

**NOTE**

Apurements et imputations dans l'Enquête famille et logements 2011

Dossier suivi par :  
Buisson Guillemette  
Tél. : 01 41 17 53 92  
Fax : 01 41 17 62 79  
Mél : [guillemette.buisson@insee.fr](mailto:guillemette.buisson@insee.fr)

Paris, le 10 avril 2014  
N° 714/DG75-F170/

Objet : **Apurements et imputations dans l'Enquête famille et logements 2011 - version du 09 avril 2014**

# Apurements et imputations de l'Enquête famille et logements 2011.

Les apurements et imputations décrits ici correspondent aux traitements mis en œuvre dans le fichier de production et de recherche d'avril 2014.

## Sommaire

Principes généraux.....	4
Les variables brutes .....	4
Principes d'apurements.....	5
La date de naissance du répondant (question 1).....	8
Composition de la fratrie du répondant (question 2).....	8
Apurement de la question 3 « Êtes-vous actuellement en couple ? » .....	10
Objectif 10	
Mise en cohérence interne à l'EFL.....	11
Mise en cohérence pour les cas que l'on ne peut trancher avec l'EAR.....	15
Variable Q3couple_R issue de ces apurements:.....	16
Passage de Q3couple_R à Q3couple.....	16
Traitements sur les personnes en couple de même sexe dans l'EFL .....	19
Apurement du sexe du répondant (Q0sexe) : .....	19
Apurement de la question 5 (Q5sexe_C):.....	19
Validation de sous-populations particulières à partir des listes A, B, C.....	20
Résultats issus de ces apurements.....	22
Correction du module couple (question 4 à 10) .....	23
Cohérence entre les questions sur le conjoint (bloc conjoint), la rupture d'union (bloc séparation) et la variable Q3couple.....	23
Date de naissance du conjoint: Q4JNAI_C, Q4MNAI_C, Q4ANAI_C (question 4) .....	23
Traitement des dates concernant la vie de couple.....	24
Règles de codage des questions Q7Pacs, Q8Mari et Q10cause_R .....	25
Correction des cohérences entre date de PACS et date de mariage .....	26
Traitement du bloc rupture d'union (Q10CAUSE_R, Q10ANNEE_S, Q10ANNEE_D).....	32
Traitement de la variable TPS_LOGR (question 9).....	32
Mise en cohérence des effectifs d'enfants déclarés (questions 11 et 12) avec les tableaux 13 et 14 ..	34
Apurements préliminaires.....	35
Mise en cohérence entre les compteurs et les enfants présents dans les tableaux.....	36
Imputations et indicateurs de « modification » des variables remplies .....	37
Enfants de couples de même sexe .....	39
Remontée des informations de la table enfant à la table individu.....	40
Apurements des caractéristiques des enfants des tableaux 13 et 14.....	41
Le sexe de l'enfant E13_14sexe .....	41
Année de naissance de l'enfant E13_14anai .....	41
Temps passé dans le logement : E13tpsigt et autres logements : E13aut_lgt E13lgt_aut_par, E13lgt_etude et E13lgt_autre.....	41
Âge à l'adoption E13_14agad.....	44
Âge de l'enfant au départ de chez son parent E14agdep .....	44
Le conjoint du répondant est-il parent de l'enfant ? (E13_14cpart) .....	45
Codage de la profession des parents (questions 17, 18, 23 et 24).....	47
Le processus de codage : .....	47
Fonctionnement de Sicare profession.....	48
Bilan du traitement.....	50
Apurement des questions q17, q18, q23 et q24.....	51
Résultats :.....	53
Etat vital et résidence des parents (questions 19, 20, 25 et 26) .....	55
Définition des variables définitives créées .....	55
Apurement Mère.....	56



Apurement Père .....	57
Recodages pour cohérence .....	58
Redressement de la non-réponse à Q19viv_m, Q19adec_m, Q25viv_p et q25adec_p.....	58
Grands-parents.....	61
Apurement des variables Q27grp_x, Q27nbpenf_x, Q27agpenf_x .....	61
Apurement des questions sur les logements dans lesquels vivent les répondants (questions 29 à 36) .....	62
Création de variables synthétiques .....	62
Correction de la question 29.....	63
Correction de la question 30.....	63
Création de variables d'aide à la décision.....	64
Traitements des cases biffées pour les question 31 à 36 .....	64
Correction de la question 31.....	64
Utilisation des précisions aux questions 32 et 36.....	65
Correction de la question 32.....	71
Correction de la question 34.....	72
Correction de la question 35.....	72
Correction de la question 36.....	72
Imputation aléatoire des données manquantes .....	74
L'année de mise en couple (Q6annee_c_I) .....	74
L'année de naissance et le sexe des enfants (E13_14anai_I et E13_14sexe_I) .....	82
Temps passé dans le logement et autre(s) logement(s) éventuel(s) des enfants (E13tpsigt_I, E13aut_lgt_I, E13lgt_aut_par_I, E13lgt_etude_I, E13lgt_autre_I).....	98
Lieu de naissance des parents (Q15 et Q21).....	119
Situation vitale des parents (Q19 et Q25) .....	129
Lieu de résidence des parents (Q20 et Q26). .....	136
Imputation aléatoire de la non-réponse des variables Q27grp_I, Q27nbpenf_I .....	153
Les autres données imputées aléatoirement (_I) .....	155
La variable E13_14VIVANT_I.....	155
Annexe.....	156
Objet : Repérage des questionnaires suite, vide, doublon, jumeau, hors champ et appariement de l'enquête Famille et logements avec l'EAR .....	157
Objet : Les enfants et les structures familiales: comparaison entre l'EFL et le Recensement.....	182
Correction des poids Zus dans EFL .....	197
1. Correction des poids individuels .....	197
2. Correction des poids ménage-famille-enfant .....	199



Le remplissage manuel du questionnaire par les enquêtés eux-mêmes, sans aide de l'agent recenseur, la saisie optique, la présence de questions filtres, la multiplicité d'informations demandées pour un sujet particulier et, enfin, l'adossement de l'EFL à l'EAR 2011, font ressortir de multiples incohérences ou anomalies et rendent les apurements nécessaires et nombreux. Ces derniers revêtent d'autant plus d'importance que l'enquête vise à étudier des configurations peu fréquentes, sur lesquelles l'apurement est susceptible d'avoir proportionnellement plus d'impact.

Les apurements ont été réalisés par l'Insee, l'Ined et la Drees. La Drees a réalisé l'apurement de la question 28, détaillé dans la note 2013\_03\_Appariement des enfants des tableaux 13 et 28 de l'EFL.

L'ensemble des autres apurements est décrit dans ce document.

## Principes généraux.

### Les variables brutes

Les variables brutes EFL peuvent être scindées en 3 catégories nécessitant des traitements différents : des variables de nombres (dates, âges, effectifs), des libellés en clair (pays professions, communes) et des cases à cocher.

Conformément au cahier des charges fourni au prestataire de la saisie optique, chaque case à cocher a une valeur allant de 0 à 3 à l'issue de cette saisie optique<sup>1</sup> :

On définit la valeur d'une case selon les spécifications suivantes :

- Les modalités à attribuer à chacune des cases de la question sont :

cas	valeur
case vide ou atteinte par un trait qui raye une partie du questionnaire	0
case cochée	1
case biffée	2
case raturée	3

Il est à noter que la procédure de saisie optique ne parvient pas toujours à faire la différence entre une case biffée et une case atteinte par un trait qui raye une partie du questionnaire.

Dans le fichier issu de la saisie, chaque question comportant des cases à cocher est codée sous la forme d'une « variable brute » unique, reprenant le détail des réponses à l'ensemble des cases à cocher. A ce stade, cette codification reprend les informations telles qu'elles figurent sur le questionnaire papier (même si elles sont incohérentes)<sup>2</sup>.

<sup>1</sup> Ces règles sont les mêmes que celles utilisées pour le RP pour la création d'une variable RP « non arbitrée ». En revanche, pour le RP le prestataire de saisie optique arbitre lui-même cette première variable afin de ne conserver que deux valeurs possibles 0 pour « case non cochée » ou 1 pour « case cochée ». Dans le cas de l'EFL cette demande n'a pas été faite afin de pouvoir analyser les données « brutes » issues de la saisie et déterminer des règles d'arbitrage plus fines.

<sup>2</sup> Globalement pour la saisie optique, le taux d'erreur doit être inférieur à 1% pour chaque variable. Par ailleurs, des contrôles de qualité supplémentaires sont imposés au prestataire sur les variables utiles à l'appariement avec le RP d'une part et le sexe du conjoint d'autre part. Pour les variables utiles à l'appariement, en particulier la date de naissance et le sexe du répondant, une vérification manuelle des toutes ces variables a été demandée pour tous les bulletins pour lesquels cet appariement ne fonctionnait pas. Pour le sexe du conjoint si la variable brute valait 01, 11, 21 ou 31, une vérification manuelle était demandée pour s'assurer qu'il ne s'agissait pas d'une erreur de saisie. In fine il devait être saisi ce qui était renseigné dans le questionnaire



Par exemple, la variable brute de la question 3 « êtes vous actuellement en couple » synthétise quatre cases à cocher, chacune pouvant prendre ces quatre valeurs, soit 256 possibilités (4\*4\*4\*4).

3 Êtes-vous actuellement en couple ?

- Oui, avec une personne qui vit dans le logement .....  1 } [Allez à la question 4](#)
- Oui, avec une personne qui vit dans un autre logement .....  2 } [Allez à la question 10](#)
- Non, mais vous avez déjà été en couple par le passé .....  3 } [Allez à la question 10](#)
- Non, vous n'avez jamais été en couple .....  4 } [Continuez page suivante, question 11](#)

Compte tenu de ces règles de saisie, il existe un nombre important de réponses différentes pour chacune des variables à cocher brutes.

Pour ces variables, une première étape consiste à réduire les possibilités de réponses différentes, en particulier quand un seul cochage est attendu. La règle suivante a été adoptée la plupart du temps (suite à des retours aux images de questionnaires notamment):

- Une variable avec une modalité à 1 et les autres à 0 est considérée comme propre.
- Dans une variable avec une modalité à 3 (raturée) et les autres à 0, la modalité à 3 est considérée comme un 1.
- Les modalités 2 et 3 sont traitées de la même façon quand elles se trouvent avec au moins une case cochée : on garde seulement la ou les cases cochées.
- La modalité 2 est considérée comme un 0 (case non cochée), ce qui est cohérent avec le fait qu'une case biffée est parfois difficile à distinguer d'un trait rayant le questionnaire. La modalité 2 n'est considérée comme un 1 que dans des cas très particuliers qui sont précisés.

Pour les nombres, le cahier des charges fourni au prestataire de la saisie optique indique les listes de valeurs que peuvent prendre les variables. Si la valeur ne correspond pas à l'une de ces valeurs possibles une vérification manuelle était demandée. Les caractères illisibles sont codés par une \*.

## Principes d'apurements.

Les étapes successives sont les suivantes :

### 1) Sélection des questionnaires conservés

- Un mineur ne doit pas répondre : les répondants mineurs n'ont pas été conservés dans le fichier (cf. la note 1290 du 11 juin 2012 en annexe)
- Les questionnaires sont destinés à un sexe donné et ne sont pas adaptés pour une personne du sexe opposé. Les rares répondants hommes à un questionnaire femme, et inversement, ne sont pas conservés<sup>3</sup> (cf. la note 1290 du 11 juin 2012 en annexe). Cela représente 455 questionnaires.

### 2) Amélioration de la cohérence interne de l'enquête (respect des champs et filtres, valeurs manquantes correspondant à des « non concernées »...) lorsque que les informations disponibles permettent de trancher (sans prise en compte à ce stade des réponses au recensement)

L'apurement consiste **avant tout à rechercher une cohérence interne au sein de chaque grand bloc de questions des questionnaires EFL**. Comme grand bloc on retient :

- Les caractéristiques du répondant (sexe, date de naissance, fratrie)
- Les informations sur sa vie de couple (questions 3 à 10)

<sup>3</sup> Comme le plan de sondage distingue zones dans lesquelles seuls les hommes sont enquêtés, et zones où seules les femmes sont enquêtées, ces répondants n'auraient pas eu de poids. Enfin, les informations qu'ils fournissent sont, le cas échéant, très proches de celles déjà données par le répondant de l'autre sexe du même ménage.



- Les informations sur ses enfants et ceux de son conjoint actuel (questions 11 à 14)
- Les informations sur ses parents (questions 15 à 26)
- Les informations sur ses petits-enfants (question 27)
- Les informations sur le mode de garde des enfants du domicile (question 28)
- Les informations sur les logements dans lesquels il vit (questions 29 à 36)

La seconde étape consiste à **rechercher une cohérence interne à l'EFL en croisant l'information des différents blocs** (en particulier le fait d'être, ou d'avoir été, en couple avec les enfants, les petits-enfants avec les enfants, les modes de garde des enfants avec les enfants).

Ces cohérences internes doivent permettre de :

- Corriger une partie de la non-réponse. Par exemple :  
Si la question 19 « votre mère est-elle vivante ? » n'est pas remplie et que l'enquêté déclare aux questions 25 et 26 que son père est vivant et vit avec sa mère, alors on considère que la mère est vivante.
- Retenir une réponse en cas de réponse ambiguë. Par exemple :  
Si la question 3 « êtes vous actuellement en couple ? » indique à la fois un cochage à « oui en couple cohabitant » et « non mais j'ai été en couple dans le passé », les éléments sur la rupture d'union et le conjoint actuel sont mobilisés pour trancher.
- Rendre les réponses compatibles avec le **respect des consignes du questionnaire**. Ces consignes sont les suivantes :
  - La question 3 « Êtes-vous en couple » est une question filtre : seules les personnes en couple doivent ensuite décrire uniquement leur conjoint actuel (questions 4 à 9) et seules les personnes qui ne sont plus en couple doivent préciser comment s'est finie leur ancienne union (question 10)<sup>4</sup>. Si une personne en couple (modalité 1 ou 2 de la question 3), fournit tout de même des éléments sur la rupture (question 10), ceux-ci sont mis en valeur manquante pour « non concernée » ; de même, si la personne n'est plus en couple, les éléments sur le conjoint actuel (questions 4 à 9) sont mis en valeur manquante pour « non concernée »<sup>5</sup> ; si la réponse est 4 l'ensemble des questions sur le conjoint et la rupture sont mise à blanc.
  - Les enfants figurant dans le tableau 13 sont les enfants du répondant ou de son conjoint actuel : les personnes qui ne sont pas actuellement en couple (q3couple=3 ou 4) ne sont donc pas concernées par la question « Son père est-il votre conjoint actuel ? » pour une femme (réciproquement sa mère pour un homme) et cette variable est mise en valeur manquante pour « non concernée ». Les enfants qui ne sont ni du répondant ni de son conjoint actuel ne font pas partie du champ et ne sont pas conservés dans la table enfant du FPR et ne sont pas pris en compte dans le décompte du nombre d'enfants de la personne (les informations brutes les concernant sont en revanche conservées à des fins de travaux méthodologiques éventuels dans les fichiers « adultes ») ;
  - Les enfants figurant dans le tableau 14 sont les enfants du répondant qu'ils soient, ou non, commun avec le conjoint actuel. Les enfants qui ne sont pas du répondant ne font pas partie du champ et ne sont pas conservés dans le FPR dans la table enfant (mais le sont dans la table adulte, cf. supra). Par rapport au tableau 13 les enfants du conjoint uniquement sont exclus du champ et ne sont pas dans le FPR. Comme dans le tableau 13 les personnes qui ne sont pas actuellement en couple (q3couple=3 ou 4) ne sont donc pas concernées par la question « Son père est-il votre conjoint actuel ? » pour une femme (réciproquement sa mère pour un homme) et cette variable est mise en valeur manquante pour « non concernée ».
  - Pour les couples de même sexe le conjoint actuel d'un homme est un homme et la réponse à la question pour les enfants « sa mère est-il votre conjoint actuel ? » est

<sup>4</sup> En effet, lors des tests sur une précédente version du questionnaire, des personnes en couple ont été gênées par des questions sur une union antérieure.

<sup>5</sup> Il s'agit par exemple de veufs/veuves décrivant leur ancien conjoint.



- mise à 2 pour non (dans les questionnaires, cette variable est généralement déjà à 2 ou non remplie) ;
- On ne demande la commune de résidence de la mère que pour un éventuel autre logement que celui du répondant. Si elle ne vit pas ailleurs que chez le répondant cette variable est mise en valeur manquante pour « non concernée ». La mère de l'enquêté peut à la fois vivre chez le répondant et vivre ailleurs.
  - On ne demande la commune de résidence du père que pour un éventuel autre logement que celui du répondant ou de la mère. Si il ne vit pas ailleurs que chez le répondant ou la mère du répondant cette variable est mise en valeur manquante pour « non concernée ». Il peut à la fois vivre avec le répondant, sa mère et dans un autre logement.
  - Les enfants du tableau 28 sont ceux du logement que le répondant et/ou son conjoint actuel soient ou non leurs parents.
  - Une seule réponse est attendue pour les questions 29, 30, 32, 34 alors que plusieurs sont possibles pour les questions 31, 35 et 36
  - La question 31 est une question filtre si on répond non à cette question les questions 32 à 36 sont mises en valeur manquantes pour « non concernée ».

### **3) Utilisation des réponses à l'enquête annuelle de recensement pour apurer l'EFL dans les cas où les réponses à cette dernière ne permettent pas de trancher.**

**Quand la cohérence interne à l'EFL ne suffit pas à trancher en cas de réponses manquantes ou incohérentes on recherche la cohérence avec le recensement.** Celui-ci est ainsi mobilisé afin de trancher entre deux interprétations possibles à partir de l'EFL. Cela revient à éviter de créer une incohérence entre le recensement et les réponses à l'EFL en l'absence de raison claire de le faire. Ce choix a été fait pour faciliter ultérieurement des travaux méthodologiques de comparaison entre les réponses à l'EAR et à l'EFL. A ce stade l'information déclarée dans le RP est récupérée sans contrôle supplémentaire (elle peut dans certains cas s'avérer aberrantes par rapports aux autres informations de l'enquête, par exemple l'année de naissance de la mère qui s'avère supérieure à l'année de naissance de l'enfant).

**4) Pour des sous-populations particulières dont les effectifs sont suffisamment restreints, l'apurement passe par des traitements au cas par cas mobilisant l'ensemble des données disponibles** (EFL, RP, listes A,B,C du questionnaire logement du recensement, voire les images des questionnaires).

### **5) Imputations après la phase d'apurement**

La phase d'apurement vise à « nettoyer » les questionnaires, en exploitant au mieux les informations fournies par l'enquêté et en ajoutant le moins d'aléa ou d'a priori possible.

Cette phase ne permet pas de reconstituer les réponses à toutes les questions, faute d'information disponible.

La non-réponse restante est imputée soit de façon déterministe soit de façon aléatoire par hot-deck. Ces travaux sont présentés dans ce document.

Des indicatrices de correction permettent d'identifier les données apurées ainsi que les données imputées. Elles sont détaillées par la suite dans ce document. Les variables correspondantes aux données imputées se termine par « \_I ».



## La date de naissance du répondant (question 1).

1 Quelle est votre date de naissance ?

<input type="text"/>					
jour		mois		année	

Cette variable a fait l'objet de consignes de saisie optique spécifique, car elle était indispensable pour permettre l'appariement des questionnaires de l'EFL avec ceux du recensement. Une vérification manuelle était demandée lors de la saisie optique pour tous les bulletins pour lesquels cet appariement ne fonctionnait pas (cf supra).

L'Ined a réalisé un apurement de la question 1

- 1. Repérage des données aberrantes, manquantes ou comportant une étoile ou un point d'interrogation (« \* », « ? ») sur les variables Q1MNAI\_X et Q1ANAI\_X.**
- 2. Utilisation des données brutes du RP pour récupérer les informations quand c'est possible**

Lorsque les dates de naissance n'étaient pas les mêmes entre le BI et EFL, il y a deux cas de figure :

- soit l'année de naissance EFL paraît cohérente par rapport aux autres dates du questionnaire et aux calculs d'âges aux événements donc on choisit de conserver cette date.
- Soit l'année de naissance EFL ne paraît pas cohérente donc on fait l'hypothèse d'un problème de saisie et on récupère la date du BI.

Finalement 444 cas ne sont pas corrigés pour le mois de naissance (Q1MNAI) qui reste en non-réponse.

Pour l'année de naissance il restait 45 cas non corrigés. Dans le FPR, l'Insee a corrigé ces non-réponses en mobilisant les données de diffusion du RP en lien avec le principe de cohérence entre les deux sources quand l'EFL à elle seule ne suffit pas à trancher.

	I_Q1anaï	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
données brutes	0	358692	99.70	358692	99.70
données apurées	1	1033	0.29	359725	99.99
données de diffusion RP	2	45	0.01	359770	100.00

## Composition de la fratrie du répondant (question 2)

L'Ined a réalisé un apurement de la question 2 : variables Q2NBFRRERES, Q2NBSOEURS, Q2NBDEMIF, Q2NBDEMIS.

