à n'accorder aucun crédit aux réponses fournies, ce qui est déontologiquement intenable et poserait la question du modèle d'imputation à retenir, tous les individus n'étant pas comparables d'un échantillon à l'autre. Notre approche par appariement est nettement plus parcimonieuse.
4. Les méthodes de calcul du score de propension et d'appariement sont paramétrables et posent un certain nombre de questions :
 - a. Le choix opéré ici d'imposer l'égalité stricte des variables socio-démographiques pour l'appariement rend la prise en compte des interactions bivariées inutile, car l'appariement est en fait direct. La conséquence est la limitation du nombre de paires concernées, ce qui occasionne une réduction du nombre d'imputations à faire, mais également une limitation de l'effet produit. Il permet toutefois de prendre à la lettre la définition de l'effet de mesure au niveau individuel et de permettre une imputation déterministe simple. Relâcher cette contrainte fait sens, car ce qui est visé est bien l'absence d'effet de mesure au niveau échantillon et non pas individu. Mais cela rend plus difficile le choix d'imputer telle valeur d'une variable de victimation pour tel individu de VVS, si un individu de VVS n'est plus associé à un individu de CVS qui lui soit identique. Il reste toutefois possible de limiter les contraintes d'égalité strictes et de choisir un taux d'imputation de façon à obtenir la bonne proportion post imputation : dans ce cas, les interactions bivariées sont a priori bienvenues [29]. Mais le choix est plus arbitraire quant aux individus sur qui faire porter l'imputation : on risque de considérer toutes les observations de VVS dans le support commun.

La méthode oscille donc entre une conception de l'effet de mesure au niveau individuel (via la définition des paires et l'imputation déterministe) ou populationnelle (en fait au niveau de l'échantillon, via l'imputation stochastique). On peut ne définir que des échantillons équilibrés en SD et non des paires de jumeaux-SD et choisir d'imputer stochastiquement les observations de VVS. Toutefois, cela rend plus compliqué l'identification des individus à imputer. La définition des paires se révèle ici très pratique.
 - b. Le second choix est celui d'apparier 1-1 sans remplacement. La contrainte 1-1 permet de définir des paires et donc facilement une imputation déterministe. L'absence de remplacement permet d'augmenter la variance des Y dans les observations CVS appariées. Cela nous paraît préférable à une répétition d'individus.

- c. La conséquence globale de ces choix est que l'effet de mesure identifié au sein des paires est très strict, car il concerne des paires d'individus et non des échantillons « proches ». La méthode ne permet donc ici qu'imparfaitement de distinguer effet de mesure et effet de sélection. Mais en relâchant les contraintes d'égalité, on peut y parvenir.
 - d. L'appariement est fait ici suivant un ordre aléatoire du score de propension. D'autres choix sont possibles, mais cela ne devrait pas modifier beaucoup les résultats au vu de la taille des échantillons considérés.
 - e. Par définition, les individus non-jumeaux SD sont potentiellement concernés par un effet de mesure, même s'ils sont pour une partie d'entre eux des contributeurs à l'effet de sélection. Pour mettre en évidence cet effet de mesure au sein des contributeurs à l'effet de sélection, il faut relâcher des contraintes d'appariement, ce qui menace l'analyse d'imprécision et de biais (les individus que l'on comparerait seraient moins ressemblants) ; mais même si les individus non-appariés étaient des femmes et les appariés tous des hommes, l'effet de mesure pourrait affecter les réponses des femmes sans qu'on puisse le mesurer dans l'échantillon. Des connaissances additionnelles sont alors nécessaires.
5. La méthode peut être étendue à tout type de protocole de collecte multimode, dès lors qu'il existe un recouvrement entre les échantillons de répondants sur les multiples modes. Si l'attribution du mode est exclusive pour certains profils de personnes enquêtées, ou si l'autosélection des répondants pour un mode donné est forte, le recouvrement sera réduit et la portée de la méthode diminuée. Mais il faut s'assurer que le choix du mode n'est pas lié à un phénomène d'autosélection lié aux variables cibles. Dans ce cas, ce choix est non-ignorable et la prise en compte de ce problème est plus complexe, voire impossible de façon correcte. Le fait que l'imputation puisse résoudre le problème n'est donc qu'apparent dans un cas aussi défavorable. Autrement dit, notre méthode ne dispense pas d'un examen attentif des données et du mécanisme de sélection.
 6. La méthode est extensible à tout type de variable Y, continue, ordonnée ou multinomiale. Seules les méthodes d'imputations varieront.
 7. La diffusion des données imputées est un problème si l'on utilise des imputations multiples. Nous n'avons utilisé cette méthode que pour estimer la variance additionnelle. Elle est ici très faible, à la fois à cause du faible nombre de valeurs imputées et des faibles prévalences de victimations. Ce pourrait aussi être le cas dans d'autres configurations : dès lors, la fourniture d'un jeu d'imputation peut être considéré acceptable.
 8. La prise en compte des poids est partielle dans la méthode : imputer les individus n'implique pas que l'égalité de distribution des victimations soit réalisée dans l'ensemble des paires de jumeaux SD (même si ce fut le cas ici, à peu près). Un algorithme complet permettrait d'identifier les individus porteurs de valeurs à imputer et de tenir compte de leur poids pour ne modifier que ceux qui impactent le plus efficacement la mesure. Il reste à définir.
 9. La méthode est utilisable après une première phase de redressement d'un échantillon sur un autre si cela a du sens : par exemple, elle peut être utilisée sur les échantillons CVS et VVS considérés indépendamment, car elle ne modifie pas les poids.

