

# LA PRECISION DE L'ESTIMATION DU NOMBRE DE COUPLES DE MEME SEXE EN FRANCE A PARTIR D'UNE ENQUETE INCLUSE DANS LE RECENSEMENT

Laurent Toulemon<sup>1</sup> ; Elisabeth Morand<sup>2</sup>, Nicolas Razafindratsima<sup>3</sup> & Romain Damian<sup>4</sup>

<sup>1</sup> *Institut national d'études démographiques (Ined),  
133, Bd DAVOUT, 75 980 PARIS Cedex 20*

*toulemon@ined.fr*

<sup>2</sup> *Ined elisabeth.morand@ined.fr*

<sup>3</sup> *Ined razafind@ined.fr*

<sup>4</sup> *Ined romain.damian@ensae.fr*

**Résumé.** Estimer le nombre de couples de même sexe est particulièrement difficile pour la statistique publique. Même en limitant l'observation aux couples corésidents, les résultats du recensement croisant le sexe des habitants et leur situation de couple ne suffisent pas, car un taux d'erreur très faible pour la question sur le sexe conduit à une proportion importante de « faux couples de même sexe » parmi les couples identifiés comme tels. Pour éviter de telles erreurs, il est donc nécessaire soit d'inclure une question (ou une modalité de réponse) spécifique pour les couples de même sexe, soit de poser explicitement une question sur le sexe du partenaire.

C'est ce qui a été fait dans l'enquête Famille et logements, une enquête réalisée dans le cadre de la vague annuelle de recensement de 2011, permettant une identification sans erreur des couples de même sexe. Cependant, l'échantillon étant constitué de groupes de districts ou d'Iris du recensement (correspondant à peu près aux secteurs des agents recenseurs), ce qui peut entraîner un effet de grappe si les couples de même sexe ont tendance à se regrouper géographiquement. Cet effet s'ajoute aux autres effets de la constitution de l'échantillon : strates par région, à probabilités inégales, échantillon inclus dans l'échantillon annuel du recensement, tirage de communes puis de grappes au sein des communes. La présente communication propose différentes méthodes pour estimer la variance du nombre de couples de même sexe déduit de l'enquête.

**Mots-clés.** Couples de même sexe, sondage, estimation de variance, effet de grappe, recensement, enquête, France

## Introduction

En 1999 l'échantillon de l'enquête Étude l'histoire familiale souffrait d'un très fort effet de grappe : un délégué sur dix, puis tirage au sort dans les secteurs d'agents recenseurs (Cassan et al., 2000). Depuis 2004, le recensement rénové de la population a remplacé le recensement exhaustif et simultané de la population, permettant ainsi à l'Insee de produire chaque année des estimations de population (Desplanques 2008). En 2011, l'effet de grappe est bien moindre : multiplication du nombre de communes participantes, grappes moins concentrées géographiquement dans les grandes communes, car fondées sur le tirage d'adresses dans un des cinq cinquièmes imbriqués (au

recensement, les adresses successives dans une rue sont affectées à des années de collecte différentes).

Par contre les populations d'intérêt sont plus sensibles à l'effet de grappe qu'en 1999, dans lesquelles les analyses se concentraient sur le passé, tandis qu'en 2011 on s'intéresse à des groupes définis par leur situation au moment de l'enquête. De plus, certains groupes comme les couples de même sexe sont connus pour se regrouper géographiquement, dans des quartiers dans lesquelles la discrimination est moins présente. L'estimation de la variance des estimations est donc cruciale pour ces groupes.

Dans cette communication, on s'intéressera plus spécifiquement aux couples de même sexe cohabitants pour lesquels on souhaite d'une part estimer la variance du nombre de couples de même sexe (en séparant hommes et femmes, avec ou sans enfant) en tenant compte de l'effet de grappe probable pour ces couples et de la variance des poids, et d'autre part valider les possibilités d'extrapolation de cette estimation aux Enquêtes annuelles de recensement et au recensement lui-même.

