

Application et amélioration de la méthode du « décompte des enfants au foyer » sur le recensement français avec le logiciel SAS

John Tomkinson*

I. Introduction

L'objectif de ce document de travail est de fournir une syntaxe commentée d'une amélioration de la méthode du « décompte des enfants au foyer » (DEF) qui cherche à mieux identifier les femmes ayant récemment eu un enfant dans les fichiers individus du recensement français. Cette amélioration cherche à pallier les « défauts » de la méthode dans sa forme « traditionnelle » qui en résultent des spécificités du recensement français, à la fois dans son questionnaire et dans son recodage.

La méthode a été développée sur le recensement 2011 et validée par une confrontation de l'Enquête annuelle de recensement à l'Enquête Famille et Logements de 2011. La méthode et les résultats détaillés sont documentés dans l'article suivant à laquelle ce document fait complément :

TOMKINSON, J. et BRETON, D., 2017, « Comment mieux identifier les mères adolescentes dans le recensement français ? Améliorations de la méthode du « décompte des enfants au foyer » », *Cahiers québécois de démographie*, 45(2), p. 269-293.

Le présent document s'inscrit dans une volonté de mise en commun des outils mis au point par les utilisateurs dans le cadre du projet Big_Stat. La syntaxe, dont figurent des extraits dans le présent document, est mise à disposition sur le site des utilisateurs du recensement <https://recensement.site.ined.fr/> ou directement sur la page <https://recensement.site.ined.fr/fr/donnees/variables-construites/> . Toute utilisation, directe ou indirecte, doit citer le présent document.

II. Données et méthode

Le recensement renouvelé

En France, depuis 2004, l'Insee a mis en place un recensement de la population sous une forme renouvelée. À la place d'une collecte exhaustive périodique, une enquête annuelle de recensement (EAR) est menée annuellement. Une EAR porte sur un cinquième des communes de moins de dix mille habitants et sur un échantillon d'adresses de communes de dix mille habitants et plus qui représentent 8 % de leur population (Lefranc, 2010 ; Toulemon, 2017). Pendant une période de cinq ans, l'ensemble des communes de moins de dix mille habitants et 40 % des habitants des communes

* Institut national d'études démographiques (Ined), Université de Strasbourg, Sage (Umr n°7363).

de dix mille habitants et plus sont enquêtés, soit environ 70 % de la population (Clanché, 2014). Ces EAR sont cumulées et leur compilation fournit les résultats du recensement pour l'année située au milieu de la période des cinq EAR¹.

Les fichiers détails du recensement sont mis à disposition et sont téléchargeables directement depuis le site web de l'Insee². Ces fichiers contiennent des microdonnées concernant plus de 20 millions d'individus qui sont représentatifs de la population en France, une ligne dans le fichier correspondant à un individu recensé.

Si l'état civil est la source incontournable pour le dénombrement de naissances, elle est insatisfaisante pour mener des études plus détaillées du profil et de l'environnement des mères compte tenu du peu de variables figurant dans le bulletin de déclaration des naissances. Les informations collectées lors de l'enregistrement d'une naissance en France et en Angleterre-Galles sont peu nombreuses, surtout en ce qui concerne les variables sociodémographiques. En France, aucune information n'est collectée sur le niveau d'éducation de la mère, sa situation de couple ou encore son mode de vie et si l'information sur la profession de la mère est demandée, celle-ci n'est pas toujours bien renseignée (Davie et Mazuy, 2010). Également, d'autres éléments du bulletin de naissance faisant défaut, tel que le rang de naissance, rendent des analyses impossibles à partir de cette source. Il est donc nécessaire à recourir à une autre source, plus riche en variables sociodémographiques. Le recensement de la population offre une alternative en considérant les enfants « d'aujourd'hui » comme les naissances « d'hier », à l'instar de la méthode du « décompte des enfants au foyer » (DEF) ou « *Own Children method* ». Le recensement a deux grands avantages par rapport à d'autres sources potentielles (par exemple, les enquêtes nationales périnatales) : (i) la richesse des informations collectées et (ii) l'exhaustivité de la population couverte, ce qui mène à une taille d'effectifs qui facilite l'analyse de sous-populations rares ou à l'échelle locale.

Le recensement français, comme ceux de nombreux autres pays, ne recueillait pas directement, jusqu'à l'EAR de 2018, le lien de parentalité entre une mère et son enfant³. Un individu du ménage doit lister chaque habitant permanent du logement et leur lien familial avec lui-même sous la forme d'une question ouverte (Figure 1). A posteriori, les liens renseignés sont recodés afin de désigner une personne de référence de la famille et du ménage par les variables « Lien à la personne de référence de la famille » (LPRF) et « Lien à la personne de référence du Ménage » (LPRM). Les limites du format de cette question sont apparentes :

¹ Par exemple, à partir des EAR de 2009 à 2013, sont publiés les résultats du recensement pour le millésime 2011.

² Le fichier détail du recensement 2011 – « Individus localisés au canton-ou-ville » – sur lequel notre présente application se base peut être téléchargé ici : https://www.insee.fr/fr/statistiques/fichier/2011208/rp2011_indcvi_txt.zip

³ Une refonte de la feuille de logement du recensement français a introduit en 2018 un système pour indiquer directement la mère (biologique ou adoptive) d'une personne si elle vit dans le logement.

- (i) nous ne disposons pas de lien direct entre chacune des personnes du ménage, ceux-ci doivent être dérivés ;
- (ii) le champ libre qui recueille les liens de relation doit être recodé et les réponses données risquent de ne pas être homogènes ;
- (iii) le recodage du recensement pose problème si la personne listée en première position est une femme. En effet, suite au recodage, la personne de référence du ménage est toujours un homme (en dehors des familles monoparentales). Les liens de cette personne avec les autres personnes du ménage, qui n'ont pas été saisis, sont recodés à partir d'un algorithme, et ce, seulement sur une partie de la population⁴.

Néanmoins, c'est sur ces deux variables dont nous reposons pour décrire les familles et les ménages au recensement et sur lesquelles se base la méthode DEF.

Figure 1 : Extrait de la feuille de logement, recensement de la population 2011 – « Liste A »

Liste A Habitants permanents du logement → Remplissez un bulletin individuel pour chacun

Inscrivez en liste A :

- les personnes qui vivent dans ce logement la plus grande partie de l'année, y compris :
 - les personnes temporairement absentes (vacances, voyage d'affaires, hospitalisation de moins d'un mois, etc.);
 - les nourrissons, même s'ils sont encore à la maternité;
 - les sous-locataires et colataires occupant une partie du logement.

Inscrivez également en liste A :

- les enfants mineurs logés ailleurs pour leurs études et dont ce logement est la résidence familiale.
- les conjoints qui ont un autre domicile pour des raisons professionnelles et qui reviennent vivre dans ce logement pour les week-ends, les vacances, etc.
- les personnes majeures qui habitent dans ce logement pour leurs études.
- les personnes qui sont présentes dans ce logement et qui n'ont pas de résidence habituelle ailleurs.
- les employés de maison, salariés et jeunes filles au pair qui habitent dans ce logement.

N'inscrivez pas les personnes à lister en liste B ou en liste C.

	Nom <small>(exemple : DUPAS, épouse MAURIN)</small>	Prénom	Lien de parenté ou relation avec la personne inscrite sur la première ligne <small>(exemples : époux, épouse, union libre, fils, fille, mère, père, sous-locataire, etc.)</small>
1			
2			
3			
4			
5			
6			
7			
8			

Si y a plus de 8 personnes, indiquez le nombre de personnes supplémentaires → et remplissez un bulletin individuel pour chacune.

Source : Insee RP 2011, feuille de logement.

La méthode DEF

La méthode DEF a été développée par Lee-Jay Cho et Wilson H. Grabill au début des années 1960 (cf. Grabill et Cho, 1965) afin de produire des mesures de fécondité différentielle à partir du recensement américain (Cho, 1986). La méthode a ensuite été affinée et a rapidement connu un réel succès. Si, à ses débuts, cette méthode a principalement été utilisée pour mesurer la fécondité dans les pays ne disposant pas d'un système fiable d'enregistrement des naissances (Cho et Man Jun Hahm, 1968 ; Retherford et Mirza, 1982 ; Ratnayake et al., 1984), ses applications se sont élargies et

⁴ Le recodage des variables « Lien avec la personne de référence de la famille » (LPRF) et « Lien avec la personne de référence du ménage » (LPRM) est effectué par une sous-traitance de l'Insee sur une partie seulement des bulletins (ceux des communes de plus de 10 000 habitants et ¼ pour les autres communes) en suivant (en grande partie) un algorithme de traitement automatique des informations saisies dans les bulletins individuels ainsi que dans la « Liste A » de la feuille de logement. Pour plus de détails sur le recodage de la variable LPRM, voir Trabut *et al.* (2015).

appliquées à des mesures différentielles de fécondité selon l'appartenance ethnique (Dugbaza, 1994 ; Dubuc, 2009), la nationalité (Desplanques, 1993), le pays de naissance (Abassi-Shavazi, 1997 ; Belanger et Gilbert, 2003), le niveau d'étude (Cicali et De Santis, 2002 ; Davie et Mazuy, 2010), le secteur d'activité des femmes (Davie et Niel, 2012), mais aussi à des données historiques (Hacker, 2003 ; Drixler, 2015). Cette méthode indirecte a pourtant été longtemps sous-utilisée en France, mais depuis la mise en accès libre des données individuelles du recensement de la population coïncidant avec le changement de méthodologie du recensement français, elle a fait l'objet d'un certain engouement (Desplanques, 2008 ; Breton, 2010 ; Davie et Mazuy, 2010 ; Kersuzan, 2010 ; Rallu, 2010 ; Temporal, 2010).

La méthode DEF s'appuie sur le principe d'une reconstitution des liens de maternité entre des enfants et des femmes recensés au sein d'un même logement. Il s'agit de calculer les taux de fécondité d'une année passée « t-a »⁵ (où t est celle du recensement), à partir du seul recensement. Le numérateur des taux de fécondité, classiquement composé des naissances par âge de la mère, est remplacé par le nombre d'enfants vivants et nés dans l'année « t-a » classé en fonction de l'âge de leur mère lors de cette même année. Classiquement la méthode DEF sous-estime les taux de fécondité par âge d'une année « t-a » du fait de la sous-estimation du nombre de naissances estimé par le nombre d'enfants d'âge « a » vivant avec leur mère. La sous-estimation – ou sous-identification des mères – est particulièrement forte pour les jeunes âges et notamment pour les mères adolescentes.

