

***Utiliser des enquêtes ouvertes sur Internet pour améliorer le redressement d'une enquête  
téléphonique probabiliste ?  
Le cas de Virage et de la population LGBT***

Auteurs : Stéphane Legleye, Ined&Université Paris-Saclay, University of Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France, Géraldine Charrance, Ined ; Christelle Hamel, Ined et l'équipe Virage

***Résumé court***

Le taux de participation aux enquêtes en population générale baisse depuis de nombreuses années, faisant craindre une augmentation parallèle des biais. Les enquêtes sur les sujets sensibles ou visant une bonne couverture de sous-populations minoritaires sont particulièrement concernées.

L'enquête Virage de l'INED, aléatoire téléphonique, porte sur les violences. Or, certaines sous-populations visées sont très concernées par les violences mais potentiellement mal couvertes : les victimes de violences graves et les lesbiennes gays bisexuels et transsexuels (LGBT) dont une fraction a un mode de vie limitant les chances de contact ou de participation à une enquête en population générale. La petite taille de ces populations conduit de plus à des effectifs réduits dans l'échantillon. On dénombre ainsi 503 personnes se définissant comme LGBT dans cet échantillon de 27 268 répondants.

Deux enquêtes de volontaires ont été réalisées sur Internet, un à destination des LGBT (n=7114 questionnaires complets sur 10612 connexions) et un à destination des victimes de violence (n=796 questionnaires complets sur 2129 connexions), appuyées par des campagnes de communication très actives sur les réseaux sociaux et avec les principales associations concernées.

Nous proposons une méthode pour : 1/ ré-estimer la taille des populations des LGBT et des victimes de violences au sein de la population cible ; 2/ intégrer une partie des questionnaires Internet LGBT ou victimes de violences sans introduire d'effet de mesure lié au mode de collecte. La présentation se concentrera sur le cas des LGBT, numériquement plus intéressant, mais l'extension de ce travail à la population des victimes de violences a été effectuée.

Mots clefs : redressement ; enquête multi-source ; LGBT ; téléphone ; aléatoire ; méthodologie

## ***Introduction***

---

L'enquête Virage a été réalisée à l'INED en 2015 après une phase de tests et de pilote en 2013 et 2014. L'objectif de l'enquête est de fournir des estimations des prévalences des violences au sein de la population générale, en assurant une bonne couverture des sous-populations particulièrement exposées, et en fournissant un niveau de détail sur la contextualisation des violences subies qui fait défaut dans les enquêtes menées traditionnellement sur le sujet. Ainsi par exemple, les violences sexuelles et au sein du couple y font l'objet de questionnements particuliers.

La méthodologie est classique : génération aléatoire de numéros de téléphone, filtrage sur les racines ARCEP (Autorité de régulation des communications et des postes) pour éliminer les plages de numéros non ouvertes à l'exploitation commerciale, passage dans l'annuaire inversé pour retrouver des adresses le cas échéant, puis filtrage des numéros non-attribués et des fax par un automate d'appel. La procédure repose ensuite sur un protocole d'appels autorisant jusqu'à 20 tentatives de contact sur un numéro, en variant les plages horaires et les jours d'appels, en semaine de 10h à 21h et le samedi de 10h à 16h. Les refus avant sélection sont rappelés deux fois et les refus des personnes sélectionnées le sont une fois. Lors de l'interrogation de la personne sélectionnée, l'équipement téléphonique et son usage sont décrits par le répondant. L'interrogation sur un téléphone mobile ne se fait que si la personne déclare ne pas avoir de ligne fixe à son domicile sur laquelle elle prend des appels lorsqu'elle ne connaît pas le numéro ; sinon, la personne est exclue. D'autres choix auraient pu être faits : soit interroger la personne quel que soit son équipement et l'usage de cet équipement (choix du Baromètre santé 2014), ce qui augmente la part des mobiles interrogés ; soit refuser les interrogations sur mobile lorsque le répondant dispose d'une ligne fixe, quel qu'en soit l'usage pour recevoir des appels, ce qui diminue drastiquement le nombre de répondants interrogés sur mobiles. Notre choix est intermédiaire et est un compromis : le questionnaire étant très long (1h02 mn en moyenne sans la phase contact), il a été jugé plus confortable pour les enquêteurs et les enquêtés de privilégier la téléphonie fixe lorsque c'était possible. Au total, 23% des interviews ont été effectuées sur mobile (calcul effectué sans pondération).

Le poids de sondage est déterminé par le nombre d'éligibles dans le ménage et le nombre de lignes fixes et mobiles permettant de joindre la personne sélectionnée. Dans l'échantillon pondéré par le poids de sondage, la proportion d'individus ne disposant que d'un mobile est de 7,4% ; or, d'après le baromètre du numérique, enquête menée en juin 2015 par le CREDOC, la proportion est estimée à 12,5% au sein de la tranche d'âge de la population cible. Le poids de sondage est alors corrigé en conséquence.

À partir de ce poids de sondage corrigé, un premier redressement de l'enquête a été effectué par calage sur marges. Compte tenu de la thématique de l'enquête et du fait que les analyses seront très souvent menées par sexe, il convient de réaliser la post-stratification par sexe, assurant une meilleure représentativité de chaque population (masculine et féminine). Celle-ci a donc été réalisée sur les critères suivantes : classe d'âge (5 modalités), niveau du diplôme le plus élevé (5 modalités), catégorie socioprofessionnelle (8 modalités), composition du ménage (5 modalités), région de résidence (22 modalités), lieu de naissance (France, étranger).

Le taux de participation dans Virage s'élève à 49%, calculé selon les recommandations de l'American Association for Public Opinion Research(1).

Pour simplifier la présentation, nous nous concentrerons sur le cas des personnes se déclarant LGBT. Une partie de notre analyse pourrait toutefois s'appliquer sans beaucoup de variation aux déclarations de violence et en particulier aux plus sensibles d'entre elles.

Dans l'enquête, l'orientation sexuelle est définie par le répondant en réponse à la question suivante : « Actuellement, vous vous définissez comme... » :

1. hétérosexuel-le (si ne comprend pas ; personnes qui ont des relations sexuelles avec des personnes du sexe opposé)
2. homosexuel-le, Gay (si homme), Lesbienne (si femmes) (si ne comprend pas ; personnes qui ont des relations sexuelles avec des personnes du même sexe qu'eux-mêmes)
3. bisexuel-le (si ne comprend pas ; personnes qui ont des relations sexuelles avec des personnes du même sexe mais aussi du sexe opposé)
4. vous ne vous définissez pas par rapport à votre sexualité
5. Autre, précisez
6. Non-réponse

Dans ce travail nous avons considéré les personnes ayant répondu les modalités (2) ou (3). Nous utilisons LGBT par abus de langage car seuls les LGB ont été inclus dans l'étude, n'ayant pas pu recueillir d'information sur le fait que les individus soient transsexuels ou non.

## *Contexte et objectifs du travail*

---

Au total, 503 individus se déclarent LGBT(T) dans l'échantillon de répondants et l'estimation de leur proportion dans la population cible s'élève à 1.56%, soit 609 410 individus sur 39 113 609 individus (chiffres obtenus après calage).

Nous faisons l'hypothèse que cette estimation est une sous-estimation pour deux raisons. D'abord, il existe un problème de mesure. La proportion d'individus déclarant ne pas vouloir répondre à la question portant sur l'orientation sexuelle, est beaucoup plus élevée que celle des personnes ayant répondu par l'affirmative (8.7% vs 1.6%). La question suscite donc des réticences et il est fort probable qu'une partie du contingent des réticents à répondre soit en fait composée de personnes qui auraient dû répondre par l'affirmative. Cette interprétation s'appuie sur deux constats. En premier lieu, la littérature sur les effets de mode de collecte insiste sur le fait que la présence d'un enquêteur, de visu ou au téléphone, diminue les chances de répondre par l'affirmative à des questions susceptibles de révéler des comportements ou des statuts stigmatisants ou déviants. Cela est notamment vrai pour les comportements et les orientations sexuelles (2-8). Le sexe de l'enquêteur peut de plus diminuer la volonté de confier de tels renseignements, ce qui est vraisemblablement le cas dans Virage pour les femmes, qui se déclarent moins souvent LGBT lorsqu'elles s'adressent à un enquêteur qu'à une enquêtrice (1.5% vs 1.0%, OR=1.48, 95%CI=[1.06 ; 2.07]), l'effet étant similaire pour les hommes, mais non significatif (1.7% vs 1.3%, OR=1.26, 95%CI=[0.93 ; 1.71]). Ces effets ne sont pas confondus par des caractéristiques de l'enquêté (âge et diplôme) ou de l'enquêteur (âge, diplôme et ancienneté dans le métier) ni par le rang d'appel<sup>1</sup> : OR=1.38, 95%CI=[0.92 ; 2.05] chez les femmes, OR=1.57, 95%CI=[1.08 ; 2.28] chez les hommes). Il est donc avéré qu'une partie des LGBT n'a pas déclaré son orientation LGBT dans l'enquête.