## **1 L'enquête Famille et logements permet, pour la première fois en France, d'estimer le nombre de couples de même sexe**

Les données sont issues de l'enquête Famille et Logement de 2011 (EFL) qui a été réalisée par l'Institut de la statistique et des études Économiques (Insee). L'Ined a initié l'enquête et participé à son élaboration (Rault et al, 2010). Dans chaque ménage recensé, une « feuille de logement », questionnaire de 4 pages, ainsi que des « bulletins individuels » de 2 pages (un par habitant du logement) sont distribués, et recueilli quelques jours plus tard par le recenseur. Les 2 400 agents recenseurs (un sur cinquante) qui ont participé à l'enquête Famille et Logements ont distribué en plus, dans chaque ménage concerné, un questionnaire spécifique supplémentaire de 4 pages à chaque adulte d'un sexe donné (né en 1992 ou avant) ; un tiers des agents recenseurs ont distribué des bulletins à tous les hommes adultes dans les ménages qu'ils recensaient, les deux autres tiers ont distribué des bulletins à toutes les femmes adultes. Les bulletins pour les hommes et les femmes sont sexués (on s'adresse à « Madame » ou « Monsieur »), mais les questions sont posées identiques. Située en début de questionnaire, la question « Êtes-vous actuellement en couple ? » prévoit deux réponses positives : « Oui, avec une personne qui vit dans le logement » et « Oui, avec une personne qui vit dans un autre logement ». En cas de réponse positive, des précisions sont demandées sur le conjoint, notamment son sexe et sa date de naissance<sup>1</sup>.

Même en limitant l'observation aux couples corésidents, les résultats du recensement croisant le sexe des habitants et leur situation de couple ne suffisent pas pour proposer une estimation fiable des couples de même sexe, car un taux d'erreur très faible pour la question sur le sexe conduit à une proportion élevée de « faux couples de même sexe » parmi les couples identifiés comme tels : il suffit d'une erreur sur le sexe d'un des partenaires d'un couple de sexe différent pour « créer » un faux couple de même sexe. Si les couples de même sexe représentent environ 0,6% de l'ensemble des couples corésidents au recensement, une erreur sur le sexe de 0,3% des personnes générerait 0,6% de personnes vivant dans un « faux couple de même sexe » et conduirait donc à doubler l'estimation du nombre de couples de même sexe (Toulemon et al., 2005 ; Banens 2013). Dans les différents pays d'Europe, des questions spécifiques sont donc introduites dans les recensements (Festy, 2007 ; Cortina, Festy 2014), tandis qu'en France le recensement reste très pauvre par

---

<sup>1</sup> Les bulletins sont accessibles sur le site de l'Insee

(<http://www.insee.fr/fr/publics/default.asp?page=communication/recensement/particuliers/efl1.htm> et <http://www.insee.fr/fr/methodes/default.asp?page=sources/sou-enq-famille-logements.htm>), et sur le site de l'Ined ([http://lili-efl2011.site.ined.fr/fr/l\\_enquete](http://lili-efl2011.site.ined.fr/fr/l_enquete)).

rapport aux autres pays d'Europe, sur ce sujet comme sur la plupart des sujets démographiques et sociaux (Festy 2012).

L'enquête Famille et logements vient donc compléter le recensement sur les questions de famille et de fécondité (Buisson, Lapinte 2013 ; Masson 2013). Concernant les couples de même sexe, l'introduction d'une question sur le sexe du conjoint était une première, et de nombreuses vérifications ont été réalisées pour éviter que des erreurs sur le sexe d'un des conjoints de couples de sexes différents conduisent à construire artificiellement des « faux couples de même sexe ». D'une part, compte tenu du mode de collecte, les deux conjoints des couples de même sexe ont rempli un bulletin de l'enquête Famille et logements, ce qui permet de comparer les informations fournies par chacun sur lui-même et sur son conjoint (sexe, date de naissance). D'autre part, tous les habitants ayant rempli un bulletin du recensement, ces informations ont pu être vérifiées grâce à l'information des bulletins, pour les deux conjoints des couples de même sexe ; enfin, dans certains cas, le retour aux images des bulletins a permis de lever les ambiguïtés (Buisson, 2014).

