



HAL
open science

Surreprésenter les populations rares grâce au tirage simultané de plusieurs enquêtes

Paul Cochet, Géraldine Charrance, Thomas Merly-Alpa, Laurent Toulemon

► **To cite this version:**

Paul Cochet, Géraldine Charrance, Thomas Merly-Alpa, Laurent Toulemon. Surreprésenter les populations rares grâce au tirage simultané de plusieurs enquêtes. 12e Colloque Francophone sur les Sondages - 2023, Mar 2023, Aubervilliers, France. pp.9. hal-04387265

HAL Id: hal-04387265

<https://hal.science/hal-04387265v1>

Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SURREPRÉSENTER LES POPULATIONS RARES GRÂCE AU TIRAGE SIMULTANÉ DE PLUSIEURS ENQUÊTES

Paul Cochet¹, Géraldine Charrance², Thomas Merly-Alpa³, Laurent Toulemon⁴

¹ *Ined - Service des Enquêtes et Sondages, France, paul.cochet@ined.fr*

² *Ined - Service des Enquêtes et Sondages, France, geraldine.charrance@ined.fr*

³ *Ined - Service des Enquêtes et Sondages, France, thomas.merly-alpa@ined.fr*

⁴ *Ined – Unité de Recherche Fécondité, Familles, Conjugalités, France, toulemon@ined.fr*

Résumé. Les enquêtes Étude des relations familiales et intergénérationnelles (Erfi) et Famille et Employeurs (FamEmp) vont chacune faire l'objet d'un tirage dans Nautile-Fidéli avec une surreprésentation de certaines sous-populations (dont les personnes en couple de même sexe). Afin d'éviter la distorsion qui surviendrait dans la base de sondage suite à ces diverses surreprésentations, l'Insee recourt à une technique de marquage consistant à doubler l'échantillon tiré d'un échantillon complémentaire de manière à ce que la réunion des deux échantillons, ainsi que la base de sondage résiduelle, aient la même structure que la base de sondage initiale. Nous proposons ici une façon de mettre à profit ce marquage afin de surreprésenter certaines populations rares. Il s'agit, par exemple, pour l'enquête Erfi d'ajouter aux couples de même sexe de l'échantillon Erfi ceux que nous pourrions capter dans l'échantillon complémentaire de l'enquête FamEmp. Sur la base de plusieurs simulations, nous montrons que cette méthode présente plusieurs vertus, permettant notamment de doubler le nombre de couples de même sexe présumés que nous pouvons attendre dans notre échantillon tout en réduisant la variance d'échantillonnage de plus d'un tiers.

Mots-clés. Marquage, échantillonnage, populations rares, partage des poids

Abstract. The Erfi and FamEmp surveys will be sampled in Nautile-Fidéli with the aim of oversampling certain sub-populations (including same-sex couples). In order to avoid the distortion that would occur in the sampling frame as a result of these various over-representations, INSEE uses a technique that consists of doubling the sample drawn with a complementary sample so that the sum of the two samples, as well as the residual sampling frame, have the same structure as the initial sampling frame. We propose here a method to take advantage of this marking technique in order to oversample certain rare populations. For instance, the Erfi survey would add to the same-sex couples of Erfi sample those that we could capture in the complementary sample of the FamEmp survey. We present several simulations, which show that this method has several benefits, allowing us to double the number of presumed same-sex couples that we can expect in our sample while reducing the sampling variance by more than one third.

Keywords. Marking, sampling, rare populations, weight sharing

Près de quinze ans après leurs premières éditions, l'Ined prépare les secondes éditions des enquêtes Étude des relations familiales et intergénérationnelles (Erfi) et Famille et Employeurs (FamEmp) pour un lancement à l'horizon 2024. Ces deux enquêtes sont menées conjointement avec une méthodologie (échantillonnage, calendrier, protocole, exploitations) et des questionnaires proches. Elles seront échantillonnées dans la base Nautile-Fidéli de l'Insee (Charrance et Merly-Alpa, 2023) : la concordance des calendriers de collecte permet un tirage simultané des deux échantillons. Nous allons mobiliser les informations sur le niveau de vie pour stratifier nos échantillons ; le niveau de vie est utilisé ici comme un moyen d'une part, de surreprésenter simplement, et sur la base d'un critère unique, certaines configurations familiales et professionnelles d'intérêt pour la recherche, et d'autre part de corriger les taux de réponses différenciés observés lors du test Erfi 2.0, la participation à l'enquête augmentant fortement avec le niveau de vie (Bouchet-Valat et al., 2022).