Deuxièmement, il existe un problème de non-réponse totale : une partie des LGBT, minoritaire sans doute, est réputée avoir un mode de vie diminuant leur accessibilité via le téléphone tandis qu'une autre, sans doute également minoritaire, s'inscrit dans un ensemble de revendications militantes de type communautaire réduisant l'acceptation des enquêtes en population générale en raison de leur caractère général et non à destinée de la communauté. Ces sous-groupes peuvent être négligeables du point de vue numérique mais leurs caractéristiques de vie les exposent sans doute plus que les autres à des violences particulières (9, 10). Aussi, les exclure de l'enquête pourrait biaiser les résultats. L'analyse montre que la déclaration d'orientation LGBT diminue avec le rang d'appel : plus il a fallu un grand nombre de tentatives pour interroger l'enquêté, moins il a tendance à se déclarer LGBT. A l'inverse, les moins de 39 ans sont surreprésentés dans les questionnaires ayant nécessité un grand nombre de tentatives d'appels. Une analyse multivariée montre que notamment parmi les 20% de répondants les plus difficiles à joindre (suivant le nombre d'appels nécessaires à la passation du questionnaire), la part de LGBT parmi les moins de 39 ans est plus réduite qu'au sein des 20% de répondants les plus faciles à joindre (OR=0.53, 95%CI=[0.28 ; 1.00]). Cela suggère bien que l'accessibilité au téléphone est réduite pour les jeunes LGBT, en tout cas pour une partie d'entre eux ; cette différence avec la population majoritaire reste cependant réduite.

Enfin, ces chiffres sont à comparer aux estimations d'autres pays, comme les USA : selon le site Wikipedia, en 2012, environ 3.8% de la population adulte est LGBT, ce qui est pratiquement le double de ce qui est mesuré dans Virage pour l'orientation sexuelle déclarée et supérieur à ce qu'on mesure pour la population qui a déclaré une expérience homosexuelle au cours de sa vie (2.5%) ([https://en.wikipedia.org/wiki/LGBT\\_demographics\\_of\\_the\\_United\\_States](https://en.wikipedia.org/wiki/LGBT_demographics_of_the_United_States)). Ces chiffres sont en accord avec ceux de l'étude de Gates, du Williams Institute, rendue publique en 2011 (11). Bien que cette estimation soit produite par un démographe lui-même transgenre et doive être considérée avec prudence, elle utilise des statistiques de six enquêtes aléatoires publiées et il est peu probable qu'elle soit grossièrement fautive. Ainsi,

---

<sup>1</sup> Ces estimations se fondent sur les questionnaires administrés par un seul enquêteur (soit 12716 femmes et 9583 hommes, soit 81% des observations). L'effet enquêteur (variance due à l'identité de l'enquêteur) étant négligeable (coefficient de corrélation intraclasse inférieur à 1%), ils ont été obtenus par des régressions logistiques classiques.

nous n'avons pas d'éléments pour expliquer pourquoi l'estimation française est près de moitié inférieure à celle des USA sauf à considérer, qu'au moins en partie, la méthodologie des enquêtes puisse y contribuer.

L'enquête LGBT Internet, en comparaison, a été diffusée très largement auprès de la communauté LGBT via un partenariat très étroit avec la quasi intégralité des associations communautaires et s'est appuyée sur des campagnes de publicité et d'information sur les supports les plus variés : flyers, affiches, stylos dans de nombreux lieux de sociabilité LGBT, bannières sur les sites Internet, annonces sur les newsletters, éditos dans les journaux et les sites internet spécialisés, relais sur les comptes facebook, etc. La collecte a duré trois mois et a fait l'objet de très nombreuses actions de communications. La couverture de la population non couverte par le téléphone est donc très bonne, tout comme celle de la population LGBT en général. Cela est particulièrement vrai pour la partie la plus inaccessible au téléphone.

L'objectif de ce travail est de corriger du seul défaut de non-réponse totale (ou du biais de couverture). La question du biais de mesure sera discutée plus loin. Nous allons d'abord recalculer l'effectif de la population LGBT puis intégrer des questionnaires Internet pour augmenter la taille de l'échantillon exploitable des LGBT dans l'enquête Virage.

### ***Calcul des marges additionnelles***

---

Dans les deux enquêtes, téléphone et internet, nous connaissons la fréquentation des lieux de rencontre LGBT des répondants. Nous connaissons également l'équipement et l'usage du téléphone des répondants. La première idée sur laquelle repose cette étude est de fournir une meilleure estimation de la taille de la population LGBT et donc d'estimer une marge pour cette sous-population pour recalculer l'enquête téléphonique. Les hypothèses sous-jacentes à ces calculs sont discutées en détail dans la section *Discussion*.

Estimons d'abord, dans l'enquête téléphonique, la proportion de LGBT fréquentant les lieux gays et lesbiens (cafés, discothèques, associations, lieux de rencontre, y compris sites sur Internet). Celle-ci s'établit à 43.5%. Il en découle une estimation de la taille de la population exposée à l'enquête Internet, qui doit être de 43.5% de la taille de la population LGBT estimée d'après l'enquête téléphonique. Nous allons compléter cette marge en considérant les individus répondants aux mêmes critères mais non joignables par téléphone dans l'enquête Internet.

Au sein de l'enquête LGBT Internet, parmi les 5790 individus dans la population cible (individus âgés de 20-69 ans résidant en métropole et résidant en logement ordinaire), 82.8% fréquentent des lieux LGBT. Parmi ces derniers, 90.1% fréquentent des sites Internet LGBT : ceux-là ont une probabilité maximale d'avoir été exposés à l'enquête Internet. Parmi ces derniers enfin, 29.1% déclarent ne pas être joignables par téléphone.

Une première estimation de la taille de la population LGBT non joignable par le téléphone s'élève donc à  $29,1\% \times 90,1\% \times 82,8\% \times 43,5\%$  soit 9.42% de la taille estimée de la population LGBT fournie par l'enquête téléphonique. Autrement dit, il faut multiplier les effectifs de LGBT estimés par l'enquête téléphonique par 1.0942, soit une augmentation de 57 430 individus.

Toutefois, nous pouvons estimer une hypothèse médiane et une hypothèse haute. Pour l'hypothèse médiane, on n'exclut pas les individus répondants qui ne fréquentent pas les sites Internet LGBT parmi ceux qui fréquentent les lieux LGBT. Le correctif à apporter à la marge est alors de  $29,5\% \times 82,8\% \times 43,5\%$  soit 10.6%. Pour l'hypothèse haute, on considère tous les individus susceptibles d'avoir été exposés à l'enquête Internet parce qu'ils fréquentent soit des lieux LGBT, soit des sites Internet LGBT, soit 97,0% de l'échantillon de répondants Internet. Parmi ces derniers, 30.4% disent ne pas être joignables au téléphone. Le correctif serait donc de  $30,4\% \times 97,0\% \times 43,5\%$  soit 12.8%, soit 78 200 personnes supplémentaires.

Ces chiffres sont bruts et ne tiennent pas compte du sexe des répondants. Or les communautés gays et lesbiennes ne sont pas structurées de la même manière, la communauté masculine étant sans doute plus visible et plus structurée, notamment sur Internet. De plus, la proportion d'homosexuels est plus importante chez les hommes que parmi les femmes, dans notre enquête comme dans celles menées dans d'autres pays (Gates, 2011). Enfin, il est probable que des différences en termes d'âges puissent également jouer. Nos calculs ont donc été affinés suivant ces variables de stratification. Le tableau 1 présente ces diverses hypothèses.