L'Insee a donc pu publier une première estimation du nombre de couples de même sexe, co-résidents au recensement, de 83 500 couples, soit 167 000 personnes (97 500 hommes et 69 500 femmes) (Buisson, Lapinte 2013).

## **2 Comment estimer la variance de cette estimation ?**

Pour connaître la précision de cette estimation, il est nécessaire d'en estimer la variance, en tenant compte des effets du plan de sondage, notamment de l'effet de grappe. L'enquête EFL a été réalisée sur un sous-échantillon de l'enquête annuelle de recensement de 2011 (EAR). Le plan de sondage est en 3 phases, à partir de l'échantillon de la vague annuelle 2011 du recensement :

- Stratification par région, avec probabilité d'inclusion décroissante avec la taille de la région, de manière à ce que la taille de l'échantillon varie proportionnellement à la racine de la population de la région. La région Île-de-France a été scindée en deux, la Ville de Paris (dans laquelle une extension a été réalisée) étant isolée ;
- Tirage de communes avec des probabilités inégales, après constitution de trois types de communes. Ce tirage compense le plus faible taux de sondage du recensement dans les grandes communes, en compensant le fait qu'une petite commune (commune de moins de 10 000 habitants) sur cinq est incluse entièrement dans la vague annuelle de 2011, tandis que dans les grandes communes environ un logement sur huit seulement est inclus. Une première série de tirage des communes avec une probabilité d'inclusion proportionnelle à sa taille conduit à construire trois types de communes (Loonis, 2010) :
  - o Systématiques : grandes communes apparaissant dans plusieurs tirages préliminaires, dans lesquelles on tire de nombreuses grappes ;
  - o Exhaustives : communes dans lesquelles il n'y a que 1 ou 2 grappes, que l'on tire entièrement ;
  - o Ordinaires : les autres, dans lesquelles on tire deux grappes ;
- Tirage d'une ou deux grappes (davantage dans les communes exhaustives) dans la commune

Chaque grappe est affectée à un sexe, c'est-à-dire que soit tous les hommes, soit toutes les femmes, âgés de 18 ans ou plus à la date d'enquête (nés en 1992 ou avant), remplissent un bulletin de l'enquête en plus des documents du recensement.

L'effet de sondage (stratification puis tirage en grappes à deux degrés, communes et zones de collecte) peut être pris en compte pour estimer la variance de l'estimation, en incluant le plan de

sondage du recensement : une petite commune sur cinq, un cinquième des « grandes adresses » et des logements neufs et un cinquième des autres logements dans les grandes communes (Ardilly, 2013).

L'effet de grappe peut être accru si les couples de même sexe se regroupent géographiquement à petite échelle, par un mécanisme de ségrégation spatiale ? Ce phénomène de regroupement à petite échelle est repérable en comptant le nombre total de couples de même sexe dans chacune des grappes ultimes de collecte, et en le comparant à un ou plusieurs autres ensembles de couples pour lesquels aucun mécanisme de ségrégation ne peut être envisagé.

Enfin, les probabilités d'inclusion inégales, indépendamment de l'effet de grappe, peuvent accroître la variance des estimations.

### **3 Méthodes d'estimation de la variance**

Nous allons utiliser quatre méthodes pour estimer la variance de l'estimation du nombre total de couples de même sexe, en séparant dans un premier temps les estimations des hommes et des femmes (les couples avec et sans enfants pourront aussi être comptés séparément).