### **1- Repérage des valeurs aberrantes et étoiles («\*»)**

Dans 24 cas NBFRRERES, NBSOEURS, NBDEMIS ou NBDEMIF sont supérieurs à 20 ou comportent une étoile. Les cas à vérifier se trouvent principalement sur la variable NBFRRERES et sur la variable NBSOEURS avec des valeurs telles que « 92 », par exemple. En regardant les images, on remarque que ce sont soit des réponses correctement déclarées, auquel cas on les laisse telles quelles, soit des erreurs de saisie où par exemple les « 0 » sont saisis en « 3 » ou en « 9 ».



**2- Correction des erreurs de saisie au cas par cas avec les images**

**3- Mise à 0 du nombre de frères, de sœurs, de demi-frères et de demi-sœurs en cas de non-réponse.**

Après analyse de la non-réponse il est apparu qu'elle correspondait massivement à des personnes qui n'étaient pas concernées. Si ce n'est pour 79 bulletins correspondant à des bulletins de personnes en couple de même sexe cohabitant créés pour pallier la non-réponse totale d'un des membres de ces couples à l'EFL (Cf. traitements des couples de même sexe.) pour lesquels la non-réponse a été imputée par hot deck (Cf. imputations)



## Apurement de la question 3 « Êtes-vous actuellement en couple ? »

### 3 Êtes-vous actuellement en couple ?

- Oui, avec une personne qui vit dans le logement .....  1
- Oui, avec une personne qui vit dans un autre logement.....  2
- Non, mais vous avez déjà été en couple par le passé.....  3
- Non, vous n'avez jamais été en couple .....  4

⇒ Allez à la question 4

⇒ Allez à la question 10

⇒ Continuez page suivante, question 11

### Objectif

Cet apurement réalisé par l'Insee avait pour objectif la création des variables Q3couple\_R (vit ou non en couple) et Q3couple (la variable détaillée correspondant au 4 modalités de la question 3)

#### Q3couple\_R :

Plutôt que de retranscrire la réponse à la **question (variable Q3couple\_X) « êtes vous en couple »** en 4 modalités (1, 2, 3 ou 4), on procède dans un premier temps à un **recodage sur deux modalités** : 1 (oui) ou 2 (non), sans distinguer s'il s'agit d'un couple cohabitant ou si l'individu non en couple l'a été dans le passé. On rajoute une troisième modalité (3) pour coder les individus pour lesquels les informations disponibles ne permettent pas de trancher entre 1 ou 2.

Dans un second temps, on affine ce premier classement en distinguant parmi les individus en couple (codé 1) ceux dont le conjoint vit dans le logement (1) et ceux pour lesquels ce n'est pas le cas (2). Là encore, la modalité 3 permet de coder les individus pour lesquels on ne peut trancher. On obtient donc un code sur deux caractères : 11, 12 ou 13.

De la même façon, on va distinguer les individus qui ne sont pas en couple selon s'ils l'ont été dans le passé (1) ou pas (2). On aura ainsi les valeurs : 21, 22 ou 23.

Q3couple : cette variable découle de la précédente :

Q3couple\_R=11 => Q3couple=1 ;

Q3couple\_R=12 => Q3couple=2 ;

Q3couple\_R=21 => Q3couple=3 ;

Q3couple\_R=22 => Q3couple=4 ;

### Mode opératoire

#### A/ Mise en cohérence interne à l'EFL

Les questions concernées par cette cohérence interne à EFL sont :

- La question Q3 (variable Q3couple\_X)
- Les questions Q4 à Q9 correspondant au bloc C pour conjoint
- La question Q10 correspondant au bloc S pour séparation.

L'apurement de la variable couple et la modification éventuelle des autres variables qui lui sont liées se fait par étapes : on définit tout d'abord ce qu'on entend par un bloc C et un bloc S renseigné, puis on examine la cohérence entre la réponse à la question Q3 « Êtes vous en couple ? » et le fait que les blocs C et S soient ou non remplis.

#### B/ Mise en cohérence pour les cas que l'on ne peut trancher avec les EAR



## Mise en cohérence interne à l'EFL

### Principes généraux :

- Pour les variables à cocher chaque modalité peut être codée 1 (cochée), 2 (rayée ou biffée), 3 (raturée) ou 0 en l'absence de ces éléments. Ces variables sont dans un premier temps recodées comme détaillé p 5 dans les principes généraux.

- Ordre de priorité dans le cas d'éléments contradictoires :

D'une façon générale, une année renseignée prime sur une case cochée. En effet, on admet qu'un individu peut se tromper de ligne au moment de cocher une case, mais il est plus difficile d'inventer une date. Concrètement si l'individu coche qu'il n'est pas marié avec son conjoint mais qu'il remplit une année de mariage, le cochage sera corrigé.

De la même manière, un bloc rempli prime sur la réponse à la question filtre Q3 : par exemple, si l'individu dit qu'il n'est plus en couple mais qu'il remplit le bloc C tandis que le bloc S est vide, on corrigera la question Q3 en indiquant qu'il est en couple.

### Création de la variable Q3couple\_R en lien avec le remplissage des blocs C et S :

1/ première étape : définir le premier caractère de Q3couple\_R, c'est-à-dire trancher sur la situation de couple du répondant : est-il ou non en couple ?

Il s'agit d'abord de définir ce qu'on entend par « **un bloc rempli** ». Les règles retenues sont les suivantes:

- le bloc séparation (S) est vide (0), semi rempli (1) ou rempli (2) :
  - S=2 si au moins une année est renseignée.
  - S=1 si les années ne sont pas renseignées mais qu'il y a une « marque » : qu'au moins une case (séparation ou décès) est cochée ou si une case est raturée et qu'il n'y a pas d'autres marquages (03 ou 30) : un marquage en « 2 » seul n'est pas considéré comme suffisant pour coder S en 1. Ce qui veut dire qu'on exclut les situations de type 02 ou 20 où le 2 correspond à un biffage. Mais on exclut aussi et surtout les cas où le 2 est une rayure (cas le plus fréquent), ce qui évite qu'un bloc rayé soit considéré comme rempli.
  - S=0 dans les autres cas.
- le bloc conjoint (C) est vide (0), semi rempli (1) ou rempli (2) :
  - C=1 : une des deux cases « pacs » ou « marié » est cochée (ou remplie de la façon suivante : pacs=10, 13, 01, 31, 03 et 21) ou la date de naissance du conjoint est remplie (au moins partiellement, parfois seulement le mois et/ou le jour) (parfois la date de naissance du conjoint est en fait la date de naissance de l'individu...) ou si le mois d'une date en lien direct avec le couple est rempli mais pas l'année.
  - C=2 : au moins une des années des dates en lien direct avec le couple (mise en couple, pacs, mariage) est remplie : on impose que le mois seul renseigné ne suffit pas car l'idée c'est ensuite de pouvoir ordonner avec une date de rupture de couple, si besoin.
  - C=0 dans les autres cas.

a/ Lorsqu'un seul bloc est rempli, c'est **le remplissage du bloc qui doit être privilégié à la réponse à la question filtre Q3** si ceux-ci ne sont pas cohérents. *Ex. la personne répond qu'elle est en couple mais ne remplit pas le bloc couple alors qu'elle renseigne le bloc Séparation. On corrigera donc sa situation en disant qu'elle n'est plus en couple.*

b/ Dans le cas où **aucun bloc n'est renseigné**, on se reporte à la réponse à la question Q3. Si celle-ci est propre (une seule case de cochée ou pas de case de cochée et une seule case de



raturée), alors on peut trancher. Si celle-ci n'est pas propre, on ne peut conclure à cette étape : il faudra (voir plus loin) faire appel aux données présentes dans l'EAR. On admet cependant pour l'instant que s'il y a deux cases de cochée et que celles-ci correspondent à la même modalité du premier caractère de Q3couple\_R (ie si COUPLE='1100' ou '0011') alors on renseigne le premier caractère de Q3couple\_R selon cette modalité.

c/ Lorsque les **deux blocs sont remplis** :

On distingue le cas des blocs semi-remplis des blocs remplis.

- Si les deux blocs sont remplis, on a donc des données relatives aux années des événements mentionnés :
  - o on compare les dates renseignées dans les deux blocs. En effet, celles-ci peuvent permettre d'ordonner les événements (mise en couple, mariage, pacs, séparation, décès) et d'en déduire dans quel ordre ceux-ci se sont passés et donc de dégager le dernier événement. Il se peut en effet, qu'un individu décrive dans le bloc S son couple précédent ou alors qu'il décrive dans le bloc C le couple qui s'est finalement séparé.
    - o On construit la variable ANNEE\_COUPLE (variable intermédiaire pour l'apurement non transmise dans le FPR) correspondant à l'année la plus récente a priori, c'est-à-dire à l'année du mariage si elle est renseignée, ou l'année de pacs ou encore l'année de mise en couple, si on les a.  
On construit aussi la variable ANNEE\_SEP (variable intermédiaire pour l'apurement non transmise dans le FPR) correspondant à la date la plus récente entre annee\_d et annee\_s.
    - o Si ANNEE\_COUPLE > ANNEE\_SEP alors le bloc S décrit le couple précédant le couple actuel et donc l'individu est bien en couple.
    - o Si ANNEE\_COUPLE < ANNEE\_SEP alors le bloc C décrit l'union de l'individu avant que celle-ci se soit rompue. L'individu n'est donc plus en couple.
    - o Dans le cas d'une égalité entre année de formation du couple et l'année de sa fin (peu d'observations) et s'il y a une unique case de cochée (94% des personnes considérées) à la question Q3, alors c'est cette case qui va déterminer si l'individu est en couple. Sinon, on ne peut trancher à cette étape.
- Si on a un seul bloc rempli et l'autre bloc semi rempli (S=1 et C=2 ou S=2 et C=1) : dans ce cas une seule date est renseignée et c'est l'événement auquel elle se réfère qui l'emporte et donc qui impacte Q3COUPLE\_R, sauf si Q3couple\_X est propre et de type '1000' ou '0100' ou '0010' : dans ce cas, on conserve l'information donnée par Q3couple\_X.
- Si aucune date n'est remplie (parmi année de mise en couple, année de mariage, année de pacs, année de séparation, année de décès), c'est-à-dire si les blocs sont vides ou semi-remplis, alors si la réponse à la question Q3 est « simple » (une unique case cochée et éventuellement une rature si on est dans le cas S=C= 0 ou 1) c'est la modalité remplie qui l'emporte, sauf si c'est '4' car normalement il ne devrait y avoir aucun bloc rempli. Dans ce cas, on ne tranche pas à cette étape.  
Dans le cas où la réponse à Q3 n'est pas « propre », on privilégie le bloc semi-rempli au bloc vide. Si le bloc semi-rempli est le bloc C, on impose aussi que les questions relatives au pacs et au mariage soient « proprement cochées » c'est-à-dire qu'il y a un unique cochage.

A l'issue de cette première étape, la variable Q3couple\_R est renseignée sur son premier caractère qui prend alors les modalités 1 ou 2, et 3 dans le cas où l'on n'a pu trancher.

Pour les situations vraiment ambiguës, on fera appel aux informations complémentaires de l'EAR.



Tableau récapitulatif :

		Bloc C rempli		
		0 = aucun marquage ou réponse aux seules questions Q5 et Q9 (1)	1 = l'une des dates du bloc est (au moins partiellement) remplie ou l'une des cases des questions Q7 et Q8 est cochée (ou biffée) (1)	2 = au moins une année relative au couple de remplie (1)
Bloc S rempli	0 = pas de marque (2)	pas de date à comparer: on se réfère à Q3couple_X qui permet de trancher si la réponse est "propre"(3)	on se réfère à Q3couple_X qui l'emporte ssi la réponse est propre (une unique case de cochée et pas de rature). Sinon c'est C (partiellement) rempli qui l'emporte (sauf si les questions pacs et mariage ne sont pas proprement cochées (4)). Si C et Q3couple_X ne sont pas "propres", on ne tranche pas.	Q3couple_R=1.
	1= au moins une marque mais pas d'année de remplie (2)	on se réfère à Q3couple_X qui l'emporte ssi la réponse est "propre" (une unique case de cochée et pas de rature ou une unique rature alors considérée comme un cochage ou éventuellement deux cochages "compatibles"). Sinon c'est S (partiellement) rempli (si une seule case de cochée pour cause) qui l'emporte. Si S et Q3couple_X ne sont pas "propres", on ne tranche pas.	pas de date à comparer: on se réfère à Q3couple_X qui permet de trancher si la réponse est "propre" et différente de '4'	Q3couple_R=1., sauf si Q3couple_X='0010' : dans ce cas, Q3couple_R=2.
	2 = une année de remplie (ou deux)	Q3couple_R=2.	Q3couple_R=2., sauf si Q3couple_X='1000' ou '0100': dans ce cas, Q3couple_R=1.	comparer les années des deux blocs, en faisant ressortir les années les plus récentes dans chaque bloc. Si celles-ci sont égales se reporter à Q3couple_X si la réponse est propre (une unique case de cochée et pas de rature). Distinguer le cas où l'on trouve une " * "dans les dates ou si seul le mois de l'évènement a été rempli : cf Q3couple_X pour trancher si réponse propre, sinon on ne tranche pas. On peut aussi essayer d'affiner et de trancher selon la place de l'*

- (1): si une ou deux cases de raturées aux questions relatives au pacs et au mariage ou si deux cases de cochée: considéré comme un marquage  
 (2): si les deux années sont vides et que Q10cause\_r\_X=33 alors S=0  
 (3): dans ce cas, on autorise une seule case raturée, qui est alors considérée comme une case cochée.  
 (4): on impose que Q7pacs\_X in ('00' '01' '10') and Q8mari\_X not in ('00' '01' '10') pour qu'in fine Q3couple\_R=1.

2/ deuxième étape : définir le deuxième caractère de Q3couple R, c'est-à-dire affiner la situation de couple du répondant (s'il est en couple, est-ce en couple cohabitant ; s'il n'est pas en couple, l'a-t-il été un jour).

Le problème se pose dans les cas où la variable Q3couple\_X n'est pas « propre » ou, si elle l'est, lorsque le choix finalement retenu pour le premier caractère de Q3couple\_R n'est pas conforme avec la valeur initiale.

On commence tout d'abord par corriger la variable Q9TPS\_LOGR\_X qui qualifie la durée de présence du conjoint dans le logement. On crée Q9TPS\_LOGR codée sur un caractère compris entre 1 et 6, correspondant à la valeur cochée de Q9TPS\_LOGR\_X. Lorsqu'il n'y a pas de case de cochée, Q9TPS\_LOGR =0 et s'il y a plusieurs cases de cochées (ou de marquées) incompatibles, Q9TPS\_LOGR=8 (modalité qui n'existe plus dans la version finale).



- si Q9TPS\_LOGR\_X contient une unique case cochée, Q9TPS\_LOGR prend la valeur de cette case.
- si Q9TPS\_LOGR\_X contient plusieurs cases cochées, Q9TPS\_LOGR prend la valeur la plus élevée.
- dans le cas où Q9TPS\_LOGR\_X = '111111' ou Q9TPS\_LOGR\_X = '100001' (ou tout autre réponse avec une croix aux modalités « 1 » et « 6 »), TPS\_LOGR=8
- si Q9TPS\_LOGR\_X = '000000' alors Q9TPS\_LOGR=0
- si pas de case de cochée, dans la majorité des cas, Q9TPS\_LOGR=0. Exception faite cependant des questionnaires où l'on trouve plusieurs cases codées à 2 : Q9TPS\_LOGR\_X=8,

a/ Les individus en couple (Q3couple\_R=1).

#### Les individus en couple cohabitant

- Pour les individus ayant répondu « clairement » à la question Q3 être en couple avec une personne qui vit dans le logement, c'est à dire avec Q3COUPLE\_X='1000' (on accepte aussi 1300, 1030, 1003, 3000) et qui déclarent en plus que le conjoint vit toute l'année dans le logement (Q9TPS\_LOGR=1), alors il y a cohérence entre Q3 et Q9 et l'on peut en conclure qu'il s'agit d'un couple cohabitant : Q3couple\_R=11.

Plus de 90 % des individus classés dans Q3couple\_R=1 sont dans ce cas.

- On trouve aussi des individus répondant « clairement » à Q3 comme étant en couple avec une personne vivant dans le logement mais qui cochent une case différente de la première à la question Q9. On décide de considérer que ces individus, dès lors que Q9TPS\_LOGR est différent de 6 (le conjoint vit rarement ou jamais dans le logement), sont effectivement en situation de couple cohabitant. En effet, il peut notamment s'agir d'un conjoint qui pour des raisons professionnelles doit habiter dans une autre ville quelques jours dans la semaine mais qui revient dans le logement chaque semaine. ) On considère la réponse à la question Q9 comme secondaire par rapport à la réponse à la question Q3 car cette question Q9 n'est pas suffisamment claire à la fois en terme de « quantité » de temps passé dans le logement (par exemple « quelques mois dans l'année » peut correspondre aussi bien à environ 3 mois qu'à 9 mois), et de « fréquence » des passages dans le logement (par exemple les situations « plutôt les vacances » et « plutôt les week-end », regroupées dans une même modalité, peuvent couvrir des situations assez différentes). Le choix de ces modalités traduit les contraintes de place disponible sur le questionnaire mais aussi le souhait de conserver des questions cohérentes avec d'autres enquêtes statistiques, en particulier le tronc commun des enquêtes ménages de l'Insee.

- Par ailleurs, lorsque Q3COUPLE\_X='1000' (ou 1300, 1030, 1003, 3000) et Q9TPS\_LOGR=6, on vérifie si la personne répondante ne se déclare par elle-même comme vivant « plus rarement » dans ce logement (variable Q29TPS, modalité 6). Si c'est le cas, on considère que le couple est bien cohabitant et que ces deux membres vivent généralement ailleurs que dans le logement où il y a eu dépôt du questionnaire

- A ces cas, on rajoute aussi les individus n'ayant pas renseigné Q3COUPLE\_X (= '0000') et qui ont Q9TPS\_LOGR =1'.

- Enfin, on compte aussi quelques individus tels que Q3COUPLE\_X = '0000' et Q9TPS\_LOGR différent de 0 ou 1 ou 8 mais correspondant à la modalité cochée de Q29TPS (=temps passé dans ce logement par le répondant) : l'individu et son conjoint habitent tous les deux dans le logement pour une durée « identique » et inférieure à un plein temps. On estime alors qu'ils vivent ensemble dans différents lieux.

#### Les individus en couple non cohabitant

- Pour les individus ayant répondu « clairement » à la question Q3 être en couple avec une personne ne vivant dans le logement, c'est à dire Q3COUPLE\_X='0100' (on accepte aussi 1300, 1030, 1003, 3000) et qui déclarent que le conjoint vit moins que « toute l'année ou presque » dans le logement (Q9TPS\_LOGR différent de 1), alors il y a cohérence entre Q3 et Q9 et l'on peut en conclure qu'il s'agit d'un couple non cohabitant : Q3couple\_R=12.

7880 individus sont dans ce cas, ce qui représente 82% des individus qui se déclarent comme étant en couple non cohabitant (Q3COUPLE\_X=0100, 1300, 1030, 1003, 3000).

- Si Q3COUPLE\_X=0100, 1300, 1030, 1003, 3000 et Q9TPS\_LOGR=1, on va regarder la réponse à la question Q29, renseignant sur la présence du répondant dans le logement. Si celle-ci est de type « une unique case de cochée et des 0 » et que la case de cochée n'est pas « toute l'année ou presque » alors Q3couple\_R=12 (58 cas).



- Si Q3COUPLE\_X=0000 et Q9TPS\_LOGR=6 alors Q3couple\_R=12.

Restent, enfin, des individus en couple, pour lesquels on n'a pu trancher s'ils étaient ou non en couple cohabitant.

Pour ces derniers, on fait appel aux informations autres disponibles dans le RP (voir plus bas).

b/ Les individus qui ne sont pas en couple (Q3couple\_R=2).

Si le bloc S est rempli (S=1 ou 2) alors Q3couple\_R =21 (même si couple='0001').

C'est plus délicat si S est vide : il peut s'agir soit de non-réponse soit d'une absence de réponse car l'individu n'est pas concerné. On regarde alors la réponse à Q3COUPLE\_X. S étant vide et Q3couple\_R=2., cela signifie que C est vide également (sinon d'après nos règles de décision, C rempli ou semi-rempli induirait Q3COUPLE\_R=1.) et donc que c'est la variable Q3COUPLE\_X qui a conduit à Q3COUPLE\_R=2. De ce fait, c'est une variable qui dans ce cas est relativement propre : plus de 98% des individus avec Q3couple\_R=2. ont Q3couple\_X=0010 ou 0001. Pour quasiment tous les autres cas, la situation est sans ambiguïté (une seule case cochée et une ou plusieurs cases raturées ; par ex. Q3couple\_X=3001 à interpréter comme 0001)

Si S=0 et si Q3COUPLE\_X n'a qu'une case de cochée, à savoir la dernière (ou pas de case cochée et la dernière case de raturée), alors Q3couple\_R=22.