**Tableau 1 : Estimations de la taille de la population LGBT selon les trois hypothèses**

	Total	Hommes	Femmes	Hommes 20-34 ans	Hommes 35-49 ans	Hommes 50-69 ans	Femmes 20-34 ans	Femmes 35-49 ans	Femmes 50-69 ans
<b>Taille Pop LGBT selon enq. Tél</b>	<b>609 410</b>	<b>324 604</b>	<b>284 807</b>	<b>103 101</b>	<b>109 805</b>	<b>111 699</b>	<b>147 175</b>	<b>92 799</b>	<b>44 833</b>
<b>% Pop LGBT</b>	<b>1,56%</b>	<b>1,70%</b>	<b>1,42%</b>	<b>1,86%</b>	<b>1,74%</b>	<b>1,54%</b>	<b>2,61%</b>	<b>1,41%</b>	<b>0,57%</b>
<b>HYPOTHESE BASSE</b>									
Enq. Tél. - % freq. LIEUX LGBT	43,50%	53,50%	32,10%	48,34%	62,79%	49,23%	29,78%	41,99%	19,02%
Enq. Web - % freq. LIEUX LGBT	82,76%	86,32%	78,88%	82,19%	87,94%	90,67%	76,65%	80,83%	87,54%
Enq. Web - Parmi freq. LIEUX, % freq. SITES	90,09%	87,58%	93,10%	91,55%	87,97%	80,94%	94,10%	92,58%	88,93%
Enq. Web - Parmi freq. LIEUX & SITES, % non joignables par tel	29,05%	28,55%	29,62%	27,15%	31,24%	26,43%	29,79%	32,34%	21,78%
Prct d'augmentation de la population LGBT	9,42%	11,55%	6,98%	9,88%	15,17%	9,55%	6,40%	10,16%	3,22%
<b>Estimation selon hypothèse basse</b>	<b>666 838</b>	<b>362 081</b>	<b>304 694</b>	<b>113283</b>	<b>126468</b>	<b>122365</b>	<b>156592</b>	<b>102229</b>	<b>46279</b>
<b>% Pop LGBT selon hypothèse basse</b>	<b>1,70%</b>	<b>1,90%</b>	<b>1,52%</b>	<b>2,04%</b>	<b>2,00%</b>	<b>1,69%</b>	<b>2,78%</b>	<b>1,56%</b>	<b>0,59%</b>
<b>HYPOTHESE MEDIANE</b>									
Enq. Tél. - % freq. LIEUX LGBT	43,50%	53,50%	32,10%	48,34%	62,79%	49,23%	29,78%	41,99%	19,02%
Enq. Web - % freq. LIEUX LGBT	82,76%	86,32%	78,88%	82,19%	87,94%	90,67%	76,65%	80,83%	87,54%
Enq. Web - Parmi freq. LIEUX, % non joignables par tel	29,54%	29,09%	30,08%	27,35%	32,08%	27,12%	30,58%	32,68%	20,88%
Prct d'augmentation de la population LGBT	10,63%	13,43%	7,62%	10,9%	17,7%	12,1%	7,0%	11,1%	3,5%
<b>Estimation selon hypothèse médiane</b>	<b>674 221</b>	<b>368 205</b>	<b>306 500</b>	<b>114 304</b>	<b>129 256</b>	<b>125 221</b>	<b>157 448</b>	<b>103 092</b>	<b>46 392</b>
<b>% Pop LGBT selon hypothèse médiane</b>	<b>1,72%</b>	<b>1,93%</b>	<b>1,53%</b>	<b>2,06%</b>	<b>2,05%</b>	<b>1,73%</b>	<b>2,79%</b>	<b>1,57%</b>	<b>0,59%</b>
<b>HYPOTHESE HAUTE</b>									
Enq. Tél. - % freq. LIEUX LGBT	43,50%	53,50%	32,10%	48,34%	62,79%	49,23%	29,78%	41,99%	19,02%
Enq. Web - % freq. LIEUX ou SITES	96,99%	97,46%	96,49%	97,63%	97,31%	97,39%	96,42%	96,74%	96,19%
Enq. Web - Parmi Freq. LIEUX ou SITES, % non joignables par tel	30,41%	29,44%	31,48%	27,62%	32,59%	27,51%	31,59%	34,60%	22,34%
Prct d'augmentation de la population LGBT	12,83%	15,35%	9,75%	13,0%	19,9%	13,2%	9,1%	14,1%	4,1%
<b>Estimation selon hypothèse haute</b>	<b>687 604</b>	<b>374 432</b>	<b>312 575</b>	<b>116 540</b>	<b>131 670</b>	<b>126 432</b>	<b>160 525</b>	<b>105 842</b>	<b>46 665</b>
<b>% Pop LGBT selon hypothèse haute</b>	<b>1,76%</b>	<b>1,96%</b>	<b>1,56%</b>	<b>2,10%</b>	<b>2,09%</b>	<b>1,75%</b>	<b>2,85%</b>	<b>1,61%</b>	<b>0,60%</b>

## Description de la population LGBT interrogée sur Internet

Dans un premier temps, comparons les distributions de quelques caractéristiques sociodémographiques des populations LGBT âgées de 20 à 69 ans des deux enquêtes. Les personnes interrogées dans l'enquête Web sont plus souvent des femmes, des personnes de moins de 40 ans, des personnes très diplômées, cadres ou professions intermédiaires, nées en France, plus souvent en couple sans enfants (tableau 2).

Nous avons également mesuré les prévalences de plusieurs types de violence dans les deux échantillons bruts, sans pondération et sans contrôle de la structure sociodémographique (tableau 3). Les personnes interrogées sur Internet rapportent plus souvent avoir été victimes de violences, quel qu'en soit le type. Parmi les personnes ayant rapporté au moins une violence, les répondants Web sont également plus nombreux à déclarer avoir subi plusieurs violences d'un même type (sauf pour les violences physiques).

Nous avons ensuite cherché à savoir si ces différences étaient dues à des structures sociodémographiques différentes ou si, dans des régressions logistiques, elles persistaient après contrôle de ces variables. En contrôlant le sexe, l'âge, le niveau de diplôme, la catégorie socioprofessionnelle, le lieu de naissance et le type de ménage, on observe un effet propre de l'échantillon sur tous les types de violences hormis les violences psychologiques. Quel que soit le type de violence, les personnes interrogées sur Internet ont tendance à rapporter plus souvent en avoir été victimes (toutes choses égales par ailleurs) que les LGBT interrogés par téléphone.

Cette étude montre donc que les répondants LGBT Internet sont effectivement plus exposés à la violence que les répondants LGBT téléphone et que cet écart persiste lorsque les caractéristiques sociodémographiques sont contrôlées. Cela confirme nos attentes. Il nous est impossible d'affirmer que l'écart en termes de violences déclarées résulte d'une authentique exposition à la violence qui serait supérieure parmi les répondants Internet ou bien d'un effet du mode de collecte ou bien d'un effet d'auto-sélection.

**Tableau 2 : Comparaison des caractéristiques sociodémographiques des échantillons LGBT téléphoniques et web (effectifs bruts, sans pondération)**

	LGBT ECH. TEL		LGBT ECH. WEB	
	Fréquence	Pourcentage	Fréquence	Pourcentage
<b>Sexe de l'enquêté</b>				
Homme	274	54,5%	3020	52,2%
Femme	229	45,5%	2770	47,8%
<b>Age de l'enquêté</b>				
20-29 ans	107	21,3%	2060	35,6%
30-39 ans	100	19,9%	1494	25,8%
40-49 ans	120	23,9%	1228	21,2%
50-59 ans	108	21,5%	682	11,8%
60-69 ans	68	13,5%	326	5,6%
<b>Diplôme</b>				
Aucun diplôme	21	4,2%	55	0,9%
BEPC/BEP/CAP	108	21,5%	353	6,1%
Baccalauréat	101	20,1%	803	13,9%
Dipl. du supérieur 1 <sup>er</sup> cycle	73	14,5%	859	14,8%
Dipl. du supérieur 2 <sup>e</sup> et 3 <sup>e</sup> cycle	200	39,8%	3720	64,2%
<b>Catégorie socioprofessionnelle</b>				
Agriculteurs exploitants	3	0,6%	14	0,2%
Artisans, commerçants et chefs d'ent.	18	3,6%	160	2,8%
Cadres et prof. Intellectuelles sup.	108	21,5%	2057	35,5%
Professions Intermédiaires	111	22,1%	1551	26,8%
Employés	100	19,9%	815	14,1%
Ouvriers	57	11,3%	235	4,1%
Retraités	58	11,5%	236	4,1%
Autres personnes sans activité prof.	48	9,5%	722	12,5%
<b>Lieu de naissance</b>				
Né en France	471	93,6%	5508	95,1%
Né à l'étranger	32	6,4%	282	4,9%

Type de ménage	LGBT ECH. TEL		LGBT ECH. WEB	
	Fréquence	Pourcentage	Fréquence	Pourcentage
Personnes seules	235	46,7%	2155	37,2%
Couples sans enfants	111	22,1%	2146	37,1%
Familles monoparentales	50	9,9%	259	4,5%
Couple avec enfants	64	12,7%	549	9,5%
Autres type de ménage	43	8,5%	681	11,8%

**Tableau 3 : Comparaison des compteurs de violences des 12 derniers mois des échantillons LGBT téléphoniques et web (effectifs bruts, sans pondération)**

	LGBT ECH. TEL		LGBT ECH. WEB	
	Fréquence	Pourcentage	Fréquence	Pourcentage
<b>Nombre de violences psychologiques au cours des 12 derniers mois</b>				
Aucune	278	55,3%	1851	32,0%
Une	101	20,1%	1235	21,3%
Plusieurs	124	24,7%	2704	46,7%
<b>Nombre de violences physiques au cours des 12 derniers mois</b>				
Aucune	467	92,8%	5164	89,2%
Une	18	3,6%	433	7,5%
Plusieurs	18	3,6%	193	3,3%
<b>Nombre de violences psychologiques à caractère sexuel au cours des 12 derniers mois</b>				
Aucune	362	72,0%	2815	48,6%
Une	94	18,7%	1922	33,2%
Plusieurs	47	9,3%	1053	18,2%
<b>Nombre de violences sexuelles au cours des 12 derniers mois</b>				
Aucune	469	93,2%	5146	88,9%
Une	27	5,4%	495	8,5%
Plusieurs	7	1,4%	149	2,6%

Les 5790 individus interrogés via l'enquête web LGBT se divisent en deux groupes :

- les joignables par téléphone (n=4036)
- les injoignables par téléphone (n=1754).

Nous avons comparé les joignables aux non joignables. L'idée est de tester la qualité de notre échantillon non aléatoire. Si l'auto-sélection pour l'enquête internet est forte et dépend du thème, et si l'enquête a sélectionné des personnes non joignables par téléphone en plus grande proportion que ce qui devrait être le cas dans la situation d'un sondage aléatoire simple, alors on pourrait s'attendre à ce que le caractère d'auto-sélection dépende du fait d'être joignable par téléphone. Les sous-échantillons présentent des différences significatives en termes de caractéristiques sociodémographiques (tableau 4). Les non joignables sont plus souvent des femmes, des personnes âgées de 30-39 ans, voire 40-49 ans. Ils comptent beaucoup moins de très diplômés (diplôme du supérieur 2<sup>e</sup> et 3<sup>e</sup> cycle), moins souvent artisans, cadres ou retraités, sont plus souvent nés en France et vivent davantage seuls ou en couple sans enfant.