Une première méthode consiste à reproduire le plan de sondage dans toute sa complexité. Les formules proposées par Pascal Ardilly permettront d'estimer la variance de l'estimation du nombre de couples. L'information sur les procédures de calage du recensement disponible sur le site de l'Insee (Insee 2014) ne renseigne pas sur la variance des estimations issues d'une EAR. Nous partons donc de l'hypothèse selon laquelle l'EAR résulte d'un tirage aléatoire de logements dans chaque région. Sous cette hypothèse il revient au même de tirer dans l'EAR ou de tirer dans l'ensemble des logements ordinaires. De nombreuses hypothèses sont utilisées dans ces estimations : constitution de strates de communes au sein des régions, selon la manière dont elles sont traitées au recensement, constitution de strates au sein des grandes communes pour isoler les « grandes adresses » et logements neufs (dont la probabilité d'inclusion est de 1/5 au lieu de 1/8), relations entre les échantillons féminins et masculins, prise en compte de la post-stratification, etc.

La deuxième méthode consiste à simplifier autant que possible la représentation du plan de sondage, et à travailler à l'échelle des grappes de collecte (secteurs d'agents recenseurs). On peut compter le nombre de couples de même sexe dans chaque grappe et simuler facilement le plan de sondage dans chaque strate : tirage de communes, puis tirage de grappes au sein des communes). Cette simplification du tirage, possible puisque toutes les personnes d'un sexe donné résidant dans une grappe sont inclus dans l'échantillon, permet d'utiliser les procédures des logiciels statistiques SAS et Stata pour calculer la variance du nombre total de couples de même, en se plaçant dans le cadre simplifié de sondages à deux degrés (communes puis grappes de collecte) à l'intérieur de strates séparées.

Deux analyses complémentaires viendront compléter et valider ces estimations. Une troisième méthode se concentre sur l'augmentation de la variance due à la dispersion des poids, importante dans l'enquête Famille, bien que le taux de sondage plus important dans les grandes communes compense un taux de sondage plus faible au recensement. Le plan de sondage est un plan complexe (Blanpain 2013). La variation des poids est une conséquence du plan de sondage, ainsi que des calages effectués sur l'échantillon. Bien qu'elle ne traduise pas la complexité du plan de sondage (on peut avoir des effets de grappe sans variation des poids, et des variations des poids sans effet de grappe), la variation des poids est un facteur d'augmentation de la variance dont on peut estimer l'effet en tant que tel. On peut ainsi estimer la variance du total en ne tenant compte que la variance des poids. Sous l'hypothèse que les poids sont indépendants de la variable d'intérêt, on utilise la formule, valable pour deux variables aléatoires indépendantes :

$$V(XY) = V(X) V(Y) + V(X) [E(Y)]^2 + V(Y) [E(X)]^2$$

En posant  $X$  le fait de vivre en couple de même sexe (la variable qui nous intéresse, de moyenne proche de 0) et  $Y = w$  le poids individuel, ramené à une moyenne de 1, on obtient :

$$V(wX) = V(X) V(w) + V(X) + V(w) [E(X)]^2$$

$$V(wX) \approx V(X) [1 + V(w)]$$

On peut donc avoir un ordre de grandeur de la perte d'efficacité de l'échantillon en calculant un effectif équivalent :  $n_{eq} = n / [1 + V(w)]$ , qui correspondrait à l'échantillon d'un tirage à probabilités égales.

La quatrième méthode s'intéresse à la variance entre les grappes de collecte du nombre de couples de même sexe. On peut modéliser le nombre de couples de même sexe comme le produit de quatre facteurs : un effet de taille (nombre de logements ou nombre d'habitants ou d'adultes), un effet de composition de la grappe (composition de la population par âge, profession, niveau de diplôme, position dans le tissu urbain, etc.), un effet dû à la corrélation entre logements entraînée par la ségrégation spatiale, et une composante aléatoire suivant une loi de Poisson. On mesure l'effet de la ségrégation spatiale comme le rapport entre le nombre de grappes incluant au moins un couple de même sexe à celui que l'on observerait en l'absence de corrélation. Pour des grappes homogènes, si le nombre moyen de couples par grappe vaut  $\lambda$ , et qu'on suppose qu'il suit une loi de Poisson, la proportion de grappes contenant au moins un couple vaut en espérance  $[1 - \exp(-\lambda)]$ . Cela permet de repérer les effets de ségrégation. Des groupes de couples de taille comparable, mais pour lesquels on est certain qu'il n'y a pas d'effet de ségrégation spatiale (par exemple des couples repérés par différentes combinaisons des mois de naissance de chacun des deux conjoints et de la différence d'âge entre conjoints), seront utilisés comme « témoins ».