Si S=0 et si Q3COUPLE\_X n'a qu'une case de cochée, l'avant-dernière, la situation est moins claire car il y a incohérence entre Q3COUPLE\_X et S qui devrait dans ce cas être renseigné. On décide de privilégier, dans ce cas, la variable Q3COUPLE\_X et donc Q3couple\_R =21, en estimant qu'il y a eu de la non-réponse totale pour le bloc S.

Pour les cas pour lesquels on n'arrive pas à trancher, Q3couple\_R=23.

On fait alors appel aux informations de l'EAR pour éclairer les situations floues de type Q3couple\_R=13, 23 ou 33.

## Mise en cohérence pour les cas que l'on ne peut trancher avec l'EAR

1/ On cherche l'existence d'un conjoint à partir des BI correspondant au logement de l'individu interrogé dans l'EFL

Pour cela, on utilise la date de naissance Q4anai\_c et le sexe du conjoint Q5sexe\_c pour repérer dans l'EAR les individus conjoints d'un répondant à l'EFL.

Plusieurs cas se présentent :

- il n'y a pas de date de naissance renseignée dans l'EFL ou la date de naissance est remplie mais on ne trouve pas d'individu correspondant dans l'EAR, soit parce qu'il n'y en a effectivement pas, soit parce que dans l'un des deux fichiers la date de naissance renseignée est soit incomplète soit erronée, et il n'y a pas un seul conjoint potentiel dans l'EAR, dans ce cas, soit il n'y a pas d'informations sur le conjoint et l'individu est classé dans en\_couple=2 ; soit il y a d'autres informations dans le bloc C qui conduisent l'individu répondant à être classé en en\_couple=1 mais l'absence de datenaiss\_cjt ne permet pas d'aller rechercher un conjoint dans l'EAR.
- La date de naissance est remplie et/ou on trouve l'individu correspondant dans l'EAR : la présence du conjoint dans le logement est confirmée.

La présence du conjoint dans l'EAR permet d'affirmer que conjoint il y a, et que celui-ci cohabite. Le contraire n'est pas toujours vrai : si on ne trouve pas de conjoint dans le logement, cela ne veut pas dire qu'il n'y en a pas. En effet, cela peut tout simplement signifier que le mois et l'année de naissance



du conjoint n'est pas suffisamment renseignée dans le bloc C de l'EFL, rendant ainsi l'appariement impossible avec l'EAR.

Un conjoint repéré dans l'EAR peut conduire à reconsidérer des cas préalablement tranchés au vue de la procédure décrite plus haut, mais non propre au départ.

2/ On regarde les informations complémentaires dans les BI des individus interrogés dans l'EFL

Le croisement de la variable Q3couple\_R avec la variable COUPLE\_X de l'EAR qui correspond à la réponse à la question « vivez-vous en couple ? » permet de valider la plupart des choix effectués. Certains cas semblent poser question cependant. Ainsi si Q3couple\_R=12 se répartit entre les deux modalités de COUPLE\_X du fait de l'interprétation non univoque de la question du recensement « vivez-vous en couple ? », le fait que des individus classés en couple cohabitant (Q3couple\_R=11) puissent se déclarer comme ne pas vivre en couple dans l'EAR est plus surprenant (1170 cas).

Côté EFL, la majorité de ces individus ont pourtant des questionnaires « propres », sans ambiguïté, où la variable Q3Couple\_X est clairement renseignée et en adéquation avec le remplissage du bloc C, et où la question Q9TPS\_LOGR est, elle aussi, cochée clairement en « toute l'année ou presque ».

Toutefois la notion de couple cohabitant est différente dans le recensement par rapport à l'EFL. Dans l'EFL c'est le point de vue de la personne qui est retenu, sans critère de durée ou de fréquence pour la présence du conjoint dans le logement.

La prise en compte de l'ensemble des informations relatives à la présence d'un conjoint côté EFL et côté EAR : Q3couple\_X, bloc C ou S rempli, un conjoint effectivement présent parmi les BI de l'EAR, les variables EAR Couple\_X, MOCO, voire MATR (une personne non célibataire ne peut, a priori, pas n'avoir jamais été en couple) permettent dans les cas les plus complexes de trancher.

### Variable Q3couple\_R issue de ces apurements:

Q3couple_R	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
11	232282	64.56	232282	64.56
12	9688	2.69	241970	67.26
13	425	0.12	242395	67.37
21	68534	19.05	310929	86.42
22	45275	12.58	356204	99.01
33	3566	0.99	359770	100.00

Un travail conjoint entre l'Ined et l'Insee a permis de valider les résultats de ces apurements.

### Passage de Q3couple\_R à Q3couple

#### 1<sup>ère</sup> étape : recodage automatique

L'essentiel des cas consiste en un recodage automatique qui permet de reprendre les modalités du questionnaire.

Q3couple\_R=11 => Q3couple=1 ;

Q3couple\_R=12 => Q3couple=2 ;

Q3couple\_R=21 => Q3couple=3 ;

Q3couple\_R=22 => Q3couple=4 ;

#### 2<sup>ème</sup> étape : répartition des « en couple sans précision » entre couple cohabitant et couple non cohabitant

425 bulletins sont classés en couple sans précision dans Q3couple\_R (Q3couple\_R=13).

Si on a retrouvé le conjoint dans le RP au cours de l'appariement du conjoint on classe ce bulletin en Q3couple=1 c'est à dire en couple cohabitant.

Si on ne retrouve pas le conjoint dans le RP à partir de l'appariement :



- et que le répondant est classé dans le RP dans les modes de cohabitation « enfant » (MOCO=11 ou 12), « parent d'une famille monoparentale » (MOCO=23), ou personne habitant seule (MOCO=32), on classe ce bulletin en Q3couple=2 c'est à dire en couple non cohabitant.
- et que le répondant est classé dans le RP dans les modes de cohabitation « couple » (MOCO=21 ou 22) ou « isolé dans un ménage de plusieurs personnes » (MOCO=31) :
  - et que le répondant déclare vivre en couple dans la variable brute du RP, on classe ce bulletin en Q3couple=1 c'est à dire en couple cohabitant.
  - et que le répondant ne déclare pas vivre en couple dans la variable brute du RP, on classe ce bulletin en Q3couple=2 c'est à dire en couple non-cohabitant.

### **3<sup>ème</sup> étape : répartition des Q3couple R=33 à partir de données annexes**

Pour 3566 bulletins on n'a pas pu décider si le répondant était en couple ou non, soit parce que les éléments sur la conjugalité étaient trop incohérents, soit en raison d'une non-réponse presque générale sur ce thème.

Si dans le RP le répondant est classé dans les modes de cohabitation « couple » (MOCO=21 ou 22), on considère alors que le répondant est en couple cohabitant, valorisant ainsi la cohérence entre les deux sources (cf. principes généraux). On code ainsi 79 bulletins (I\_Q3couple=3).

Pour les personnes qui ne sont pas en couple d'après MOCO (2552 cas avec I\_Q3couple=3):

- Les personnes de 75 ans ou plus, dont la réponse à la question vivez-vous en couple dans le RP n'est pas oui, sont classées en Q3couple=3 (couple dans le passé). En cette situation est la plus répandue parmi les répondants.
- Les personnes de moins de 75 ans classées en « enfant » dans le mode de cohabitation RP (MOCO=11 ou 12) :
  - qui répondent oui à la question vivez-vous en couple sont classées en Q3couple=2 couple non-cohabitant, ce qu'indique le RP.
  - qui ne répondent pas oui à la question vivez-vous en couple sont classées en Q3couple=4 c'est-à-dire jamais en couple.
- Les personnes de moins de 75 ans classées en « parent de famille monoparentale » dans le mode de cohabitation RP (MOCO=23) :
  - qui répondent oui à la question vivez-vous en couple sont classées en Q3couple=2 couple non-cohabitant, ce qu'indique le RP.
  - qui ne répondent pas oui à la question vivez-vous en couple sont classées en Q3couple=3 c'est-à-dire en couple par le passé.
- Les personnes de moins de 75 ans classées en « personnes seules » (MOCO=32) :
  - qui répondent oui à la question vivez-vous en couple sont classées en Q3couple=2 couple non-cohabitant.
  - qui ont moins de 25 ans et ne répondent pas oui à la question vivez-vous en couple sont classées en Q3couple=4 c'est-à-dire jamais en couple.
- Les personnes de moins de 75 ans classées en « personnes isolés » (MOCO=31) :
  - qui répondent oui à la question vivez-vous en couple sont classées en Q3couple=1 couple cohabitant.
  - qui ont moins de 25 ans et ne répondent pas oui à la question vivez-vous en couple sont classées en Q3couple=4 c'est à dire jamais en couple.

Les personnes âgées entre 25 et 75 ans classées en « personnes seules » ou « isolées » (MOCO=32 ou 31) qui ne répondent pas oui à la question vivez-vous en couple sont réparties aléatoirement en Q3couple=3 ou 4 par hot-deck selon le mode de cohabitation et l'âge (1359 cas correspondant à I\_Q3couple=4).



I_Q3couple	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
0	342669	95.25	342669	95.25
1	1983	0.55	344652	95.80
2	11128	3.09	355780	98.89
3	2631	0.73	358411	99.62
4	1359	0.38	359770	100.00

#### 4<sup>ème</sup> étape : des validations au cas par cas.

L'apurement a jusqu'ici essentiellement mobilisé les informations sur la conjugalité : des contrôles de cohérence à partir des données apurées des enfants, et plus généralement de l'ensemble des données EFL, ont fait ressortir quelques cas posant questions, de même le travail d'apurement sur les couples de même sexe ont fait remonter des questions sur cette variable. L'utilisation de l'ensemble des données EFL des données RP ainsi que des listes A,B,C ont alors permis de corriger à la marge ces cas (Cf. traitement des couples de même sexe, apurement enfant...).

Table de Q3couple\_R par Q3couple

Q3couple_R	Q3couple				Total
	1	2	3	4	
11	232145	13	61	63	232282
	64.53	0.00	0.02	0.02	64.56
	99.94	0.01	0.03	0.03	
	99.77	0.13	0.09	0.13	
12	53	9579	25	31	9688
	0.01	2.66	0.01	0.01	2.69
	0.55	98.87	0.26	0.32	
	0.02	98.81	0.04	0.07	
13	325	79	8	13	425
	0.09	0.02	0.00	0.00	0.12
	76.47	18.59	1.88	3.06	
	0.14	0.81	0.01	0.03	
21	61	0	68473	0	68534
	0.02	0.00	19.03	0.00	19.05
	0.09	0.00	99.91	0.00	
	0.03	0.00	97.49	0.00	
22	16	0	0	45259	45275
	0.00	0.00	0.00	12.58	12.58
	0.04	0.00	0.00	99.96	
	0.01	0.00	0.00	95.96	
33	79	23	1666	1798	3566
	0.02	0.01	0.46	0.50	0.99
	2.22	0.64	46.72	50.42	
	0.03	0.24	2.37	3.81	
Total	232679	9694	70233	47164	359770
	64.67	2.69	19.52	13.11	100.00



## Traitements sur les personnes en couple de même sexe dans l'EFL

Les personnes en couple de même sexe dans l'EFL se repèrent en utilisant :

- les réponses à la question 3 : « Êtes vous actuellement en couple ? » : les personnes ayant répondu « oui avec une personne qui vit dans le logement » et « oui avec une personne qui vit dans un autre logement » sont considérées comme des personnes en couple ;
- parmi les personnes en couple, les hommes (respectivement les femmes) qui répondent « un homme » (respectivement « une femme ») à la question 5 : « votre conjoint/ami est : un homme ? une femme ? » sont considérés comme étant en couple homosexuel.

C'est une approche individuelle des personnes en couple de même sexe, et non une approche du couple au sein d'un ménage. Toutefois, dans un deuxième temps pour les logement avec au moins une personne en couple de même sexe se déclarant en couple cohabitant, une vérification particulière a été faite afin de vérifier qu'il y avait bien dans ce logement deux personnes se déclarant en couple de même sexe. Cette vérification est rendue possible par la spécificité de la collecte EFL consistant à interroger toutes les personnes majeures d'un sexe donné dans un logement.

### Apurement du sexe du répondant (Q0sexe) :

Un appariement a été effectué entre les répondants EFL et les bulletins individuels du recensement sur la base du numéro de logement, du sexe, et de la date de naissance, puis en levant progressivement certaines contraintes. Dans 1236 cas, le sexe de l'EFL, déduit du type de bulletin EFL rempli : homme ou femme, est différent du sexe de l'EAR: 322 femmes EFL sont des hommes dans le RP et 914 hommes EFL des femmes RP.

Parmi eux, la proportion qui se déclare en couple de même sexe est forte (35 %) :

En cas d'incohérence sur le sexe du répondant dans l'EFL et l'EAR, si la personne se déclare en couple de même sexe dans l'EFL mais qu'il n'y a pas une autre personne en couple de même sexe dans le logement, on garde la variable sexe de l'EAR. Dans tous les autres cas (couple de personnes de sexes différents et célibataire), on garde la variable sexe de l'EFL.

Cette correction concerne 454 bulletins EFL. Ces bulletins sont ensuite considérés comme hors champ (hommes qui ont répondu à un questionnaire femme, et inversement, ou hommes appartenant à un iris femme, qui sont dans un logement où des femmes ont répondu) et ne sont pas retenus dans la base EFL.

Au final, pour 782 bulletins pour lesquels le sexe du répondant est différent de celui présent dans le recensement, on conserve le sexe renseigné dans l'EFL.

### Apurement de la question 5 (Q5sexe C):

Pour le sexe du conjoint si la variable brute valait 01, 11, 21 ou 31, une vérification manuelle était demandée pour s'assurer qu'il ne s'agissait pas d'une erreur de saisie. In fine il devait être saisi ce qui était renseigné dans le questionnaire (Cf. principes généraux).

Pour repérer, **avant apurement de la question 5**, les couples de même sexe on retient parmi les personnes en couple après apurement (Q3couple vaut « 1 » ou « 2 ») les cas pour lesquels la variable brute Q5couple\_C\_X prend la valeur : « 01 ».

Parmi les personnes en couple, on trouve ainsi, en pondéré avant apurement, 276 962 personnes en couple de même sexe, soit 0,9% des personnes en couple.

De plus :

- 610 524 personnes en couple ne répondent pas à la question 5 sur le sexe du conjoint, soit 1,9% des personnes en couple.
- 95 053 personnes en couple n'ont pas une réponse claire à la question (c'est à dire valant « 10 » ou « 01 »).



### **1<sup>ère</sup> étape : comparaison avec le sexe du conjoint renseigné dans les bulletins individuels du recensement.**

On a cherché à appairer, pour l'ensemble des personnes en couple, le conjoint de l'enquête EFL, lorsqu'il est mentionné, avec l'un des répondants au recensement du logement en fonction de la date de naissance du conjoint renseignée et de son sexe, puis en levant progressivement certaines contraintes (logement et date de naissance complète mais sans le sexe, date de naissance incomplète et sexe...). Cet appariement permet ensuite de comparer le sexe du conjoint issu de l'EFL et celui issu du RP.

Pour 95% des personnes en couple EFL, le sexe du conjoint est le même dans l'EFL que dans la variable brute du recensement.

Dans cette sous population « non ambiguë », on dénombre 160 000 répondants EFL en couple de même sexe, soit 0,5% des ces personnes en couple.

Dans 4% des cas le sexe du conjoint est renseigné dans l'EFL mais on ne retrouve pas l'information dans le recensement. Cela correspond essentiellement au cas des couples non cohabitants EFL, pour lesquels, logiquement, on ne retrouve pas le conjoint dans le logement. A ce stade, on garde l'information EFL.

Il y a un nombre assez faible de cas où le sexe du conjoint est renseigné dans le recensement mais pas dans l'EFL (76 cas soit 0,03% des personnes en couple). A ce stade, on récupère l'information RP.

Pour 1 414 personnes en couple dans l'EFL le sexe du conjoint est différent entre l'EFL et le recensement. Dans ce cas on considère que ces personnes sont toutes en couple avec une personne de sexe différent. En effet la probabilité d'être en couple de même sexe est faible (moins d'1%). Par ailleurs 83% d'entre elles se déclarent mariées à la question 8 de l'EFL, limitant encore la probabilité d'être en couple de même sexe. Enfin, le retour aux listes A, B, C dans environ 300 de ces cas (dont l'exhaustivité des 113 cas qui ne se déclaraient pas mariés) n'a pas permis de retrouver de personnes en couple de même sexe.

Dans 668 cas (0,3% des personnes en couple) le sexe du conjoint n'est disponible ni dans le recensement ni dans l'EFL.

### **2<sup>ème</sup> étape utilisation de la question 8 de l'EFL : Vous-êtes vous mariés ensemble ?**

Quand le sexe du conjoint n'est disponible que dans le recensement, ou que dans l'EFL ou bien n'est disponible ni dans le recensement ni dans l'EFL et que le répondant se déclare marié à la question Q8 on considère, après retour à quelques images et listes A, B, C que cette personne est en couple avec une personne de sexe différent. On a ainsi pu changer le sexe du conjoint déclaré uniquement dans l'EFL ou uniquement dans l'EAR quand la personne se déclarait également mariée. Ces cas de couples de même sexe étaient, en effet, invalidés par les listes A, B, C.

### **3<sup>ème</sup> étape : utilisation des listes A, B, C pour vérifier la présence de couple de même sexe parmi les 272 cas de non-réponse EFL et RP sur le sexe du conjoint.**

## **Validation de sous-populations particulières à partir des listes A, B, C**

### **1<sup>ère</sup> étape : Les personnes en couple avec enfant.**

Le retour aux listes A, B, C a permis de vérifier les cas de personnes se déclarant en couple avec une personne du même sexe et déclarant des enfants dans les tableaux 13 et 14, ou déclarant des enfants dans le logement aux questions 11 et 12. Les effectifs sont en effet assez réduits pour être vérifiés de façon systématique.

Cela a conduit à :

- Mettre en cohérence, dans le cas de couple cohabitant, les déclarations des deux membres du couple de même sexe sur leurs enfants et ceux de leur conjoint. En effet le nombre



d'enfants déclarés dans le bulletin EFL du premier membre du couple doit être déclaré en enfant du conjoint dans celui du deuxième membre du couple, et inversement.

- Éliminer des erreurs sur la variable couple : certains enfant vivant avec leur parent se déclarant en couple avec eux ou inversement, certains colocataires déclarant la date de naissance de leur colocataire à la question 4....
- Éliminer des enfants du tableau T13, qui n'y ont pas leur place, en lien avec le problème repéré dans le point précédent : certains enfants déclarent, par exemple, leur parent dans la partie sur le couple et vont inscrire leurs frères et sœurs, voire eux-mêmes, dans la partie relative aux enfants.
- Corriger quelques cas non détectés d'erreurs sur le sexe du conjoint.

### **2<sup>ème</sup> étape : Une seule personne (avec ou sans enfant) en couple cohabitant dans le logement.**

De même ces listes A, B, C ont été mobilisées pour vérifier les cas où dans un logement une seule personne se déclare en couple cohabitant (Q3couple= « 1 ») avec une personne du même sexe. Logiquement, les deux membres d'un couple de même sexe devraient répondre à l'enquête.

Cela a conduit à :

- Générer des bulletins EFL pour les cas de couple de même sexe avéré d'après les listes A, B, C (les listes A,B,C indiquent un lien de couple : concubin, Pacsé, conjoint... les prénoms sont des prénoms d'un même sexe et il y a un BI pour les deux membres du couple) où seulement un des membres du couple est répondant EFL (79 bulletins ont ainsi été créés). Les données relatives à la vie de couple et aux enfants sont issues des données du répondant. On leur affecte également les poids du répondant. Dans l'EFL l'ensemble des individus d'un ménage ont le même poids. Les réponses sur les parents, les petits enfants, les frères et sœurs et les logements sont mises en non-réponse. On peut repérer ces bulletins générés via leur CABEFLA qui commence par un 2 ou par la variable I\_CABBI qui prend la modalité « bulletin créé ».
- Éliminer des erreurs sur la variable couple (certains enfants vivant avec leur parent se déclarant en couple avec celui-ci ou inversement, certains répondant renseignant la date de naissance de leur colocataire frère ou sœur à la question 4...). Sont conservées en couple de même sexe les personnes qui dans les listes A,B,C indiquent un lien de couple (Pacsé, conjoint...) ou dans les cas où le lien n'est pas explicite quand les deux membres du couple se déclarent en couple dans le RP à la variable vivez-vous en couple. Les autres ne sont pas considérées comme étant en couple.
- Corriger quelques cas non détectés d'erreurs sur le sexe du conjoint.