En plus de ce premier critère de surreprésentation, commun aux deux enquêtes, d'autres critères de surreprésentation propres à chaque enquête sont mobilisés. Ainsi l'enquête Erfi ambitionne de surreprésenter les couples de même sexe là où l'enquête FamEmp projette de surreprésenter les bénéficiaires de l'Allocation d'Education de l'Enfant Handicapé (AEEH).

Pour limiter la charge de collecte sur les individus, l'Insee a pour pratique d'exclure de la base de tirage les individus récemment tirés dans une enquête. Or, si aucune mesure n'était prise, ces diverses surreprésentations résulteraient, après tirage, en une distorsion de la base de sondage résiduelle du point de vue des caractéristiques sociodémographique des individus qui la composent. C'est pour éviter cela que l'Insee recourt à une technique de « marquage », aussi appelée disjonction des échantillons : chaque échantillon tiré se voit doublé d'un autre échantillon, qualifié d'« échantillon complémentaire », de manière à ce que la réunion de ces deux échantillons (et donc la base de sondage résiduelle) ait la même structure que la base de sondage initiale (Paliot, 2021).

Cette communication se propose de formaliser, sur le plan théorique, une manière de mettre à profit l'étape de marquage en repérant dans le complémentaire tiré de chaque enquête des populations d'intérêt rares. De façon plus concrète, cela se traduit par la volonté de capter les couples de même sexe dans le complémentaire de l'enquête FamEmp afin de les ajouter à ceux tirés dans l'échantillon de l'enquête Erfi et inversement, seront captés les bénéficiaires de l'AEEH dans le complémentaire de Erfi en vue de leur intégration à l'échantillon de FamEmp. Nous expliquerons la mise en œuvre de cette méthode et présenterons l'effet sur la variance d'estimation, à l'aide de simulations.

Pour simplifier la présentation, nous nous focaliserons sur la captation des couples de même sexe dans l'échantillon complémentaire de FamEmp, l'opération consistant en la captation des bénéficiaires de l'AEEH dans le complémentaire de l'enquête Erfi étant l'opération miroir de celle décrite dans la suite de ce document.

1 Mise en œuvre de la procédure et résultats attendus

1.1 Principe général

Notons G l'échantillon uniforme marqué pour Erfi et G' l'échantillon mis en collecte, qui surreprésente les couples de même sexe et les déciles extrêmes de niveau de vie. La probabilité pour tout individu d'être dans G' va dépendre des probabilité d'inclusion π_G (constante) et $\pi_{G'/X}$ (qui sont déterminées par le plan de sondage de G') :

$$\pi_{G'} = \pi_G \times \pi_{G'/X}$$

De la même façon, on note F l'échantillon uniforme marqué pour l'enquête FamEmp et F' l'échantillon mis en collecte qui surreprésente les bénéficiaires de l'AAE et les déciles extrêmes de niveau de vie. Nous désirons ajouter à G' les couples de même sexe qui sont dans l'échantillon complémentaire de F' dans F (noté \bar{F}'), c'est-à-dire ceux qui sont présents dans l'échantillon uniforme marqué mais qui ne sont pas mis en collecte pour FamEmp.

La probabilité d'inclusion $\pi_{F/\bar{F}'}$ qu'un individu X soit dans le complémentaire \bar{F}' peut s'écrire de la façon suivante :

$$\pi_{F/\bar{F}'} = \pi_F \times (1 - \pi_{F'/X})$$

Les poids obtenus à partir de G' sont extrapolables à l'ensemble des couples de même sexe cohabitants, tout comme le sont les poids obtenus sur \bar{F}' . Pour tenir compte de ces deux manières d'inclure les couples de mêmes sexe (à partir de G' et \bar{F}') et éviter des doubles comptes (Favre-Martinoz et Gros, 2017), il est nécessaire de recourir à une méthode de partage des poids. Plusieurs méthodes peuvent être envisagées (elles sont discutées en partie 2.1.3), mais elles reposent toutes sur la nécessité de compter le « nombre de liens », c'est-à-dire le nombre de façons d'atteindre un individu. Il faut donc, pour la mise en œuvre d'une telle opération, avoir à disposition une indicatrice indiquant si l'individu X peut théoriquement être capté à la fois dans l'échantillon G' (c'est toujours le cas) et dans l'échantillon complémentaire \bar{F}' (ce n'est le cas que pour les strates de F' non sélectionnées exhaustivement dans F).