## 4 Premiers résultats

Le repérage des grappes de collecte a posteriori pose un certain nombre de questions pratiques. Nous n'avons pas encore constitué le fichier définitif nettoyé des grappes de collecte, nettoyé des erreurs de codage et incluant les changements de dernière minute opérés lors de la collecte et imparfaitement documentés. Les résultats ci-dessous, calculés sur des fichiers provisoires non validés, sont donc fournis à titre d'ordre de grandeur.

**Nombre de grappes avec au moins un couple de même sexe.** On compte dans le fichier provisoire 557 couples de même sexe (299 couples de femmes et 258 d'hommes)<sup>2</sup> répartis dans 876 grappes d'hommes et 1 521 grappes de femmes (Tableau 1). Au total, 435 grappes de collecte incluent au moins un couple de même sexe. Sans tenir compte de l'effet de taille ou de composition des zones, on peut estimer le nombre de grappes dans lesquelles on aurait dû trouver au moins un couple de même sexe à 495. La perte de précision due à la ségrégation est donc limitée, la variance étant augmentée de 14% (variance en  $1/n$ ) à cause de l'hétérogénéité des grappes de collecte.

---

<sup>2</sup> Après validation de 25 cas douteux qui avaient été écartés dans un premier temps, l'estimation finale du nombre de couples co-résidents sera établie à partir de 582 couples de même sexe (309 couples de femmes et 273 couples d'hommes), portant l'estimation en population de 167 à 172 000 personnes (101 000 hommes et 72 000 femmes) vivant en couple de même sexe et recensées dans le même logement.

**Tableau 1. Répartition des grappes de collecte selon le nombre de couples de même sexe**

	Hommes		Femmes		Ensemble	
		Poisson		Poisson		Poisson
0	695	652,5	1267	1249,6	1962	1902,1
1+	181	223,5	254	271,4	435	494,9
1	134	192,2	217	245,6	351	437,8
2+	47	31,3	37	25,8	84	57,1
2	33	28,3	32	24,1	65	52,4
3	5	2,8	2	1,6	7	4,4
4	6	0,2	3	0,1	9	0,3
5	1	0,0	0	0,0	1	0,0
6	0	0,0	0	0,0	0	0,0
7	2	0,0	0	0,0	2	0,0
Total	876	876	1521	1521	2397	2397
Inflation	1,23		1,07		1,14	

Note : Les colonnes « Poisson » résultent d'une estimation fondée sur l'hypothèse d'homogénéité des grappes, pour chaque sexe séparément. La colonne de droite additionne les estimations pour les grappes « Hommes » et « Femmes ». Le facteur d'inflation est estimé comme l'inverse du rapport du nombre de grappes avec un couple de même sexe à celui estimé sous l'hypothèse d'une répartition poissonnienne.

**Variation des poids.** Avec un coefficient de variation de l'ordre de 0,6 (tableau 2), la variance des pondérations ramenées à une moyenne 1 est de l'ordre de 0,3, ce qui correspond à augmentation de la variance du total de l'ordre de 30%. Pour les couples de même sexe, la variation des poids est plus importante, et l'augmentation de la variance qui en résulte est de l'ordre de 50% pour les hommes et 40% pour les femmes.