Dans notre cas, l'intérêt est moins la mesure de la fécondité durant l'adolescence – l'enregistrement des naissances dans les deux pays est de bonne qualité – que d'identifier les femmes ayant eu un enfant afin de disposer des informations sociodémographiques sur ces mères pour connaître les profils sociodémographiques des mères au moment où elles ont donné naissance à leur enfant. Pour ce faire, nous devons nous approcher le plus possible du moment de la maternité pour limiter des changements de situations (mode de vie, situation d'emploi, ...) entre ce moment et celui du recensement. Nous nous intéresserons donc aux femmes ayant eu un enfant dans l'année « t-1 ». Afin de construire leurs profils, il nous faudra donc d'abord les identifier dans les recensements de la population à partir des principes de la méthode DEF⁶. Nous allons montrer que la méthode DEF, telle qu'elle est, nécessite d'être réajustée afin de nous permettre d'identifier au mieux les femmes ayant donné naissance, particulièrement à des âges jeunes.

⁵ Habituellement, « a » prend une valeur entre 1 et 15 selon l'application (Cho, 1986).

⁶ Nous appelons abusivement ici le principe d'appariement des enfants à leur mère « la méthode DEF » alors que cette méthode vise à calculer des taux de fécondité par âge.

III. Application

La méthode DEF repose sur deux variables clés : « Lien à la personne de référence de la famille » (LPRF) et « Lien à la personne de référence du ménage » (LPRM), dont les modalités sont présentées dans le Figure 2. Il suit que la méthode repose sur les concepts de famille et ménage tels qu'ils soient définis dans le recensement (cf. Figure 3).

Figure 2 : Modalités des variables « Lien à la personne de référence de la famille » (LPRF) et « Lien à la personne de référence du ménage » (LPRM)

Lien à la personne de référence de la Famille (LPRF)		Lien à la personne de référence du ménage (LPRM)	
0	Hors famille	1	Personne de référence du ménage
1	Personne de référence de la famille	2	Conjoint de la personne de référence du ménage
2	Conjoint de la personne de référence de la famille	3	Enfant de la personne de référence du ménage ou de son conjoint
3	Enfant de la famille	4	Petit-enfant
Z	Hors logement ordinaire	5	Ascendant
		6	Autre parent
		7	Ami
		8	Pensionnaire ou sous-locataire
		9	Domestique ou salarié logé
		Z	Hors logement ordinaire

Source : Insee RP 2011, dictionnaire des variables.

Figure 3 : Insee définitions de « Famille » et « Ménage »

Famille	Ménage
<p>« Une famille est la partie d'un ménage comprenant au moins deux personnes et constituée :</p> <ul style="list-style-type: none"> • soit d'un couple vivant au sein du ménage, avec le cas échéant son ou ses enfant(s) appartenant au même ménage ; • soit d'un adulte avec son ou ses enfant(s) appartenant au même ménage (famille monoparentale). <p>Pour qu'une personne soit enfant d'une famille, elle doit être célibataire et ne pas avoir de conjoint ou d'enfant faisant partie du même ménage.</p> <p>Un ménage peut comprendre zéro, une ou plusieurs familles. »</p>	<p>« Un ménage, au sens du recensement de la population, désigne l'ensemble des personnes qui partagent la même résidence principale, sans que ces personnes soient nécessairement unies par des liens de parenté. Un ménage peut être constitué d'une seule personne. [...] »</p>

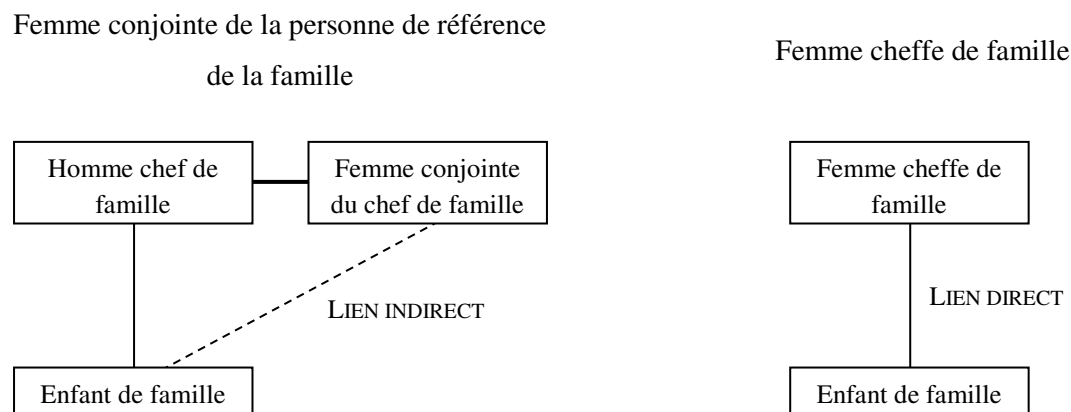
Source : <https://www.insee.fr/fr/metadonnees/definitions>.

Une fois le fichier téléchargé, il peut être importé via la syntaxe incluse en annexe (Extrait A. 1) après avoir spécifié sa location et le répertoire où enregistrer les bases produites par le programme SAS. La syntaxe présentée correspond au champ d'enfants nés dans l'année précédant le recensement et résidant en France métropolitaine. L'utilisateur a le libre choix à la modifier selon ses besoins. Au préalable de l'application de la méthode DEF nous créons des identifiants pour chaque individu, famille et ménage (cf. Extrait A. 2 en annexe).

1. Application « traditionnelle » de la méthode DEF

Conformément à la méthode DEF, nous apparions une mère à ses enfants en utilisant la variable LPRF. La mère de l'enfant peut être directement identifiée si elle est la cheffe de famille, ou indirectement si elle est la conjointe du chef de famille (pas de différenciation possible entre les familles recomposées et les familles « nucléaires ») (Figure 4). Pour chaque mère identifiée, nous déterminons son âge au moment de la naissance de l'enfant (variable AGE_MERE_NAIS) en soustrayant l'âge de l'enfant (dans cette application égal toujours à 1 an) de son âge. La syntaxe de ce processus est donnée en Extrait 1.

Figure 4 : Schéma d'appariement d'une mère à ses enfants



Source : <https://www.insee.fr/fr/metadonnees/definitions>

Extrait 1 : Appariement d'une mère à ses enfants et calcul de son âge à la naissance

```

/***** Appariement d'une mère à ses enfants - méthode "traditionnelle" *****/
/* Etablir une base de femmes âgées de 16 ans et plus en différence de millésimes
et vivant en ménage ordinaire */
data femmes_rp ;
set rp ;
where catpc="0" and sexe="2" and 16 le aged_num ;

/* Etablir une base d'enfants âgés de 1 an en différence de millésimes */
data enfs_1_rp (keep = idfam idmen idenf) ;
set rp ;
where catpc="0" and aged_num=1 and inai notin ("4", "5", "6") and lprf="3" ; /* NB.

```

```

champ = enfants nés en france métropolitaine */
rename individu_id=idenf ;

/* Etablir une base de mères potentielles */
data meres_rp ;
set femmes_rp ;
where lprf in ("1","2") ;

/* Dans le cas d'un couple lesbienne, nous faisons l'hypothèse que la mère soit la
plus âgée des deux femmes */
proc sort data= meres_rp ; by idfam descending aged_num ;
proc sort nodupkeys data=meres_rp ; by idfam ;
proc sort data=enfs_1_rp ; by idfam ;

data meres_enfs_1_rp enfs_non_attaches (keep = idfam idmen idenf) ;
merge meres_rp (in=a) enfs_1_rp (in=b) ;
if a=b then output meres_enfs_1_rp ;
else if a ne b and b=1 then output enfs_non_attaches ;
by idfam ;

/* Calculer l'âge à la naissance pour les mères reliées à leur(s) enfant(s) */
data meres_enfs_1_rp ;
set meres_enfs_1_rp ;
age_mere_nais = aged_num - 1 ;
run ;

```

Les résultats de cet appariement initial soulèvent trois problèmes : (i) des enfants non recensés ; (ii) des enfants non reliés à leur mère ; et (iii) une distribution de naissances par âge biaisée. Il y a un déficit de 40 mille enfants recensés par rapport aux naissances enregistrées à l'état civil dans l'année t-1⁷, soit 5,0 % des naissances. Le sous-dénombrement des plus jeunes enfants est régulièrement observé aux recensements français (Coeffic, 1993 ; Desplanques, 2008 ; Toulemon, 2017) ainsi que dans d'autres pays (O'Hare, 2009). Ce phénomène reste relativement difficile à expliquer. En France, les hypothèses portent sur une sous-déclaration de plus jeunes enfants due aux biais de collecte (manque d'espace dans les bulletins du recensement ; imputations en cas de non-réponse) ainsi que d'une surdéclaration à l'état civil des enfants nés de mères résidant en France (Toulemon, 2017). De toute manière, cette absence d'enfant n'est ni liée à la méthode DEF ni possible à corriger. Également, il y a presque cinq mille enfants vivant « hors logement ordinaire » (ex. dans des établissements sociaux et sanitaires) pour qui nous ne disposons d'aucune information sur leurs liens familiaux et donc ne pouvons pas appairer à leurs mères.