Tableau 1 : marges des différents échantillons en pourcentage (pondérés par les poids de sondage)

	Âge				Sexe				Diplôme				Taille d'UU				Type de ménage				CSP				Type de logement				Statut d'occupation			
	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS	RRP	CVS	VVS	CVS VVS
1	16	8	9	8	48	44	48	46	25	28	17	23	16	17	15	16	19	37	34	35	1	1	1	1	38	46	32	39	38	43	34	38
2	23	22	19	21	52	56	52	54	9	8	8	8	29	32	30	31	9	8	3	6	4	3	3	3	63	54	68	61	62	57	66	62
3	17	16	18	17					22	22	23	22	30	30	30	30	28	27	25	26	9	9	11	10								
4	16	17	19	18					18	16	19	17	25	22	25	23	45	28	38	33	14	13	16	14								
5	14	17	18	18					12	11	15	13									17	16	16	16								
6	15	21	17	19					14	15	18	16									13	11	10	11								
7																					26	34	31	33								
8																					17	12	12	12								

RRP=EAR 2013

Note : 0.6% des CSP ont été recodées dans CVS, contre 5.0% dans VVS.

1-8 : marges des variables socio-démographiques SD :

Âge : 1=14-24 ; 2=25-39 ; 3=40-49 ; 4=50-59 ; 5=60-69 ; 6=70+

Sexe : 1=homme ; 2=femme

Diplôme : 1=Aucun diplôme ou CEP ; 2=Brevet ; 3=CAP, BEP ; 4=Bac ; 5=Bac+2 ; 5=Bac+3+

TUU : 1=Paris ; 2≥ 100 000 h ; 3=2 000-100 000 h ; 4=rural

Type de ménage : 1=personne seule ; 2=famille monoparentale ; 3=couple sans enfant ; 4=reste

CSP : 1=agriculteur ; 2=artisans commerçants ; 3=cadres et professions intellectuelles supérieures ; 4=professions intermédiaires ; 5=employés ; 6=ouvriers ; 7=retraités ; 8=autres inactifs.

Type de logement : 1=maison ; 2=appartement

Statut d'occupation : 1=propriétaire ; 2=locataire.