### **Traitement particulier des enfants**

**Dans le cas des personnes en couple de même sexe avec enfant, les compteurs du nombre d'enfants du répondant et ceux de son conjoint ne sont pas nécessairement en cohérence avec le nombre d'enfants dans la table enfant.** En effet, un enfant déclaré par deux adultes répondant en couple de même sexe n'apparaîtra, in fine, qu'une fois dans la base enfant. Il sera associé à un unique CABEFLA. Il sera, en revanche, inclus dans les compteurs des deux adultes répondants, même pour celui dont le CABEFLA n'a pas été retenu dans le compteur Q11nb\_enf (et éventuellement Q11nb\_enf\_lgt) pour celui dont le CABEFLA est retenu et dans le compteur Q12nb\_enf\_C de l'autre. Il aura aussi un CABEFL\_CA rempli. Cet identifiant (CABEFL\_CA) permet, ainsi, de retrouver les données individuelles redressées du deuxième membre du couple. Toutes les données brutes sont, quant à elles, disponibles dans le FPR dans la base individu.

Si les enfants sont, le plus souvent, présents sur les bulletins des deux membres du couple, c'est à dire doublonnés, ils sont dans la grande majorité des cas déclarés en tant qu'enfant d'un seul membre du couple (à la fois en Q11enfant, Q11nb\_enf, Q12enfant\_c, Q12nb\_enf\_c et E13\_14part et E13\_14cpart). Pour éviter les doublons dans la table enfant on retient toute l'information du répondant parent que l'on garde dans son bulletin (CABEFL A). L'information du répondant non-parent n'est pas conservée si ce n'est le nombre d'enfant du conjoint.

Sur les 154 cas de personnes en couple de même sexe résidant avec des enfants :



- Dans 96 cas il n'y a pas d'ambiguïté pour le rattachement d'un enfant à un parent puisqu'une seule des personnes en couple se déclare parent au niveau des compteurs et du tableau des enfants. De plus, dans 80 de ces cas on a bien la symétrie entre le nombre d'enfants du conjoint du premier membre du couple et le nombre d'enfant du deuxième membre du couple, et inversement.
- Dans 14 cas il ne peut y avoir d'ambiguïté car la personne répondante est en couple non cohabitant (un seul répondant)
- Dans 16 cas une seule des personnes en couple cohabitant est répondante l'autre a été créée à partir des réponses du répondant (cf. point précédent : une seule personne en couple cohabitant dans le logement) : pas de doublons d'enfants dès le départ.
- Dans seulement 28 cas il y avait des doublons dans les compteurs et le tableau des enfants. Dans ce cas la priorité est donnée aux réponses « Êtes-vous sa mère (son père)? ». Quand un seul membre du couple répondait oui à cette question on lui affectait les enfants dans les compteurs ainsi que dans la table enfants. Il ne reste in fine que 8 cas pour lesquels cette règle ne fonctionnait pas. Pour ces 8 cas les enfants ont été affectés au hasard à l'un des deux membres du couple.

## **Résultats issus de ces apurements.**

On passe ainsi de l'estimation brute de 276 962 personnes en couple de même sexe ( 0,9% des personnes en couple) à 204 965 personnes (soit 0,6% des personnes en couple).

120 378 sont des hommes (59%) et 84 588 des femmes (41%).

173 076 (84%) se déclarent en couple cohabitant (dans un même logement si un de membres du couple se déclare en couple cohabitant et l'autre en couple non cohabitant nous avons conservé cette différence de perception).

90 443 sont pacsés (44%).

18 564 résident avec au moins un enfant (9%) dans 81% des cas se sont des femmes.

Ces résultats diffèrent légèrement en volume de ceux publiés dans l'Insee première N° 1435 de février 2013, Le couple dans tous ses états, *Non-cohabitation, conjoints de même sexe, Pacs...*

Par rapport à cette publication quelques apurements complémentaires ont été apportés. En particulier des bulletins EFL pour les conjoints des personnes en couple de même sexe cohabitant, avéré dans les listes A, B, C et avec un BI existant, ont été créés.



## Correction du module couple (question 4 à 10)

L'Ined a réalisé l'apurement des questions 4 à 10 en lien avec la variable apurée Q3couple.

### Cohérence entre les questions sur le conjoint (bloc conjoint), la rupture d'union (bloc séparation) et la variable Q3couple.

La question Q3 est une question filtre. Si l'enquêté répond 1 ou 2 il doit compléter le bloc conjoint et ne rien remplir dans le bloc séparation, si la réponse est 3 il doit compléter le bloc séparation et pas le bloc conjoint, si la réponse est 4 il ne doit remplir ni le bloc conjoint ni le bloc séparation.

A partir de la variable apurée Q3couple, on crée donc les variables corrigées de chaque bloc pour mettre à vide les réponses du bloc qui ne doit pas être traité :

- Les personnes en couple (Q3couple=1 ou 2) ne sont donc pas concernées par la question 10, les variables de la question 10 sont donc recodées à valeur manquante (correspondant à « non concernée »). Au sein du bloc conjoint, la non-réponse est codée « Z » pour les variables caractère et 9 pour les variables numériques et les dates (9, 99 ou 9999 selon la longueur des variables).
- Les personnes pas en couple (Q3couple=3 ou 4) ne sont donc pas concernées par le bloc conjoint, les variables de ce bloc sont donc à valeur manquante.
  - Pour les personnes en couple par le passé (Q3couple=3) : dans le bloc séparation, la non-réponse est codée « Z » pour les variables caractère et 9 pour les variables numériques et les dates (9, 99 ou 9999 selon la longueur des variables).
  - Les personnes jamais en couple (Q3couple=4) ne sont donc pas concernées par la question 10, les variables de la question 10 sont donc recodées à valeur manquante.

### Date de naissance du conjoint: Q4JNAI C, Q4MNAI C, Q4ANAI C (question 4)

#### **1- On repère les dates aberrantes et les étoiles («\*»)**

Les dates aberrantes sont définies ainsi, année de naissance (Q4ANAI\_C\_X) avant 1912 ou après 1992. Les dates comportant des étoiles sont les suivantes : "19\*\*", "19\*4", "19\*7", "19\*8", "192\*", "193\*", "194\*", "195\*", "196\*", "197\*", "198\*", "199\*"

#### **2- Correction des jours et mois de naissance aberrants**

Pour les cas où le jour de naissance=0 ou pour lesquels le mois de naissance= 0, 19 ou 24, on regarde ce qu'on a dans le BI du conjoint : on conserve alors l'information du BI.

#### **3- Correction des années de naissance aberrantes**

Pour les années de naissance inférieures à 1911 (enquêtés plus que centenaires) et supérieures à 1993 (moins de 18 ans atteints en 2011), on reprend l'année de naissance indiquée dans le BI. Malgré cette correction, il reste des cas aberrants.

#### **4- Correction des dates « étoilées »**

- Pour les cas où le chiffre des unités manque, on recode par 5. Ex : 192\* devient 1925.
- Pour les cas où une décennie manque, on recode par la décennie de naissance d'ego, en faisant l'hypothèse d'une homogamie selon la classe d'âge.



- Avec ces corrections, il reste un cas non corrigé, ANAI\_C\_I=19\*5, dont nous regardons l'image, il s'agit d'une erreur de saisie où ANAI\_C\_I est en réalité égal à 1989.

#### **5- Correction des années de naissance aberrantes**

- On convertit les variables intermédiaires ANAI\_C\_I et Q4ANAI\_C\_X (date de naissance d'ego) en variables numériques : ANAICJT et ANAIEGO (respectivement années de naissance du conjoint et d'ego).
- On crée la variable ECARTNAIS mesurant l'écart de naissance entre ego et son conjoint.
- On regarde les identifiants des personnes (157 cas) dont les conjoints sont codés comme étant nés avant 1911 ou après 1994.
- On code les années de naissance codées à blanc en 9999
- Après avoir regardé les 157 images au cas par cas, on établit que les dates aberrantes sont soit des erreurs de saisie, soit des erreurs de réponse avec des individus écrivant 1999 au lieu de 1969, soit des cas corrects.
- On corrige les erreurs de saisie avec les images
- On corrige les erreurs de réponse en recodant la mauvaise décennie par la décennie de naissance d'ego.
- Les cas corrects, jugés ainsi en raison d'une proximité chronologique avec la date de naissance du répondant (contrairement aux erreurs de réponse), sont maintenus tels quels en notant que la plupart des conjoints sont décédés (question 10) et que parfois il s'agit simplement d'un conjoint très âgé ou très jeune (1995).

#### **6- Correction du mois et du jour de naissance du conjoint**

Les mois de naissance « étoilés » ou aberrants sont: "0\*", "00", "1\*", "19", "24", "\*1", "\*6", "14". On regarde les identifiants des cas où le mois de naissance est aberrant. On recode ensuite, les jours et mois de naissance du conjoint, JNAICJT\_I et MNAICJT\_I (préalablement convertis en variables numériques-étape année de naissance du conjoint) qui sont à blanc en 99.

Après avoir regardé les images, nous notons que les cas où le jour de naissance est égal à « \*1 », « \*6 » ou « \*1 » sont en réalité des non-réponses, on les recode donc également en 99. Sur 7 bulletins avec des jours et mois de naissance défectueux, on a uniquement la date de naissance d'après les images, on recode donc les jours et mois en 99. Pour les cas où le mois de naissance est égal à « 14 », il s'agit de deux erreurs de saisie corrigées (mois de naissance=11 et 9). Enfin, pour les cas. On constate également une erreur de syntaxe, où la personne a écrit 1 au lieu de 01. Enfin, les trois derniers cas sont des inversions entre le jour et le mois de naissance (comme dans les formulaires anglo-saxons où le mois vient avant le jour). En corrigeant les mois de naissance des conjoints, nous avons éliminé les jours défilants, donc la dernière étape ne consiste qu'à rebaptiser les variables en Q4JNAI\_C, Q4MNAI\_C, Q4ANAI\_C.

#### **7- Utilisation des données RP pour la non-réponse.**

Dans les cas rares de non-réponse sur au moins un des éléments (jour, mois année) de la question 4 à l'EFL mais avec un appariement avec le RP la donnée brute RP est récupérée.

Finalement il reste 364 bulletins avec de la non-réponse sur l'année de naissance du conjoint.

### **Traitement des dates concernant la vie de couple**

**Les variables dates : Q6MOIS\_C, Q6ANNEE\_C, Q7MOISPACS, Q7ANNEE\_PACS, Q8MOIS\_MARI, Q8ANNEE\_MARI**

Pour les valeurs manquantes ou aberrantes, on essaie de récupérer les données dans le bloc.

#### **La date de mise en couple (question 6) :**

- On repère les données aberrantes, manquantes ou comportant une étoile



- On crée les variables C\_MOIS\_C et C\_ANNEE\_C corrigées à partir des variables d'origine.

#### Corrections :

Quand il n'y a pas de mois de mise en couple ou que celui ci comporte une étoile (\*) ou est hors bornes : si l'année de mise en couple est renseignée et égale à l'année de mariage, on copie le mois de mariage dans le mois de mise en couple. Si l'année de mise en couple est renseignée et égale à l'année de PACS, on copie le mois de PACS dans le mois de mise en couple.

Quand l'année de mise en couple commence par 18xx, on remplace automatiquement par 19xx.

Quand il y a une étoile (\*) dans la date de mise en couple et que le mois et année de mariage ou PACS sont renseignés, on récupère le chiffre manquant quand les autres sont égaux.

#### **La date de PACS (question 7) :**

- On repère les données aberrantes, manquantes ou comportant une étoile
- On crée les variables C\_MOIS\_PACS et C\_ANNEE\_PACS corrigées à partir des variables d'origine.

#### Corrections :

Si le mois de PACS n'est pas renseigné ou est inférieur à 12 et que l'année de PACS est 1999, on met automatiquement C\_MOIS\_PACS=12.

Si l'année de PACS comporte une étoile (\*) mais qu'il y a un mois de PACS égale au mois de mise en couple et que les chiffres de l'année de PACS sans étoile sont égaux à ceux de mise en couple, on copie l'année de mise en couple.

Si l'année de PACS=199\* on corrige 1999

Si le mois de PACS est égal à 0 ou supérieur à 12 on supprime l'information.

#### **La date de mariage (question 8) :**

- On repère les données aberrantes, manquantes ou comportant une étoile
- On crée les variables C\_MOIS\_MARI et C\_ANNEE\_MARI corrigées à partir des variables d'origine.

#### Corrections :

Si le mois de mariage n'est pas renseigné ou comporte une étoile ou est aberrant mais que le mois de mise en couple est renseigné et que les années sont identiques, on recopie le mois de mise en couple.

Si l'année de mariage comporte une étoile mais que les autres données sont renseignées et identiques à la date de mise en couple, on recopie l'année de mise en couple.

Quand l'année de mariage commence par 18xx, on remplace automatiquement par 19xx.

Après ces corrections, il reste quelques erreurs : un mois de mise en couple=0 et une année de mariage=2017. On recode, après un retour aux images du questionnaire, l'année de mariage en 2011 et le mois de mise en couple en 99 (non-réponse).

**Note :** Ces dates corrigées ont ensuite été retravaillées par la suite avec des calculs d'âge du répondant et du conjoint à la mise en couple, au PACS et au mariage.

### **Règles de codage des questions Q7Pacs, Q8Mari et Q10cause R**

On applique les règles de codage suivantes à la variable brutes Q7PACS\_X (question7 : « vous êtes vous pacés ensemble ? ») :



Règles de codage appliquées à la variable Q7PACS\_X, notre nouvelle variable intermédiaire s'appelle PACSENS

PACS \ PACSENS	Non-réponse	1 « oui »	2 « non »	8 « Incompatible »	9 « À classer »
00	X				
01			X		
03			X		
10		X			
12					X
13		X			
20					X
30		X			
31			X		
33				X	
32					X
23					X
22					X
11				X	

Nous faisons de même avec la variable Q8MARI\_X rebaptisée temporairement MARIAG (question 8), puis avec la variable Q10CAUSE\_R (question 10) rebaptisée temporairement ROMPUNION.

## **Correction des cohérences entre date de PACS et date de mariage**

Il doit exister une cohérence entre le mariage et le PACS. Une personne peut être pacsée et par la suite mariée avec le même conjoint, l'inverse n'étant pas possible. Pour corriger ces variables on regarde l'ensemble des questions 7 et 8. Il doit y avoir une cohérence entre les dates. Autrement dit, pour une personne déclarant par exemple être mariée et avoir été Pacsée, il faut que l'année de PACS soit antérieure à l'année de mariage, et que l'année de PACS soit valide (= > 1999).

Il faut recoder quelques non-réponses en réponses négatives, en effet, certaines personnes mariées ont bien répondu à la question sur le mariage mais pas du tout à celle sur le PACS, et les modalités de PACS et d'ANNEE\_PACS ont donc été recodées 9 et 9999 alors qu'au fond ce serait plutôt 2 pour non et valeur manquante pour « non concernée ».

### **1- on crée des variables indiquant si le mois et l'année de chaque événement sont renseignés**

Afin de pouvoir créer une variable concaténée, il nous fallait dans un premier temps savoir si les mois et les années étaient renseignés. Ainsi lorsque le mois ou l'année est renseigné, la nouvelle variable créée prend la valeur « 1 »; le cas échéant elle prend la valeur « 0 ».

Les variables sont les suivantes : Moispacs (mp), Anneepacs (ap), Moismari (mm), anneemari (am).

### **2- on crée une variable concaténée (indic4)**

La variable 'indic4' est la variable concaténée des variables suivantes : Pacsens+mariag+mois PACS+année PACS +mois mariage +année mariage.

La création de cette variable nous permet d'avoir une visibilité de tous les cas rencontrés pour le traitement de ce module, il semblait judicieux de créer cette variable qui permet un traitement plus spécifique et automatisé. Elle permet donc de prendre des décisions de corrections en voyant les incohérences pacs/mariage dans leur ensemble



indic4	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
000000	120074	33.38	120074	33.38
000001	448	0.12	120522	33.51
000010	15	0.00	120537	33.51
000011	3287	0.91	123824	34.43
000100	9	0.00	123833	34.43
000101	7	0.00	123840	34.43
000110	1	0.00	123841	34.43
000111	4	0.00	123845	34.43
001000	5	0.00	123850	34.43
001010	1	0.00	123851	34.43
001011	5	0.00	123856	34.43
001100	58	0.02	123914	34.45
001111	26	0.01	123940	34.46
010000	3077	0.86	127017	35.31
010001	2317	0.64	129334	35.96
010010	73	0.02	129407	35.98
010011	67578	18.79	196985	54.77
010100	21	0.01	197006	54.77
010101	12	0.00	197018	54.77
010110	4	0.00	197022	54.78
010111	42	0.01	197064	54.79
011000	15	0.00	197079	54.79
011010	2	0.00	197081	54.79
011011	127	0.04	197208	54.83
011100	133	0.04	197341	54.86
011101	3	0.00	197344	54.86
011110	2	0.00	197346	54.87
011111	247	0.07	197593	54.93
020000	771	0.21	198364	55.15
020001	8	0.00	198372	55.15
020010	2	0.00	198374	55.15
020011	46	0.01	198420	55.16
020100	13	0.00	198433	55.17
021000	1	0.00	198434	55.17
021100	82	0.02	198516	55.19
021110	1	0.00	198517	55.19
080000	4	0.00	198521	55.19
080001	2	0.00	198523	55.19
080011	8	0.00	198531	55.19
090000	338	0.09	198869	55.29
090001	3	0.00	198872	55.29
090011	31	0.01	198903	55.30
100000	153	0.04	199056	55.34
100001	2	0.00	199058	55.34
100010	1	0.00	199059	55.34
100011	17	0.00	199076	55.35
100100	70	0.02	199146	55.37
100101	1	0.00	199147	55.37
100111	8	0.00	199155	55.37
101000	5	0.00	199160	55.37
101011	1	0.00	199161	55.37
101100	1471	0.41	200632	55.78
101101	1	0.00	200633	55.78
101110	2	0.00	200635	55.78
101111	20	0.01	200655	55.79
110000	68	0.02	200723	55.80
110001	17	0.00	200740	55.81
110010	2	0.00	200742	55.81
110011	377	0.10	201119	55.91
110100	6	0.00	201125	55.92
110101	24	0.01	201149	55.92
110111	132	0.04	201281	55.96
111011	39	0.01	201320	55.97
111100	21	0.01	201341	55.98
111101	10	0.00	201351	55.98
111111	1717	0.48	203068	56.46
120000	301	0.08	203369	56.54
120001	2	0.00	203371	56.54
120011	10	0.00	203381	56.54
120100	197	0.05	203578	56.60
121000	29	0.01	203607	56.61
121100	8429	2.34	212036	58.95
121101	1	0.00	212037	58.95
121110	3	0.00	212040	58.95
121111	2	0.00	212042	58.95
181100	2	0.00	212044	58.95
181111	1	0.00	212045	58.95
190100	1	0.00	212046	58.95



191100	9	0.00	212055	58.95
191111	1	0.00	212056	58.96
200000	1194	0.33	213250	59.29
200001	117	0.03	213367	59.32
200010	4	0.00	213371	59.32
200011	1337	0.37	214708	59.69
200100	1	0.00	214709	59.69
200111	1	0.00	214710	59.69
201011	1	0.00	214711	59.69
201100	7	0.00	214718	59.70
201111	1	0.00	214719	59.70
210000	2523	0.70	217242	60.40
210001	1853	0.52	219095	60.91
210010	54	0.02	219149	60.93
210011	85588	23.79	304737	84.72
210100	11	0.00	304748	84.72
210101	6	0.00	304754	84.73
210110	2	0.00	304756	84.73
210111	38	0.01	304794	84.74
211000	6	0.00	304800	84.74
211001	2	0.00	304802	84.74
211010	1	0.00	304803	84.74
211011	156	0.04	304959	84.78
211100	147	0.04	305106	84.82
211101	1	0.00	305107	84.82
211111	162	0.05	305269	84.87
220000	51867	14.42	357136	99.29
220001	14	0.00	357150	99.29
220010	3	0.00	357153	99.29
220011	119	0.03	357272	99.33
220100	5	0.00	357277	99.33
221000	13	0.00	357290	99.33
221011	1	0.00	357291	99.33
221100	32	0.01	357323	99.34
280000	15	0.00	357338	99.35
280001	5	0.00	357343	99.35
280011	21	0.01	357364	99.35
290000	49	0.01	357413	99.37
290001	2	0.00	357415	99.37
290011	67	0.02	357482	99.39
291100	1	0.00	357483	99.39
800000	4	0.00	357487	99.39
800011	1	0.00	357488	99.39
801100	3	0.00	357491	99.39
810000	3	0.00	357494	99.39
810011	49	0.01	357543	99.40
810111	1	0.00	357544	99.40
811011	2	0.00	357546	99.40
811111	5	0.00	357551	99.41
820000	16	0.00	357567	99.41
820100	1	0.00	357568	99.41
821100	8	0.00	357576	99.41
880000	1	0.00	357577	99.41
881111	1	0.00	357578	99.41
890000	1	0.00	357579	99.41
900000	1096	0.30	358675	99.72
900001	1	0.00	358676	99.72
900011	13	0.00	358689	99.72
901000	1	0.00	358690	99.72
901100	1	0.00	358691	99.72
910000	28	0.01	358719	99.73
910001	11	0.00	358730	99.73
910011	247	0.07	358977	99.80
910111	1	0.00	358978	99.80
911100	1	0.00	358979	99.80
911111	1	0.00	358980	99.80
920000	28	0.01	359008	99.81
920100	1	0.00	359009	99.81
921100	8	0.00	359017	99.81
980000	1	0.00	359018	99.81
981100	1	0.00	359019	99.81
990000	654			

### 3- On crée les variables C\_PACS et C\_MARI selon les règles de décision suivantes :

- Si le cochage d'un événement et les dates associées sont cohérents on n'effectue aucune modification.