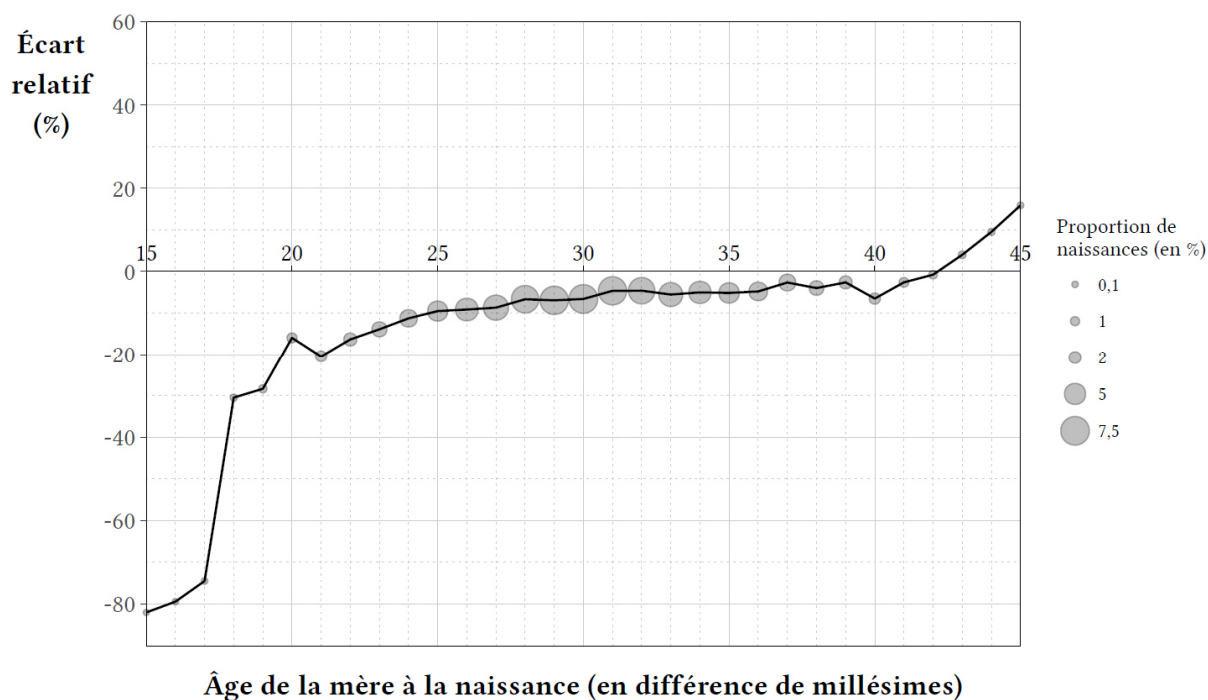
Pour les 13 mille enfants de 1 an soit non associés à une mère soit « hors famille », notre amélioration tente de leur appairer à une mère en proposant une méthode alternative (cf. étapes C et D décrites dans la section suivante).

Enfin, le Figure 5 montre la distribution de naissances par âge de la mère qui résulte de l'application « traditionnelle ». Nous constatons une forte sous-identification de mères âgées de moins

⁷ Le fichier détail du recensement cumulant les EAR de 2009 à 2013, nous faisons l'hypothèse que les enfants recensés à 1 an sont représentatifs des naissances enregistrées sur la période 2008-2012.

de 25 ans, et surtout moins de 20 ans, ce qui est contrebalancé par une surestimation aux âges élevés (à partir de 42 ans). C'est ce biais que nous corrigeons dans les étapes A et B de notre méthode améliorée.

Figure 5 : Écart relatif entre le nombre de naissances selon l'âge de la mère renvoyé par la méthode DEF contre le nombre enregistré à l'état civil



Note : La taille des cercles représente la proportion de naissances ayant eu lieu à l'âge x (selon l'état civil).

Source : Insee RP 2011 (données pondérées), état civil 2008-2012 – Naissances : tabulation sur mesure, Insee (producteur), ADISP-CMH (diffuseur).

2. Méthode améliorée

Nous avons observé que l'application « traditionnelle » résulte dans une distribution de naissances par âge déformé au détriment de jeunes mères. Après une observation des structures de ménages, il s'avère que les jeunes mères – vivant souvent sous le toit de leurs parents – ne sont rarement codées en tant que cheffe de famille ou conjoint du chef de famille, soit en raison des erreurs de déclaration ou de recodage, le critère d'inclusion comme mère potentielle dans l'application « traditionnelle ».

Étape A : Le cas des mères les plus âgées avec une autre mère potentielle dans la famille (0,5 % des enfants recensés)

Nous faisons l'hypothèse que la sous-estimation de naissances aux jeunes âges et la surestimation aux âges élevés résultent en partie de fausses attributions d'enfants à leur grand-mère dans les ménages « complexes » - des ménages multigénérationnels où la mère de l'enfant vit avec sa propre

mère. Nous cherchons pour les enfants originalement liés à une mère âgée de 42 ans ou plus à la naissance, une autre mère alternative au sein du ménage. Nous considérons que les mères ayant moins de 42 ans à la naissance ont été correctement appariées à leurs enfants. L'étape A résume les cas où il y a une autre mère potentielle dans le ménage (femme âgée de 15 ans ou plus au moment de la naissance de l'enfant). L'Extrait 2 détaille l'identification des familles dans lequel une mère alternative existe.

Extrait 2 : Identification des familles avec une autre mère potentielle

```

data meres_ge_42_rp meres_enfs_pot0_rp ;
set meres_enfs_1_rp ;
if age_mere_nais ge 42 then output meres_ge_42_rp ;
else output meres_enfs_pot0_rp ;
run ;

/* Isoler les familles ou il y a une mère de 42 ans ou plus */;
data list_idfam_meres_ge_42_rp (keep = idfam) ;
set meres_ge_42_rp ;
run ;

proc sort data=list_idfam_meres_ge_42_rp ; by idfam ; run ;
proc sort data=rp out=rp_sortidfam ; by idfam ; run ;

data familles_meres_42_plus_rp ;
merge list_idfam_meres_ge_42_rp (in=a) rp_sortidfam (in=b) ;
if a=1 ;
by idfam ;
run ;

/* Ne garder que les mères potentielles dans ces familles */
data femmes_fam_42_plus_rp ;
set familles_meres_42_plus_rp ;
where catpc="0" and sexe="2" and 16 le aged_num ;
run ;

/* Classer les femmes au sein d'une même famille par âge : de la plus âgée à la
plus jeune */
proc sort data=femmes_fam_42_plus_rp ; by idfam descending aged_num ; run ;

data femmes_fam_42_plus_rp (rename= (count=rang_femme)) ;
set femmes_fam_42_plus_rp ;
count + 1 ;
by idfam ;
if first.idfam then count = 1 ;
run ;

/* Identifier les familles ou il y a plus d'une mère possible */;
data multi_meres_potentielles_rp (keep = idfam) ;
set femmes_fam_42_plus_rp ;
where rang_femme = 2 ;
run ;

proc sort data=multi_meres_potentielles_rp nodupkeys ; by idfam ; run ;
proc sort data=meres_ge_42_rp; by idfam; run ;

data meres_ge_42_pota_rp meres_ge_42_potb_rp ;
merge multi_meres_potentielles_rp (in=a) meres_ge_42_rp (in=b);
if a=b then output meres_ge_42_pota_rp ;
else output meres_ge_42_potb_rp ;
by idfam;
run ;

```

Une fois ces familles identifiées, nous examinons la fratrie de l'enfant et les définir comme les fratries « continues » ou « discontinues ». Cette étape consiste à trouver l'enfant le plus jeune de la famille en dehors des enfants âgés de 1 an. Si cet enfant est âgé de 16 ans ou plus, nous définissons la fratrie comme « discontinue », s'il a moins de 16 ans, la fratrie est dite « continue » (Extrait 3).

Extrait 3 : Définir le type de fratrie

```
/***** Décrire la fratrie dans les familles avec une autre mère potentielle *****/
proc sort data=meres_ge_42_pota_rp ; by idfam;

data fams_meres_ge_42_pota_rp ;
merge meres_ge_42_pota_rp (in=a) rp_sortidfam (in=b);
if a=1;
by idfam;

/* Trouver le plus jeune enfant de la famille ayant plus de 1 an */;
data enfs_meres_ge_42_pota_rp ;
set fams_meres_ge_42_pota_rp ;
where lprf="3" and aged_num ne 1 ;

proc sort data=enfs_meres_ge_42_pota_rp ; by idfam aged_num;
proc sort data=enfs_meres_ge_42_pota_rp nodupkeys out=enf_plus_jeune2 ; by idfam ;

/* On détermine une "fratrie discontinue" si ce deuxième plus jeune a 15 ans ou
plus en plus du plus jeune enfant (i.e. a 16 ans ou plus) */
data enf_frat_discon ;
set enf_plus_jeune2 ;
where aged_num ge 16;

/* Créer les base des familles de fratrie "continue" et "discontinue" */
data idfam_meres_frat_discon (keep = idfam);
set enf_frat_discon ;

data meres_pota12_rp meres_pota34_rp ;
merge idfam_meres_frat_discon (in=a) meres_ge_42_pota_rp (in=b);
if a=b then output meres_pota12_rp ; /* <-- fratries "discontinues" */
      else output meres_pota34_rp ; /* <-- fratries "continues" */
by idfam;
run ;
```

Dans les fratries « discontinues », nous réattribuons l'enfant à une autre mère potentielle si, et seulement si, la mère attribuée à l'origine n'est pas en couple. Dans le cas échéant, nous réattribuons l'enfant à une mère alternative et s'il existe plus d'une autre mère potentielle au sein de la famille, nous donnons la priorité (i) à celle(s) ayant déclarée(s) vivre en couple ; et (ii) en cas de choix entre plusieurs mères alternatives ayant le même statut de couple, à celle qui est la plus âgée (Extrait 4).