**Tableau 2. Variation des poids et augmentation de la variance qui en résulte**

	Ensemble de l'échantillon			Répondants vivant en couples de même sexe	
	Hommes	Femmes	Ensemble	Hommes	Femmes
Effectif	121 312	238 458	359 770	516	598
Moyenne m	187,7	105,1	132,9	184,9	116,2
Ecart-type s	102,8	60,2	86,6	130,8	71,3
Coefficient de variation (s/m)	0,55	0,57	0,65	0,71	0,61
Variance des poids standardisés V	0,30	0,33	0,42	0,50	0,38
Inflation de variance (1+V)	1,30	1,33	1,42	1,50	1,38
Racine(1+V)	1,14	1,15	1,19	1,22	1,17

Les deux facteurs d'augmentation de la variance que sont, d'une part, la corrélation entre individus due à la ségrégation spatiale à petite échelle et, d'autre part, la variance des poids, sont estimés aux tableaux 1 et 2. On ne peut prétendre que ces deux facteurs résument les effets du plan de sondage : d'une part la corrélation entre individus peut opérer à des échelles plus vastes que la grappe de collecte, et la comparaison avec un modèle poissonnien ne résume pas les effets de grappe ; d'autre part, la variation des poids n'est qu'un indicateur très indirect et imparfait de la complexité du plan de sondage. Cependant, le tableau 3 propose une première estimation de l'ordre de grandeur de l'inflation de la variance due à ces deux composantes.

**Tableau 3. Premiers ordres de grandeur de la variance des estimations du nombre de couples de même sexe**

	couples d'hommes	couples de femmes	Couples de même sexe
Nombre	258	299	557
Augmentation de la variance due à l'effet de grappe	1,23	1,07	
Augmentation due à la dispersion des poids	1,50	1,38	
Augmentation totale de la variance	1,85	1,47	
Augmentation des écart-types	1,36	1,21	
Variance poissonnienne	258	299	557
Variance augmentée	478	440	918
Ecart-type poissonien $s_1$	16,1	17,3	33,4
Ecart-type augmenté $s_2$	21,9	21,0	42,8
Largeur relative de l'intervalle de confiance à 95% (estimation = 100)			
$2 s_1/n$ (%)	12,5	11,6	12,0
$2 s_2/n$ (%)	16,9	14,0	15,4

Note : Les chiffres de la colonne « couples de même sexe » sont déduits de l'addition des colonnes « hommes » et « femmes », sous l'hypothèse que les deux échantillons sont indépendants (addition des variances).

Pour les hommes et encore plus pour les femmes, l'effet de ségrégation à l'intérieur des grappes apparaît faible par rapport à l'effet de variance des poids. La combinaison de ces effets conduit à augmenter les variances de 85% et 47% pour les hommes et les femmes respectivement. En estimant la variance à partir d'un modèle poissonnien simple (variance = espérance), on arrive à des intervalles de confiance de plus ou moins 16,9% pour les hommes (au lieu de 12,5% pour un effectif identique issu d'un tirage aléatoire simple). Pour les femmes la largeur de l'intervalle passe de 11,6% à 14%. On en déduit une estimation pour l'ensemble des couples de même sexe, hommes et femmes, de l'ordre de plus ou moins 15%.

## Conclusion

Ces deux premières estimations sont plutôt pessimistes : l'estimation de l'effet de ségrégation spatiale ne tient pas compte de la taille ni de la composition différentes des grappes, et la variance des poids est sûrement compensée par l'effet de stratification de l'échantillon et la post-stratification. Elles conduisent au total à des imprécisions modérées, avec des intervalles de confiance de plus ou moins 17% pour les hommes et plus ou moins 14% pour les femmes. Ces premiers résultats sont donc rassurants quant à l'ampleur de la variance des estimations issues de l'enquête Famille et logements. La communication reprendra ces calculs plus rigoureusement, mais il semble que l'estimation des couples de même sexe, dont la validité s'appuie sur un protocole de collecte et d'apurement qui limite beaucoup les possibilités d'erreurs, soit également assez précise, parce que fondée sur des échantillons de couples de taille limitée mais issus d'un très vaste échantillon.

Par ailleurs, le fichier de l'enquête peut servir de base d'apprentissage pour définir des règles d'identification des couples de même sexe au sein des ménages, à partir de l'information disponible dans le recensement. En utilisant des méthodes d'exploration des données (*data mining*), il est possible d'en inférer des estimations pour l'ensemble du territoire et pour des sous-ensembles mal couverts par l'enquête EFL, à partir de différentes vagues annuelles du recensement ou du recensement complet (regroupement de 5 vagues annuelles).