A contrario, les écarts en termes de prévalences de violences sont assez peu marqués entre ces deux populations. Les non joignables sont très légèrement moins souvent victimes de toute forme de violences, hormis les violences psychologiques (tableau 5). En contrôlant les variables sociodémographiques, il n'y a plus d'effet du fait d'être joignable ou non par téléphone sur les occurrences de violence (de tout type). Les différences en termes de prévalences ne sont qu'un effet de structure sociodémographique. Les deux sous-échantillons ont donc des comportements très proches. Les personnes non joignables n'ont pas un profil atypique et semblent représenter assez justement les répondants de l'enquête Internet.

Malgré les profils très proches entre joignables et injoignables, nous avons décidé de réaliser deux appariements distincts avec les questionnaires téléphoniques, afin d'assurer la présence de joignables et d'injoignables. Ces appariements ont pour objectif d'augmenter la taille de l'échantillon LGBT mais l'introduction des individus ayant répondu sur Internet et n'étant pas joignables par téléphone permet également de corriger le défaut de couverture de l'enquête téléphonique.

**Tableau 4 : Comparaison des caractéristiques sociodémographiques des répondants LGBT Web selon s'ils sont joignables par téléphone ou non (effectifs bruts, sans pondération)**

	LGBT WEB Joignables par tel.		LGBT WEB Non joignables par tel.	
	Fréquence	Pourcentage	Fréquence	Pourcentage
<b>Sexe de l'enquêté</b>				
Homme	2139	53,0%	881	50,2%
Femme	1897	47,0%	873	49,8%
<b>Age de l'enquêté</b>				
20-29 ans	1473	36,5%	587	33,5%
30-39 ans	983	24,4%	511	29,1%
40-49 ans	832	20,6%	396	22,6%
50-59 ans	501	12,4%	181	10,3%
60-69 ans	247	6,1%	79	4,5%
<b>Diplôme</b>				
Aucun diplôme	30	0,7%	25	1,4%
BEPC/BEP/CAP	224	5,6%	128	7,3%
Baccalauréat	535	13,3%	268	15,3%
Dipl. du supérieur 1 <sup>er</sup> cycle	569	14,1%	289	16,5%
Dipl. du supérieur 2 <sup>e</sup> et 3 <sup>e</sup> cycle	2678	66,4%	1044	59,5%
<b>Catégorie socioprofessionnelle</b>				
Agriculteurs exploitants	12	0,3%	2	0,1%
Artisans, commerçants et chefs d'ent.	123	3,0%	37	2,1%
Cadres et prof. Intellectuelles sup.	1454	36,0%	603	34,4%
Professions Intermédiaires	1045	25,9%	506	28,8%
Employés	529	13,1%	286	16,3%
Ouvriers	155	3,8%	80	4,6%
Retraités	176	4,4%	60	3,4%
Autres personnes sans activité prof.	542	13,4%	180	10,3%
<b>Lieu de naissance</b>				
Né en France	3821	94,7%	1685	96,1%
Né à l'étranger	215	5,3%	69	3,9%
<b>Type de ménage</b>				
Personnes seules	1477	36,6%	678	38,7%
Couples sans enfants	1459	36,1%	687	39,2%
Familles monoparentales	200	5,0%	59	3,4%
Couple avec enfants	396	9,8%	153	8,7%
Autres type de ménage	504	12,5%	177	10,1%

**Tableau 5 : Comparaison : Comparaison des compteurs de violences des 12 derniers des répondants LGBT Web selon s'ils sont joignables par téléphone ou non (effectifs bruts, sans pondération)**

	LGBT WEB Joignables par tel.		LGBT WEB Non joignables par tel.	
	Fréquence	Pourcentage	Fréquence	Pourcentage
<b>Nombre de violences psychologiques au cours des 12 derniers mois</b>				
Aucune	1299	32,2%	552	31,5%
Une	884	21,9%	351	20,0%
Plusieurs	1853	45,9%	851	48,5%
<b>Nombre de violences physiques au cours des 12 derniers mois</b>				
Aucune	3593	89,0%	1571	89,6%
Une	307	7,6%	126	7,2%
Plusieurs	136	3,4%	57	3,2%
<b>Nombre de violences psychologiques à caractère sexuel au cours des 12 derniers mois</b>				
Aucune	1946	48,2%	869	49,5%
Une	1354	33,5%	568	32,4%
Plusieurs	736	18,2%	317	18,1%
<b>Nombre de violences sexuelles au cours des 12 derniers mois</b>				
Aucune	3574	88,6%	1572	89,6%
Une	362	9,0%	133	7,6%
Plusieurs	100	2,5%	49	2,8%



## ***Intégration des questionnaires Web dans l'enquête téléphonique***

---

Le deuxième objectif de cette étude est donc l'intégration des questionnaires Internet à l'enquête téléphonique, afin d'accroître le nombre de questionnaires LGBT soumis à analyse. Le problème de l'intégration des questionnaires Internet est celui, classique, que pose la comparaison des questionnaires obtenus via un autre mode de collecte et via un processus de sélection sur la base du volontariat. La question de l'effet mode de collecte a été évoquée plus haut ; celle de la sélection s'y ajoute. En effet, les LGBT ayant rempli le questionnaire par Internet ont probablement des caractéristiques qui les distinguent fortement de la population interrogée dans l'enquête téléphonique, en particulier ceux que l'on souhaite intégrer parce qu'ils n'étaient pas joignables dans l'enquête téléphonique, c'est-à-dire pratiquement pas couverts par le dispositif d'enquête. Cette préoccupation est une précaution générale, car nous avons vu que dans notre cas, les joignables ne diffèrent pas des injoignables. Toutefois, les LGBT Internet présentent globalement des caractéristiques sociodémographiques et de violences déclarées très différentes de leurs homologues interrogés au téléphone. Il n'est pas dans notre intention de tenter de démêler les effets de mode de collecte des effets de sélection mais de consentir à sélectionner des questionnaires qui soient comparables à ceux recueillis par téléphone du point de vue des violences déclarées. En effet, dans notre cas, intégrer tous les questionnaires collectés par Internet reviendrait à augmenter les prévalences des violences subies au sein de la sous-population des LGBT de l'enquête principale par effet de collecte et effet de sélection, alors qu'une telle correction n'est pas possible pour la population majoritaire. Un exemple récent montre que certains types de violence notamment les rapports sexuels forcés sont effectivement rapportés plus couramment dans des panels internet de volontaires que dans des enquêtes téléphoniques aléatoires (12).

### ***Stratégie retenue pour la sélection et l'intégration des questionnaires Internet***

L'intégration des questionnaires Internet a pour but d'augmenter la taille de la sous-population LGBT dans l'enquête téléphonique, ce qui doit permettre des analyses plus fines de ce groupe minoritaire. Le choix doit s'effectuer en considérant trois contraintes. D'abord, la volonté de ne pas introduire de questionnaire aux réponses trop atypiques aux questions de violences qui sont le cœur de l'enquête Virage, afin d'exclure tout individu dont les réponses pourraient résulter d'une situation réelle trop exceptionnelle ou d'une volonté délibérée d'exagération. Le risque est important dans la mesure où l'enquête Internet est sur la base du volontariat. Ensuite, il faut considérer que l'introduction de questionnaires Internet peut biaiser à la hausse l'estimation des violences subies comme nous l'avons vu précédemment. Finalement, il faut aussi garder à l'esprit que la population majoritaire ne peut bénéficier d'une procédure similaire : introduire des questionnaires LGBT Internet déclarant des violences nombreuses biaiserait la comparaison entre population majoritaire et population LGBT.

Pour cette raison, nous avons retenu pour l'appariement des questionnaires semblables aux questionnaires téléphoniques du point de vue des indicateurs synthétiques de violence déclarées, du sexe et de l'âge. Le choix de ces variables reste matière à discussion, mais il importe de ne pas retenir uniquement des individus présentant exactement toutes les caractéristiques des répondants de l'enquête téléphonique, sans quoi l'opération reviendrait à dupliquer les observations déjà obtenues au téléphone. Il faut donc veiller à se prémunir contre un effet de mode (des violences rapportées notablement différentes) mais laisser entrer en ligne de compte une certaine diversité sociodémographique. Cela est d'autant plus important que le sous-échantillon de répondants téléphoniques est très réduit. En revanche, comme nous ne voulons pas introduire de distorsion des relations entre variables sociodémographiques et variables de violences entre les deux échantillons, nous avons choisi de tenir compte des interactions entre celles-ci.

Pour cela, nous utiliserons la procédure de modélisation de score de propension et d'appariement sur celui-ci décrite par Rosenbaum et Rubin. Dans le cas d'une population de LGBT internet de taille infinie, elle assure que les questionnaires Internet appariés aux questionnaires téléphoniques seront totalement comparables à ces derniers pour les variables introduites dans le calcul du score de propension (Rosenbaum and Rubin, 1983, 1984). Ici, la population Internet est de taille limitée, mais la procédure assure le maximum de comparabilité, qui sera vérifiée a posteriori. De plus, le calage effectué par la suite sur les échantillons appariés permettra de redresser les éventuelles disparités sociodémographiques.