### **Non-réponse aux variables pacs et mari**

- Si on possède une date (soit sur le pacs soit sur le mariage): on met « oui » à l'événement concerné et non-réponse à l'autre.
- Lorsqu'on possède une année de PACS (et un mois), et que l'année de PACS est inférieure à 1999 on considère que l'enquêté n'est pas pacsé.
- Si seul un mois de PACS est renseigné: on choisit de mettre en non-réponse.
- Les cas où l'on possède un mois de PACS avec une date complète de mariage, on met le bloc PACS en non-réponse et on considère que la personne est bien mariée.

### **Enquêtés renseignant être uniquement Pacsés**

- Lorsque la personne coche uniquement qu'elle est Pacsée mais n'indique aucune date (ni mois, ni année) on met les dates en non-réponse ou « non concernées » selon c\_couple.
- Lorsque la personne indique une année et mois de mariage: on recopie la date de mariage à la date de PACS et on met non-réponse au mariage.
- La personne qui ne déclare qu'une année (ou un mois) de PACS lorsque celle-ci est correcte on considère bien que la personne est Pacsée, si cette date est incohérente on considère qu'elle n'est pas pacsée.
- Si il y a une date complète de PACS et une date complète de mariage renseignée : on considère, si la date de PACS est correcte, que la personne est bien Pacsée et non mariée, si ce sont exactement les mêmes dates et que la date de PACS est correcte, on décide de ne garder que le PACS.

### **Enquêtés renseignant être uniquement mariés**

- Les règles de décisions pour les cas ci-dessus (enquêtés qui cochent être seulement pacsés) sont adaptées aux enquêtés qui cochent seulement être mariés. Elles gardent la même logique de correction.

Quand il y a deux dates (une date de pacs et une date de mariage):

- Si la date du PACS est plus récente que la date de mariage : on considère que la personne n'est pas Pacsée.
- Si la date du PACS est inférieure à 1999 : on supprime l'information et on considère que la personne n'est pas Pacsée.
- Si la date de PACS est postérieure à 1999 et que la date de mariage est postérieure à la date de PACS : on garde les deux dates on considérant que la personne s'est Pacsée puis mariée

### **Enquêtés renseignant être Pacsés et être mariés (pacs=1 et mari=1)**

- Si les dates sont renseignées, il doit y avoir une chronologie. Il faut que la date de mariage soit après la date de PACS (et que le PACS soit égale ou supérieur à 1999). Si les dates sont cohérentes, on garde les deux événements. Sinon, on garde le mariage.
- S'il y a seulement une date de PACS, il faut que la date de PACS soit correcte. Si ce n'est pas le cas, on considère que la personne a voulu renseigner une année de mariage, on met donc « non » à PACS et on transfère la date de PACS dans celle du mariage si celle-ci était en non réponse.

### **Enquêtés renseignant ne pas être Pacsés et être mariés**

- Si le mois et l'année de mariage sont renseignées correctement, on garde les informations.
- Si l'enquêté indique une date de PACS, si la date de PACS est supérieure à 1999 et qu'elle est inférieure à la date de mariage on considère que la personne est Pacsée ; à l'inverse on considère bien que la personne n'est pas Pacsée et qu'elle est juste mariée.
- Si la personne ne renseigne aucune date de mariage mais une année (et un mois) de PACS : on choisit de reprendre la date de PACS pour la recopier dans la date de mariage (en considérant que la personne s'est trompée d'encadré)



### **Enquêtés renseignant ne pas être mariés et être Pacsés**

- Les règles sont les mêmes (à l'inverse) que pour les enquêtés déclarant ne pas être pacsés mais mariés.

### **Enquêtés renseignant ne pas être Pacsés et ne pas être mariés**

- Si la personne renseigne une date de mariage : on considère qu'elle est mariée
- Si la personne renseigne une date de PACS et que l'année est correcte : on considère qu'elle est Pacsée.

Malgré ce traitement, il reste 500 cas où C\_PACS=0. Ce sont tous des PACS=« 00 ». Y-a-t-il pour ces cas-là une date de PACS ? La majorité n'ont pas de date de pacs, donc on les recode comme 9, pour les cas où il y a bien une date indiquée pour le PACS, on regarde si il y a une donnée pour le mariage. S'il y a une date de pacs, on regarde s'il ne s'agit pas plutôt d'une date de mariage, si la personne n'a rien coché à mariage ou a coché « non » à mariage (la majorité), on recode pacs=1 sinon, on recode pacs=9. Malgré ces traitements, il reste encore quelques cas de c\_pacs et de c\_mari à 0, on recode ces cas-restants à 9.

De même, 1215 n'ont rien répondu à la question sur le mariage. On fait le même traitement que ci-dessus. La majorité d'entre elles n'ont mis aucune date de mariage non plus, donc on les recode en 9. Pour celles qui ont indiqué une date de mariage, on regarde s'il ne s'agit pas plutôt d'une date de PACS. Si la personne n'a rien coché à mariage ou a coché « non » à mariage (la majorité), on recode pacs=1, sinon, on recode pacs=9. Malgré ces traitements, il reste encore quelques cas de c\_pacs et de c\_mari à 0, on recode ces cas-restants à 9

Pour les cas où la personne est mariée avant 1999 on considère qu'elle n'est pas concernée par la question du PACS puisque le PACS existe depuis 1999 uniquement.

**Note** : Malgré toutes, ces corrections, nous avons découvert des incohérences lors de l'analyse des données. Des corrections au cas par cas ont été effectuées.

Après le traitement des différentes situations rencontrées, il n'existe plus de cas ambigus. Les dates de PACS et de mariage suivent une chronologie. Pour les cas où la personne est mariée avant 1999 on considère que ce n'est pas un PACS puisque le PACS existe depuis 1999.

#### **4- Traitement de la non-réponse via l'EAR.**

Dans le FPR les variables Q7pacs et Q8mari reprennent C\_Pacs et C\_mari avec un traitement supplémentaire effectué par l'Insee de la non-réponse. On peut repérer ces traitements à partir des indicatrices I\_Q7pacs et I\_Q8mari.

#### **Q8mari.**

Ainsi, pour 713 bulletins (I\_q8mari=3) l'utilisation des variables brutes du RP a permis de corriger 713 cas.

Si l'état matrimonial légal dans le recensement était marié la non-réponse à Q8mari est mise à 1 (le répondant et son conjoint sont mariés ensemble) sinon elle est mise à 2.

99 cas (I\_q8mari=4) proviennent de la même règle de décision à partir de la variable de diffusion du RP sur l'état matrimonial légal.



Q8mari		I_Q8mari					
Fréquence							
Pourcentage							
Pctage en ligne							
Pctage en col.	0	1	2	3	4	Total	
1	165488	123	9662	449	97	175819	
	68.28	0.05	3.99	0.19	0.04	72.54	
	94.12	0.07	5.50	0.26	0.06		
	72.92	60.00	67.03	62.97	97.98		
2	61451	82	4752	264	2	66551	
	25.35	0.03	1.96	0.11	0.00	27.46	
	92.34	0.12	7.14	0.40	0.00		
	27.08	40.00	32.97	37.03	2.02		
Total	226939	205	14414	713	99	242370	
	93.63	0.08	5.95	0.29	0.04	100.00	

Valeur(s) manquante(s) = 117400

### Q7Pacs.

La non-réponse à Q7Pacs est considérée comme du non, c'est-à-dire la personne n'est pas Pacsée. En effet ces effectifs pondérés sont alors cohérents avec ceux calculés en partant des résultats publiés dans l'Insee première N° 1336 - FÉVRIER 2011 Un million de pacsés début 2010 que l'on fait évoluer à partir des chiffres du ministère de la justice.

Ce traitement concerne 3968 bulletins pour le Pacs, soit 1,6% des personnes en couple.

Q7pacs		I_Q7pacs			
Fréquence					
Pourcentage					
Pctage en ligne					
Pctage en col.	0	1	2	Total	
1	12236	330	0	12566	
	5.05	0.14	0.00	5.18	
	97.37	2.63	0.00		
	7.81	0.40	0.00		
2	144511	81325	3968	229804	
	59.62	33.55	1.64	94.82	
	62.88	35.39	1.73		
	92.19	99.60	100.00		
Total	156747	81655	3968	242370	
	64.67	33.69	1.64	100.00	

Valeur(s) manquante(s) = 117400

**A partir de ces variables Q7pacs et Q8mari, on crée les modalités non-réponse (99 et 9999) et « non concernée » (valeur manquante) pour les variables Q7mois\_pacs, Q7annee\_pacs, Q8mois\_mari et Q8annee\_mari.**

Par exemple, si Q7pacs=2, Q7mois\_pacs= et Q7annee\_pacs= .



## Traitement du bloc rupture d'union (Q10CAUSE\_R, Q10ANNEE\_S, Q10ANNEE\_D)

### 1- On applique les règles de codage à la variable cause\_r

rompunion	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
0	289765	80.56	289765	80.56
1	38640	10.74	328405	91.30
2	28640	7.96	357045	99.26
9	2646	0.74	359691	100.00

### 2- On crée une variable concaténée (indicrupture)

Elle comporte la variable ROMPUNION (issue des règles de codage), ASEPAPAR (égale à 0 si il n'y a pas de date de séparation/ 1 si il y a une date) et ADECES (égale à 0 si il n'y a pas de date de décès/ 1 si il y a une date). Celle-ci nous permet de traiter le bloc dans sa globalité.

### 3- On crée la variable Q10CAUSE\_R, variable corrigée

#### Règles de décision :

- Si Q10CAUSE\_R\_X était non renseigné mais qu'il y a une date, on met Q10CAUSE\_R à 1 ou 2 selon la date renseignée (une date de séparation ou de décès).
- S'il y a 2 dates, une date de séparation et une date de décès :

Si une seule case est cochée, on garde la date liée au cochage : si la séparation est cochée, on garde la date de séparation et inversement pour le décès.

Si les deux cases sont cochées :

Si Q10ANNEE\_S= Q10ANNEE\_D, on garde le décès.

Si l'année de séparation est différente de l'année de décès, on garde la séparation. Soit la personne parle de deux unions et la séparation vient après le décès du conjoint d'une union précédente, soit c'est la même union et la séparation est la cause de rupture du couple, le décès vient après.

- Si les deux cases sont cochées et qu'il n'y a qu'une date on garde le cochage lié à la date.
- S'il y a une incohérence entre le cochage et l'information d'une date (8 cas dans le pilote), on garde le cochage et on déplace la date. Exemple, si l'enquêté a coché décès mais a indiqué une date de séparation, on laisse décès et on déplace la date.

### 4- On crée les variables Q10ANNEE\_S et Q10ANNEE\_D

**Note** : Ces dates corrigées ont ensuite été retravaillées par la suite avec des calculs d'âge du répondant à la séparation ou au décès du conjoint.

## Traitement de la variable TPS\_LOGR (question 9)

#### Cas simples :

On recode les cas biffés (où il n'y a que des 0 et des 2) en 999999. On regardera ensuite avec la variable q3couple si ce sont des non-réponses ou des « non concernées ».

Ex : 200002 → 999999

Tous les cas où seule la première case est cochée sont recodés en 1.

Ex : 100200 → 1

Pour les autres cas où une seule case est cochée, la modalité correspondant à cette case est retenue.

Ex : 001300 → 3



### Cas complexes :

Pour les cas où plusieurs cases sont cochées sans ambiguïté, on recode par les modalités correspondantes.

Ex : 013100 →24,  
311110 →2345.

La question 9 n'acceptant pas de réponses multiples, il faut recoder ces cas de réponses multiples dans des modalités uniques. Après discussion avec les concepteurs du questionnaire, nous avons pris la décision de recoder tous les cas multiples en 5 (« Une partie du temps, en alternance »). On rebaptise la variable corrigée, C\_TPS\_LOGR.

### **Traitement d'une partie de la non-réponse.**

La variable Q9tps\_logr reprend cette variable et intègre un traitement partiel de la non-réponse. Ainsi cette variable est renseignée à 1 dans le cas de couple cohabitant aussi bien dans l'EFL que d'après le mode de cohabitation du RP (MOCO=21 ou 22). Ce traitement permet de recoder 11 103 réponses en modalité 1 de la question 9. On repère ces bulletins à partir de I\_Q9TPS\_LOGR=2

#### Procédure FREQ

Q9tps_logr	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
1	231342	95.45	231342	95.45
2	2850	1.18	234192	96.62
3	438	0.18	234630	96.81
4	722	0.30	235352	97.10
5	2108	0.87	237460	97.97
6	3960	1.63	241420	99.61
Z	953	0.39	242373	100.00

Valeur(s) manquante(s) = 117397

I_Q9tps_logr	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
0	229321	94.61	229321	94.61
1	1949	0.80	231270	95.42
2	11103	4.58	242373	100.00

Valeur(s) manquante(s) = 117397



## Mise en cohérence des effectifs d'enfants déclarés (questions 11 et 12) avec les tableaux 13 et 14

Réalisée par l'Insee, la table **enfant** est créée à partir des tableaux 13 et 14 du questionnaire. Chaque enfant inscrit dans l'un des deux tableaux du questionnaire correspond alors à une ligne de cette table. On trouve dans cette table toutes les informations qui lui sont liées (date de naissance, lien de filiation avec le répondant ou le conjoint de celui-ci, temps de présence dans le logement, ...) ainsi que les informations issues des questions 11 et 12 sur le nombre d'enfants (total et présents dans le logement) du parent répondant et du conjoint de celui-ci.

Les enfants attendus dans ces tableaux sont :

- les enfants du répondant
- les enfants du conjoint du répondant s'ils sont présents dans le logement.

Il faut alors contrôler dans quelle mesure les informations diverses, dont on dispose concernant les enfants du répondant, ceux de son conjoint, et parfois d'autres enfants du logement sont cohérentes entre elles. Si ce n'est pas le cas, il faut alors apporter les corrections adéquates et compléter les variables manquantes qu'on est en mesure de remplir grâce aux autres données.

### Rappel de la signification des variables utilisées :

Variables dites de comptage ou compteurs d'enfants après apurement :

E11nb\_enf : nombre d'enfants du répondant

E11nb\_enf\_lgt : nombre d'enfants du répondant dans le logement

E12nb\_enf\_c : nombre d'enfants du conjoint (hors ceux en commun)

E12nb\_enf\_lgt\_c : nombre d'enfants du conjoint (hors ceux en commun) dans le logement

Ces variables sont suffixées par \_X

Variables à cocher des tableaux, permettant de définir la filiation des enfants mentionnés :

E13\_14PART : réponse à la question « Êtes-vous son père /mère ? ». Il s'agit de la variable finale, codée sur un caractère : 1=oui / 2=non.

Cette même variable dans sa forme initiale, issue de la saisie optique, s'appelle E13\_14PART\_X. Elle est codée sur deux caractères. Le premier correspond à la modalité « oui » et le second à la modalité « non ».

Ainsi E13\_14PART\_X='10' correspond à E13\_14PART=1, s'il n'y a pas eu de correction suite au repérage d'incohérence.

Cette variable se décline sur le même modèle pour la question « Son père/mère est-il votre conjoint actuel ? » : E13\_14CPART et E13\_14CPART\_X.

La variable LOCA, non présente initialement, indique si l'enfant apparaît dans le tableau T13 (présent dans le logement) ou dans le tableau T14 (hors logement). LOCA = « logemen » ou « horslog ».

La principale cohérence à vérifier est celle entre les compteurs des questions 11 et 12 (E11nb\_enf, E11nb\_enf\_lgt et E12nb\_enf\_c, E12nb\_enf\_lgt\_c) et les lignes des tableaux 13 et 14. Cette vérification se fait à plusieurs niveaux :

- Nombre total d'enfants du champ : cohérence globale entre le nombre total de lignes remplies et le nombre d'enfants du répondant (E11NB\_ENF\_X) plus le nombre d'enfants du conjoint dans le logement (E12NB\_ENF\_LGT\_C\_X) puisque les enfants du conjoint hors logement ne sont pas censés apparaître dans les tableaux.
- Répartition des enfants entre ceux vivant dans le logement et ceux ne vivant pas dans le logement : cohérence entre les compteurs logement (ie les nombres d'enfants dans le logement E11NB\_ENF\_LGT\_X et E12NB\_ENF\_LGT\_C\_X) et le nombre de lignes du tableau T13, d'une part, et le solde du nombre total d'enfants du répondant moins le nombre d'enfants du répondant dans le logement avec le nombre de lignes du tableau T14, d'autre part.
- Répartition des enfants entre égo et son conjoint : l'adéquation entre les compteurs d'égo (variables E11) et les lignes correspondant aux enfants d'égo (identifiées par la variable



E13\_14PART\_X=10 signifiant que l'enfant déclaré est considéré par égo, d'après son cochage initial, comme étant le sien) d'une part, et le compteur d'enfants du conjoint dans le logement (E12\_NB\_ENF\_LGT\_C\_X) et les lignes correspondant aux enfants du conjoint (identifiées par les variables E13\_14CPART\_X=10 et E13\_14PART\_X=01 : l'enfant déclaré n'est pas celui du répondant mais est celui de son conjoint), d'autre part.

Ces contrôles de cohérence vont conduire non seulement à modifier des variables mal renseignées ou non remplies mais aussi à éliminer des observations ou au contraire à en rajouter.

## **Apurements préliminaires**

### **- Traiter les cases à cocher pas claires**

Avant de pouvoir comparer les différents éléments, on procède à un nettoyage de premier niveau des variables à cocher E13\_14PART\_X et E13\_14CPART\_X.

Théoriquement, ces variables sont censées prendre trois valeurs, selon la case qui est cochée : '01', '10' ou '00' si la personne n'est pas concernée.

Lorsqu'en plus d'un cochage il y a un biffage (codé 2) ou une rature (codée 3), on met en place quelques règles de priorité. Ainsi si E13\_14PART\_X=12 alors on considère que la réponse est de type « 10 ».

### **- Vérifier que des enfants ne sont pas comptés deux fois :**

- d'une part, au sein d'un même questionnaire, il faut s'assurer qu'un même enfant n'est pas mentionné plusieurs fois, notamment dans le tableau 13 et dans le tableau 14. On observe plusieurs cas de figure :
  - Un enfant apparaît dans le premier tableau et dans le second tableau : même prénom et même année de naissance.
  - Deux enfants avec le même prénom apparaissent, mais les dates de naissance sont différentes, et on est dans le cas d'une famille recomposée : il faut alors s'assurer que les deux enfants ne sont pas frères et sœurs. (il y a en effet quelques cas - rares - ou deux enfants ayant le même prénom sont mentionnés. L'un étant l'enfant du répondant et l'autre celui du conjoint).
  - Le prénom ou l'année de naissance des enfants ne sont pas renseignés, ce qui rend l'identification des doublons impossibles en direct, mais la mise en cohérence avec les compteurs laissent à penser qu'il s'agit d'enfants doublonnés.

En pratique, on repère des suspicions de doublons en cherchant les lignes d'enfants ayant le même prénom et/ou la même année de naissance. On va les considérer comme doublons uniquement si la cohérence globale entre les compteurs et les lignes des tableaux n'est pas vérifiée et si le fait de ne conserver qu'une seule occurrence permet de retrouver les informations données par les compteurs. Autrement dit, si on a, par exemple, deux uniques observations enfant, l'une dans le T13 et l'autre dans le T14, avec E11NB\_ENF\_X=2 et E11NB\_ENF\_LGT\_X=1 (et E12NB\_ENF\_LGT\_C\_X=0), alors, même s'ils sont codés comme doublons potentiels, les deux occurrences « suspectes » seront conservées puisqu'en accord avec les informations des compteurs.