Extrait 4 : Réattribution des mères dans les cas de fratries « discontinues »

```
/***** Réattribuer les mères dans les cas de fratries "discontinues" *****/
```

```

/** Les mères originalement attribuées ne changent pas si elles se déclarent vivre
en couple */
data meres_enfs_pota1_rp ;
set meres_pota12_rp ;
where couple = "1" ;

/** Si la mère originalement attribuée ne se déclare pas de vivre en couple, on
réattribue l'enfant à la plus âgée de ses filles */
data meres_pota2_rp (keep = idfam) ;
set meres_pota12_rp ;
where couple = "2" ;

proc sort data=meres_pota2_rp ; by idfam;
proc sort data=femmes_rp ; by idfam;

data meres_oss_pota2_rp ;
merge meres_pota2_rp (in=a) rp_sortidfam (in=b where=(catpc="0" and sexe="2" and 16
le aged_num));
if a=1 ;
by idfam ;

/* Classer les femmes au sein d'une famille (i) d'abord par son statut de couple,
(ii) puis par âge du plus âgée au plus jeune */
proc sort data=meres_oss_pota2_rp ; by idfam lprf couple descending aged_num ;

data meres_oss_pota2_rp ;
set meres_oss_pota2_rp ;
rang_femme + 1 ;
by idfam ;
if first.idfam then rang_femme = 1;

/* Attribuer l'enfant à la plus âgée de ses filles */
data meres_alt_pota2_rp ;
set meres_oss_pota2_rp ;
where rang_femme = 2 ;

proc sort data=meres_alt_pota2_rp ; by idfam;
proc sort data=enfs_1_rp ; by idfam;

data meres_enfs_pota2_rp ;
merge meres_alt_pota2_rp (in=a) enfs_1_rp (in=b) ;
by idfam ;
if a=b ;
age_mere_nais = aged_num - 1 ;
run ;

```

Dans les fratries « continues », si la mère d'origine est âgée de 50 ans ou plus au moment de la naissance nous réattribuons l'enfant à une autre mère potentielle dans la famille, en raisonnant que la mère initialement attribuée est la grand-mère de l'enfant et il est donc nécessaire à le réattribuer à une de ses filles (Extrait 5). Nous employons les mêmes conditions de priorité que celles décrites au-dessus s'il existe plusieurs mères alternatives potentielles.

Extrait 5 : Réattribution des mères dans les cas de fratries « continues »

```

***** Réattribuer les mères dans les cas de fratries "continues" *****/

/** Les mères originalement attribuées ne changent pas si elles ont moins de 50
ans */
data meres_enfs_pota3_rp ;

```

```

set meres_pota34_rp ;
where age_mere_nais lt 50 ;

/** Si la mère originalement attribuée a 50 ans ou plus, on réattribue l'enfant à
la plus âgée de ses filles ***/
data fams_meres_enfs_pota4_rp (keep = idfam) ;
set meres_pota34_rp ;
where age_mere_nais ge 50 ;

proc sort data=fams_meres_enfs_pota4_rp ; by idfam;

data meres_alt_pota4_rp ;
merge fams_meres_enfs_pota4_rp (in=a) rp_sortidfam (in=b where=(lprf="3" and
catpc="0" and sexe="2" and 16 le aged_num)) ;
by idfam ;
if a=1 ;

/* Classer les femmes au sein d'une famille (i) d'abord par son statut de couple,
(ii) puis par âge du plus âgée au plus jeune */
proc sort data=meres_alt_pota4_rp ; by idfam couple descending aged_num ;
proc sort data=meres_alt_pota4_rp nodupkeys ; by idfam ;

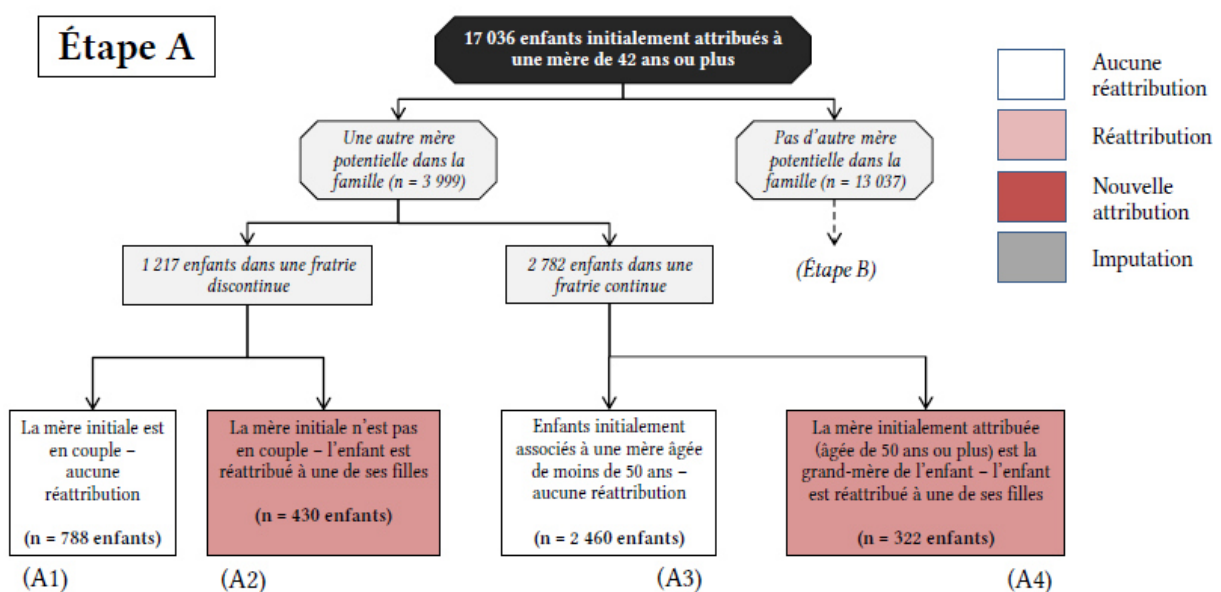
proc sort data=enfs_1_rp ; by idfam;

data meres_enfs_pota4_rp ;
merge meres_alt_pota4_rp (in=a) enfs_1_rp (in=b) ;
by idfam ;
if a=b ;
age_mere_nais = aged_num - 1 ;
run ;

```

Le processus de l'étape A est résumé dans le Figure 6. Cette étape réattribue 18,8 % (n = 752 / 3 999) des enfants initialement attribués à une mère de 42 ans ou plus à la naissance et où il existe une autre mère potentielle dans la famille.

Figure 6 : Méthode de réattribution des enfants initialement attribués à une mère de 42 ans ou plus, à une mère alternative potentielle



Source : Tomkinson et Breton, 2016.

Étape B : Les cas des ménages composés d'une mère âgée, sans autre mère potentielle (1,7 % des enfants recensés)

Pour les enfants attribués à une mère âgée de 42 ans ou plus à la naissance, mais pour qui il n'existe aucune autre mère potentielle au sein de leur famille, la grande majorité a été associée à une mère de moins de 50 ans, et nous jugeons que ces enfants l'ont été avec raison à leur mère et qu'aucune correction ne peut être envisagée. Ces mères restent donc inchangées (Extrait 6).

Extrait 6 : Les mères originalement attribuées ne changent pas si elles ont moins de 50 ans

```
/***** Réattribuer les enfants dans les familles sans autre mère potentielle
apparente *****/

/** Les mères originalement attribuées ne changent pas si elles ont moins de 50
ans ***/
data meres_enfs_potb1_rp meres_potb2345_rp (keep = idfam individu_id) ;
set meres_ge_42_potb_rp ;
if age_mere_nais lt 50 then output meres_enfs_potb1_rp ; else output
meres_potb2345_rp ;
run ;
```

Pour les enfants attribués à une mère âgée de 50 ans ou plus, nous observons l'écart d'âge entre ces mères et leurs conjoints. Si le chef de famille est significativement plus jeune que sa conjointe – de 18 ans ou plus en moins –, nous posons l'hypothèse d'un mauvais recodage de la variable LPRF : que l'homme « chef de famille » est en fait le fils de la femme attribuée comme mère de l'enfant et que, par conséquent, l'enfant n'est que le petit-enfant de cette dernière et enfant du premier. Dans ces cas, nous imputons l'âge à la naissance de la mère de l'enfant en fonction du celui du père, en appliquant la différence d'âge médiane entre les hommes et les femmes en couple cohabitant avec au moins un enfant (Extrait 7). Cette imputation ne nous permet donc pas d'avoir les caractéristiques individuelles de la mère de l'enfant.

Extrait 7 : Imputation de l'âge de la mère de l'enfant dans le cas d'un conjoint âgé de 18 ans ou plus en moins de la mère originalement attribuée

```
/** Dans les familles où la mère originalement attribuée a 50 ans ou plus,
calculer la différence d'âge entre les chefs de famille et leurs conjointes***/

proc sort data=meres_potb2345_rp ; by idfam ;
proc sort data=familles_meres_42_plus_rp ; by idfam ;

data fam_meres_potb2345_rp ;
merge meres_potb2345_rp (in=a) familles_meres_42_plus_rp (in=b) ;
by idfam ;
if a=1 ;

/* Calculer la différence d'âge entre les chefs et leurs conjointes */

data chefs_h (rename=(aged_num=age_chef))
conjointes (rename=(aged_num=age_conjointe)) ;
```

```

set fam_meres_potb2345_rp (keep = idfam lprf sexe aged_num ipondi2) ;
if lprf="1" and sexe="1" then output chefs_h ;
else if lprf="2" and sexe="2" then output conjointes ;

proc sort data=chefs_h ; by idfam ;
proc sort data=conjointes ; by idfam ;

data age_diff (keep = idfam age_diff age_chef age_conjointe ipondi2) ;
merge chefs_h (in=a) conjointes (in=b) ;
by idfam ;
if a=b;
age_diff = age_chef - age_conjointe ;

/**/ Si le chef de famille a 18 ans ou plus en moins de sa conjointe on impute
l'âge de la mère de l'enfant en fonction de l'âge de celui-ci */

proc sort data=fam_meres_potb2345_rp ; by idfam ;
proc sort data=age_diff ; by idfam ;

data agediff_meres_potb2345_rp ;
merge fam_meres_potb2345_rp (in=a) age_diff (in=b) ;
by idfam ;
if a=1 ;

proc sort data=agediff_meres_potb2345_rp nodupkeys ; by idfam ;

data imputer_potb2 (keep = idfam ipondi2 age_h) men_potb345_rp (keep = idfam idmen)
;
set agediff_meres_potb2345_rp ;
rename age_chef=age_h ;
if age_diff ne . and age_diff le -18 then output imputer_potb2 ;
else output men_potb345_rp ;

proc sort data=imputer_potb2 ; by idfam ;
proc sort data=enfs_1_rp ; by idfam;

data enfs_imputer_potb2 ;
merge imputer_potb2 (in=a) enfs_1_rp (in=b) ;
by idfam ;
if a=b ;

/* Ecart d'âge médian entre homme et femme dans les couples cohabitant avec un
enfant */

data moco22_h (rename=(aged_num=age_h)) moco22_f (rename=(aged_num=age_f)) ;
set rp (keep = idfam aged_num ipondi2 moco sexe) ;
if moco="22" and sexe="1" then output moco22_h ;
else if moco="22" and sexe="2" then output moco22_f ;

proc sort data=moco22_h ; by idfam ;
proc sort data=moco22_f ; by idfam ;

data age_diff_moco22 (keep = idfam age_diff_moco22 age_h age_f ipondi2) ;
merge moco22_h (in=a) moco22_f (in=b) ;
by idfam ;
if a=b;
age_diff_moco22 = age_h - age_f ;

proc means data=age_diff_moco22 missing nway median n;
class age_h ;
var age_diff_moco22 ;
output out=median_age_gap_moco22 (drop = _type_ _freq_) median=median n=n;
weight ipondi2 ;
where -30 le age_diff_moco22 le 30 ;

data median_age_gap_moco22 ;
set median_age_gap_moco22 ;