### ***Calcul d'un score de propension***

A partir de cette étape jusqu'à la fin du processus d'appariement, les opérations sont menées sur deux bases :

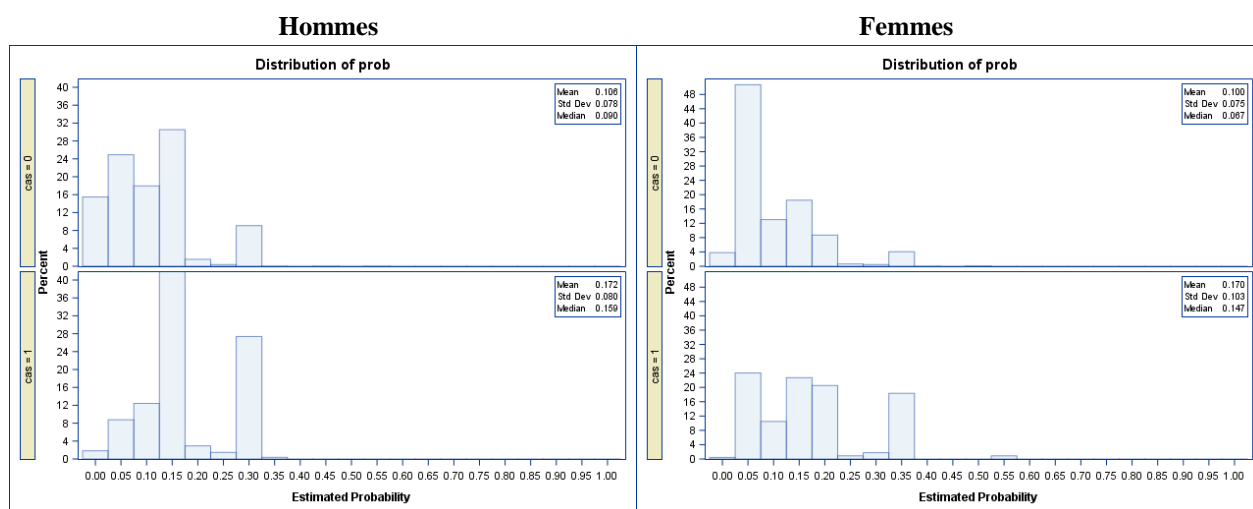
- les questionnaires LGBT Tel (T) + les questionnaires Web joignables par téléphone (WJT)
- les questionnaires LGBT Tel (T) + les questionnaires Web non joignables par téléphone (WNJT)

Dans un premier temps, on réalise une régression logistique par sexe pour calculer un score de propension à répondre par téléphone plutôt que sur Internet. Les variables explicatives mobilisées sont :

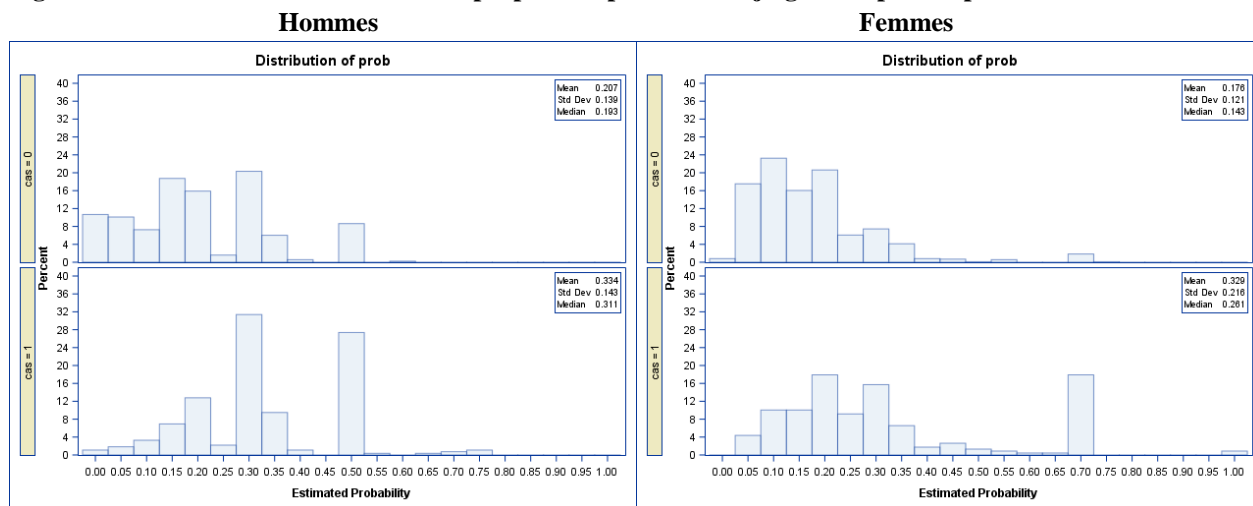
- la classe d'âge  
(« 20-34 ans », « 35-49 ans », « 50-69 ans »)
- le nombre de violences psychologiques au cours des 12 derniers mois  
(« Aucune », « Une », « Deux », « Trois », « Quatre », « Cinq ou plus »)
- le nombre de violences physiques au cours des 12 derniers mois  
(« Aucune », « Une », « Deux », « Trois ou plus »)
- le nombre de violences psychologiques à caractères sexuel au cours des 12 derniers mois  
(« Aucune », « Une », « Deux », « Trois ou plus »)
- le nombre de violences sexuelles au cours des 12 derniers mois  
(« Aucune », « Une », « Plusieurs »)
- les quatre interactions entre les variables de violences et la variable d'âge

Avant de procéder à l'appariement, regardons la distribution des scores de propension dans les deux échantillons (cas=0 : échantillon web / cas=1 : échantillon téléphonique). Les scores moyens sont effectivement différents dans les deux échantillons, mais les déviations standards restent proches. La superposition des scores sera donc partielle ce qui a pour conséquence qu'une partie seulement des LGBT Internet pourra être appariée.

**Figures 1 et 2 : Distribution des scores de propension pour joignables par téléphone**



**Figure 3 et 4 : Distribution des scores de propension pour les non joignables par téléphone**



### *Appariement des observations*

On procède ensuite à l'appariement un à un sur le score de propension : les LGBT interrogés sur Internet joignables et non joignables sont appariés séparément aux LGBT interrogés par téléphone. On répète l'opération d'appariement trois fois de suite car on désire introduire plusieurs observations web pour une observation téléphonique. Les questionnaires LGBT téléphoniques sont donc appariés un nombre variable de fois (entre 1 et 3 fois). Nous avons retenu les questionnaires web appariés sur au moins 2 décimales.

**Tableau 6 : Taille des échantillons LGBT et effectifs de questionnaires appariés sur deux décimales**

	Taille totale échantillon			Nb observations retenues pour l'échantillon (*)		
	Hommes	Femmes	Total	Hommes	Femmes	Total
Echantillon Tel	274	229	503	274	229	503
Echantillon Web joignables	2 139	1 897	4 036	778	619	1397
Echantillon Web non joignables	881	873	1 754	528	458	986
<b>Total</b>	<b>3 294</b>	<b>2 999</b>	<b>6 293</b>	<b>1580</b>	<b>1 306</b>	<b>2886</b>

(\*) Toutes les observations de l'échantillon de l'enquête téléphonique sont conservées. Pour les observations de l'échantillon web, seules sont retenues les observations appariées sur au moins deux décimales à des questionnaires téléphoniques.

Grâce à l'introduction de questionnaires de l'enquête Web LGBT, on passe d'un échantillon de 503 LGBT à 2886, soit une multiplication par plus de 5,5 des effectifs de questionnaires LGBT.

### *Calcul du pseudo-poids de sondage des questionnaires Internet*

L'enquête Web étant une enquête non probabiliste (participation sur la base du volontariat), il n'est pas possible de calculer des poids de sondage. Nous avons donc opté pour le calcul de poids de sondage « fictifs ». Les opérations se déroulent en trois temps. En premier lieu, nous avons calculé le poids de sondage des questionnaires web LGBT joignables par téléphone (g). Nous souhaitons que la somme des poids de sondage des LGBT reste inchangée, de ce fait, on redistribue le total sur davantage de questionnaires. On peut donc ensuite calculer un nouveau poids de sondage moyen ((b)/(f)). Dans un deuxième temps, il faut réduire le poids de sondage des questionnaires téléphoniques LGBT. Pour cela, on applique un coefficient correcteur (j), qui correspond au ratio de l'ancien effectif de questionnaires LGBT sur le nouveau (sans prendre en compte les non joignables), c'est-à-dire ((c)/(f)).

Pour finir, pour les questionnaires Web non joignables par téléphone, nous nous sommes basés sur les résultats de la ré-estimation de la taille de la population LGBT grâce à la prise en compte de l'erreur de couverture de l'enquête téléphonique. Pour chaque sous-groupe sexe/ classe d'âge, nous disposons d'une taille estimée de la population LGBT non joignable par téléphone. Par un simple produit en croix, on calcule la somme des poids de sondage des questionnaires Web LGBT non joignables ((k)\*(b)/(a)) que l'on divise par le nombre de questionnaires (l) pour obtenir un poids de sondage individuel (n).