1.2 Résultats attendus sur le nombre de couples cohabitants de même sexe par la réunion des différents échantillons

D'après les simulations réalisées et les résultats du test de l'enquête Erfi 2.0 (Bouchet-Valat et al., *op. cit.*), afin d'atteindre l'objectif fixé de 10 000 répondants, nous devrions tirer un échantillon G de 139 110 individus. En outre, nous estimons la part d'individus appartenant à couple de même sexe au sens de Fidéli¹ à 0,9% environ sur l'ensemble du fichier. Nous serions donc en mesure de capter, à travers notre seul échantillon G , 1 252 couples de même sexe présumés (sans prendre en compte les effets croisés entre le fait d'être en couple de même sexe et le décile de niveau de vie: corrélation, participation différentielle à l'enquête). Il s'agit en réalité d'une estimation haute, l'échantillon de 139 110 individus comprenant plusieurs échantillons de réserve qui ne seront mis en exploitation que si le terrain est amené à s'essouffler avant que nous n'atteignons l'objectif de 10 000 répondants. En ne tenant pas compte de ces échantillons de réserve, on attend 1011 individus vivant en couple de même sexe dans l'échantillon G .

A ce millier d'individus échantillonnés dans G et vivant en couple de même sexe présumé nous allons ajouter les individus issus de \bar{F}' : celui-ci étant d'environ 119 836 individus, ce qui nous conduirait, avec le même taux de 0,9%, à ajouter 1 079 individus vivant en couple de même sexe (là aussi, sans tenir compte des liens entre cette caractéristique et les surreprésentations à l'œuvre dans F').

Ainsi, en mettant à profit l'échantillon complémentaire dans le cadre du tirage d'échantillons de deux enquêtes en simultané, nous sommes en mesure de doubler la taille du

¹ Est défini comme appartenant à un couple de même sexe tout individu pour qui la variable lien_familial est égale à 10, 21 ou 22 et pour qui il existe un autre individu dans le même logement avec lien_familial = 10, 21 ou 22 qui est du même sexe. Nous parlons ici de « couples de même sexe au sens de Fidéli » ou encore de « couples de même sexe présumés » car la variable d'identification des couples de même sexe est imparfaite, captant notamment bon nombre de personnes en collocation. D'après des analyses réalisées sur les données d'Epicov, sur l'ensemble des couples de même sexe repérés par notre variable d'identification, 53% sont effectivement en couple de même sexe au moment de l'enquête.

domaine couples de même sexe au sein de notre échantillon sans augmenter la taille de l'échantillon de marquage (échantillon de collecte + échantillon complémentaire) équilibré du point de vue des caractéristiques sociodémographiques des individus qui la composent.

Tableau 1: Estimations du nombre de couples de même sexe dans chaque échantillon

	Echantillon total marqué pour Erfi (G)	Echantillon marqué non mobilisé pour FamEmp (\bar{F}')
Strate	n	n
<i>Décile de niveau de vie manquant</i>	16 530	12 094
<i>Décile 1</i>	12 258	-
<i>Décile 2</i>	12 258	5 235
<i>Décile 3</i>	12 258	8 640
<i>Décile 4</i>	12 258	11 051
<i>Décile 5</i>	12 258	12 856
<i>Décile 6</i>	12 258	13 707
<i>Décile 7</i>	12 258	13 869
<i>Décile 8</i>	12 258	13 980
<i>Décile 9</i>	12 258	14 056
<i>Décile 10</i>	12 258	14 104
<i>Ménage percevant l'AAEH*</i>		244
Ensemble	139 110	119 836
Part de CMS	0,90%	0,90%
Nombre de CMS	1 252	1 079

*Strate spécifique à l'enquête FamEmp

2 Simulation du gain en précision sur le domaine couples de même sexe lié au tirage dans le complémentaire de FamEmp

Nous nous attachons dans cette partie à reproduire, de façon simplifiée, les conditions du tirage de l'enquête Erfi afin notamment d'évaluer le gain en précision de l'estimation auquel nous pouvons nous attendre en ajoutant un millier de couples de même sexe présumés via le complémentaire de FamEmp. Comme nous l'avons vu précédemment, l'échantillon de l'enquête Erfi contient approximativement un millier de couples de même sexe présumés auxquels nous ajoutons un millier de couples de même sexe grâce au complémentaire de l'enquête FamEmp (cf. tableau 1). Cependant, la variance liée à ce tirage pourrait être légèrement supérieure à celle d'un tirage direct de ces 2000 individus, car une partie de l'échantillon provient d'un mécanisme complexe. Nous cherchons ici à estimer la part d'augmentation de variance liée à cette méthode, pour en estimer le *design effect*.