## Bibliographie

Ardilly P., 2013, « Enquête Famille et Logement 2011. Calcul de précision », Note interne Insee.

Banens M., 2013, “Sex miscoding and same-sex couple estimates. Evidence from the French 2011 census and family survey”, presented at the IUSSP International Population Conference, Busan, South Korea, August 26-31, 2013, 19 p.

[http://www.banens.fr/docs/Busan\\_Sex\\_miscoding.pdf](http://www.banens.fr/docs/Busan_Sex_miscoding.pdf)

Blanpain N., 2013, « Calcul des pondérations de l'enquête Famille et Logements », note interne Insee N° 137/DG75-F170, diffusée avec le fichier de production et de recherche de l'enquête Famille et logements.

Buisson G., Lapinte A., 2013, « Le couple dans tous ses états. Non-cohabitation, conjoints de même sexe, Pacs... », *Insee première*, 1435, 4p.

<http://www.insee.fr/fr/ffc/ipweb/ip1435/ip1435.pdf>

Buisson G., 2014, Note Insee N° 714/DG75-F170, « Apurements et imputations dans l'Enquête famille et logements 2011 », Note diffusée avec le fichier de production et de recherche de l'enquête Famille et logements.

Cassan F., Héran F., Toulemon L., 2000, « Étude de l'histoire familiale : l'édition 1999 de l'enquête Famille », *Courrier des statistiques*, n° 93, p. 19-23.

[http://www.insee.fr/fr/themes/document.asp?reg\\_id=0&id=852](http://www.insee.fr/fr/themes/document.asp?reg_id=0&id=852)

Cortina C., Festy P., 2014, “Identification of same-sex couples and families in censuses, registers and surveys”, Families and Societies Working paper series, 8,

[http://www.familiesandsocieties.eu/?page\\_id=131](http://www.familiesandsocieties.eu/?page_id=131).

Desplanques G., 2008, « Avantages et incertitudes des enquêtes annuelles de recensement en France », *Population*, 63, 3, p. 477-502. Voir aussi [http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/doc/traitement\\_donnees\\_rp.htm](http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/doc/traitement_donnees_rp.htm)

Festy P., 2007, “Enumerating same-sex couples in censuses and population registers”, *Demographic Research*, 17(2), p 339-368.

<http://www.demographic-research.org/volumes/vol17/12/17-12.pdf>

Festy P., 2012, Questions sur la famille et la résidence dans les recensements européens 2010-2011. Complimentary material. Working group on the census form,

[http://www.cnis.fr/cms/Accueil/activites/Organisation/Groupe\\_de\\_travail/Groupe\\_de\\_travail\\_actifs/GT\\_evolution\\_questionnaire\\_recensement/calendrier\\_GT\\_evolution\\_questionnaire?reunion=99383](http://www.cnis.fr/cms/Accueil/activites/Organisation/Groupe_de_travail/Groupe_de_travail_actifs/GT_evolution_questionnaire_recensement/calendrier_GT_evolution_questionnaire?reunion=99383) and

[http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/groupe\\_actifs/questionnaire\\_RP/document\\_s\\_complementaires/DC\\_2011\\_2e\\_reunion\\_GT\\_RP\\_famille\\_RP\\_europeen\\_famille\\_residence.PDF](http://www.cnis.fr/files/content/sites/Cnis/files/Fichiers/groupe_actifs/questionnaire_RP/document_s_complementaires/DC_2011_2e_reunion_GT_RP_famille_RP_europeen_famille_residence.PDF)

Insee, 2014, Notes sur « La précision des résultats du recensement » et « Les pondérations », disponibles sur la page « Conseils pour l'utilisation des résultats du recensement » à l'adresse électronique [http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/doc/conseils\\_utilisation.htm](http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/doc/conseils_utilisation.htm)

Loonis V., 2010, « Mise à disposition de l'échantillon communal de l'Enquête Famille et Logements 2011 », note interne Insee, N° 348/DG75-F410 et « Modifications apportées à l'échantillon initial de l'Enquête Famille et Logements », note interne Insee N° 1134/DG75-F410, en annexe de (Blanpain 2013).