**Tableau 7 : Pseudo-poids de sondage des questionnaires Web**

	Hommes 20-34 ans	Hommes 35-49 ans	Hommes 50-69 ans	Femmes 20-34 ans	Femmes 35-49 ans	Femmes 50-69 ans
<b>Situation initiale</b>						
Taille pop. LGBT joignables (a)	103 101	109 805	111 699	147 175	92 799	44 833
Somme poids de sondage Q. Tel. LGBT (b)	94535,09	117804,54	149954,86	102511,35	93581,90	69896,29
Eff. Q. Tel. LGBT (c)	68	91	115	82	86	61
Poids de sondage moyen des Q. tel. LGBT (d)	1390,22	1294,56	1303,96	1250,14	1088,16	1145,84
<b>Calcul poids de sondage Q. Web LGBT joignables par tél.</b>						
Eff. Q. Web – Joignables (e)	204	262	312	233	159	127
Total Q. Joignables (Tel. + Web) (f)	272	353	427	315	245	188
Nouveau Poids de sondage moyen des Q. joignables (g)	347,56	333,72	351,18	325,43	381,97	371,79
<b>Calcul coefficient de correction pour les poids de sondage des Q. Tel. LGBT</b>						
Ancien poids de sondage moyen (h)	1390,22	1294,56	1303,96	1250,14	1088,16	1145,84
Nouveau Poids de sondage moyen (i)	347,56	333,72	351,18	325,43	381,97	371,79
Coef. Correcteur poids Q. Tel (j)	0,25	0,26	0,27	0,26	0,35	0,32
<b>Calcul poids de sondage des Q. Web LGBT non joignables par tél.</b>						
Taille pop. LGBT non joignables (k)	11 203	19 451	13 522	10 273	10 293	1 559
Eff. Q. Web - Non joignables (l)	151	218	159	213	182	45
Somme poids Q. web non joignables (m)	10272,50	20867,64	18152,80	7155,60	10379,88	2429,97
Poids Q. Web non joignables (n)	68,03	95,72	114,17	33,59	57,03	54,00

Les questionnaires web non joignables par téléphone ont en moyenne un poids 4,7 fois inférieur aux questionnaires téléphoniques LGBT et web joignables par téléphone.

### **Calage de l'échantillon composite**

En plus des marges définies pour le premier calage (classe d'âge, diplôme, lieu de naissance, catégorie socioprofessionnelle, type de ménage et région de résidence), nous avons ajouté une marge correspondant à la taille des populations LGBT par tranche d'âge (effectifs correspondant au calcul de l'hypothèse médiane). Comme précédemment, le calage est réalisé par sexe. Cette marge confond les joignables et les injoignables, qui sont par définition presque indiscernables du point de vue des variables de sexe, d'âge et de violence introduites dans le score de propension.

**Tableau 8 : Marges additionnelles concernant la population LGBT**

	Hommes		Femmes	
LGBT 20-34 ans	114304	0,6%	157448	0,8%
LGBT 35-49 ans	129256	0,7%	103092	0,5%
LGBT 50-69 ans	125221	0,7%	46392	0,2%
Non LGBT	18730447	98,1%	19707449	98,5%
Total	19099228	100,0%	20014381	100,0%

Regardons comment évoluent les statistiques de poids entre avant et après le calage. La somme des poids sur l'échantillon total est fixe et vaut 39 113 609. Nous proposons au final une troncature des poids les plus bas et les plus élevés (1% aux extrêmes), afin de contenir les ratios des poids max et min et le coefficient de variation.

**Tableau 9 : Statistiques des poids avant et après calage pour l'échantillon composite**

	Indicateurs	Tel. Non LGBT	Tel. LGBT	Web LGBT joignables Tel	Web LGBT Non joignables Tel	Total
<b>Taille échantillon</b>	Eff.	26 725	503	1397	986	29 651
<b>Poids de sondage</b>	Moyenne	1 437,9	332,2	330,1	70,2	1321,5
	Min	729,8	181,9	271,3	31,0	31,0
	Max	12 986,2	1 404,1	371,8	114,2	12 986,2
	Max/Min	17,8	7,7	1,4	3,7	418,9
	Coef. var	57,5	51,6	9,3	41,8	65,3
<b>Poids après calage</b>	Moyenne	1 436,1	360,4	305,9	68,1	1319,1
	Min	145,8	46,7	54,2	10,8	10,8
	Max	30 134,1	2 401,3	910,7	296,1	30 134,1
	Max/Min	206,7	51,4	16,8	27,4	2790,2
	Coef. var	92,4	77,7	51,3	57,2	99,5
<b>Poids après calage tronqué (1% des extrêmes inf. et sup.)</b>	Moyenne	1 435,5	367,3	311,8	73,5	1319,1
	Min	148,0	47,8	55,0	47,3	47,3
	Max	6 069,2	2457,3	932,0	303,0	6 069,2
	Max/Min	41,0	51,4	16,9	6,4	128,3
	Coef. var	82,8	77,7	51,2	49,2	89,9

Note : La faible augmentation des poids minimum et maximum de certaines catégories après troncature est due à la dilatation des poids rendue nécessaire pour obtenir le véritable effectif de la population.

A cause de la troncature des poids, certaines caractéristiques sociodémographiques peuvent présenter au final une distribution un peu éloignée de la cible. Les écarts les plus importants (pourcentage d'évolution par rapport à la cible supérieur à 5%) sont signalés par (\*) dans le tableau 10. Le seul vrai défaut porte sur le déficit de personnes vivant dans d'autres types de ménages, les autres écarts restent marginaux.

**Tableau 10 : Distribution des variables de calage après calage et troncature des poids**

	Hommes			Femmes		
	Cible	Tel + Web après calage et troncature	Pourcentage d'évolution (avant/après troncature)	Cible	Tel + Web après calage et troncature	Pourcentage d'évolution (avant/après troncature)
<b>Classe d'âge</b>						
20-29 ans	19,14	18,93	-0,01	18,83	18,47	-0,02
30-39 ans	20,58	20,67	0,00	20,2	20,36	0,01
40-49 ans	22,33	22,57	0,01	21,97	22,11	0,01
50-59 ans	20,81	20,63	-0,01	21,16	21,06	0,00
60-69 ans	17,14	17,19	0,00	17,84	17,99	0,01
<b>Diplôme</b>						
Aucun diplôme	18,25	17,6	-0,04	19,83	19,59	-0,01
BEPC/BEP/CAP	33,97	33,98	0,00	26,85	26,98	0,00
Baccalauréat	18,57	18,68	0,01	19,89	19,89	0,00
Dipl. du supérieur 1 <sup>er</sup> cycle	12,87	13,09	0,02	17,18	17,11	0,00
Dipl. du supérieur 2 <sup>e</sup> et 3 <sup>e</sup> cycle	16,34	16,65	0,02	16,25	16,43	0,01
<b>Lieu de naissance</b>						
En France	86,09	86,45	0,00	85,69	85,84	0,00
A l'étranger	13,91	13,55	-0,03	14,31	14,16	-0,01
<b>Catégorie socio-professionnelle</b>						
Agriculteurs exploitants	1,72	1,75	0,02	0,6	0,6	0,00
Artisans, commerçants et chefs d'ent.	6,46	6,28	-0,03	2,47	2,46	0,00
Cadres et prof. Intellectuelles sup.	14,51	14,77	0,02	9,29	9,32	0,00
Professions Intermédiaires	17,67	17,87	0,01	19,14	19,15	0,00
Employés	9,91	10,03	0,01	31,64	31,71	0,00
Ouvriers	26,91	26,59	-0,01	6,2	6,1	-0,02
Retraités	15,84	15,9	0,00	15,84	15,43	-0,03
Autres personnes sans activité prof.	6,98	6,81	-0,02	15,39	15,22	-0,01

	Hommes			Femmes		
	Cible	Tel + Web après calage et troncature	Pourcentage d'évolution (avant/après troncature)	Cible	Tel + Web après calage et troncature	Pourcentage d'évolution (avant/après troncature)
<b>Type de ménage</b>						
Personnes seules	17,28	17,66	0,02	15,81	16,04	0,01
Couples sans enfants	26,63	26,72	0,00	27,45	27,57	0,00
Familles monoparentales	4,87	4,96	0,02	10,23	10,33	0,01
Couple avec enfants	43,13	43,09	0,00	39,19	39,19	0,00
Autres type de ménage	8,09	7,56	<b>-0,07 (*)</b>	7,32	6,87	<b>-0,06 (*)</b>
<b>Région</b>						
Île-de-France	19,08	19,05	0,00	19,63	19,49	-0,01
Champagne-Ardenne	2,12	2,16	0,02	2,08	2,11	0,01
Picardie	3,06	2,97	-0,03	3	3	0,00
Haute-Normandie	2,9	2,95	0,02	2,91	2,92	0,00
Centre	3,99	3,92	-0,02	3,95	4	0,01
Basse-Normandie	2,3	2,33	0,01	2,25	2,28	0,01
Bourgogne	2,57	2,58	0,00	2,52	2,47	-0,02
Nord-Pas-de-Calais	6,37	6,34	0,00	6,38	6,37	0,00
Lorraine	3,79	3,8	0,00	3,7	3,75	0,01
Alsace	3,03	2,94	-0,03	2,97	2,99	0,01
Franche-Comté	1,88	1,91	0,02	1,81	1,73	-0,04
Pays de la Loire	5,7	5,76	0,01	5,55	5,59	0,01
Bretagne	5,09	5,11	0,00	4,93	4,96	0,01
Poitou-Charentes	2,78	2,81	0,01	2,76	2,77	0,00
Aquitaine	5,17	5,15	0,00	5,21	5,19	0,00
Midi-Pyrénées	4,64	4,61	-0,01	4,55	4,62	0,02
Limousin	1,15	1,14	-0,01	1,14	1,15	0,01
Rhône-Alpes	10,03	10,14	0,01	9,95	9,95	0,00
Auvergne	2,15	2,17	0,01	2,09	2,08	0,00
Languedoc-Roussillon	4,15	4,15	0,00	4,28	4,32	0,01
Provence-Alpes-Côte d'Azur	7,54	7,51	0,00	7,84	7,73	-0,01
Corse	0,51	0,51	0,00	0,51	0,51	0,00
<b>Statut LGBT par âge</b>						
LGBT - 20-34 ans	0,6	0,61	0,02	0,79	0,81	0,03
LGBT - 35-49 ans	0,68	0,69	0,01	0,52	0,53	0,02
LGBT - 50-69 ans	0,66	0,67	0,02	0,23	0,24	0,04
Non LGBT	98,07	98,02	0,00	98,47	98,43	0,00