Tableau 2 : synthèse des écarts de distributions des variables de calage dans les trois échantillons pondérés par le poids de sondage (différences standardisées de niveau variable d) et R-indicateurs relativement au recensement

	CVS	VVS	Agrégation
sexe	9.4	9.7	9.3
age	7.2	0.4	4.5
Diplôme	4.0	8.7	3.5
Tuu	4.1	0.9	2.5
Typmen	20.0	19.8	19.7
CSP	5.8	6.6	5.7
Typlog	16.8	10.5	6.1
Stoc	9.3	9.8	1.8
Moyenne des d	9.6	8.3	6.6
R^1 - sans interaction	0.81	0.82	0.78
R^2 - avec interactions	0.76	0.63	0.69

R^1 et R^2 : cf. section Procédures

Tableau 3 : distribution des poids des différents échantillons

	Variable	Min	Max	CV	Sum
CVS seul	pnorm	0.1	6.7	59.8	14585.0
VVS seul	pnorm	0.1	9.0	66.2	12896.0
CVS-VVS : VVS	pcalnorm	0.03	21.7	113.2	14848.3
CVS-VVS : VVS	pcalnorm	0.05	4.0	61.2	12632.7
CVS-VVS : CVS-VVS	pcalnorm	0.03	21.7	93.5	27481.0

Tableau 4 : Synthèse des associations entre les variables Y et les variables sociodémographiques (SD) et auxiliaires (X)

		vavn	vsvn	vphyn	vlogn	vvehn	men	Nb d>10	
		Df	d	d	d	d	d		
mar_age	Age	6	14.8	10.9	29.3	2.5	21.5	21.7	5
mar_sexe	Sexe	2	6.2	1.7	14.4	1.1	7.5	6.6	1
mar_diplometrnum	Diplôme	6	11.7	15.3	8.7	7.7	15.6	14.8	4
mar_tuutrnum	Taille UU	4	22.2	17.2	7.8	13.8	22.7	7.1	4
mar_typ_mennum	Type ménage	4	8.2	6.0	15.3	6.4	19.3	15.8	3
mar_cspr	Groupe social	8	8.1	8.8	16.6	6.1	13.6	11.9	3
mar_typlogx	Logement	2	30.0	17.3	28.7	10.6	24.4	12.8	6
mar_stocx	Statut occ.	2	18.0	3.3	43.5	18.4	16.5	15.7	5
Insecurdrb	Insécurité domicile	2	32.6	25.8	28.3	42.2	14.7	36.1	6
insecurqrb	Insécurité quartier	2	64.6	38.1	41.7	36.1	30.6	52.4	6
policer	Opinion police	5	19.9	8.6	15.4	11.3	12.6	18.8	6
polpresr	Présence police	5	18.6	9.7	16.4	17.6	11.2	15.8	6

Base : CVS-VVS agrégé avec le poids de sondage

Tableau 5 : Ecart de réponses aux variables cibles Y au sein des paires CVS-VVS

Nombre de différences	Fréquence	Pourcentage
0	4967	68.48
1	1650	22.75
2	500	6.89
3	101	1.39
4	27	0.37
5	6	0.08
6	2	0.03
≥1	2286	31.52
N	7253	

Tableau 6 : Taux d'imputations dans VVS

Valeurs imputées par individu	Imputations déterministes		Imputations stochastiques	
	Fréquence	%	Fréquence	%
0	10610	82.3	11865	92.0
1	1650	12.8	826	6.4
2	500	3.9	169	1.3
3	101	0.8	31	0.2
4	27	0.2	5	0.0
5	6	0.0		
6	2	0.0		
≥1	2286	17.7	1031	8.0
N	12896		12896	
Détail par variable	Fréquence	%	Fréquence	%
Vavn	211	1.6	121	0.9
Vsvn	583	4.5	366	2.8
Vphyn	283	2.2	76	0.6
Vlogn	703	5.5	297	2.3
Vvehn	582	4.5	56	0.4
Men	741	5.7	361	2.8
Total	3103		1277	
Taux (valeurs)		4.0		1.7
N	12896 × 6		12896 × 6	

Vavn : vol avec violence ; vsvn : vol sans violence ; vphyn : violence physique ; vlogn : cambriolage ; vvehn : vol de véhicule ; men : menace

Tableau 7 : Prévalence des victimations étudiées dans CVS, VVS, l'agrégation CVS-VVS avant et après imputation