Pour cela nous avons créé une base de sondage fictive contenant un million d'individus sur lesquels trois variables sont renseignées :

1. Une variable catégorielle A prenant la valeur 1 pour 50% de l'échantillon, le reste prenant la valeur 2. Cette variable A nous servira par la suite de variable de stratification au même titre que les déciles pour le tirage de l'enquête.

2. Une variable catégorielle B prenant la valeur 1 pour 10% de l'échantillon², les 90% restant prenant la valeur 2. Cette variable est assimilable à la variable d'identification des couples de même sexe présumés qui sera à construire pour le tirage des deux enquêtes.
3. Une variable d'intérêt continue X attribuant à chaque individu une valeur aléatoire selon une loi log-normale de paramètres $Log - N(1700,1500)$ ³

Le plan de sondage de l'enquête FamEmp est un tirage stratifié selon A pour lequel on souhaite obtenir 15 000 individus ayant la modalité A1 et 3 750 ayant la modalité A2. Pour gérer le marquage, on tire un échantillon uniforme de taille $n = 30\,000$. Le complémentaire pour le marquage est composé des 11 250 individus A2 non sélectionnés pour l'enquête, et ne contient donc aucun individu A1.

Le plan de sondage de l'enquête Erfi est un tirage stratifié selon B , qui conduit à avoir 1 000 individus B1⁴. Il est tiré dans la base après marquage de l'enquête FamEmp, c'est-à-dire dans la base de sondage dont on a retiré l'échantillon uniforme de taille 30 000 obtenu à l'étape précédente.

2.1 Méthode et paramètres testés

2.1.1 Scénarios de dépendance entre variables

Afin d'anticiper les différentes configurations possibles, plusieurs scénarios ont été testés faisant varier la dépendance entre les variables. Ces scénarios sont au nombre de cinq :

1. Absence de corrélation entre les variables (scénario « indépendance »)
2. Corrélation entre A et X (scénario « A_X ») : pour ce faire une variable Z a été créée en ajoutant à X une valeur définie aléatoirement selon une loi normale $N(0,100)$. La valeur de A a ensuite été définie à 1 pour toutes les observations pour lesquelles $Z_i < \text{médiane}(Z)$ et 2 pour celles au-dessus de la médiane.
3. Corrélation entre A et B (scénario « A_B ») : les probabilités conditionnelles $P(A1/B1)$ et $P(A2/B1)$ ont respectivement été définies à 20% et 80%. Les probabilités $P(A1/B2)$ et $P(A2/B2)$ sont en revanche inchangées et restent à 50%.
4. Corrélation entre B et X (scénario « B_X ») : les valeurs X_i attribués aux individus B1 ont été modifiées de manière à ce qu'elles suivent une loi log-normale prenant pour paramètres $Log - N(2300,1300)$; les valeurs X des individus B2 conservent la distribution initiale.
5. Corrélation entre A , B et X (scénario « AB_X ») : ce scénario est une combinaison des scénarios 2 et 4. Les valeurs X_i ont été altérées selon le même processus que dans le scénario 4 suite à quoi nous avons ajouté du bruit à partir d'une variable Z qui a ensuite

² La prévalence de 10% utilisée ici diffère de la proportion de la population en couple de même sexe au sens de la base de sondage, estimée à 0,9%. Nous avons fait ce choix pour pouvoir réaliser des tirages de taille suffisante sans devoir simuler une population trop grande, ce qui aurait largement ralenti les simulations et ralenti la convergence des estimations de variance. Une vérification ex post avec des taux proches de 1% a cependant été réalisée, et conduit à des résultats sensiblement équivalents

³ Ce choix de modélisation et de paramètres s'est fait de manière à ce que nous obtenions une distribution proche de ce que nous pouvons observer empiriquement dans les enquêtes sur des variables telles que les revenus mensuels déclarés.

⁴ On ne s'intéressera pas dans la suite de ce texte aux individus ayant la modalité B2, qui sont les individus qui ne sont pas couples de même sexe.

servi au recodage de A selon les mêmes conditions que dans le scénario 2. Il s'agit probablement du scénario le plus réaliste.