Masson L., 2013, « Avez-vous eu des enfants ? Si oui, combien ? », *France, portrait social* - Insee Références - Édition 2013, p. 93-109.

[http://www.insee.fr/fr/ffc/docs\\_ffc/FPORSOC13g\\_D1\\_fecondite.pdf](http://www.insee.fr/fr/ffc/docs_ffc/FPORSOC13g_D1_fecondite.pdf)

Rault W., Mazuy M., Rivière A., Toulemon L., 2010, « L'enquête Famille et logements associée au recensement de 2011 » in Tremblay Marie-Ève, Lavallée Pierre, El Haj Tirari Mohammed, *Pratiques et méthodes de sondage, Actes du colloque Sondages 2010*, p. 113-117.

[http://www.ined.fr/fichier/t\\_telechargement/59703/telechargement\\_fichier\\_fr\\_actes\\_pages\\_ok.pdf](http://www.ined.fr/fichier/t_telechargement/59703/telechargement_fichier_fr_actes_pages_ok.pdf)

Toulemon L., Vitrac J., Cassan F., 2005, Le difficile comptage des couples homosexuels d'après l'enquête EHF, in Lefèvre C. et Fillon A. (ed.), Ch. 32 in *Histoires de familles, histoires familiales. Les résultats de l'enquête Famille de 1999*, INED, Cahier 156, p. 589-602.

[http://www.ined.fr/lili\\_efl2010/cahier\\_ined\\_156/ci\\_156\\_partie\\_9.32.pdf](http://www.ined.fr/lili_efl2010/cahier_ined_156/ci_156_partie_9.32.pdf).

## THE ACCURACY OF THE ESTIMATE OF THE NUMBER OF SAME-SEX COUPLES IN FRANCE BASED ON A SURVEY INCLUDED IN THE CENSUS

Laurent Toulemon<sup>1</sup> ; Elisabeth Morand<sup>2</sup>, Nicolas Razafindratsima<sup>3</sup> & Romain Damian<sup>4</sup>

<sup>1</sup> *Institut national d'études démographiques (Ined),  
133, Bd DAVOUT, 75 980 PARIS Cedex 20  
toulemon@ined.fr*

<sup>2</sup> *Ined elisabeth.morand@ined.fr*

<sup>3</sup> *Ined razafind@ined.fr*

<sup>4</sup> *Ined romain.damian@ensae.fr*

**Abstract.** Enumerating same-sex couples is a challenge for official statistics, as well as studying the demographics of these couples is a challenge for research. Despite these couples gained in many European countries the right to an official recognition through a specific status or the right to marriage, their identification in censuses and surveys is very difficult. Even if the description is restricted to coresident couples, census data based on cross-tabulation on questions on couple situation and sex do not suffice, as a very limited number of errors in the sex of one or the other member of different-sex couples leads to a strong proportion of “false same-sex couples” among identified “same sex-couples”. In order to avoid such errors, it is necessary to include specific categories in the “couple” question and/or to explicitly ask a question about the sex of the partner.

This was done in France for the first time in a one-percent survey, the Family and dwellings survey, which took place within the 2011 census, allowing a precise enumeration of same-sex couples. As the sampling unit was the enumerator zone, the sample suffers from large cluster effects, if same sex couples are grouped in some neighborhoods. Taking these effects into account is necessary to estimate the variance of the counts. Furthermore, the variance of the counts is also a function of the sampling frame: regional strata, with unequal probabilities, subsample of the yearly wave of the census survey, selection of municipalities, and of clusters within municipalities. Several estimates of the variance of the number of same sex couples are presented in the paper.

**Keywords.** Same-sex couples, sampling, variance estimate, cluster effect, census, survey, France.