### Mesure du pourcentage d'efficacité des échantillons

Pour juger de l'efficacité d'un redressement, on peut calculer la statistique suivante, qui donne une idée de la perte de précision due à la variance des poids (après redressement) d'un échantillon donné de taille n : plus S est proche de 1, meilleur est le redressement.

$$S = 100 \times \frac{(\sum \text{poids})^2}{n \sum \text{poids}^2}$$

**Tableau 11 : Pourcentage d'efficacité et taille des échantillons efficaces après calage et troncature des poids**

	Echantillon enq. téléphonique			Ech. Enq. Téléphonique + LGBT web appariés		
	LGBT	Non LGBT	Total	LGBT	Non LGBT	Total
Taille de l'échantillon	503	26 725	27 268	2 886	26 765	29 651
% efficacité	59,42%	55,27%	55,22%	57,98%	59,32%	55,33%
Taille ech. efficace	299	14 771	15 057	1 673	15 877	16 404

La grande dispersion des poids dans la population LGBT après intégration des questionnaires web LGBT conduit à une taille efficace assez réduite pour cette sous-population. Toutefois, globalement, la procédure n'altère pas l'efficacité du

redressement sur l'ensemble de l'échantillon, et conduit tout de même à une multiplication par 5,6 de la taille de l'échantillon efficace des LGBT (soit une augmentation de 459 %).

Comme le montre les prévalences dans le tableau ci-dessous, l'introduction des questionnaires web ne modifie que très marginalement les prévalences de violence sur la population cible. A contrario, les écarts sont un peu plus marqués si l'on se limite à la sous-population des LGBT. Si les prévalences des violences psychologiques à caractère sexuel ou non sont identiques, on constate une plus faible déclaration de violences physiques et sexuelles. Ce point est discuté plus loin.

**Tableau 12 : Distributions des compteurs de violences dans l'échantillon téléphonique et dans l'échantillon composite et dans le sous-échantillon LGBT après calage et troncature des poids etn**

	Population totale		Population LGBT	
	Ech. tel calé tronqué	Ech. composite calé tronqué	Ech. tel calé tronqué	Ech. composite calé tronqué
<b>Violences psychologiques au cours des 12 derniers mois</b>				
Aucune	69,48	69,45	52,82	51,58
1	14,73	14,74	20,09	20,78
2	6,35	6,33	10,99	10,11
3	3,87	3,86	5,65	5,83
4	1,99	2,02	4,56	4,77
5 ou plus	3,58	3,61	5,88	6,93
<b>Violences psychologiques à caractère sexuel au cours des 12 derniers mois</b>				
Aucune	86,35	86,35	69,44	69,3
1	11,33	11,35	20,84	20,49
2	1,86	1,84	7,7	8,18
3 ou plus	0,46	0,46	2,02	2,02
<b>Violences physiques au cours des 12 derniers mois</b>				
Aucune	95,41	95,46	92,11	93,31
1	2,91	2,9	3,78	3,58
2	1,12	1,08	2,54	1,81
3 ou plus	0,56	0,56	1,57	1,29
<b>Violences sexuelles au cours des 12 derniers mois</b>				
Aucune	98,09	98,09	92	93,61
1	1,69	1,7	6,41	5,15
Plusieurs	0,22	0,21	1,6	1,23
<b>Orientation sexuelle</b>				
Hétérosexuels	98,43	98,23	0	0
Homosexuels/bisexuels	1,57	1,77	100	100

## Discussion

Les problèmes posés par cette méthode sont principalement de cinq ordres : 1/ la difficulté d'appréciation de la justesse de l'ampleur de la correction des marges ; 2/ la difficulté d'intégrer les populations transsexuelles et transgenre ; 3/ la difficulté de représentation de la population LGBT non couverte par l'enquête téléphonique ; 4/ la difficulté du choix des variables introduites dans la modélisation du score de pension ; 5/ la difficulté de calcul d'une variance associée.

La première difficulté peut être décomposée. D'abord, dans l'enquête téléphonique aléatoire, nous estimons la proportion d'individus LGBT fréquentant des lieux de rencontre LGBT. La marge additionnelle sera calculée pour cette sous-population. Comme ce critère est restrictif, car il est certain qu'une très large part des LGBT ne fréquentant pas les lieux LGBT a été exposée à l'enquête Internet, il en résulte une sous-estimation probable de la taille de la marge additionnelle. Nous faisons ensuite l'hypothèse (H1) que cette proportion de 43.5% de personnes fréquentant les lieux de rencontre LGBT est bien mesurée. Cela est vrai si l'enquête téléphonique ne comporte pas de biais de couverture ni ne souffre d'erreur de non-réponse de la population LGBT concernée, et que la mesure de cette variable est bien faite. Il est impossible de vérifier l'étendue de ces défauts potentiels, mais en tout état de cause, s'il existe un défaut de couverture ou de non-réponse, il est probable qu'il conduise à une sous-estimation : les personnes les plus militantes ou les plus injoignables par téléphone en raison de leur mode de vie devraient être sous-représentées dans l'échantillon de répondants. Ainsi, nous avons donc deux arguments en faveur d'une sous-estimation de la marge additionnelle : l'un définitionnel, l'autre méthodologique. Il existe également un troisième argument indirect, qui est l'erreur de mesure :

nous avons rappelé dans l'introduction qu'il existe des éléments probants étayant l'hypothèse d'une sous-déclaration de l'orientation LGBT d'une large partie de la population dans l'enquête téléphonique. Notre méthode de calcul d'une marge ne prend pas ces éléments en compte : le résultat numérique sera donc a priori sous-estimé.

Le calcul repose quant à lui sur une autre hypothèse. Dans l'enquête Internet, nous identifions les LGBT fréquentant les lieux LGBT et qui éventuellement, fréquentent aussi des sites LGBT. Parmi ceux-ci, nous repérons ceux qui sont injoignables au téléphone, soit 29.5% pour ceux qui fréquentent les lieux et 29.1% pour ceux qui fréquentent lieux et sites. Nous faisons donc l'hypothèse (H2) que la proportion de personnes injoignables parmi les LGBT fréquentant lieux ou sites est bien estimée. Autrement dit, (H2) stipule que ces proportions proches de 29% sont bien celles qu'on observerait dans la population des LGBT. Nous ne pouvons assurer que cela est bien le cas. Néanmoins, le détail du tableau 1 montre que les proportions de LGBT de l'enquête Internet se déclarant injoignables au téléphone varient très peu avec le sexe et l'âge ainsi qu'avec le fait de fréquenter ou non des sites LGBT. C'est un argument de poids en faveur de la fiabilité de la mesure. Nous pouvons en effet redouter que cette caractéristique d'injoignabilité varie nettement avec le rapport à l'homosexualité, donc au genre ainsi qu'avec le rapport à Internet et donc l'âge. Ce n'est pas le cas. De plus les tableaux 2 et 3 montrent que la population LGBT interrogée sur Internet varie somme toute assez peu de celle interrogée par téléphone : elle est plus jeune et diplômée mais aussi plus souvent victime de violences, sans que les écarts soient massifs. Par exemple, la part des LGBT fréquentant des lieux LGBT s'élève à 43% dans l'enquête téléphonique et 84% dans l'enquête Internet : le sens de l'écart était prévisible et correspond à nos hypothèses de départ concernant l'existence d'une frange de LGBT plus mobiles et moins joignables par téléphone, donc non couverte par l'enquête téléphonique. Mais la proportion dans l'enquête Internet n'est pas de 100%, ce qui aurait été un argument fort en faveur d'un biais massif de recrutement. Finalement, il faut rappeler que l'exposition à l'enquête LGBT fut très bonne et entretenue durant toute la collecte par de nombreuses actions de communication, mobilisant des canaux bien plus diversifiés que les supports physiques distribués dans les lieux LGBT. Pour 61% des questionnaires Internet, l'association ou le support media ayant aiguillé sur l'enquête sont connus : Facebook arrive en tête avec 43% des provenances connues. Internet et les réseaux sociaux sont donc sans doute beaucoup plus efficaces que les lieux physiques pour faire connaître l'enquête. Cela peut expliquer la jeunesse relative et le niveau d'éducation supérieur des LGBT recrutés par Internet.