- d'autre part, au sein d'un même logement, lorsqu'il y a plusieurs individus répondants. Les enfants qui doivent apparaître sont ceux du répondant et du conjoint de celui-ci, donc, à part dans le cas de parents en couple homosexuels (voir plus bas, point sur ce sujet et la partie traitements des couple de même sexe), un enfant ne devrait théoriquement pas apparaître dans deux questionnaires. Dans la plupart des cas, il s'agit de jeunes majeurs répondants qui inscrivent dans les tableaux leurs frères et sœurs lorsqu'ils habitent avec eux chez leurs parents.

Une fois les enregistrements doublonnés repérés, il s'agit de définir celui qu'on conserve et celui (ceux) qu'on élimine. La grande majorité des cas de doublons d'enfants correspond à deux situations :

- l'enfant vit une partie du temps chez le parent déclarant, une partie du temps ailleurs et le répondant a voulu donner le maximum d'informations en l'inscrivant aussi dans le tableau T14 afin de renseigner les informations relatives à son deuxième lieu de résidence



- l'enfant ne vit plus chez son parent mais le répondant l'a inscrit dans le tableau 13 avant de se rendre compte qu'il fallait le mettre dans le tableau 14. Les deux lignes sont alors remplies.

Le repérage d'enfant doublonné est une indication pour l'étape suivante, à savoir la mise en cohérence entre les compteurs et les enfants présents dans les tableaux. Lorsqu'on a l'adéquation entre ces informations, on n'élimine pas d'enfants considérés comme « doublons potentiels ».

## **Mise en cohérence entre les compteurs et les enfants présents dans les tableaux**

On dispose de deux types d'informations sur le nombre d'enfants :

- celles qu'on récupère grâce au remplissage des tableaux 13 et 14 et aux variables E13\_14PART\_X et E13\_14\_CPART\_X

- celles qu'on obtient directement à partir des compteurs d'effectifs remplis : E11NB\_ENF, E11NB\_ENF\_LGT, E11NB\_ENF\_C, E11NB\_ENF\_LGT\_C et en complément E11ENFAN et E11ENFAN\_C

Une même information apparaît ainsi sous deux formes dans la base. Pour comparer ces deux types de données, on synthétise les informations relatives aux enfants au niveau du répondant. On cherche à estimer dans quelle mesure ces informations de nature différente se recoupent ou au contraire présentent des décalages. Ces contrôles se font par étapes :

### A- cohérence sur le nombre d'enfants total

Le nombre d'enfants total apparaissant dans les tableaux doit correspondre à la somme du nombre d'enfants total du répondant (E11NB\_ENF) et du nombre d'enfants dans le logement du conjoint éventuel de celui-ci (E11NB\_ENF\_LGT\_C). Lorsque l'égalité n'est pas respectée, deux cas de figures existent alors :

- il y a plus d'enfants dans les tableaux que ce qui est attendu d'après les compteurs. On trouve notamment qu'un certain nombre d'enfants présents qu'une partie du temps dans le logement ne sont pas inclus au départ dans les compteurs logements et figurent toutefois dans le T13. En effet, la question relative au temps de présence passé dans le logement (E13\_TPSLGT) conduit le répondant à inclure dans le T13 des enfants qui ne sont présents qu'une petite partie du temps, mais qu'il avait initialement jugés comme ne faisant pas partie du logement (malgré l'indication en petit sous le compteur logement). On vérifie notamment que les enfants apparaissant dans les tableaux sont bien enfants du répondant ou du conjoint et qu'il n'y a pas de doublon. On s'assure aussi que les enfants du conjoint ne figurent pas dans le T14.
- il y a moins d'enfants dans les tableaux que ce qui est attendu d'après les compteurs : on vérifie notamment que les enfants communs du répondant et de son conjoint n'ont pas été comptés en double dans les compteurs de l'un et de l'autre. En effet, dans un certain nombre de bulletins, on constate que les enfants eus en commun avec le conjoint sont inclus dans les compteurs du répondant ET dans ceux de son conjoint. Parfois aussi, les enfants du conjoint sont inclus dans le compteur du répondant car il les considère sûrement un peu comme les siens, surtout s'il les a élevés.

Si l'incohérence persiste, on fait plutôt confiance aux lignes des tableaux, en se disant qu'il y a plus de chance qu'il y ait une erreur de saisie des compteurs qu'un oubli d'enfant ou un enfant inventé. Cependant, si les tableaux sont complètement vides alors que des enfants y sont attendus d'après les compteurs, on suppose qu'il s'agit de non-réponse. On va alors s'appuyer sur les données de l'EAR et des BI remplis pour corriger la non-réponse dans le tableau 13. On crée aussi, si besoin, des lignes dans le tableau 14, en ne remplissant que les variables LOCA (qui indique si l'enfant réside dans le logement ou non) et E13\_14PART et E13\_14CPART.



B- une fois le nombre d'enfants global mis en cohérence entre les tableaux et les compteurs, on s'assure qu'ils sont correctement distribués entre le T13 et le T14, au niveau global :  $E11NB\_ENF\_LGT+E12NB\_ENF\_LGT\_C$ =nombre d'enfants apparaissant dans le T13 et  $E11NB\_ENF - E11NB\_ENF\_LGT$ =nombre de lignes dans le T14.

C- enfin, on vérifie la bonne répartition entre les enfants du répondant et ceux de son conjoint qui peut conduire à la modification des variables E13\_14PART et E13\_14CPART ou des compteurs : Les deux variables sont abordées dans leur globalité pour s'assurer de la cohérence avec les compteurs. On distingue trois types d'enfants : ceux communs entre le répondant et son conjoint, ceux du seul répondant, et ceux du seul conjoint. Cela nécessite en amont d'avoir « rendu propres » ces deux variables.

La modification de ces variables se fait d'après plusieurs critères :

- rendre la réponse « propre » (certains cochages initiaux sont incompatibles ; le répondant ne peut, par exemple, pas être à la fois le parent et ne pas être le parent)
- faire en sorte que le nombre d'enfants du répondant (identifiés par  $E13\_14PART=1$ ) et le nombre d'enfants du conjoint ( $E13\_14PART=2$  et  $E13\_14CPART=1$ ) soient compatibles avec les compteurs.
- La cohérence avec la variable Q3couple qui précise la situation de couple du répondant. En effet, on ne peut avoir d'enfants déclarés comme enfants « du répondant et de son conjoint » alors que le répondant déclare ne pas vivre en couple.
- La cohérence entre l'année de naissance de l'enfant et l'année de naissance du déclarant. Si l'écart d'âge est trop faible (ou trop grand avec la mère), voire négatif, on suppose qu'il s'agit soit d'un problème de remplissage des années de naissance de l'un ou l'autre des individus (répondant ou enfant), soit d'un lien de filiation qui n'a pas lieu d'être. La mise en adéquation avec les compteurs permet généralement de trancher (entre doit on supprimer la ligne qui ne correspond pas à un enfant du déclarant ou la laisser en modifiant une des années de naissance ?).
- La cohérence avec l'année de mise en couple : un enfant n'est, a priori, sauf cas particulier, pas enfant du conjoint du répondant si la mise en couple a eu lieu après la naissance de l'enfant.
- La cohérence entre les années de naissance des enfants : sauf cas particulier, un enfant né après un frère/sœur, enfant des deux membres du couple, sera lui aussi enfant des deux membres du couple.

Si certaines conditions doivent impérativement être vérifiées, comme le fait qu'un enfant ne peut pas être enfant du conjoint du répondant si le répondant n'est pas en couple, d'autres sont plus des critères pour trancher quand on hésite entre deux situations mais ne nécessitent pas une correction de la variable si elle est cohérente, par ailleurs, avec les autres contraintes et qu'elle est sans ambiguïté au départ. C'est le cas par exemple du critère de cohérence sur les années de naissance des enfants.

Par ailleurs, les comparaisons de dates doivent se faire avec prudence car celles-ci ne sont pas toujours correctement saisies. Afin de renforcer les critères faisant appel aux années, on vérifie dans la mesure du possible, qu'il y a adéquation entre les informations de l'EFL et celles issues de l'EAR.

## **Imputations et indicateurs de « modification » des variables remplies**

A - Lignes enfants rajoutées et lignes enfants supprimées

Dans la table enfant, on dénombre **609 398** observations, dont **46 409** non présentes initialement et rajoutées au vu des informations annexes (compteurs et données de l'EAR). Par ailleurs, **18 895** observations ont été supprimées de la table initiale (doublons ou enfants ne correspondant pas au champ demandé).



Les imputations sont de deux natures différentes :

- 5 179 correspondent à des informations tirées de l'EAR. Il s'agit d'enfants, tous vivant dans le logement, estimés comme manquants dans les tableaux T13 et T14 au vu des compteurs et qu'on a effectivement retrouvés côté EAR. Dans certains cas les compteurs sont vides mais la variable indiquant que le répondant ou son conjoint a eu des enfants est cochée à « oui ». Les lignes ont été rajoutées en récupérant les informations disponibles dans l'EAR (notamment l'année de naissance et les liens de filiations).
- La grande majorité des lignes enfants imputées (41 230 observations), pour l'essentiel imputées dans le tableau 14, ont été rajoutées sans appui sur l'EAR. Cela a été fait lorsque :
  - ➔ Le tableau 14 est complètement vide alors que la différence entre les compteurs nombre d'enfants total du répondant et nombre d'enfants dans le logement (ce dernier étant cohérent avec ce qu'on trouve dans le T13) est positif. On estime alors que le T14 est sujet à une non-réponse totale. On va donc créer les lignes manquantes. A ce stade, on n'« invente » aucune information puisque seules les variables `loca='horslog'` et `E13_14PART='1'` sont remplies.
  - ➔ Le tableau 14 est « saturé », c'est à dire que les 7 lignes mises à disposition du répondant ont été remplies alors que le différentiel des compteurs indique qu'il y a plus de 7 enfants hors logement. On suppose alors qu'il y a eu un oubli de remplissage d'un bulletin suite<sup>6</sup> (ou un problème d'appariement). Cette hypothèse est confortée par le fait qu'on observe un petit pic à 7 dans la répartition des personnes selon le nombre de lignes d'enfants dans le tableau 14. On décide alors de rajouter des lignes d'enfants pour compenser ce décalage.
  - ➔ Les répondants avaient indiqué (initialement et de façon claire) être grands-parents. Pour l'ensemble de ces personnes, dès lors qu'elles avaient au moins 36 ans, on a modifié `E11NB_ENF` (tout en laissant à 0 le compteur `E11NB_ENF_LGT` ainsi que les compteurs relatifs aux enfants du conjoint). On a d'abord regardé par tranches d'âge quinquennales le nombre d'enfants moyen des personnes ayant répondu à la question. On a ensuite, pour chacun des individus concernés et selon les tranches d'âge, affecté à `E11NB_ENF` la valeur de l'entier le plus proche (en pratique 1, 2 ou 3). Pour finir, on a rempli les informations `loca='horslog'` (on a en effet supposé que si l'individu était grand-parent, ses enfants ne vivaient plus avec lui) ainsi que `E13_14PART='10'`. Cette opération a conduit à modifier 1030 bulletins EFL et à rajouter 2684 lignes enfants.

## B - Indicatrice de modification des compteurs des questions 11 et 12

Les indicatrices de modification des compteurs prennent trois modalités (variables non diffusées, mais reconstituables par comparaison entre les données brutes et redressées) :

- 0 : pas de modification
- 1 : la valeur initiale, manquante, a été remplie
- 2 : la valeur initiale a été modifiée.

Champ des enfants :

	I_E11NB_ENF	I_E11NB_ENF_LGT	I_E12NB_ENF_C	I_E12NB_ENF_LGT_C
0 : Identique	92,6	57,4	33,6	30,2
1 : Vide	4,2	40,7	65,5	68,8
2 : Modifié	3,2	1,9	0,9	1,0
	100	100	100	100

Lecture : 92,6% des enfants de la table finale apparaissent sur un bulletin où le compteur du nombre d'enfant du répondant n'a pas été changé par rapport à sa valeur initiale.

<sup>6</sup> Quand le répondant avait plus de 7 enfants ayant quitté son domicile, il devait remplir dans un autre bulletin EFL les enfants qui ne pouvaient pas tenir dans les 7 lignes du tableau 14. Ce deuxième bulletin qui ne contient que les informations sur les enfants supplémentaires s'appelle un bulletin suite.



On constate un fort taux de non-remplissage des compteurs initiaux autres que celui du nombre d'enfant total du répondant (première colonne des tableaux). Ce non-remplissage est particulièrement important pour les compteurs des enfants du conjoint (I\_E11NB\_ENF\_C et I\_E12NB\_ENF\_LGT\_C). Dans la plupart des cas, cette non-réponse initiale se corrige facilement car elle correspond à un effectif nul.

Au niveau de l'ensemble des compteurs, on crée un indicateur synthétique I\_E11\_12ENF de modification des compteurs. On dénombre alors :

- 16 % d'enfants qui n'ont subi aucun changement de compteur.
- 78 % d'enfants pour lesquels au moins un compteur a été rempli car vide initialement, et dont les autres n'ont pas été modifiés.
- 2 % d'enfants pour lesquels les compteurs d'enfants globaux (E11\_NB\_ENF et E12\_NB\_ENF\_C) n'ont pas été modifiés, contrairement aux compteurs d'enfants dans le logement (E11\_NB\_ENF\_LGT et E12\_NB\_ENF\_LGT\_C).
- et enfin, 4 % d'enfants pour lesquels les compteurs globaux NB\_ENF et E12\_NB\_ENF\_C ont été modifiés.

Si la part des enfants avec au moins un compteur vide au départ est si élevée c'est que la plupart du temps, la partie concernant le conjoint est vide lorsque celui-ci n'a pas d'enfant issu d'une précédente union (ce qui se corrige très facilement par un 0).

On crée aussi un indicateur synthétique I13\_14parent, de modification des variables E13\_14PART et E13\_14CPART, traitant conjointement les deux variables.

6 % des enfants sont dans le cas d'une variable E13\_14PART\_X !! E13\_14CPART\_X non propre initialement mais corrigeable sans difficulté, par exemple du type : '0112', c'est-à-dire avec deux cochages bien identifiés et un biffage sur la dernière case. On interprète alors ce codage comme '0110' : l'enfant est celui du conjoint mais pas du répondant lui-même. On inclut aussi dans cette catégorie, les situations où le répondant dit que l'autre parent de l'enfant est le conjoint actuel (E13\_14CPART\_X =10) alors que le répondant n'est pas en couple. Il s'agit souvent d'union rompue relativement récemment et/ou de personnes âgées). On compte 83 % d'enfants pour lesquels les deux variables E13\_14PART et E13\_14CPART n'ont pas été modifiées (inclus dans cette catégorie les cas où le répondant dit que l'autre parent de l'enfant n'est pas le conjoint actuel (E13\_14CPART\_X =01) alors qu'il n'est pas en couple, donc « non concernée » par cette question. On considère alors que E13\_14CPART\_X=00.

## Enfants de couples de même sexe

Contrairement aux autres enfants, les enfants de couples de même sexe peuvent apparaître sur deux bulletins, lorsque les deux membres du couple sont répondants. On ne conserve cependant qu'une observation pour un enfant donné dans la table enfant.

Néanmoins, pour ces enfants, lorsque le deuxième parent (ou beau-parent) est répondant et habite dans le même logement (couple cohabitant), on introduit, en plus des variables CABEFLA et CABBIA correspondant aux identifiants EFL et EAR du parent répondant, une variable CABEFLA\_C. Cette variable correspond au CABEFLA du bulletin du deuxième parent répondant (et donc de même sexe) de l'enfant s'il est présent dans le logement. Pour les autres enfants, c'est-à-dire ceux vivant avec un seul parent ou avec deux parents de sexe différent, cette variable CABEFLA\_C est vide. Pour tous les enfants dont le parent répondant est en couple cohabitant, que ce soit ou non avec une personne de même sexe, on a CABBIA\_C correspondant au cabbi (de l'EAR) du conjoint du (beau-)parent répondant.

Un enfant déclaré par deux adultes répondants, en couple de même sexe sera, en revanche, inclus dans les compteurs **des deux** adultes répondants, même pour celui dont le CABEFLA n'a pas été retenu.

Pour ces personnes en couple de même sexe avec enfants, on ne trouve pas de lignes dans la table enfant associées au CABEFLA du 2<sup>ème</sup> (beau-)parent, même si les compteurs logements de ce répondant sont positifs.



Enfin, un travail de validation au cas par cas a été fait, en s'appuyant sur l'ensemble des données de l'EFL, de l'EAR et des listes ABC de l'EAR, afin de s'assurer de la fiabilité des données relatives aux enfants des couples de même sexe (cf. la partie traitements sur les couples de même sexe).

### **Remontée des informations de la table enfant à la table individu.**

Les variables E11 et E12 relatives aux nombres d'enfants sont récupérées dans les variables correspondantes Q11 et Q12 de la table individu.

Quand il n'y a pas d'observation dans la table enfant :

- les variables Q11 de la table individu indiquent que le répondant n'a pas d'enfant (y compris en cas de non-réponse initiale).
- la variable (Q12nb\_enf\_lgt\_C) sur le nombre d'enfants du conjoint dans le logement est, comme pour le répondant, égale à 0 (y compris en cas de non-réponse initiale).

La variable (Q12nb\_enf\_C) sur le nombre total d'enfant du conjoint, pour laquelle on ne dispose pas d'information auxiliaire de vérification, correspond à la valeur de la variable brute Q12nb\_enf\_c\_X. On vérifie simplement qu'elle est supérieure à Q12nb\_enf\_lgt\_C.

Quand cette variable n'est pas nulle Q12enfan\_C est égale à 1, et à 0 sinon.



## Apurements des caractéristiques des enfants des tableaux 13 et 14.

Une fois la table enfant réalisé par l'Insee, l'Ined a apuré les caractéristiques des enfants de cette table.

### Le sexe de l'enfant E13 14sexe

1<sup>ère</sup> étape : recodage de la variable brute :

Si E13\_14sexe\_X prend les valeurs "01", "03", "21", "31", "2", "02", "23" on recode E13\_14sexe à 2.

Si E13\_14sexe\_X prend les valeurs "10", "30", "12", "13", "1", "20", "32" on recode E13\_14sexe à 1.

Si E13\_14sexe\_X vaut "00", " " on recode E13\_14sexe en Z indiquant la non réponse.

Les autres cas sont codés en inconnu à ce stade.

2<sup>ème</sup> étape : utilisation d'une liste de prénoms, pour les prénoms sans ambiguïté uniquement, pour coder en 1 ou 2 une partie de la non-réponse et des cas inconnus. Le sexe à partir du prénom n'a pas été utilisé dans les cas où l'appariement enfant permettait de récupérer le sexe sur le BI.

3<sup>ème</sup> étape : utilisation des données brutes du RP des enfants appariés pour corriger une partie de la non réponse résiduelle et des cas inconnus.

4<sup>ème</sup> étape : les cas restant inconnus sont considérés comme de la non réponse.

Résultats :	E13_14sexe	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
	1	291419	47.82	291419	47.82
	2	275541	45.22	566960	93.04
	Z	42438	6.96	609398	100.00

### Année de naissance de l'enfant E13 14anai.

1<sup>ère</sup> étape : On récupère automatiquement l'année de naissance de l'enfant à partir de la variable brute du RP quand elle est en non-réponse dans l'EFL, avec une étoile ou inférieure à 1900.

2<sup>ème</sup> étape : si la date de naissance est différente entre le RP et l'EFL et que l'une des sources seulement indique que l'enfant a moins de 4 ans et que l'enfant est présent dans le tableau 28 on retient l'année comprise entre 2007 et 2011.

3<sup>ème</sup> étape : si la date de naissance est différente entre le RP et l'EFL on retient l'année qui permet de rendre cohérent l'écart d'âge avec le ou les parents. Si les deux années sont cohérentes, on garde l'année EFL.

### Temps passé dans le logement : E13tpsigt et autres logements : E13aut lgt E13lgt aut par, E13lgt etude et E13lgt autre.

Les enfants du tableau 14 ne sont pas concernés par ces questions. Pour ces enfants ces variables sont de ce fait en valeur manquante.

1<sup>ère</sup> étape : recodage de la variable E13tpsigt\_X brute :

Si E13tpsigt\_X prend les valeurs "001", "301", "031", "201", "021", "003", "002", "203", "221", "223", "231", "331" on recode E13tpsigt à 3.



Si E13tpsigt\_X prend les valeurs "010","310","013","210","012","030","020","232","312" on recode E13tpsigt à 2.

Si E13tpsigt\_X prend les valeurs "100","130","103","120","102","300","200","122","133","320" on recode E13tpsigt à 1.

Si E13tpsigt\_X vaut "000"," " on recode E13tpsigt en Z indiquant la non réponse.

Les autres cas sont codés en inconnu à ce stade.