2.1.2 Méthodes de tirage

Afin de quantifier le gain en termes de précision du complémentaire de FamEmp, nous avons comparé les résultats obtenus sur l'estimation de la moyenne de X sur le domaine B1, et sa variabilité, selon que nous mobilisons l'échantillon Erfi seul (ce qui conduit à 1 000 couples de même sexe) ou l'échantillon Erfi auquel s'ajoute les couples de même sexe captés dans complémentaire de FamEmp (ce qui conduit à environ 2 000 couples de même sexe).

La méthode de tirage utilisée est le tirage systématique, trié selon A . Nous introduisons par ailleurs deux variantes pour le tirage de FamEmp : tirage systématique trié seulement selon A (dit « sans tri ») ou tirage systématique trié suivant A et B (dit « avec tri »). Cela permet d'évaluer si la prise en compte de la caractéristique B dans le tirage de la seconde enquête a un effet sur la précision finale obtenue.

2.1.3 Méthodes de pondération

Deux méthodes de partage de poids ont été utilisées pour calculer les poids de sondage.

Une première façon de procéder, qualifiée de partage des poids « classique », consiste à diviser par 2 le poids de sondage d_k des individus captés à la fois dans l'échantillon marqué Erfi et dans le complémentaire de FamEmp (autrement dit, s'il possède la modalité A2).

Les résultats obtenus selon cette première méthode seront comparés avec celle du partage des poids dit « pondéré » où la pondération des liens est adaptée en utilisant le coefficient de répartition optimal (correspondant aux tailles respectives des échantillons) ; cela conduit à attribuer le même poids à tous les individus B1 d'une même strate A , quel que soit leur échantillon d'origine.

2.2 Résultats

Les résultats qui suivent sont basés sur 10 000 tirages réalisés pour chaque croisement scénario \times méthodes de tirage \times méthodes de pondération, pour lesquels nous avons calculé la moyenne et la variance de X . Les résultats de ces différentes simulations sont reportés dans le tableau 2.

Aucun biais n'est introduit par le cumul des individus FamEmp avec partage des poids, quels que soient les scénarios, méthodes de pondération ou encore méthodes de tirage retenus, ce qui est conforme à l'attendu.

L'ajout du complémentaire FamEmp augmente la précision de l'estimation sur le domaine couples de même sexe: nous estimons que la variance peut être réduite *a minima* de 31 %, même si ce gain est inférieur aux 50%⁵ que l'on pourrait théoriquement attendre en doublant la taille de l'échantillon de couples de même sexe ; le *design effect* de la méthode de tirage mise en œuvre est donc de 1,38. Ce gain est conservé dans le scénario le moins favorable, lorsqu' A , B et X sont tous liés entre eux.

Quant à la méthode de tirage, il semblerait que le tirage systématique selon B n'a aucun impact sur la précision de l'estimation. Il n'est donc pas nécessaire d'opérer un quelconque tri sur la variable d'identification des couples de même sexe pour le tirage de l'échantillon FamEmp.

⁵ Ou plus précisément 53% compte tenu des paramètres utilisés pour la simulation.

Enfin, de façon assez prévisible, il semble préférable de recourir à la méthode du partage des poids « pondéré » pour le calcul des poids de sondage, celle-ci offrant une meilleure précision que la méthode du partage des poids « classique ».

Conclusion

La méthode proposée ici, consistant à mobiliser l'échantillon de marquage dans le cas du tirage simultané de plusieurs enquêtes afin de capter des populations d'intérêt rares, présente plusieurs vertus. Tout d'abord, elle permet d'augmenter le nombre d'unités dans les domaines d'intérêt de façon conséquente sans que cela nécessite de tirer directement un sur-échantillon et sans déséquilibrer la base de sondage pour des tirages ultérieurs. Dans l'exemple de l'enquête Erfi, nous pouvons par exemple espérer capter un millier supplémentaire de couples de même sexe présumés à l'aide du complémentaire de l'enquête FamEmp par rapport au millier que nous aurions obtenu en mobilisant le seul échantillon Erfi.

Par ailleurs, nos simulations, même si elles sont réalisées dans un cadre simplifié, montrent que la complexité de la méthode mise en œuvre ne neutralise pas le gain en précision de nos estimations attendu en doublant la taille de l'échantillon : on observe ainsi une réduction de la variance d'échantillonnage estimée à un tiers minimum dans le cas de Erfi, ce qui n'est pas négligeable.