	Référence		Avant correction					Après correction déterministe				Après correction stochastique			
	CVS seul (calé)		VVS		Agrégation			Agrégation				Agrégation			
	%	Stderr	%	Stderr	%	Stderr	VR2	%	Stderr	VR1	VR2	%	Stderr	VR1	VR2
vavn	0.99%	0.08%	2.88%	0.15%	1.63%	0.08%	1.64	1.33%	0.07%	0.82	1.34	1.33%	0.07%	0.82	1.34
vsvn	2.78%	0.14%	6.44%	0.22%	4.16%	0.12%	1.50	3.35%	0.11%	0.80	1.21	3.32%	0.11%	0.80	1.20
vphyn	2.20%	0.12%	2.90%	0.15%	2.28%	0.09%	1.04	2.20%	0.09%	0.96	1.00	2.17%	0.09%	0.95	0.99
vlogn	3.74%	0.16%	6.44%	0.22%	4.89%	0.13%	1.31	4.17%	0.12%	0.85	1.11	4.31%	0.12%	0.88	1.15
vvehn	4.52%	0.17%	5.89%	0.21%	5.06%	0.13%	1.12	4.86%	0.13%	0.96	1.08	4.97%	0.13%	0.98	1.10
men	4.53%	0.17%	7.68%	0.23%	5.48%	0.14%	1.21	4.88%	0.13%	0.89	1.08	4.79%	0.13%	0.87	1.06
Moyenne	3.13%		5.37%		3.92%		1.25	3.46%		0.88	1.11	3.48%		0.89	1.11

