

**Utiliser des enquêtes ouvertes sur Internet pour améliorer le redressement d'une enquête téléphonique probabiliste ?**

**Le cas de VIRAGE et de la population LGBT**

**Stéphane Legleye<sup>1,2</sup>, Géraldine Charrance<sup>1</sup>, Christelle Hamel<sup>1</sup>**

**1 Institut national d'études démographiques**

**2 Univ. Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, *Villejuif*, France**

- VIRAGE, ce n'est pas UNE enquête mais DES enquêtes :
  - Enquête principale, téléphonique, aléatoire
  - Enquêtes ouvertes, sur Internet, auprès des populations LGBT & Victimes de violences
  - Enquêtes exhaustive sur Internet auprès des étudiants de 4 universités (Paris 1, Paris 7, Brest, Strasbourg)

**Objectif:** Estimer des prévalences des violences en population générale

- en assurant **une bonne couverture des sous-populations particulièrement exposées,**
- et en fournissant un **bon niveau de détail sur la contextualisation** des violences subies

Interrogation sur un grand nombre de types de violences (physiques, psychologiques, sexuelles...) et dans de nombreuses sphères (couple, travail, famille)

- Population âgée entre 20 et 69 ans, résidant en France métropolitaine en ménages ordinaires
- Enquête réalisée entre février et novembre 2015,
- Taux de réponse de 49% selon calcul AAPOR
- 27 268 questionnaires collectés

- Méthodologie de collecte classique
  - Génération aléatoire de numéros, filtrage ARCEP, annuaire inversé, test tonalité avec automate
  - Vérification de l'éligibilité des lignes : fixe ou mobile ne prenant pas d'appels sur sa ligne fixe lorsqu'il ne connaît pas le numéro appelant
  - Tirage d'un individu dans le ménage selon la méthode Kish
  - Jusqu'à 20 tentatives d'appels avant sélection, 20 après sélection
  - 2 rappels après refus avant sélection, 1 rappel après refus après sélection

# Contexte et objectifs de l'étude

---

- Population d'intérêt : Personnes se déclarant homosexuels ou bisexuels que nous appellerons par abus de langage LGBT
- 503 individus (sur 27 268) se déclarent LGBT dans l'enquête principale
- D'après l'enquête principale, on estime que la **proportion de LGBT** parmi les 20-69 ans est de **1,56% (609 410 individus)**
- Hypothèse : cette proportion est sous-estimée en raison :
  - D'une erreur de mesure
  - Biais de non-réponse totale

# Contexte et objectifs de l'étude (2)

---

- Pour corriger ces biais et améliorer les possibilités d'études de cette population :
  - Calcul d'une marge de calage LGBT
  - Introduction de questionnaires collectés via une enquête ouverte à destination de la population LGBT

# L'enquête Internet

---

- 5790 questionnaires dans la cible sur 7114 complets
- Comparaison des LGBT téléphone et Internet
  - Répondants internet sont plus souvent : femmes, jeunes, diplômés, cadres sup., nées en France et des personnes en couple sans enfants
  - Répondants Internet ***rapportent plus souvent des violences, et en en plus grand nombre***
  - En contrôlant les variables SD, maintien d'un effet propre de l'échantillon pour les violences rapportées



# Calcul de marges LGBT

---

- Informations requises dans les 2 enquêtes :
  - Fréquentation des lieux LGBT
    - Notée FreqT pour l'enquête tel. (43,5%)
    - Notée FreqW pour l'enquête Internet (82,8%)
  - Non joignabilité par téléphone
    - Notée NJT pour l'enquête tel. (0%)
    - Notée NJW pour l'enquête Internet (29,1%)

On estime donc que la population LGBT non couverte par l'enq. tel. est égale à 10,5% ( $FreqT * FreqW * NJW$ )

Soit 63 990 individus

# Introduction de quest. Issus d'une enq. ouverte

---

- Différences avérées (effet du mode de collecte et de l'auto-sélection)
  - Se prémunir contre un effet de mode mais laisser entrer en ligne de compte une certaine diversité sociodémographique ?
- ➔ Sélectionner des quest. comparables à ceux collectés par tél. (pour éviter les effets de mode) du point de vue du sexe, de l'âge et des violences déclarées

# Score de propension & appariement

---

- On modélise, pour chaque sexe et selon le fait d'être joignable par téléphone ou non (pour les web), la propension à répondre par tél. plutôt que sur Internet en fonction :
  - de la classe d'âge
  - des compteurs de violences au cours des douze derniers mois (psychologiques, physiques, psychologiques à caractère sexuel, sexuelles au cours des 12 derniers mois)
  - des 4 interactions entre la classe d'âge et les compteurs de violences
- Puis on apparie les questionnaires un à un sur le score de propension

# Résultats de l'appariement

---

- Les questionnaires LGBT collectés par téléphone sont appariés entre 1 et 3 fois à un questionnaire web (sur 2 décimales)
- Sur les 5790 quest. Web, 2383 sont ainsi appariés (soit 41%), 1397 joignables et 986 non joignables
- On passe de 503 questionnaires LGBT à 2886 (soit une multiplication par 5,7)

# Calcul de pseudo-poids de sondage

---

- Enquête web : non probabiliste – pas de probabilité de sondage
- Pour les questionnaires web joignables par tél. :
  - on ne veut pas augmenter le poids total de la pop. LGBT joignables par téléphone
  - on divise donc la somme des poids par le nouvel effectif de questionnaires LGBT, on obtient le poids individuel d'un quest. web joignables
  - on applique un coef. correcteur au poids de sondage des quest. tél.