La part des LGBT dans la population d'autres pays comparables, comme les USA, est nettement plus élevée que ce que nous pouvons estimer directement dans Virage (3.8% vs 1.6% pour l'orientation sexuelle et 2.5% pour la déclaration d'une expérience homosexuelle) ([Gates, 2011](#)). La correction fondée sur l'hypothèse médiane que nous proposons conduit à réviser à la hausse nos estimations initiales de 10.6% ce qui porte le taux de LGBT à 1.7% et celui de la déclaration d'une expérience homosexuelle à 2.8%. Les chiffres sont donc vraisemblables.

La deuxième difficulté est l'identification de la population Trans, i.e. transsexuelle et transgenre, dans nos calculs. Dans l'enquête téléphonique, nous ne questionnons pas les individus sur leur identité de genre, mais uniquement sur leur orientation sexuelle. Il n'est pas donc possible d'identifier la population Trans dans l'enquête téléphonique. Nous l'avons donc exclue de notre critère LGBT, dans l'enquête téléphonique comme dans l'enquête Internet. Il en résulte une sous-estimation évidente sur le principe. Une étude récente menée par le Williams Institute de l'université de Californie dans les 10 États des États-Unis où une loi impose la production d'une pièce d'identité avec photographie pour exercer son droit de vote aux élections, estime qu'environ 85 000 transgenres sont en âge de voter dans ces États (dont 24000 transgenres pourraient être privés de la possibilité de voter en raison d'incompatibilité entre papiers d'identité et leur identité sexuelle pour une question de sexe et d'apparence)([Brown and Herman, 2014](#)). Le rapport de ce nombre à la population de ces états en âge de voter, soit environ 39 000 000 d'individus, permet d'estimer la proportion de personnes transgenres concernée par cette impossibilité à 0.2%. Une autre statistique du Williams Institute de 2011 suggère que la proportion de transgenres aux USA est de l'ordre de 0.3% ([Gates, 2011](#)). Si l'ordre de grandeur peut être appliqué à la France, alors nous devrions encore augmenter notre correctif pour l'estimation de la taille de la marge des LGBT en France.

La troisième difficulté tient au fait que nous intégrons des questionnaires Internet alors que ceux-ci ne reposent pas sur un processus de sélection aléatoire, à la suite d'une opération de modélisation puis d'appariement sur score de propension. La volonté est clairement de se placer dans la situation contrefactuelle où les LGBT non joignables par téléphone auraient été contactés par téléphone, c'est-à-dire que tout effet mode de collecte aurait été gommé. Si la démarche est légitime pour ne pas introduire d'effet de mode de collecte pour la mesure des violences au sein de la population LGBT alors que la démarche est impossible à conduire pour la population majoritaire, elle compromet



évidemment la prise en compte de la diversité de la population répondante à l'enquête internet, qu'elle ne représente pas adéquatement. Les calculs présentés dans ce document montrent en effet clairement que la population LGBT internet candidate à l'appariement est nettement plus exposée à certains types de violence que la population finalement sélectionnée. Malgré la meilleure estimation de la taille de la population LGBT par adjonction d'un pan de la population LGBT non couverte par l'enquête téléphonique, notre méthode n'estime donc pas mieux les violences auxquelles elle est soumise. Elle conduit à une sous-estimation naturelle de celles-ci qui est intrinsèque à la méthodologie téléphonique et qui s'applique à tous les segments de la population cible, ce qui était voulu.

Les mêmes critiques peuvent être portées à l'encontre de la population LGBT interrogée sur Internet joignable par téléphone. Pour celle-ci, de plus, la question de la double probabilité d'inclusion demeure : en effet, les LGBT interrogés sur Internet joignables par téléphone avaient deux possibilités d'être contacté et de répondre à l'enquête, par Internet et par téléphone. Toutefois, le problème est symétrique pour les LGBT interrogés par téléphone ; en effet, ceux-ci étaient certainement exposés à l'enquête Internet en raison de sa très bonne couverture, de sa visibilité médiatique et de sa publicité dans les réseaux et médias sociaux communautaires. Par conséquent, il semble légitime de considérer que les poids des LGBT téléphone et Internet joignables au téléphone sont proches. C'est l'hypothèse qui a été retenue ici. Peut-on alors mieux estimer les violences subies par la population LGBT ? En raison du caractère non aléatoire de l'enquête internet LGBT et de l'effet du mode de collecte, l'estimation du biais est difficile.

Précisons à ce propos que les questionnaires Internet joignables et injoignables sont introduits suivant le même procédé dans l'enquête téléphonique. L'analyse préparatoire avait montré qu'ils étaient presque indiscernables du point de vue des violences subies dès lors qu'on contrôlait leurs caractéristiques sociodémographiques dans des analyses multivariées ; le procédé d'appariement les confond de plus en pratique avec les questionnaires téléphoniques du point de vue des variables introduites dans le modèle de score de propension.

La quatrième difficulté est classique, et tient au choix des variables introduites dans le modèle de score de propension présidant à l'appariement. Nous nous sommes restreints au sexe, à trois tranches d'âge, aux compteurs de 4 types de violences et aux interactions entre la tranche d'âge et ces compteurs pour des raisons de simplicité et la faiblesse des effectifs de questionnaires LGBT téléphoniques. Des variables importantes et liées aux violences, comme le diplôme, la PCS, la vie en couple etc. ne sont donc pas contrôlées. C'est probablement cela qui explique la baisse relative des prévalences des violences physiques et sexuelles mesurée chez les LGBT internet relativement aux LGBT téléphone, dans l'échantillon composite calé. Une révision du modèle pourrait permettre d'y remédier.

Enfin, relativement à la question 5 concernant la variance, il est à noter que la procédure d'intégration des questionnaires LGBT internet réduit la variabilité des variables d'intérêt et que l'estimation de la marge de la population non couverte par le téléphone conduit à des poids très réduits pour les questionnaires Internet intégrés relativement (ratio de presque 5 relativement aux poids des questionnaires LGBT téléphone), il est peu probable que la variance globale en soit augmentée. Là aussi, des travaux ultérieurs tenteront d'apporter des éléments de réponse à cette question.

## Références

1. AAPOR (2011) *Standard Definitions: Final Dispositions Of Case Codes And Outcome Rates For Surveys* (AAPOR).
2. FIRDION, J.-M. & RIANDEY, B. (1993) Vie personnelle et enquête par téléphone : l'exemple de l'enquête ACSF, *Population*, 48, 1257-1280.
3. BECK, F., GUIGNARD, R. & LEGLEYE, S. (2014) Does computer survey technology improve reports on alcohol and illicit drug use in the general population? A comparison between two surveys with different data collection modes in France, *PLoS One*, 9, e85810.
4. DE LEEUW, E. D. (2008) Choosing the method of data collection, in: De Leeuw, E. D., Hox, J. J. & Dillman, D. A. (Eds.) *International handbook of survey methodology*, pp. 117-135 (New York, Lawrence Erlbaum Associates).
5. KANN, L., BRENER, N. D., WARREN, C. W., COLLINS, J. L. & GIOVINO, G. A. (2002) An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents, *J Adolesc Health*, 31, 327-35.
6. BECK, F. & PERETTI-WATEL, P. (2002) The Impact of Data Collection Methodology on the Reporting of mode de Illicit Drug Use by Adolescents, *Population*, 57, 571-591.
7. GMEL, G. (2000) The effect of mode of data collection and of non-response on reported alcohol consumption: a split-sample study in Switzerland, *Addiction*, 95, 123-34.
8. TOURANGEAU, R. & SMITH, T. W. (1996) Asking Sensitive Questions: The Impact Of Data Collection Mode, Question Format, And Question Context, *Public Opinion Quarterly*, 60, 275-304.
9. BECK, F., FIRDION, J., LEGLEYE, S. & SCHILTZ, M. (2010) Suicide et minorités sexuelles : état des lieux des connaissances et perspectives, Rapport au Comité d'observation et de prévention du suicide (COPS) [Suicide and sexual minorities: review of literature and hints for prevention], pp. 128 (Saint-denis, INPES).
10. LEGLEYE, S., BECK, F., PERETTI-WATEL, P., CHAU, N. & FIRDION, J. M. (2010) Suicidal ideation among young French adults: association with occupation, family, sexual activity, personal background and drug use, *J Affect Disord*, 123, 108-15.
11. GATES, G., J (2011) How many people are lesbian, gay, bisexual or transgender?, pp. 8 (Los Angeles, Williams Institute).
12. LEGLEYE, S., CHARRANCE, G. & RAZAFINDRATSIMA, N. (2015) The use of a non-probability Internet panel to monitor sexual and reproductive health in the general population, *Sociological Methods and Research*, Online advance publication.
13. ROSENBAUM, P. R. & RUBIN, D. B. (1984) Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, *Journal of the American Statistical Association*, 79, 516-524.
14. ROSENBAUM, P. R. & RUBIN, D. B. (1983) The central role of propensity scores in observational studies for causal effects, *Biometrika*, 70, 45-55.
15. BROWN, T. N. & HERMAN, J. L. (2014) Voter ID laws and their added costs for transgender voters, pp. 1-24 (Williams Institute).