Vavn : vol avec violence ; vsvn : vol sans violence ; vphyn : violence physique ; vlogn : cambriolage ; vvehn : vol de véhicule ; men : menace

VR1 : ratio par rapport à l'agrégation non imputée

VR2 : ratio par rapport à CVS calé seul (CVS classique)

Tableau 8 : effet « causal » de la réponse à Internet obtenu par régression sur l'agrégation CVS-VVS

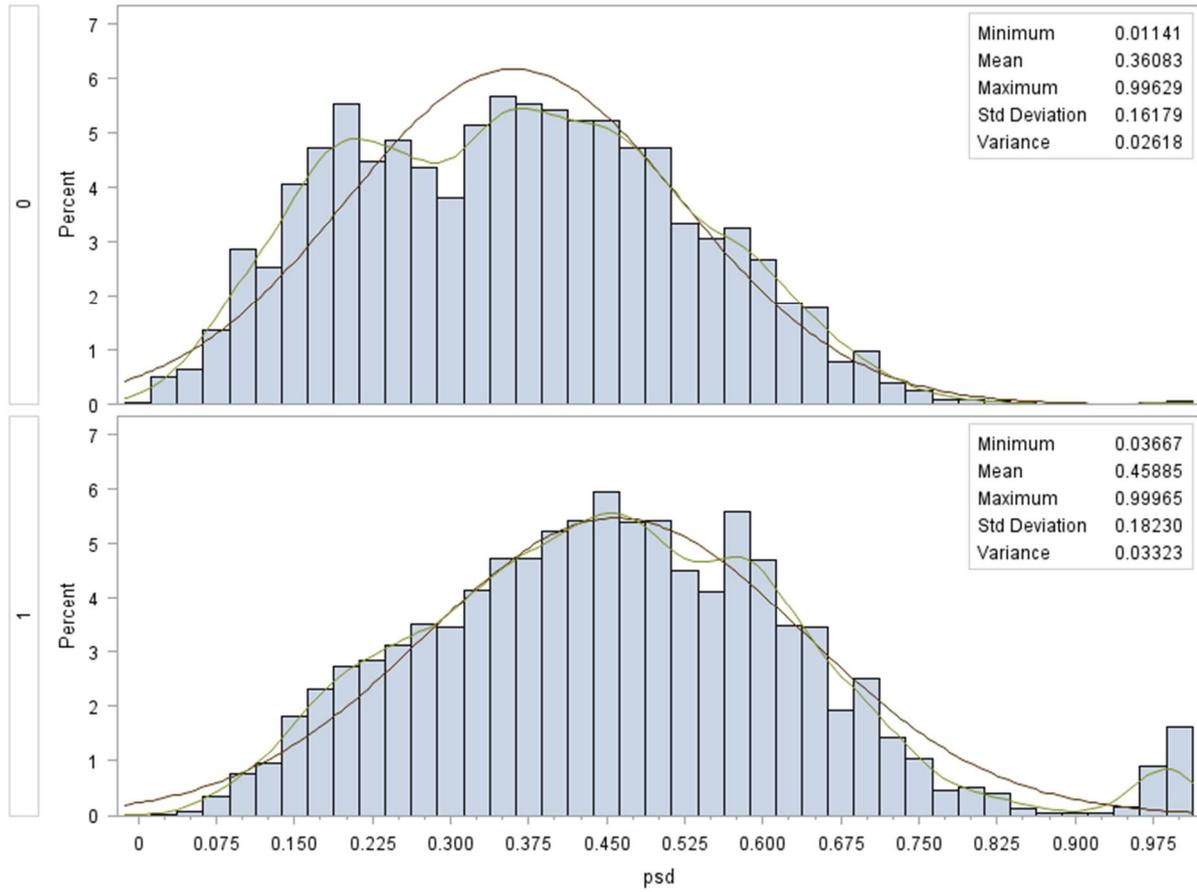
Dependent	Pré imputation			Post imputation							
	%	StdErr	P	Déterministe				Stochastique			
				%	StdErr	P	R*	%	StdErr	P	R*
vavn	1.53%	0.16%	0.000	0.81%	0.15%	0.989	0.53	0.02%	0.16%	0.897	0.01
vsvn	3.55%	0.25%	0.000	1.62%	0.23%	0.323	0.46	0.11%	0.27%	0.679	0.03
vphyn	0.65%	0.19%	0.000	0.49%	0.19%	0.121	0.75	0.23%	0.22%	0.294	0.36
vlogn	3.05%	0.27%	0.000	0.91%	0.26%	0.670	0.30	0.46%	0.34%	0.176	0.15
vvehn	1.51%	0.26%	0.000	0.50%	0.28%	0.454	0.33	0.34%	0.35%	0.333	0.22
men	2.67%	0.29%	0.000	1.54%	0.28%	0.253	0.58	0.01%	0.34%	0.978	0.00
MOYENNE	2.16%			0.98%			0.45	0.20%			0.09