```

```

age_mere_nais = age_h - median - 1 ;

proc sort data=enfs_imputer_potb2 ; by age_h ;
proc sort data=median_age_gap_moco22 ; by age_h ;

data imputation_age_potb2_rp ;
merge enfs_imputer_potb2 (in=a) median_age_gap_moco22 (in=b) ;
by age_h ;
if a=1 ;
run ;

```

Dans les cas où l'écart d'âge entre les membres du couple est cohérent, nous raisonnons que ce couple s'agit des grands-parents de l'enfant en faisant l'hypothèse d'un mauvais recodage des liens familiaux qui attribue l'enfant à sa grand-mère. Nous cherchons donc la présence d'une autre mère potentielle au sein du ménage et réattribuons l'enfant dans le cas affirmative avec les mêmes priorités dans le cas de plusieurs alternatives (Extrait 8).

Extrait 8 : Imputation de l'âge de la mère de l'enfant dans le cas d'un conjoint âgé de 18 ans ou plus en moins de la mère originalement attribuée

```

/**/ Si l'écart d'âge est moins de 18 ans, on raisonne que l'enfant vit chez ses
grands-parents ***/

/* Y a-t-il une autre femme dans le ménage de ces familles ? Si oui, on lui
attribue l'enfant */

proc sort data=men_potb345_rp ; by idmen ;
proc sort data=rp out=rp_sortidmen ; by idmen ;

data fem_men_potb3_rp men_potb45_rp (keep = idfam) ;
merge men_potb345_rp (in=a) rp_sortidmen (in=b where=(sexe="2" and 16 le aged_num
le 50)) ;
by idmen ;
if a=b then output fem_men_potb3_rp ;
else if a=1 and a ne b then output men_potb45_rp ;

/* Classer les femmes au sein d'un ménage (i) d'abord par son statut de couple,
(ii) puis par âge du plus âgée au plus jeune */
proc sort data=fem_men_potb3_rp ; by idmen couple descending aged_num ;
proc sort data=fem_men_potb3_rp nodupkeys out=meres_alt_potb3_rp ; by idmen ;

proc sort data=men_potb345_rp ; by idfam ;
proc sort data=enfs_1_rp ; by idfam ;

data enfs_imputer_potb3 ;
merge men_potb345_rp (in=a) enfs_1_rp (in=b) ;
by idfam ;
if a=b ;

proc sort data=enfs_imputer_potb3 ; by idmen ;
proc sort data=meres_alt_potb3_rp ; by idmen ;

data meres_enfs_potb3_rp ;
merge meres_alt_potb3_rp (in=a) enfs_imputer_potb3 (in=b) ;
by idmen ;
if a=b ;
age_mere_nais = aged_num - 1 ;
run ;

```

S'il n'y a pas de mère alternative au sein du ménage, nous cherchons la présence d'un père alternatif, que nous qualifions comme un homme âgé entre 16 et 52 ans⁸. Si un tel homme n'existe pas, aucune réattribution ne peut être affectée et l'enfant reste attribué à la mère originalement identifiée. Lorsqu'un père potentiel existe⁹, nous imputons, comme précédemment, l'âge de la mère en fonction de l'âge de celui-ci (Extrait 9).

Extrait 9 : Recherche d'un père potentiel au sein du ménage et imputation de l'âge de la mère dans le cas échéant

```
/* S'il n'y a pas d'autre femme dans le ménage, on cherche un fils âgé de 18 ans ou plus */  
  
proc sort data=men_potb45_rp ; by idfam ;  
  
data fam_potb4_rp (keep = idfam) fils_potb5_rp ;  
merge men_potb45_rp (in=a) rp_sortidfam (in=b where=(lprf="3" and sexe="1" and aged_num ge 18)) ;  
by idfam ;  
if a=b then output fils_potb5_rp ;  
else if a=1 and a ne b then output fam_potb4_rp ;  
  
/* S'il n'y a pas de fils âgé de 18 ans ou plus, on ne réattribue pas l'enfant */  
  
proc sort data=fam_potb4_rp ; by idfam ;  
proc sort data=meres_ge_42_potb_rp ; by idfam ;  
  
data meres_enfs_potb4_rp ;  
merge fam_potb4_rp (in=a) meres_ge_42_potb_rp (in=b) ;  
by idfam ;  
if a=b ;  
  
/* S'il existe un fils âgé de 18 ans ou plus, on raisonne qu'il est le père de l'enfant et on impute l'âge de la mère en fonction de son âge */  
  
/* Dans le cas de plusieurs fils on garde celui le plus probable d'être le père */  
proc sort data=fils_potb5_rp ; by idfam couple descending aged_num ;  
proc sort data=fils_potb5_rp nodupkeys ; by idfam ;  
  
proc sort data=enfs_1_rp ; by idfam ;  
  
data fils_enfs_potb5_rp (rename=(aged_num=age_h)) ;  
merge fils_potb5_rp (in=a) enfs_1_rp (in=b) ;  
by idfam ;  
if a=b ;  
  
proc sort data=fils_enfs_potb5_rp ; by age_h ;  
proc sort data=median_age_gap_moco22 nodupkeys ; by age_h ;  
  
data imputation_age_potb5_rp ;  
merge fils_enfs_potb5_rp (in=a) median_age_gap_moco22 (in=b) ;
```

⁸ Ces âges pour un homme correspondent à une conjointe âgée entre 16 et 49 ans selon la distribution de différence des âges au sein des couples cohabitants avec au moins un enfant recensés en 2011.

⁹ De façon identique à la recherche des mères alternatives, s'il existe plusieurs pères potentiels, nous priorisons celui en couple et le plus âgé.

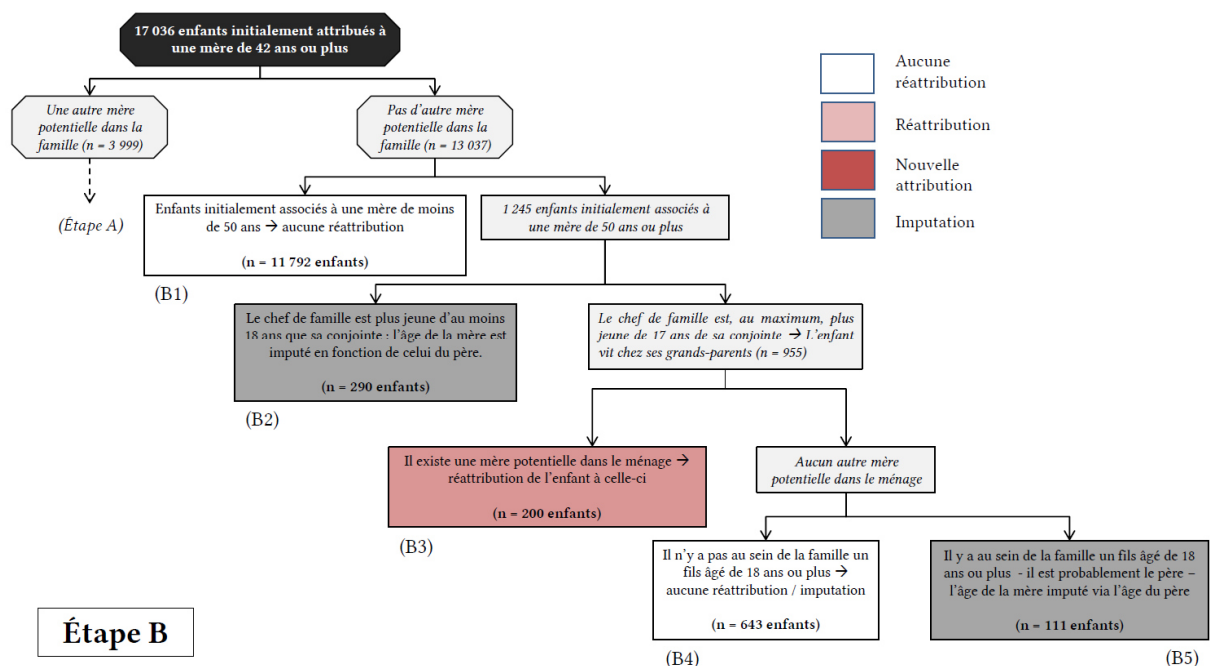

```

by age_h ;
if a=1 ;
run ;

```

Le Figure 7 résume les processus de l'étape B. Finalement, seuls 1,5 % des enfants attribués initialement par la méthode DEF à une mère âgée de 42 ans ou plus, et vivant dans une famille n'ayant aucune mère alternative ont été réattribué à une nouvelle mère lors de cette étape. En outre, dans 3,5 % de ces cas nous avons dû imputer l'âge de la mère à partir de l'âge du père probable, opération qui rend impossible toute analyse différentielle sur la mère, faute de pouvoir la décrire.

Figure 7 : Méthode de réattribution des enfants initialement attribués à une mère de 42 ans ou plus, sans mère alternative potentielle au sein de la famille



Source : Tomkinson et Breton, 2016.