De nouveaux obstacles théoriques pourront se poser après la collecte, notamment si le comportement de non-réponse n'est pas le même pour les individus en couple de même sexe provenant de strates différentes. La collecte des enquêtes Erfi et FamEmp aura lieu en 2023 et 2024, et une analyse détaillée des comportements de réponse des individus sera faite à cette occasion.

Tableau 2 - Synthèse des résultats obtenus par simulation

	Moyenne		Variance		% réduction variance par rapport à Erfi seul		
	PP "classique"	PP "pondéré"	PP "classique"	PP "pondéré"	PP "classique"	PP "pondéré"	Gain PP "pondéré" / Gain PP "classique"
Indépendance							
Erfi + Complémentaire FamEmp avec tri (2000 couples de même sexe)	2060,1	2060,1	1034,9	986,8	32%	35%	1,10
Erfi + Complémentaire FamEmp sans tri (2000 couples de même sexe)	2060,2	2060,3	1009,0	946,0	33%	38%	1,12
Erfi seul (1000 couples de même sexe)	2060,2	2060,2	1517,3	1517,3			
A_B							
Erfi + Complémentaire FamEmp avec tri (2000 couples de même sexe)	2060,3	2060,3	699,9	633,2	52%	57%	1,09
Erfi + Complémentaire FamEmp sans tri (2000 couples de même sexe)	2060,2	2060,1	689,6	603,6	53%	59%	1,11
Erfi seul (1000 couples de même sexe)	2060,4	2060,4	1472,0	1472,0			
A_X							
Erfi + Complémentaire FamEmp avec tri (2000 couples de même sexe)	2059,9	2060,3	962,9	415,5	35%	72%	2,06
Erfi + Complémentaire FamEmp sans tri (2000 couples de même sexe)	2059,9	2060,3	1000,5	409,9	32%	72%	2,23
Erfi seul (1000 couples de même sexe)	2060,0	2060,0	1480,4	1480,4			
B_X							
Erfi + Complémentaire FamEmp avec tri (2000 couples de même sexe)	2419,9	2419,8	933,4	884,5	33%	36%	1,11
Erfi + Complémentaire FamEmp sans tri (2000 couples de même sexe)	2419,7	2419,6	958,2	929,9	31%	33%	1,07
Erfi seul (1000 couples de même sexe)	2419,9	2419,9	1386,6	1386,6			
AB_X							
Erfi + Complémentaire FamEmp avec tri (2000 couples de même sexe)	2409,3	2409,4	932,3	881,9	33%	37%	1,11
Erfi + Complémentaire FamEmp sans tri (2000 couples de même sexe)	2409,3	2409,5	952,1	921,2	32%	34%	1,07
Erfi seul (1000 couples de même sexe)	2409,1	2409,1	1394,1	1394,1			

Note : \bar{X} selon les scénarios vaut 2 060, (indépendance ; A_B ; A_X), 2 420 (B_X) ou 2 409 (AB_X)

Bibliographie

BOUCHET-VALAT, Milan, BONDON, Marine, BREDAPOPA, Ruxandra, CHARRANCE, Géraldine, COCHET, Paul, MARKOU, Efi et TOULEMON, Laurent. Recourir aux incitations financières dans une grande enquête par questionnaire : les enseignements du test Erfi 2.0. *Rétribuer, inciter, dédommager : Formes et implications de l'argent dans les enquêtes en sciences sociales*. Aubervilliers. 2022.

CHARRANCE, Géraldine et MERLY-ALPA, Thomas. Les coordonnées de contact dans Fidéli-Nautile : Quels enseignements en tirer pour la conduite d'enquête, *12^e colloque international francophone sur les sondages, 2023*, Aubervilliers.

FAVRET-MARTINOZ, Cyril et GROS, Emmanuel. La méthode du partage des poids. *Département des méthodes statistiques, Insee*. [en ligne]. 2017. Disponible à l'adresse: <https://www.insee.fr/fr/statistiques/fichier/2838097/7-methode-du-partage-des-poids.pdf>

PALIOD, Nicolas. Disjonction d'échantillons d'individus dans les tirages de l'Insee. In : *11^e colloque international Francophone sur les sondages*. [en ligne]. 2021. Bruxelles, Disponible à l'adresse : https://drive.google.com/file/d/1s_XHXzewzQJyMphWXMweKTmj7bL52xLf/view?usp=sharing&usp=embed_facebook