2<sup>ème</sup> étape : recodage de la variable brute E13autlgt\_X :

La variable brute E13autlgt\_X peut prendre plusieurs réponses possibles si l'enfant cohabitant vit dans plusieurs autres logements. Afin de prendre en compte cette possibilité cette variable brute conduit à la création de 4 variables apurées. Une première variable E13aut\_lgt indiquant si oui ou non l'enfant cohabitant vit aussi dans un autre logement. Puis, dans le cas où cette variable indique que oui (E13aut\_lgt=1), 3 variables renseignant la nature de ces autres logements : E13lgt\_aut\_par si c'est chez son autre parent E13lgt\_etude si c'est pour ses études et E13lgt\_autre quand il vit ailleurs pour une autre raison.

Si E13autlgt\_X prend les valeurs "1000","1300","1030","1003","1200","1020","1002","3000","2000","1222","1023" alors :

- E13aut\_lgt vaut 2 indiquant que l'enfant ne vit pas dans un autre logement.
- E13lgt\_aut\_par, E13lgt\_etude et E13lgt\_autre sont en valeurs manquantes pour « non concernée ».

Si E13autlgt\_X prend les valeurs "0100","3100","0130","0103","2100","0120","0102","0300","0200" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 1.
- E13lgt\_etude vaut 2.
- E13lgt\_autre vaut 2.

Si E13autlgt\_X prend les valeurs "0010","3010","0310","0013","2010","0210","0012","0031","0020","0030","0313" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 2.
- E13lgt\_etude vaut 1.
- E13lgt\_autre vaut 2.

Si E13autlgt\_X prend les valeurs "0001","3001","0301","0031","2001","0201","0021","0003","0002","0203","0221","0223" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 2.
- E13lgt\_etude vaut 2.
- E13lgt\_autre vaut 1.

Si E13autlgt\_X prend les valeurs "0111","0222" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 1.
- E13lgt\_etude vaut 1.
- E13lgt\_autre vaut 1.

Si E13autlgt\_X prend les valeurs "0011","0022","0033" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 2.
- E13lgt\_etude vaut 1.
- E13lgt\_autre vaut 1.

Si E13autlgt\_X prend les valeurs "0101","0202","0303" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 1.
- E13lgt\_etude vaut 2.



- E13lgt\_autre vaut 1.

Si E13autlgt\_X prend les valeurs "0110","0112","0113","0220" alors :

- E13aut\_lgt vaut 1 indiquant que l'enfant vit dans un autre logement.
- E13lgt\_aut\_par vaut 1.
- E13lgt\_etude vaut 1.
- E13lgt\_autre vaut 2.

Si E13autlgt\_X vaut "0000", " " on code E13aut\_lgt, E13lgt\_aut\_par, E13lgt\_etude et E13lgt\_autre en Z indiquant la non-réponse.

Dans les autres cas ces 4 variables sont codées en inconnu à ce stade.

3<sup>ème</sup> étape : utilisation de la réponse à la question sur les autres logements pour corriger la réponse au temps passé dans le logement.

Si la variable E13tpslgt est codée à l'issue de la première étape en non-réponse (« Z ») ou en réponse inconnue et que E13aut\_lgt indique qu'il n'y a pas d'autre logement (E13aut\_lgt=2) alors E13tpslgt vaut 1 c'est à dire que l'enfant vit tout le temps dans ce logement.

Si la variable E13tpslgt est codée à l'issue de la première étape en réponse inconnue et que E13aut\_lgt indique qu'il y a un autre logement (E13aut\_lgt=1) alors :

- E13tpslgt vaut 2 si E13tpslgt\_X prend l'une des valeurs suivantes "020", "220", "330", "110", "112"
- E13tpslgt vaut 3 si E13tpslgt\_X prend l'une des valeurs suivantes "002","222","331", "011","101","111", "202"
- E13tpslgt vaut 4 c'est à dire une partie du temps seulement, sans précision dans les autres cas.

4<sup>ème</sup> étape : utilisation de la réponse à la question sur le temps passé dans le logement pour corriger la réponse à la question sur les éventuels autres logements.

Si la variable E13aut\_lgt est codée à l'issue de la deuxième étape en non-réponse (« Z ») ou en réponse inconnue et que la variable E13tpslgt indique que l'enfant vit tout le temps dans le logement (E13tpslgt=1) alors E13aut\_lgt indique qu'il n'y a pas d'autre logement (E13aut\_lgt=2) et les 3 variables sur la nature d'un éventuel autre logement sont mises en valeur manquante pour « non concernée ».

Si la variable E13aut\_lgt est codée à l'issue de la deuxième étape en non-réponse (« Z ») et que la variable E13tpslgt n'indique pas que l'enfant vit tout le temps dans le logement (E13tpslgt différent de 1) alors E13aut\_lgt indique qu'il y a un autre logement (E13aut\_lgt=1) et les 3 variables sur la nature d'un éventuel autre logement sont mises en non-réponse c'est à dire codées « Z ».

Si la variable E13aut\_lgt est codée à l'issue de la deuxième étape réponse inconnue et que la variable E13tpslgt n'indique pas que l'enfant vit tout le temps dans le logement (E13tpslgt différent de 1) alors E13aut\_lgt indique qu'il y a un autre logement (E13aut\_lgt=1) et :

- Si E13autlgt\_X prend l'une des valeurs suivantes "1100","2320", "2300","2200" alors E13lgt\_aut\_par vaut 1, E13lgt\_etude vaut 2 et E13lgt\_autre vaut 2.
- Si E13autlgt\_X prend l'une des valeurs suivantes "1210", "1012", "1010", "2030", "2020" alors E13lgt\_aut\_par vaut 2, E13lgt\_etude vaut 1 et E13lgt\_autre vaut 2.
- Si E13autlgt\_X prend l'une des valeurs suivantes "2021","1001", "1031", "2221","2223", "2203", "2023","2002" alors E13lgt\_aut\_par vaut 2, E13lgt\_etude vaut 2 et E13lgt\_autre vaut 1.
- Si E13autlgt\_X prend l'une des valeurs suivantes "2222" alors E13lgt\_aut\_par vaut 1, E13lgt\_etude vaut 1 et E13lgt\_autre vaut 1.
- Si E13autlgt\_X prend l'une des valeurs suivantes "3011","2022" alors E13lgt\_aut\_par vaut 2, E13lgt\_etude vaut 1 et E13lgt\_autre vaut 1.
- Si E13autlgt\_X prend l'une des valeurs suivantes "2202" alors E13lgt\_aut\_par vaut 1, E13lgt\_etude vaut 2 et E13lgt\_autre vaut 1.
- Si E13autlgt\_X prend l'une des valeurs suivantes "3110", "2112","2220" alors E13lgt\_aut\_par vaut 1, E13lgt\_etude vaut 1 et E13lgt\_autre vaut 2.



5<sup>ème</sup> étape : mise en cohérence finale entre les informations sur le temps passé dans le logement et un éventuel autre logement.

Les cas résiduels de réponse inconnue sont mis en non-réponse pour les 5 variables (E13tpsigt, E13aut\_lgt, E13lgt\_aut\_pat, E13lgt\_etude et E13lgt\_autre). En cas d'incohérence entre E13tpsigt et E13aut\_lgt on considère que la réponse à E13aut\_lgt l'emporte c'est à dire :

- Si le répondant a répondu que l'enfant vit tout le temps dans le logement mais qu'il déclare aussi qu'il vit ailleurs on considère que l'enfant ne vit pas tout le temps dans le logement et on code E13tpsigt=4, soit une partie du temps seulement, sans précision.
- Si le répondant a répondu que l'enfant ne vit pas tout le temps dans le logement mais qu'il déclare aussi qu'il ne vit pas ailleurs (E13aut\_lgt=2) on considère que l'enfant vit tout le temps dans le logement.

6<sup>ème</sup> étape : utilisation de l'appariement des enfants avec l'EAR.

- Si les variables sur le temps passé dans le logement et un éventuel autre logement sont en non-réponse et que l'enfant n'est pas apparié avec l'EAR on considère que le logement n'est pas le logement principal de l'enfant et donc E13tpsigt=4, E13aut\_lgt=1 : il y a un autre logement. Toutefois les 3 variables sur la nature de cet autre logement sont laissées en non-réponse.

## Âge à l'adoption E13\_14agad.

Au départ 1,1% des enfants ont une valeur renseignée dans la variable âge à l'adoption.

Toutefois cet âge à l'adoption paraît, dans certains cas, incohérent, notamment quand l'âge à l'adoption est supérieur à l'âge de l'enfant. Des cas de 00 semblent aussi être, en réalité, de la non-réponse.

Les règles de décision retenues, issues du retour aux images, sont les suivantes :

- Si l'âge à l'adoption est supérieur de plus d'un an à l'âge de l'enfant on considère que c'est une erreur que l'enfant n'a pas été adopté et donc E13\_14agad est mis en valeur manquante pour « non concernée ».
- Si l'âge à l'adoption est, à plus ou moins un an, le même que l'âge de l'enfant
  - S'il y a trois enfants ou plus dont l'âge à l'adoption est, à plus ou moins un an, l'âge de l'enfant et que ce nombre d'enfants adoptés est égal au nombre total d'enfant du répondant on considère que c'est une erreur et que les enfants n'ont pas été adoptés. Il s'agit de cas où les enquêtés ont renseigné l'âge de leurs enfants et non des âges à l'adoption.
  - Si il y a deux enfants ou plus dont l'âge à l'adoption est, à plus ou moins un an, l'âge de l'enfant et que les enfants ont plus de 11 ans on considère que c'est une erreur et que les enfants n'ont pas été adoptés. Si les enfants sont âgés de 10 ans ou moins on garde l'information liée à l'adoption.
  - Dans les autres cas, si l'enfant à plus de 21 ans on considère que c'est une erreur et que les enfants n'ont pas été adoptés. On garde l'information sur l'adoption sinon.
- Pour les enfants du tableau 14, si l'âge à l'adoption est égal à l'âge au départ du domicile on considère que l'âge à l'adoption est une erreur et que les enfants n'ont pas été adoptés

On supprime ainsi environ 700 enfants adoptés et la part des enfants adoptés parmi l'ensemble des enfants et de 1%.

## Âge de l'enfant au départ de chez son parent E14agdep

Les enfants du tableau 13 ne sont pas concernés par cette question. Pour ces enfants la variable E14agdep est en valeur manquante pour « non concernée ».



Les règles de décision retenues, issues du retour aux images, sont proches de celles mises en place pour l'âge à l'adoption si ce n'est que les corrections transforment la réponse en non-réponse car tous les enfants du tableau 14 sont concernés. Ce sont les suivantes :

- Si l'âge au départ de chez son parent est supérieur de plus d'un an à l'âge de l'enfant on considère que c'est une erreur et E14agdep est mis en non-réponse (99).
- Si l'âge au départ de chez son parent est, à plus ou moins un an, le même que l'âge de l'enfant, que tous les enfants du répondant présentent cette caractéristique et que le répondant a trois enfants ou plus, on considère que c'est une erreur et E14agdep est mis en non-réponse (99).
- Si l'âge de départ est 0 an et que le répondant a trois enfants ou plus tous partis à 0 an on considère que c'est une erreur et E14agdep est en non-réponse (99).

On ajoute ainsi environ 800 non-réponses à la variable E14agdep (le taux de non-réponse pour cette variable est de 37% : 124 186 bulletins)

### **Le conjoint du répondant est-il parent de l'enfant ? (E13 14cpart).**

Après la mise en cohérence entre les enfants présents dans les tableaux 13 et 14 et les questions 11 et 12 des nombres d'enfants du répondant et de son conjoint dans le logement et hors logement, il ne peut pas y avoir de non-réponse à la question « le répondant est-il parent de l'enfant ? ». En revanche il peut y avoir de la non-réponse à cette même question pour son conjoint.

Pour pallier cette non-réponse :

- Si l'information sur l'année de mise en couple du répondant et l'année de naissance de l'enfant sont renseignées :
  - si l'enfant est né avant la mise en couple il n'est pas considéré comme enfant du conjoint, si l'enfant est né après il est considéré comme enfant commun aux deux membres du couple.
- Si l'année de mise en couple du répondant ou l'année de naissance de l'enfant est manquante :
  - Dans le tableau 13, si l'enfant réside principalement dans ce logement on dispose de l'information saisie dans la liste A de la feuille logement du recensement. Les liens renseignés dans cette liste sont alors utilisés pour repérer des beaux-enfants.
  - Dans le tableau 13, si l'enfant ne réside pas principalement dans le logement où que les liens renseignés dans la liste A de la feuille logement n'indiquent pas clairement le nombre d'enfants communs et/ou de beaux enfants :
    - Si l'année de mise en couple avec son conjoint actuel est renseignée et que le répondant s'est mis en couple après l'âge de 45, l'enfant n'est pas considéré comme enfant du couple actuel. C'est le cas le plus souvent hors données manquantes.
    - Dans les autres cas l'enfant est considéré comme enfant du couple actuel ; situation la plus fréquente.
  - Dans le tableau 14,
    - Si l'année de mise en couple avec son conjoint actuel est renseignée et supérieure à 1992, que le répondant avait plus de 25 ans lors de cette mise en couple l'enfant n'est pas considéré comme enfant du couple actuel. En effet la plupart des enfants mineurs (donc nés après 1992) qui ne vivent pas avec leurs deux parents ont leurs parents séparés. Cette condition assure ainsi de ne pas imputer trop d'enfants mineurs ne vivant pas avec leur deux parents vivant eux-même ensemble.
    - Si l'année de mise en couple avec son conjoint actuel est renseignée et inférieure à 1992, que le répondant est un homme qui avait plus de 45 ans lors de cette mise en couple l'enfant n'est pas considéré comme enfant du couple actuel. Cette condition assure ainsi de ne pas imputer d'enfants trop jeunes par rapport à l'âge du répondant.
    - Si l'année de mise en couple avec son conjoint actuel est renseignée et inférieure à 1992, que le répondant est une femme qui avait plus de 39 ans lors de cette mise en couple l'enfant n'est pas considéré comme enfant du



couple actuel. Cette condition assure ainsi de ne pas imputer d'enfants trop jeunes par rapport à l'âge du répondant.

- Si l'année de mise en couple avec son conjoint actuel est renseignée et inférieure à 1992, que le répondant avait entre 15 ans et 45 ans lors de cette mise en couple et qu'il a moins de 60 ans l'enfant n'est pas considéré comme enfant du couple actuel.
- Dans les autres cas il est considéré comme enfant du couple actuel ; situation la plus fréquente.



## Codage de la profession des parents (questions 17, 18, 23 et 24)

En 2010, lors de l'enquête pilote, il y avait 33 000 libellés de profession à coder dont 20% ont été envoyés en reprise. Il est apparu que nombre d'entre eux l'étaient inutilement.

Un certain nombre de reprises manuelles étaient excessives et, notamment, pour un même couple libellé\*statut ces reprises conduisaient à des codes PCS très différents, voire même des groupes sociaux différents (exemple pour SNCF on trouve comme groupe social en reprise manuelle 4, 5 ou 6 ou pour cadre supérieur de santé les groupes 3 et 4).

Un traitement particulier a donc été développé par l'Insee pour les CS des parents afin de pallier ce problème de qualité et limiter le nombre de reprises manuelles.

### Le processus de codage :

- Détection des libellés incodables** à partir de la liste validée par l'expert variable Sicore. Parmi les 561 158 libellés renseignés presque **10 %** s'avèrent, après analyse des experts Sicore et de l'équipe projet, incodables. Ces codes ne seront pas envoyés en reprise. Un indicateur permet de les repérer, ils sont par ailleurs ventilés en 4 catégories :
  - **N'a jamais travaillé** : le questionnaire indique que la personne n'a jamais travaillé (par exemple A ELEVE AVEC AMOUR SES 5 ENFANTS, SANS PROFESSION) environ **20 500**
  - **Profession non déclarée** : le libellé indique que la personne a pu travailler, mais aucune information de métier n'est déclarée (par exemple RETRAITE APRES 40 ANS DE TRAVAIL, NE SAIS PLUS, CHERCHE UN EMPLOI...) environ **9 000**
  - **Codage impossible** : une profession est déclarée mais le libellé est trop pauvre pour pouvoir coder en groupe social (par exemple FONCTIONNAIRE, AGENT SNCF) environ **22 500**
  - **Aide-familiale** : par exemple TRAVAILLAIT AVEC SON MARI environ **500**.
- Premier codage Sicore (81% soit environ 451 500 libellés codés)**. Un indicateur permet de les repérer. De même des indicateurs permettent de savoir quelles variables annexes ont manqué à Sicore pour coder la catégorie sociale parfaitement.
- Codage automatique** à partir d'une table de passage libellé\*statut et catégorie sociale (CS en 18 postes ou groupe social selon les cas) validée par l'expert variable (**2% soit presque 11 000 libellés codés**). Un indicateur permet de les repérer.
- Appariement entre le libellé de la profession de l'EFL non encore codé et des libellés du RP niveau CS 18 postes**. Si pour un couple libellé\*statut de l'EFL on retrouve toujours la même catégorie sociale en CS 18 postes au RP 2007 (avec une fréquence d'apparition supérieure ou égale à 5) cette catégorie est affectée dans l'EFL (**moins de 1%, soit moins de 4 500 libellés codés**). Un indicateur permet de les repérer.
- L'appariement entre les libellés encore non codés de l'EFL et du RP** est effectué une seconde fois au **niveau du groupe social**. Si pour un couple libellé\*statut de l'EFL on retrouve toujours le même groupe social au RP 2007 (avec une fréquence d'apparition supérieure ou égale à 5) ce groupe est affecté dans l'EFL (**moins de 2 000 libellés codés**). Un indicateur permet de les repérer.
- 2<sup>ème</sup> codage Sicore sur la base du premier mot significatif du libellé** (cas des professions multiples). C'est le principe retenu dans le cadre de la reprise RP. Toutefois si le libellé indique une succession de profession (présence d'un puis), on privilégie le dernier mot significatif (**4% soit presque 20 500 libellés codés**).



7. **Appariement entre le premier mot (ou dernier) du libellé de la profession de l'EFL non encore codé et le libellé du RP niveau CS 18 postes. (environ 1 000 libellés codés).** Un indicateur permet de les repérer.
8. **Appariement entre le premier mot (ou dernier) du libellé de la profession de l'EFL non encore codé et le libellé du RP niveau groupe social. (moins de 1 000 libellés codés).** Un indicateur permet de les repérer.
9. **Reprise manuelle des 18 000 libellés restants.** Seule une occurrence par couple profession statut est envoyée en reprise.

Le principe de la reprise manuelle est le suivant :

- a. La consigne est de coder sur 4 positions quand ce codage est possible, en CS 18 sinon, voire en groupe social quand le libellé ne permet pas d'aller au-delà.
- b. La consigne est de différencier les non codés dans les mêmes quatre catégories que dans l'étape 1 de détection des libellés incodables.

## **Fonctionnement de Sicore profession**

### **1. Sicore code en PCS (4 positions)**

Il n'existe pas de bases de connaissances Sicore pour coder uniquement en niveau CS. Sicore code en PCS, sauf dans le cas où la variable annexe STRE est activée pour coder en CS les anciennes professions des chômeurs et des retraités. On obtient alors, pour ces derniers, des codes de la forme 7700 ou 4700 et des codes PCS pour les actifs occupés (voir variable annexe STRE).

### **2. Processus de codification dans Sicore**

Dans un premier temps, le libellé d'enquête (réponse de l'enquêté) sera comparé aux **libellés de référence** dont Sicore dispose. Si le libellé d'enquête ressemble à un des libellés de référence, Sicore poursuivra le processus de codification à l'aide des variables annexes. La liste des libellés de référence utilisée par Sicore est alimentée avec les libellés cités dans la nomenclature et les libellés rencontrés le plus souvent dans les réponses aux enquêtes.

Dans un deuxième temps, on va utiliser les informations complémentaires pour pouvoir déterminer la correspondance avec un code précis de la nomenclature. Un même libellé de profession peut en effet renvoyer à différents codes de la nomenclature. Dans le cas de la profession, quatorze variables annexes peuvent être utilisées.

Exemple (simplifié) du processus de codification  
codification du libellé "VENDEUR DE MEUBLES"

Si non-ressemblance avec un libellé de référence => échec

Si ressemblance avec un libellé de référence => poursuite avec utilisation des variables annexes et de leurs règles logiques :

Si statut d'indépendant alors :

Si emploie moins de 10 salariés alors code PCS = 223A

Si emploie 10 salariés ou plus alors code PCS = 233C

Si statut de salarié alors :

Si qualification d'agent de maîtrise alors code PCS = 462B

Si qualification d'employé alors code PCS = 554B

On remarque que selon les modalités des variables annexes, pour un même libellé, le codage est différent dès le niveau CS (deux premières positions du code).