Étape C : Le cas des enfants de familles non attribuables à une femme de la famille (0,75 % des enfants recensés)

Il en résulte de l'application « traditionnelle » de la méthode DEF qu'une minorité d'enfants vivant en famille (0,8 %) ne soient pas reliés à une mère en absence d'une femme codée en tant que cheffe de famille ou conjointe du chef. Plusieurs situations peuvent en être à l'origine : soit l'enfant vit chez son père qui est non cohabitant avec la mère ; soit chez son père qui est séparé de la mère ; soit un mauvais recodage du sexe d'une mère monoparentale. Une autre hypothèse serait celle d'un mauvais codage du lien avec la personne de référence du ménage ou de la famille. Ainsi la mère serait présente dans le ménage, mais non codée comme telle.

Nous cherchons à relier les enfants à une mère dans un premier temps en cherchant une femme dans le ménage âgée de 16 à 50 ans au moment du recensement. Nous désignons celle-ci comme la mère de l'enfant. Lorsqu'il n'y a pas de mère potentielle au sein du ménage, nous imputons l'âge de la mère via l'âge du père de l'enfant (le chef de famille) (Extrait 10).

Extrait 10 : Réattribution ou imputation de l'âge de la mère pour les enfants initialement non attribués à une mère

```

/***** Peut-on trouver une mère pour les enfants non rattachés dans un premier
temps ? *****/

/* S'il y a une autre femme dans le ménage de ces enfants, on lui attribue l'enfant
*/

data meres_potc1_rp ;
set femmes_rp ;
where aged_num le 50 ;

/* Classer les femmes au sein d'un ménage (i) d'abord par son statut de couple,
(ii) puis par âge du plus âgée au plus jeune */
proc sort data=meres_potc1_rp ; by idmen couple descending aged_num ;
proc sort data=meres_potc1_rp nodupkeys ; by idmen ;

proc sort data=enfs_non_attaches ; by idmen ;

data meres_enfs_potc1_rp enfs_potc2x_rp (keep = idfam idmen idenf) ;
merge enfs_non_attaches (in=a) meres_potc1_rp (in=b) ;
by idmen ;
if a=b then output meres_enfs_potc1_rp ;
else if a=1 and a ne b then output enfs_potc2x_rp ;

data meres_enfs_potc1_rp ;
set meres_enfs_potc1_rp ;
age_mere_nais = aged_num - 1 ;

/* Sinon on impute l'âge de la mère en fonction l'âge du père de l'enfant */

proc sort data=enfs_potc2x_rp ; by idfam ;

data peres_enfs_potc2_rp (rename=(aged_num=age_h)) enfs_potcx_rp (keep = idfam
idmen idenf) ;
merge enfs_potc2x_rp (in=a) rp_sortidfam (in=b where=(lprf="1" and sexe="1" and 16
le aged_num le 52)) ;
by idfam ;
if a=b then output peres_enfs_potc2_rp ;
else if a=1 and a ne b then output enfs_potcx_rp ;

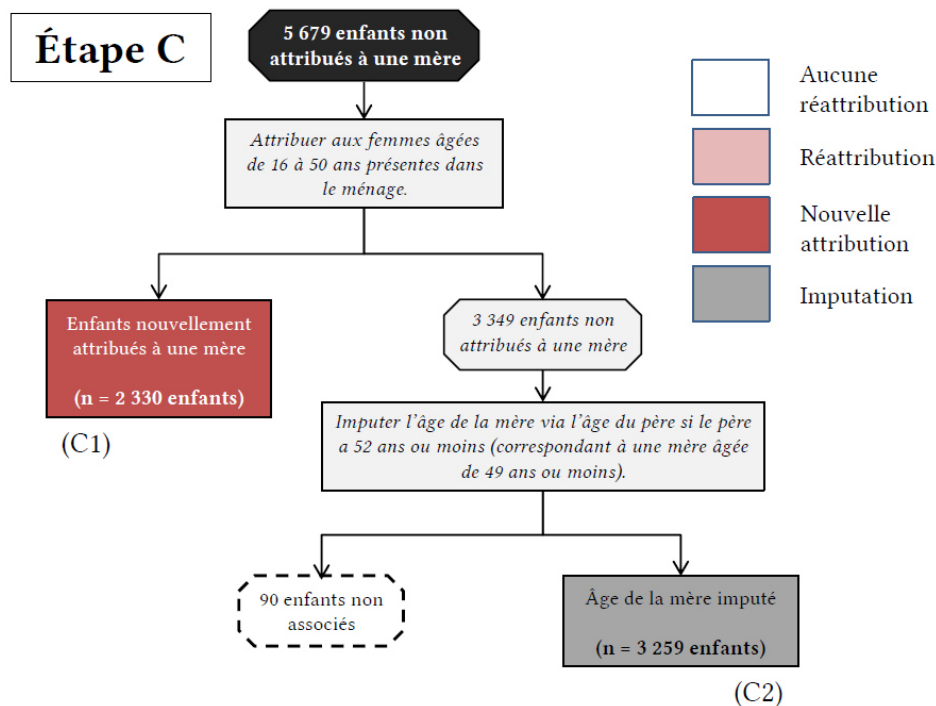
proc sort data=peres_enfs_potc2_rp ; by age_h ;
proc sort data=median_age_gap_moco22 nodupkeys ; by age_h ;

data imputation_age_potc2_rp ;
merge peres_enfs_potc2_rp (in=a) median_age_gap_moco22 (in=b) ;
by age_h ;
if a=1 ;

```

Le Figure 8 présente l'étape C, processus qui permet de « récupérer » des naissances « écartées » par l'application « traditionnelle » de la méthode DEF. Cette récupération est possible dans la majorité des cas des enfants non reliés à leurs mères dans un premier temps.

Figure 8 : Méthode de réattribution des enfants vivant en famille, initialement non attribués à une mère



Source : Tomkinson et Breton, 2016.

Étape D : Le cas des enfants d'un ménage codés « hors famille » (0,99 % des enfants recensés)

Afin d'attribuer les enfants « hors famille » à une mère, nous posons d'abord l'hypothèse que la mère de l'enfant est présente dans le ménage sans l'être codée en tant que telle. Il existe des enfants codés comme enfants de la personne de référence du ménage ou de son conjoint, ce qui implique qu'ils devraient également être codés de la même manière pour ce qui est de leur lien à la personne de référence de la famille. Nous tentons d'associer dans un premier temps les enfants hors famille à une femme cheffe du ménage ou conjointe du chef âgée entre 16 et 50 ans au moment du recensement.

Ensuite, pour les enfants qui ont un lien de grand-parentalité avec le chef du ménage, nous cherchons à les attribuer à une femme, enfant du chef de ménage, et donc la mère probable de l'enfant. Enfin, ayant exploré les structures des ménages des enfants ayant un lien d'« Autre parent » avec le chef du ménage, une mère potentielle est présente dans le ménage dans la grande majorité des cas et donc nous attribuons ces enfants à une femme du ménage âgée entre 16 et 50 ans, toujours en utilisant les mêmes critères de priorité (Extrait 11).

Extrait 11 : Attribution de la mère pour les enfants « hors famille »

```
/* **** Peut-on trouver une mère pour les enfants hors famille ? **** */

/* Etablir une base de ménages avec un enfant âgé de 1 an en différence de
millésimes hors famille */
data enfs_horsfam_1_rp (keep = idfam idmen idenf lprm_enf) ;
set rp ;
where catpc="0" and aged_num=1 and inai notin ("4","5","6") and lprf="0" ; /* NB.
champ = enfants nés en france métropolitaine */
rename individu_id=idenf lprm=lprm_enf ;

/* (i) Attribuer les enfants "Enfant de la personne de référence du ménage ou de
son conjoint" aux femmes "Personne de référence du ménage / Conjoint de la personne
de référence du ménage" */
proc sort data=enfs_horsfam_1_rp ; by idmen ;

data meres_enfs_potd1_rp enfs_potd1x (keep = idenf idfam idmen) ;
merge enfs_horsfam_1_rp (in=a where=(lprm_enf="3")) rp_sortidmen (in=b
where=(catpc="0" and sexe="2" and lprm in ("1","2") and 16 le aged_num le 50)) ;
by idmen ;
if a=b then output meres_enfs_potd1_rp ;
else if a=1 and a ne b then output enfs_potd1x ;

data meres_enfs_potd1_rp ;
set meres_enfs_potd1_rp ;
age_mere_nais = aged_num - 1 ;

/* (ii) Attribuer les enfants "Petits-enfants" aux femmes "Enfant de la personne de
référence du ménage ou de son conjoint" */
data meres_potd2_rp ;
set femmes_rp ;
where lprm="3" and aged_num le 50 ;

proc sort data=meres_potd2_rp ; by idmen couple descending aged_num ;
proc sort data=meres_potd2_rp nodupkeys ; by idmen ;

proc sort data=enfs_horsfam_1_rp ; by idmen ;

data meres_enfs_potd2_rp enfs_potd2x (keep = idenf idfam idmen) ;
merge enfs_horsfam_1_rp (in=a where=(lprm_enf="4")) meres_potd2_rp (in=b) ;
by idmen ;
if a=b then output meres_enfs_potd2_rp ;
else if a=1 and a ne b then output enfs_potd2x ;

data meres_enfs_potd2_rp ;
set meres_enfs_potd2_rp ;
age_mere_nais = aged_num - 1 ;

/* (iii) Attribuer les enfants "Autre parent" à une femme du ménage */
data meres_potd3_rp ;
set femmes_rp ;
where aged_num le 50 ;

proc sort data=meres_potd3_rp ; by idmen couple descending aged_num ;
proc sort data=meres_potd3_rp nodupkeys ; by idmen ;

proc sort data=enfs_horsfam_1_rp ; by idmen ;

data meres_enfs_potd3_rp enfs_potd3x (keep = idenf idfam idmen) ;
merge enfs_horsfam_1_rp (in=a where=(lprm_enf="6")) meres_potd3_rp (in=b) ;
by idmen ;
if a=b then output meres_enfs_potd3_rp ;
else if a=1 and a ne b then output enfs_potd3x ;
```

```
data meres_enfs_potd3_rp ;
set meres_enfs_potd3_rp ;
age_mere_nais = aged_num - 1 ;
```

Pour les enfants hors famille qui ne sont toujours pas liés à leur mère après le processus décrit ci-dessus, nous répétons ces étapes, cette fois-ci en cherchant le père de l'enfant ayant pour objectif d'imputer l'âge de la mère à partir de l'âge du père (Extrait 12).