Vavn : vol avec violence ; vsvn : vol sans violence ; vphyn : violence physique ; vlogn : cambriolage ; vvehn : vol de véhicule ; men : menace * : Ratio par rapport à la valeur pré-imputée

Régression linéaire ajustée sur le décile de score de propension, les 8 variables de calage et les rangs des quatre variables d'opinions

La régression pour les imputations stochastiques est faite sur les 5 imputations, la variance tient en tient compte.

Figure 1 : Histogramme de répartition du score de propension



Bibliographie **Calibri 13**

1. Jonas, M. *7 things I learnt about survey response rates*. 2016 [cited 2018 April]; Available from: natcen.ac.uk/blog/7-things-i-learnt-about-survey-response-rates.
2. Czajka, J.L. and A. Beyler, *Declining response rate in federal surveys: trends and implications*. 2016, U.S. Department of Health & Human Services. p. 86.
3. Pew Research Center, *What low response rates mean for telephone surveys*. 2017.
4. Skalland, B., *An alternative to the response rate for measuring a survey's realization rate of the target population*. Public Opinion Quarterly, 2011. **75**(1): p. 89-98.
5. Davern, M.D., et al., *Are lower response rates hazardous to your health survey? An analysis of three state health surveys*. Health Serv Res, 2010. **45**(5): p. 1324-1344.
6. Groves, R.M., *Nonresponse Rates And Nonresponse Bias In Household Surveys*. Public Opinion Quarterly, 2006. **70**(5): p. 646-675.
7. Davern, M.E., *Nonresponse Rates are a Problematic Indicator of Nonresponse Bias in Survey Research*. Health Serv Res, 2013. **48**(3): p. 905-912.
8. Groves, R.M. and E. Peytcheva, *The impact of nonresponse rates on nonresponse bias a meta-analysis*. Public Opinion Quarterly, 2008. **72**(2): p. 167-189.
9. Gautier, A., ed. *Baromètre santé médecins généralistes [Health barometer among general practitioners]*. Baromètre santé. 2011, INPES: Saint-Denis. 266.
10. Legleye, S., et al., *Surveying end-of-life medical decisions in France: evaluation of an innovative mixed-mode data collection strategy*. Interactive journal of medical research, 2016. **5**(1): p. 8.
11. Legleye, S., et al., *A randomized trial of survey participation in a national random sample of general practitioners and gynecologists in France*. Rev Epidemiol Sante Publique, 2014. **62**(4): p. 249-55.
12. Verger, P., et al., *Attitudes et pratiques des médecins généralistes de ville relatives à la vaccination en général et à celle contre la grippe A/H1N1 en 2009. Une étude auprès d'un panel national dans trois régions*. Etudes et Résultats, 2011. **770**.
13. Olson, K., J. Smyth, and H.M. Wood, *Does giving people their preferred survey mode actually increase survey participation rates? An Experimental Examination*. 2012, University of Nebraska: Lincoln.
14. Smyth, J., K. Olson, and A. Kasabian, *The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement*. Survey Research Methods, 2014. **8**(3): p. 137-152.
15. Biemer, P.P., *Total survey error, Design, implementation, and evaluation*. Public Opinion Quarterly, 2010. **74**(5): p. 817-848.
16. Groves, R.M. and L.E. Lyberg, *Total survey error: past, present and future*. Public Opinion Quarterly, 2010. **74**(5): p. 849-879.
17. Razafindranovona, T., *Exploitation de l'enquête expérimentale Vols, violence et sécurité*, in *Document de travail*. 2016, Insee: Paris. p. 48.
18. Fortin, N., T. Lemieux, and S. Firpo, *Decomposition methods in econometrics*, in *Handbook of Labor Economics*, D. Card and O. Ashenfelter, Editors. 2011, Elsevier North Holland: Cambridge, MA. p. 1-102.
19. Nopo, H., *Matching as a tool to decompose wage gaps*, in *Discussion paper series*. 2004, IZA: Bonn.
20. Imbens, G.W. and D.B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. 2015, Cambridge, MA: Cambridge University Press
21. Crump, R.K., et al., *Moving the Goalposts: Addressing Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand*, in *Discussion paper series*. 2006, Institute for the Study of Labor (IZA): Bonn, Germany.
22. Morgan, S.L. and C. Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 1st ed. Analytical Methods for Social Research, ed. M.R. Alvarez, N.L. Beck, and L.W. Lawrence. 2007: Cambridge University Press.
23. Austin, P.C., *Double propensity-score adjustment: A solution to design bias or bias due to incomplete matching*. Statistical methods in medical research, 2014. **0**(0): p. 1-33.
24. Park, S., J.K. Kim, and S. Park, *An imputation approach for handling mixed-mode surveys*. The annals of applied statistics, 2016. **10**(2): p. 1063-1085.
25. Suzer-Gurtekin, Z.T., *Investigating the Bias Properties of Alternative Statistical Inference Methods in Mixed-Mode Surveys*. 2013, University of Michigan.
26. Deville, J. and C.-E. Sarndal, *Calibration estimators in survey sampling*. Journal of the

- American Statistical Association, 1992. **87**(418): p. 376-382.
27. Myers, J.A., et al., *Effects of Adjusting for Instrumental Variables on Bias and Precision of Effect Estimates*. American Journal of Epidemiology, 2011. **174**(11): p. 1213-1222.
 28. Vannieuwenhuyze, J. and G. Loosveldt, *Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects*. Sociological Methods and Research, 2013. **42**(1).
 29. Rubin, D.B., *The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials*. Statistics in Medicine, 2007. **26**: p. 20-36.
 30. Bethlehem, J., F. Cobben, and B. schouten, *Des indicateurs de la représentativité aux enquêtes*. Techniques d'Enquêtes, 2009(Recueil du symposium 2008 de Statistique Canada): p. 1-10.
 31. Austin, P.C. and E.A. Stuart, *Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies*. Stat Med, 2015. **34**(28): p. 3661-79.
 32. Bross, I.D.J., *How to use riddit analysis*. Biometrics, 1958. **14**(1): p. 18-38.
 33. Hayes, L.J. and G. Berry, *Sampling variability of the Kunst-Mackenbach relative index of inequality*. J Epidemiol Community Health, 2002. **56**(10): p. 762-5.