# Calcul de pseudo-poids de sondage

---

- Pour les questionnaires web non joignables par tél. :
  - on souhaite que la somme des poids de sondage des questionnaires LGBT non joignables soit proportionnelle à la taille de la population non joignable
  - Et que les poids individuels soient du même ordre de grandeur que pour la population LGBT joignable par tél.

# Calage sur marges

---

- Calage réalisé par sexe sur les caractéristiques suivantes :
  - Classe d'âge
  - Diplôme
  - Catégorie socioprofessionnelle
  - Type de ménage
  - Lieu de naissance
  - Région de résidence
  - Statut LGBT \* classe d'âge

# Résultats du calage sur marges

Indicateurs	Poids de sondage	Poids après calage	Poids après calage tronqué (1% max & min)
Moyenne	1321,5	1319,1	1319,1
Min	31,0	10,8	47,3
Max	12 986,2	30 134,1	6 069,2
Max/Min	418,9	2790,2	128,3
Coef. var	65,3	99,5	89,9

## Mesure du prct d'efficacité des échantillons

En termes de taille d'échantillon efficace, on passe, pour la pop. LGBT de 300 personnes à 1675 (eff. Multipliés par 5,5).

Sur la pop. Totale, la progression est de 8%.



# Résultats du calage sur marges

- Evolution des prévalences de violences**

	Population totale			Population LGBT		
	Ech. tel calé tronqué	Ech. composite calé tronqué	ecarts en pts de prct	Ech. tel calé tronqué	Ech. composite calé tronqué	ecarts en pts de prct
Violences psychologiques au cours des 12 derniers mois						
Au moins une	30,52	30,55	0,03	47,18	48,42	1,24
Violences psychologiques à caractère sexuel au cours des 12 derniers mois						
Au moins une	13,65	13,65	0,00	30,56	30,7	0,14
Violences physiques au cours des 12 derniers mois						
Au moins une	4,59	4,54	-0,01	7,89	6,69	-1,2
Violences sexuelles au cours des 12 derniers mois						
Au moins une	1,91	1,91	-0,01	8,00	6,39	-1,61
Orientation sexuelle						
Homo/bisexuels	1,57	1,77	-0,01	100	100	0

Différences minimales sur la population cible

Plus marquées sur la pop. LGBT, hausse des violences psy, baisse des violences physiques et sexuelles

- ***6 difficultés associées à cette méthode***

- Apprécier la justesse de l'ampleur de la correction de la marge
- Choix des variables de modélisation du score de propension
- Donner un poids aux questionnaires appariés
- Calcul d'une variance associée
- Intégrer les pop. transexuelles et transgenres dans la méthode
- Représenter la pop. LGBT non couverte par l'enq. tél.

# Discussion (2)

---

## **1) Apprécier la justesse de l'ampleur de la correction de la marge**

Marge calculée à partir de la prop. de LGBT fréquentant des lieux LGBT., or même des LGBT ne fréquentant pas de lieux physiques LGBT ont pu être exposés à l'enquête

Sous-estimation par définition

Quid de la justesse de l'estimation de la proportion de non joignables par téléphone ?

Un élément rassurant : faible variation avec le sexe et l'âge

## Discussion (3)

---

### **2) Choix des variables de modélisation du score de propension**

Théoriquement on peut rendre les questionnaires Internet comparables aux questionnaires téléphone pour les variables d'intérêt, les variables sociodémographiques et leur liaison.

Mais restriction au sexe, trois classes d'âge, 4 compteurs de types de violences

Nous ne contrôlons pas toutes les car. sociodémographiques qui pourraient être liées aux violences

La structure de corrélation Violences / Variables de calage est partiellement celle de l'échantillon téléphonique

→ Le calage restitue une différence

Des travaux ultérieurs tenteront d'apporter des éléments de réponse

## Discussion (4)

---

### 3) Choix du poids des questionnaires Internet appariés (avant calage)

L'équipement téléphonique n'est pas questionné de façon identique pour l'enquête téléphonique et pour les joignables Internet ; il ne peut pas être pris en compte pour les injoignables

Choix retenus :

→ Joignables : redistribution des poids pour maintenir une somme de poids stable sur les joignables

Laisse de côté l'équipement téléphonique - pourrait être celui de l'homologue téléphone ?

→ Injoignables : Ratio par rapport à la marge additionnelle

# Discussion (5)

---

## 4) Calcul d'une variance associée

Procédure utilisée réduit la variabilité des variables d'intérêt  
Estimation de la marge non couverte par le tel. conduit à de très faibles poids pour les questionnaires appariés

Il est peu probable que la variance globale soit nettement augmentée (sous l'hypothèse que les biais de sélection sont gommés)

*Des travaux ultérieurs tenteront d'apporter des éléments de réponse*

## Discussion (6)

---

### **5) Intégrer les pop. transsexuelles et transgenres**

Dans l'enq. tél., pas de questions sur les identités de genre, uniquement sur l'orientation sexuelle

→ Sous estimation de la pop. LGBT

→ Aux Etats-Unis, la prop. De transgenres se situe entre 0,2% et 0,3% selon plusieurs études de Williams Institute

# Discussion (7)

---

## 6) Représenter la pop. LGBT non couverte par l'enq. Tél.

Processus non aléatoire + modélisation + appariement sur score de propension = situation contrefactuelle où les LGBT auraient été contactés par tel, tout effet mode de collecte aurait été gommé.

→ Avantage : ne pas introduire d'effet mode de collecte

→ Inconvénient : ne représente pas la diversité de la pop. sur internet.

Les répondants internet candidats à l'appariement sont plus exposés à certaines violences que la pop. finalement sélectionnée

Les prévalences de violences au sein de cette sous-population ne sont donc pas mieux estimées (sélection sur score)