Extrait 12 : Imputation de l'âge de la mère pour les enfants « hors famille »

```
/* (iv) Pour les enfants non rattachés à une mère du ménage on cherche à imputer
son âge en fonction l'âge du père de l'enfant */

proc sort data=enfs_potd1x ; by idmen ;

data peres_enfs_potd4a_rp (rename=(aged_num=age_h)) ;
merge enfs_potd1x (in=a) rp_sortidmen (in=b where=(sexe="1" and lprm in ("1","2")
and 16 le aged_num le 52)) ;
by idmen ;
if a=b ;

proc sort data=enfs_potd2x ; by idmen ;

data peres_potd4b_rp ;
set rp ;
where sexe="1" and lprm="3" and 16 le aged_num le 52 ;

proc sort data=peres_potd4b_rp ; by idmen couple descending aged_num ;
proc sort data=peres_potd4b_rp nodupkeys ; by idmen ;

data peres_enfs_potd4b_rp (rename=(aged_num=age_h)) ;
merge enfs_potd2x (in=a) peres_potd4b_rp (in=b) ;
by idmen ;
if a=b ;

proc sort data=enfs_potd3x ; by idmen ;

data peres_potd4c_rp ;
set rp ;
where sexe="1" and 16 le aged_num le 52 ;

proc sort data=peres_potd4c_rp ; by idmen couple descending aged_num ;
proc sort data=peres_potd4c_rp nodupkeys ; by idmen ;

data peres_enfs_potd4c_rp (rename=(aged_num=age_h)) ;
merge enfs_potd3x (in=a) peres_potd4c_rp (in=b) ;
by idmen ;
if a=b ;

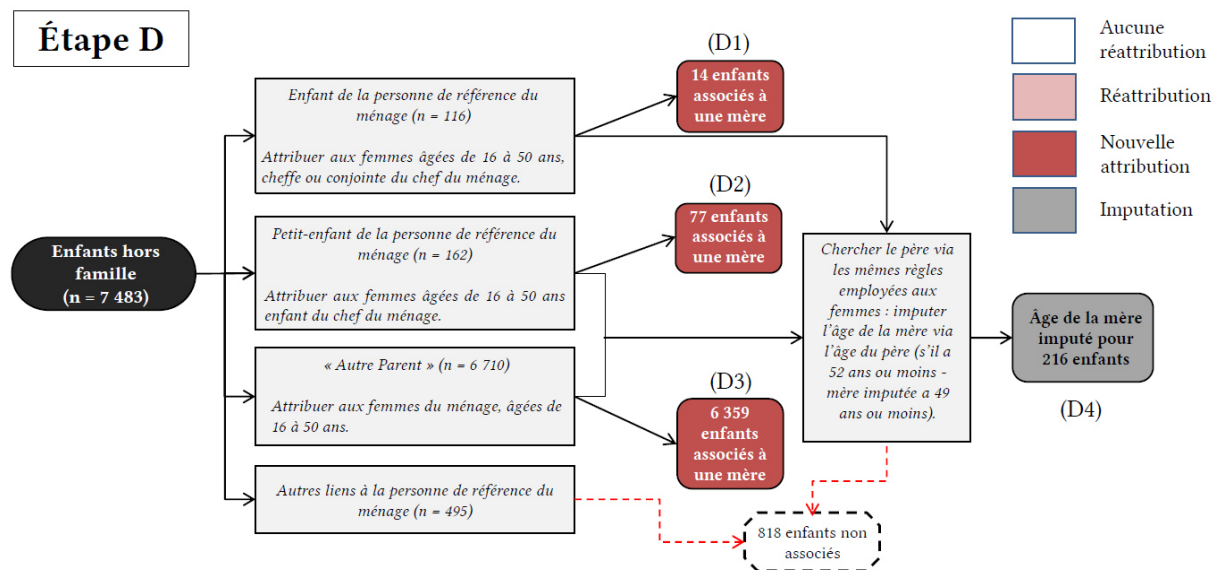
data peres_enfs_potd4_rp ;
set peres_enfs_potd4a_rp peres_enfs_potd4b_rp peres_enfs_potd4c_rp ;

proc sort data=peres_enfs_potd4_rp ; by age_h ;
proc sort data=median_age_gap_moco22 nodupkeys ; by age_h ;

data imputation_age_potd4_rp ;
merge peres_enfs_potd4_rp (in=a) median_age_gap_moco22 (in=b) ;
by age_h ;
if a=1 ;
```

Le Figure 9 présente la logique de l'étape D, ce qui permet à attribuer des enfants hors famille à une mère.

Figure 9 : Méthode d'attribution des enfants hors famille à une mère



IV. Résultats et validation de la méthode améliorée

Les différentes améliorations que nous avons apportées augmentent la proportion d'enfants recensés associés à une mère (la proportion des enfants non associés passe de 1,6 % à 0,1 %). Tomkinson et Breton (2016) montrent la réduction de la forte sous-identification des jeunes mères et discutent les raisons pour lesquelles la malgré les améliorations offertes, quelques mères restent inidentifiables.

Une confrontation de l'Enquête famille et logements¹⁰ (EFL) à notre algorithme d'identification des mères permet d'évaluer sa qualité. Le Figure 10 montre la proportion de mères correctement identifiées par étape. Dans l'ensemble, l'algorithme identifie correctement la mère de l'enfant dans 93 % des cas. Concernant les étapes C1 et A3 dans lesquelles l'identification est d'une moindre qualité (74 % et 76 % respectivement), après observation de la structure des ménages des enfants de ces exemples, les liens erronés de l'étape C1 ne peuvent pas être corrigés de manière automatique. Concernant ceux de l'étape A3, ces enfants sont faussement attribués à leur grand-mère, aucune réattribution n'étant effectuée, car la mère initiale, en réalité la grand-mère des enfants, était âgée de

¹⁰ Un échantillon de 238 000 femmes et 121 000 hommes couplé à l'EAR de 2011. Entre autres, l'enquête se renseigne sur la situation familiale des personnes et de leurs enfants. Chaque personne devait lister les enfants cohabitant ou non et la question « Êtes-vous sa mère ? » a été posée à toutes les femmes par rapport à chaque enfant vivant dans le ménage qui permet donc une identification de manière directe de la mère (biologique ou adoptive) d'un enfant.

moins de 50 ans. Abaisser le seuil de réattribution à un âge plus jeune entraînerait un nombre trop important de très jeunes mères.

Figure 10 : Évaluation par l'EFL de notre méthode d'identification des mères au recensement

REATTRIBUTIONS / NOUVELLES ATTRIBUTIONS			AUCUNE REATTRIBUTION		
Étape	Nb de cas vérifiables	% vraie mère identifiée	Étape	Nb de cas vérifiables	% vraie mère identifiée
A2	2	100	A1	7	100
A4	2	100	A3	17	76
C1	19	74	B1	45*	98
D3	38	100	B4	7	100
B3	3	100			
Ensemble	64	92	Ensemble	122	93
TOTAL	186	93			

Note : * = un échantillon aléatoire des 45 cas sur les 91 cas retrouvés a été effectué compte tenu de la lourdeur de la vérification manuelle.

Source : Tomkinson et Breton, 2016.

V. Les bases de données produites par le programme *rpdef.sas*

Le programme *rpdef.sas* produit la table *meres_tmoins1.sas7bdat*. Cette table contient les informations du recensement pour les femmes identifiées comme ayant eu un enfant dans l'année qui a précédé la collecte. En outre, nous avons créé les variables suivantes :

- IDENF – l'identifiant de l'enfant qui lui a été attribué ;
- AGE_MERE_NAIS – son âge à la naissance de l'enfant ;
- POT – l'étape du processus qui a attribué l'enfant à sa mère.

Le code suivant (

Extrait **13)** créer une base *femmes_def.sas7bdat* qui comprend toute femme recensée en ménage ordinaire âgée entre 16 et 50 ans au moment du recensement (entre 15 et 49 ans dans l'année $t-1$) et créer les variables suivantes :

- MERE – égal à 1 si la femme a eu un enfant dans l'année $t-1$, 0 autrement ;
- NAIS_POIDS – égal au nombre d'enfants eus dans l'année $t-1$ multiplié par le poids de la femme (IPONDI2).

Cette deuxième variable est à utiliser pour les analyses reposant sur le calcul des taux (numérateur). Notons que les imputations ne figurent pas dans ces bases, les informations relatives à la mère n'étant évidemment pas disponibles.

Extrait 13 : Création de la base *femmes_def.sas7bdat*

```
/***** Créer la base "femmes_def" *****/  
  
proc sort data=def.meres_tmoins1 out=meres_tmoins1 ; by individu_id ;  
  
data meres_tmoins1 ;  
set meres_tmoins1 ;  
  nbnaiss + 1 ;  
  by individu_id ;  
  if first.individu_id then nbnaiss = 1 ;  
  
proc sort data=meres_tmoins1 ; by individu_id descending nbnaiss ;  
proc sort data=meres_tmoins1 nodupkeys ; by individu_id ;  
  
data femmes_def (drop = rang_femme lprm_enf pot nbnaiss) ;  
set meres_tmoins1 (where=(16 le aged_num le 50))  
  femmes_rp (where=(16 le aged_num le 50)) ;  
if age_mere_nais ne . then mere = 1 ; else mere = 0 ;  
if mere = 1 then do ; nais_poids = nbnaiss * ipondi2 ; end ;  
  
proc sort data=femmes_def ; by individu_id descending mere ;  
proc sort data=femmes_def nodupkeys out=def.femmes_def ; by individu_id ;  
  
proc freq data=def.femmes_def ;  
table mere / missing ;  
run ;
```

VI. Résumé

Le programme *rpdef.sas* présenté dans ce document de travail permet de mieux identifier les mères dans les fichiers détails du recensement. Développé sur le fichier détail de 2011, il est transférable à d'autres fichiers du recensement. Les bases produites ouvrent des possibilités d'analyse sur les populations rares, la fécondité différentielle ou à des échelons géographiques fines.

Références

- ABBASI-SHAVAZI, M., 1997, « An assessment of the own-children method of estimating fertility by birthplace in Australia », *Journal of the Australian Population Association*, 14(2), p. 167-185.
- BELANGER, A. et GILBERT, S., 2003, « La fécondité des immigrantes et de leurs filles nées au Canada », *Rapport sur l'état de la population du Canada*, 2002, n° 91-202, p. 135-161.
- BRETON, D., 2010, « La fécondité avant 25 ans ne baisse plus en France depuis 15 ans », in : HECHT, J., BERGOUIGNAN, C., PRIOUX, F., BLAYO, C. et PARANT, A. (eds.), *La fécondité - Représentation, causalité, prospective - Actes du XV^e colloque national de démographie*, Paris : CUDEP, pp. 52-74.
- CHO, L.-J., 1986, *The own-children method of fertility estimation*, Honolulu : Population Institute, 188 p.
- CHO, L.-J. et MAN, J.-H., 1968, « Recent change in fertility rates of the Korean population », *Demography*, 5, p. 690-698.
- CICALI, D. et DE SANTIS, G., 2002, « Estimating cohort fertility levels and differentials by socioeconomic characteristics with the own-children method and pseudo panels », *Genus*, 58(2), p. 35-61.
- CLANCHE, F., 2014, « Focus ? Le nouveau recensement en continu : une méthode originale, des résultats nombreux et faciles d'accès », *Informations sociales*, 2014/3 (n° 183), p. 22-25.
- COEFFIC, N., 1993, « L'enquête post-censitaire de 1990. Une mesure de l'exhaustivité du recensement », *Population*, 48(6), p. 1655-1681.
- DAVIE, E. et MAZUY, M., 2010, « Fécondité et niveau d'études des femmes en France à partir des enquêtes annuelles de recensement », *Population*, 65(3), p. 475-511.
- DAVIE, E. et NIEL, X., 2012, « Mesurer et étudier la fécondité selon le rang de naissance : élaborer une statistique de nombre de naissances et d'âge à l'accouchement par rang », *Insee Document de travail n° F1205*, 57 p.
- DESPLANQUES, G., 1993, « Mesurer les disparités de fécondité à l'aide du seul recensement », *Population*, 48(6), p. 2011-2023.
- DESPLANQUES, G., 2008, « Avantages et incertitudes des enquêtes annuelles de recensement en France », *Population*, 63(3), p. 477-501.
- DRIXLER, F., 2015, « Conjuring the Ghosts of Missing Children : A Monte Carlo Simulation of Reproductive Restraint in Tokugawa Japan », *Demography*, 52(2), p. 667-703.
- DUBUC, S., 2009, « Application of the Own-Children Method for estimating fertility by ethnic and religious groups in the UK », *Journal of Population Research*, 26(3), p. 207-225.
- DUGBAZA, T., 1994, « Recent Trends and Differentials in Aboriginal and Torres Strait Islander Fertility, 1981-1991 », *ABS Demography Working Paper*, n° 1994/1.

- GRABILL, W. et CHO, L.-J., 1965, « Methodology for the Measurement of Current Fertility From Population Data on Young Children », *Demography*, 2, p. 50-73.
- HACKER, J., 2003, « Rethinking the "early" decline of marital fertility in the United States », *Demography*, 40(4), p. 605-620.
- KERSUZAN, C., 2010, « La méthode du décompte des enfants au foyer : évaluation des estimations de fécondité générale et de fécondité par rang dérivées des données censitaires », in : HECHT, J., BERGOUIGNAN, C., PRIOUX, F., BLAYO, C. et PARANT, A. (eds.), *La fécondité - Représentation, causalité, prospective - Actes du XVe colloque national de démographie*, Paris : CUDEP, pp. 30-49.
- LEFRANC, C., 2010, « La diffusion des résultats du recensement rénové de la population », *Courrier des statistiques*, n° 129, p. 1-6.
- O'HARE, W., 2009, « Why Are Young Children Missed So Often in the Census ? », *Kids Count Working Paper*, 15 p.
- RALLU, J., 2010, « Fécondité et origine du conjoint des migrants de première et deuxième générations », in : HECHT, J., BERGOUIGNAN, C., PRIOUX, F., BLAYO, C. et PARANT, A. (eds.), *La fécondité - Représentation, causalité, prospective - Actes du XVe colloque national de démographie*, Paris : CUDEP, pp. 120-132.
- RATNAYAKE, K., RETHERFORD, R. et SIVASUBRAMANIAM, S., 1984, *Fertility Estimates for Sri Lanka Derived from the 1981 Census*, Matara, Sri Lanka : Department of Geography, Ruhuna University ; Honolulu : East-West Population Institute, East-West Center, 63 p.
- RETFERFORD, R. et MIRZA, M., 1982, « Evidence of age exaggeration in demographic estimates for Pakistan », *Population Studies*, 36(2), p. 257-270.
- TEMPORAL, F., 2010, « Migrations et fécondité dans les Doms : tendances passées et perspectives », in : HECHT, J., BERGOUIGNAN, C., PRIOUX, F., BLAYO, C. et PARANT, A. (eds.), *La fécondité - Représentation, causalité, prospective - Actes du XVe colloque national de démographie*, Paris : CUDEP, pp. 123-141.
- TOMKINSON, J. et BRETON, D., 2017, « Comment mieux identifier les mères adolescentes dans le recensement français ? Améliorations de la méthode du « décompte des enfants au foyer » », *Cahiers québécois de démographie*, 45(2), p. 269-293.
- TOULEMON, L., 2017, « Undercount of young children and young adults in the new French census », *Statistical Journal of the IAOS*, 33(2), p. 311-316.
- TRABUT, L., LELIEVRE, E. et BAILLY, E., 2015, « Famille et recensement font-ils bon ménage ? », *Population*, 70(3), p. 637-666.

Annexes

Extrait A. 1: Syntaxe pour l'importation du fichier texte vers SAS

```
libname def 'c:\votre_repertoire' ;

/***** Importer le fichier téléchargé *****/
data import ;
infile "c:\votre_repertoire\fd_indcvi_2011.txt"
firstobs=2 delimiter = ';' dsd lrecl=32767 missover;
informat
cantville $5.
nummi $8.
achlr3 $1.
aemmr $1.
aged $3.
ager20 $2.
agerev $3.
agerevq $3.
anai $4.
anemr $2.
apaf $1.
arm $5.
ascen $1.
bain $1.
bati $1.
catiris $1.
catl $1.
catpc $1.
chau $1.
chfl $1.
chos $1.
clim $1.
cdbl $1.
couple $1.
csl $1.
cuis $1.
dept $3.
derou $1.
dipl $2.
dnai $3.
eau $1.
egoul $1.
elec $1.
empl $2.
etud $1.
garl $1.
hlml $1.
iletud $1.
ilt $1.
immi $1.
inai $1.
inatc $1.
infam $1.
inper $2.
inperf $2.
ipondi $17.
iris $9.
lienf $1.
lprf $1.
lprm $1.
matr $1.
metrodom $1.
moco $2.
modv $2.
nal7 $2.
```

```
na5 $2.  
naidt $2.  
nbpi $2.  
ne18fr $1.  
ne24fr $1.  
ne3fr $1.  
ne6fr $1.  
nenfr $1.  
nperr $1.  
numf $1.  
oridt $2.  
rech $1.  
region $2.  
sani $1.  
sanidom $2.  
sexe $1.  
sfm $2.  
statr $1.  
stocd $2.  
surf $1.  
tact $2.  
tactd16 $3.  
tactmf $2.  
tactpf $2.  
tp $1.  
trans $1.  
triris $6.  
typc $1.  
typfc $1.  
typl $1.  
typmc $1.  
typmr $2.  
voit $1.  
wc $1. ;
```

```
input  
cantville $  
nummi $  
achlr3 $  
aemmr $  
aged $  
ager20 $  
agerev $  
agerevq $  
anai $  
anemr $  
apaf $  
arm $  
ascen $  
bain $  
bati $  
catiris $  
catl $  
catpc $  
chau $  
chfl $  
chos $  
clim $  
cmbl $  
couple $  
csl $  
cuis $  
dept $  
derou $  
dipl $  
dnai $  
eau $  
egoul $
```

```

elec $
empl $
etud $
garl $
hlml $
iletud $
ilt $
immi $
inal $
inatc $
infam $
inper $
inperf $
ipondi $
iris $
lienf $
lprf $
lprm $
matr $
metrodom $
moco $
modv $
nal7 $
na5 $
naidt $
nbpi $
ne18fr $
ne24fr $
ne3fr $
ne6fr $
nenfr $
nperr $
numf $
oridt $
rech $
region $
sani $
sanidom $
sexe $
sfm $
statr $
stocd $
surf $
tact $
tactd16 $
tactmf $
tactpf $
tp $
trans $
triris $
typc $
typfc $
typl $
typmc $
typmr $
voit $
wc $ ;
run ;

```

Extrait A. 2 : Préparation de la base pour l'application

```

data rp ;
set import ;
where region notin ("01","02","03","04") ; /* NB. champ = individus résidant en
France métropolitaine */

```

```

/* Ecarter les variables dont nous n'avons pas besoin */
drop achlr3 ager20 agerevq anai arm ascen bain bati catiris chau chfl chos clim
cmb1 cuis derou eau egoul elec garl iletud ilt metrodom na17 na5 naidt nbpi ne18fr
ne24fr ne3fr ne6fr oridt rech sani sanidom statr surf tactd16 tycp typl typmc typmr
wc ;

/* Transformer les variables "ipondi" et "aged" en variable numerique */
virgule = index(ipondi,',') ;
if virgule ne 0 then do ;
substr(ipondi,virgule,1) = '.' ; end ;
ipondi2=ipondi*1 ;

aged_num=1*aged ;

/* Créer des identifiants d'individu, de famille et de ménage */
individu_id=_n_ ;
idfam=cats(cantville, nummi, numf) ;
idmen=cats(cantville, nummi) ;

run ;

```