### Listes des 14 variables annexes

Nom Sicore	Contenu	Commentaire
STRE	l'occupation d'un emploi	distinction actifs occupés, retraités, chômeurs
STATUT	le statut dans l'emploi	distinction entre salariés et indépendants
PUB	le statut de l'employeur	distinction entre public et privé
CPF	la position professionnelle ou qualification	distinction entre ouvriers, ingénieurs...
NBS	le nombre de salariés employés	pour les indépendants, distinction artisans, commerçants, de chefs d'entreprise
NAF2	l'activité de l'employeur	code NAF sur 2 positions
NAF	l'activité de l'employeur	code NAF sur 4 positions
FN	la fonction	fabrication, installation, vente, secrétariat...
T	la taille de l'entreprise	distinction grande, moyenne, petite
SP	le statut d'apprenti	certaines codes PCS font explicitement référence au statut d'apprenti
S	le sexe	utilisation marginale pour quelques professions
DEP	le département	pour les professions de l'agriculture
OPA	l'orientation principale agricole	pour les professions de l'agriculture
SAU	la surface agricole utilisée	pour les professions de l'agriculture

**L'absence d'une variable annexe a peu d'effet sur le taux de codage (% de codages réussis) mais peut avoir des effets non négligeables sur la qualité des codages (% de bien codés parmi les codés).**

En l'absence d'information précise, on appliquera un codage par défaut que l'on peut traduire par "sans autres indications, il est le plus probable que le code soit le suivant". Ce codage par défaut a été décidé soit au regard des instructions contenues dans la nomenclature, soit en fonction des résultats d'enquêtes.

Dans l'enquête famille et logements seulement 3 des 14 variables annexes sont disponibles : le sexe, le département de résidence et le statut dans l'emploi.

La variable statut utilisée par Sicore est issue des questions Q18 et Q24 pour respectivement la mère et le père.

La règle de codage de cette variable est la suivante :

Règle de transcodage de Q18STATUT\_M\_X et Q24STATUT\_P\_X en STATUT\_SIC (X un entier valant 0, 2 ou 3):

Saisie RP	Variable annexe SICORE	Libellé
Q18STATUT_M_X ou Q24STATUT_P_X =	STATUT_M_SIC ou STATUT_P_SIC =	
1XXX	3	Salarié
1XX1	3	Salarié
11XX	3	Salarié
1X1X	3	Salarié
X1XX	1	Indépendant
XX1X	1	Indépendant
X11X	1	Indépendant
X1X1	1	Indépendant
XX11	1	Indépendant
Autres	*	Manquant



Ainsi pour les variables Q18statut\_M\_X et Q24statut\_P\_X valant « 1330 », « 1001 », « 1120 », « 1310 » par exemple le statut codé pour le codage Sicore vaut 3=salarié. Pour celles codées « 2122 », « 3010 », « 3110 », « 0110 », « 0011 » par exemple le statut utilisé par Sicore est 1=indépendant.

La modalité 2 de SICORE : « salarié de sa propre entreprise, gérant mandataire » n'est pas utilisée.

## **Bilan du traitement**

### La profession de la mère.

Dans 121 000 cas environ aucun libellé n'est rempli. Dans 68% de ces situations la variable n'a jamais travaillé est cochée.

13% des libellés renseignés ne sont pas codables par Sicore, principalement car ils ne renvoient pas à une profession (plus des 3/4) : dans 2/3 des cas cela renvoie à la non-activité de la mère (mère au foyer).

Quand le libellé de profession est rempli, dans 78% des cas Sicore parvient à coder un code profession (étape 2).

Le codage par Sicore sur le premier mot significatif du libellé uniquement permet de coder 3% des libellés.

3% sont également envoyés en site de reprise.

Les cinq autres étapes ont un impact très marginal : elles ne permettent de coder que les 3% restant.

numéro de l'étape	libellé de l'étape	Répartition des bulletins EFL	Précision au sein de l'étape	Répartition des bulletins EFL
01	libellé incodable	13%	n'a jamais travaillé une profession trop floue un libellé sans profession aide-familiale	63% 22% 13% 2%
02	SICORE	78%	réussite avec utilisation des variables annexes réussite mais au moins une variable annexe nécessaire était à valeur manquante	18% 82%
03	codage automatique par l'expert profession	1%		
04	codage en CS à partir du RP	1%		
05	codage en GS à partir du RP	0%		
06	SICORE sur premier mot significatif	3%	réussite avec utilisation des variables annexes réussite mais au moins une variable annexe nécessaire était à valeur manquante	6% 94%
07	codage en CS à partir du RP sur premier mot significatif	0%		
08	codage en GS à partir du RP sur premier mot significatif	0%		
SR	envoi en site de reprise	3%		
Ensemble des libellés renseignés		100%		

### La profession du père.

Dans 45 000 cas environ aucun libellé n'est rempli. C'est beaucoup moins que pour les mères. Cependant, contrairement aux mères dans seulement 7% de ces situations la variable n'a jamais travaillé est cochée. Dans 60% des cas le statut dans l'emploi n'est pas renseigné lui aussi.

6% des libellés renseignés ne sont pas codables par Sicore, principalement car le libellé renseigné est trop flou (plus des 3/4).

Quand le libellé de profession est rempli, dans 82% des cas Sicore parvient à coder un code profession (étape 2).

Le codage par Sicore sur le premier mot significatif du libellé uniquement permet de coder 4% des libellés et 4% sont envoyés en site de reprise.



Les cinq autres étapes ont, pour la profession du père comme pour celle de la mère, un impact très marginal.

numéro de l'étape	libellé de l'étape	Répartition des bulletins EFL	Précision au sein de l'étape	Répartition des bulletins EFL
01	libellé incodable	6%	n'a jamais travaillé une profession trop floue un libellé sans profession aide-familiale	1% 76% 23% 0%
02	SICORE	82%	réussite avec utilisation des variables annexes réussite mais au moins une variable annexe nécessaire était à valeur manquante	14% 86%
03	codage automatique par l'expert profession	2%		
04	codage en CS à partir du RP	1%		
05	codage en GS à partir du RP	0%		
06	SICORE sur premier mot significatif	4%	réussite avec utilisation des variables annexes réussite mais au moins une variable annexe nécessaire était à valeur manquante	10% 90%
07	codage en CS à partir du RP sur premier mot significatif	0%		
08	codage en GS à partir du RP sur premier mot significatif	0%		
SR	envoi en site de reprise	4%		
Ensemble des libellés renseignés		100%		

## **Apurement des questions q17, q18, q23 et q24**

Les questions Q17 et Q18 permettent de savoir si la mère n'a jamais travaillé et si ce n'est pas le cas la profession exercée et le statut de cet emploi.

Les questions 23 et 24 sont les mêmes pour le père.

### **Pour la profession du père et de la mère :**

121 000 bulletins environ, soit un tiers de bulletins n'ont pas de codage de la profession pour la mère car aucun libellé de profession n'a été renseigné par l'enquêté. C'est le cas de 45 000 bulletins, 12%, pour la profession du père.

**Après le passage Sicore et des traitements présentés précédemment environ 34 000 bulletins ont un groupe social incodable pour la mère, 14% des libellés renseignés, et 23 000 pour le père 7% des libellés renseignés.**

Parmi ces incodables les traitements précédents permettent de distinguer 4 groupes : le libellé indique que le père ou la mère n'a jamais travaillé, étaient des aidants familiaux, le libellé est une profession trop floue pour être codée en groupe social, le père ou la mère ont travaillé mais aucune profession n'est indiquée.

Cette indication va nous permettre d'apurer les variables « n'a jamais travaillé » et statut de l'emploi.

### **Avant apurement :**

25% des bulletins indiquent que la mère n'a jamais travaillé et 1% que le père n'a jamais travaillé.

Par ailleurs certains bulletins indiquent que la mère ou le père n'a jamais travaillé alors qu'un groupe social et un statut ont été codés, le statut est non renseigné alors qu'une profession a été renseignée...

**1ère étape : utilisation des traitements pour le codage des PCS pour corriger n'a jamais travaillé et statut.**



Quand le groupe social de la mère ou du père est classé en incodable avec la précision n'a jamais travaillé, on code n'a jamais travaillé à 1 et le statut à blanc.

Quand le groupe social de la mère ou du père est classé en incodable avec la précision aidants familiaux on code n'a jamais travaillé à 0 et le statut en aide familiale non rémunéré.

Quand le groupe social de la mère ou du père est classé en incodable avec la précision profession floue, on code n'a jamais travaillé à 0.

Quand le groupe social a été codé, on code n'a jamais travaillé à 0.

### **2ème étape : traitement des cas avec plusieurs cochages à la variable statut.**

Quand on a un statut coché et ne sais pas on privilégie le statut renseigné à ne sais pas.

Certains groupes sociaux ou catégorie sociale n'ont qu'un statut possible. Quand plusieurs statuts sont renseignés, on retient celui qui est cohérent avec le groupe social ou la catégorie sociale. Par exemple la catégorie sociale cadre d'entreprise (36) est une catégorie de salariés.

Si on tenant compte de la catégorie sociale plusieurs statut restent renseignés, on classe le statut en « ne sait pas » puisqu'on ne peut pas en privilégier un plutôt qu'un autre.

### **3ème étape : traitement des incohérences entre n'a jamais travaillé, statut et catégorie sociale.**

Quand la variable n'a jamais travaillé est cochée et qu'un statut est coché, on retient n'a jamais travaillé et on met à « 0000 » le statut. En effet les cas avec un libellé vont dans le sens de « n'a jamais travaillé ».

Quand le groupe social est renseigné ou non renseigné mais avec l'indication qu'un libellé de profession a été renseigné mais trop flou pour être codé et quand le statut est non renseigné, on le code en ne sait pas.

Quand le groupe social est employé ou ouvrier (5 ou 6) normalement le statut ne peut être que salarié. Les mères employées ou ouvrières dont le statut n'est pas salarié sont principalement assistantes maternelles (50%), femmes de ménage ou aides à domicile (25%).

Pour les pères (beaucoup plus rare : moins de 1 000 cas seulement), il s'agit d'ouvriers agricoles ou d'ouvriers industriels.

Par ailleurs plus d'un tiers des cas étaient ambigus sur le choix du statut (« 1100 » ou « 1010 »).

On décide donc de recoder les statuts des groupes sociaux employé et ouvrier en salarié.

Quand le groupe social est agriculteur exploitant ou indépendants (1 ou 2) le statut ne peut pas être salarié.

Pour les mères dans cette situation (seulement 1 400 cas environ) les libellés associés correspondent dans 90% des cas à des statuts d'indépendants (gérante de société, chef d'entreprise...). Pour les pères la situation est la même (environ 2 500 cas).

On décide donc de recoder les statuts des groupes sociaux employé et ouvrier en indépendants ou/et aidants selon les différents cas de statut initialement renseignés. Par exemple « 1100 » deviendra « 0100 », « 1000 » deviendra « 0110 »...

In fine on retient un codage de statut le plus proche du RP avec :

salarié codé 10

non salarié codé 2\*:

ici 21 non salarié indépendant

22 non salarié aides familiaux

20 non salarié sans précision

30 ne sait pas

40 non concernée car n'ont jamais travaillé

ZZ pour la non-réponse.



## Résultats :

Q17njt_M	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
0	254956	70.87	254956	70.87
1	104735	29.11	359691	99.98
Z	79	0.02	359770	100.00

Q17gs_M	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
1	28011	10.98	28011	10.98
2	18429	7.23	46440	18.21
3	7354	2.88	53794	21.09
4	31151	12.21	84945	33.31
5	87261	34.22	172206	67.52
6	33941	13.31	206147	80.83
Z	48888	19.17	255035	100.00

Valeur(s) manquante(s) = 104735

Q18statut_M	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
10	169725	66.55	169725	66.55
20	7589	2.98	177314	69.53
21	36106	14.16	213420	83.68
22	9129	3.58	222549	87.26
30	6564	2.57	229113	89.84
ZZ	25922	10.16	255035	100.00

Valeur(s) manquante(s) = 104735

Q23njt_P	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
0	356495	99.09	356495	99.09
1	3196	0.89	359691	99.98
Z	79	0.02	359770	100.00

Q23gs_P	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
1	38000	10.66	38000	10.66
2	38322	10.75	76322	21.40
3	30134	8.45	106456	29.86
4	36239	10.16	142695	40.02
5	30902	8.67	173597	48.68
6	118799	33.32	292396	82.00
Z	64178	18.00	356574	100.00

Valeur(s) manquante(s) = 3196



Q24statut_ p	Fréquence	Pourcentage	Fréquence cumulée	Pctage. cumulé
10	234385	65.73	234385	65.73
20	8933	2.51	243318	68.24
21	76530	21.46	319848	89.70
22	395	0.11	320243	89.81
30	8290	2.32	328533	92.14
ZZ	28041	7.86	356574	100.00

Valeur(s) manquante(s) = 3196



## Etat vital et résidence des parents (questions 19, 20, 25 et 26)

APUREMENT des variables Q19viv\_m\_X, Q19adec\_m\_X, Q20res\_m\_X, Q25viv\_p\_X, Q25adec\_p\_X et Q26res\_p\_X réalisé par l'Insee

### Définition des variables définitives créées

Ces variables seront renommées à la fin du programme d'apurement avec des noms plus explicites avant d'être imputées puis introduites dans le FPR, mais on gardera dans ce document les noms qu'elles ont dans le programme SAS (le nom définitif est entre parenthèses).

#### - pour la mère :

- 1 variable viv\_m pour caractériser l'état vital :

**vivm (Q19viv\_m)**

1 : vivante  
2 : décédée  
Z : non renseigné

- 2 variables res\_m1 et res\_m2 pour caractériser sa résidence lorsqu'elle est vivante :  
**res\_m1 (Q20res\_lgt\_M) : la mère vit dans le logement de l'enquêté**

blanc : mère dcd (vivm =2)  
0 : non, ne vit pas dans le lgt  
1 : oui, vit dans le lgt  
Z : non renseigné\*  
*\* res\_m non renseigné ou viv\_m non renseigné*

**res\_m2 (Q20res\_aut\_M) : la mère vit ailleurs que dans le logement de l'enquêté**

blanc : mère dcd  
00 : ne vit pas ailleurs  
10 : vit ailleurs\*\* et sait où ( précisé ou pas)  
01 : vit ailleurs\*\* et ne sait pas où  
ZZ : non renseigné\*  
*\* res\_m non renseigné ou viv\_m non renseigné*  
*\*\*ailleurs : pas dans le logement de ego*

#### - pour le père :

- 1 variable viv\_p pour caractériser l'état vital :

**vivp (Q25viv\_p)**

1 : vivant  
2 : décédé  
Z : non renseigné

- 3 variables res\_p0, res\_p1 et res\_p2 pour caractériser sa résidence lorsqu'il est vivant :

**res\_p0 (Q26res\_avec\_m\_P) : le père vit avec la mère)**

blanc : père dcd ou mère dcd  
0 : non, ne vit pas avec la mère  
1 : oui, vit avec la mère



Z : non renseigné\*

\* *res\_p non renseigné ou viv\_p non renseigné*

**res\_p1 (Q26res\_lgt\_P) : le père vit dans le logement de l'enquêté)**

blanc : père dcd

0 : non, ne vit pas dans le lgt

1 : oui, vit dans le lgt (avec ou sans la mère)

Z : non renseigné\*

\* *res\_p non renseigné ou viv\_p non renseigné*

**res\_p2 (Q26res\_aut\_P) : le père vit ailleurs c'est-à-dire ni avec la mère, ni dans le logement de l'enquêté**

blanc : père dcd

00 : ne vit pas ailleurs\*\*

10 : vit ailleurs\*\* et sait où ( précisé ou pas)

01 : vit ailleurs\*\* et ne sait pas où

ZZ : non renseigné\*

\* *res\_p non renseigné ou viv\_p non renseigné*

\*\* *ni avec la mère, ni dans le logement de l'enquêté*

## Apurement Mère

- **Q19VIV M et Q19ADEC M**

- si l'année de décès de la mère est inférieure ou égale à l'année de naissance d'ego et si la mère n'est pas décédée : on met l'année de décès à blanc  
sinon, si l'année de décès de la mère est supérieure à l'année de naissance d'ego on dit que la mère est décédée

- si 1900 < année de naissance d'ego <= 1930 ou si année de décès renseignée ou si mentionné « décédé » dans comr\_m : on dit que la mère est décédée

- si viv\_m indéterminé (00, 02 ..11) et si quelque chose (sauf NSP) dans le lieu (com, dep, dom) on dit qu'elle est vivante

- **Q20res\_lgt\_m et q20res\_aut\_m:**

- si aucune case et aucun lieu remplis : on initialise res\_mr1 et res\_mr2 respectivement à Z et ZZ

- si res\_mr2='10 and comr\_mr='NSP' alors on met res\_mr2=01

- si la mère est vivante :

**1- si res\_mr2 =00 (pas rempli)**

- si res\_mr1=1 (vit dans le logt)

- si même commune que ego (ou comr\_m=MLGT) on met géo à blanc et on laisse res\_mr2=00 (25914 cas)

- sinon, si autre commune, ou dept ou dom rempli on met res\_mr2=10 (res\_mr1=1)(1078 cas)

- sinon, si commune= Ne sait pas on met res\_mr2=01

- sinon on laisse res\_mr2=00

- si res\_mr1=0

- si un lieu indiqué (commune sauf NSP, dept ou dom) on met res\_mr2=10 (6102 cas)

- sinon, si commune= Ne sait pas on met res\_mr2=01

- sinon on laisse res\_mr2=00



## 2- si res\_mr2 =10 (ailleurs)

- si comr\_m=MLGT on met res\_mr2=00 (8 cas)
- sinon, si commune= Ne sait pas on met res\_mr2=01
- sinon on laisse res\_mr2=10

### 3- si res\_mr2 =01 (ne sait pas où)

- si commune autre que Ne sait pas on met res\_mr2=10 (36 cas)
- sinon on laisse res\_mr2=01

## Apurement Père

### - Q25VIV P et Q25ADEC P : comme pour la mère

En plus :

- si viv\_mr='00' et que le père est vivant et vit avec la mère, alors on dit que la mère est vivante
- si viv\_pr='00' et si le père vit avec la mère et que la mère est vivante, alors on dit que le père est vivant

### - Q26res lgt p, Q26res avecm p et Q26res aut p :

- si aucune case et aucun lieu remplis : on initialise res\_pr0 res\_pr1 et res\_pr2 respectivement à Z,Z et ZZ

- si le père est vivant :

### 1- si res\_pr2 =00 (pas rempli)

- si res\_pr0=0 et res\_pr1=0 (ne vit pas avec la mère et ne vit pas dans le logt)
  - si un lieu indiqué (commune autre que NSP, IDEM,MLGT, dept ou dom) alors on met res\_pr2=10
  - sinon, si commune= Ne sait pas on met res\_pr2=01
  - sinon on laisse res\_pr2=00
- si res\_pr0=0 et res\_pr1=1 (ne vit pas avec la mère et vit dans le logt)
  - si même commune que ego (ou comr\_m=MLGT) on met géo à blanc et on laisse res\_pr2=00
  - sinon, si autre commune, ou dept ou dom rempli on met res\_pr2=10
  - sinon, si commune= Ne sait pas on met res\_pr2=01
  - sinon on laisse res\_pr2=00
- si res\_pr0=1 et res\_pr1=0 (vit avec la mère et ne vit pas dans le logt)
  - si même commune que la mère (ou comr\_pr=IDEM) on met géo à blanc et on laisse res\_pr2=00
  - sinon, si autre commune que la mère ou dept ou dom rempli on met res\_pr2=10
  - sinon, si commune= Ne sait pas on met res\_pr2=01
  - sinon on laisse res\_pr2=00
- si res\_pr0=1 et res\_pr1=1 ( vit avec la mère et vit dans le logt)
  - si même commune que ego (ou comr\_m=MLGT) ou même commune que la mère on met géo à blanc et on laisse res\_mr2=00
  - sinon, si autre commune, ou dept ou dom rempli on met res\_pr2=10
  - sinon, si commune= Ne sait pas on met res\_pr2=01
  - sinon on laisse res\_pr2=00

### 2- si res\_pr2 =10 (ailleurs)

- si comr\_m=MLGT on met res\_pr2=00
- sinon, si commune= Ne sait pas on met res\_pr2=01



- sinon on laisse res\_pr2=10

**3- si res\_pr2 =01 (ne sait pas où)**

- si commune autre que Ne sait pas on met res\_pr2=10 (36 cas)

- sinon on laisse res\_pr2=01

## Recodages pour cohérence

Lorsque la mère est vivante :

- si elle habite le logt seulement (res\_mrrr1=1 et res\_mrrr2=00)

- si le père habite avec la mère on dit qu'il habite dans le logement (→res\_prrr1=1)

- si le père habite dans le logement on dit qu'il habite avec la mère (→res\_prrr0=1)

- si res\_mrrr1 et res\_mrrr2 ne sont pas renseignés et que le père vit avec la mère on remplit la question sur le logement d'ego de la mère comme le père (res\_mrrr1=res\_prrr1), mais on laisse res\_mrrr2 non rempli (ZZ)

- si la mère habite ailleurs seulement (res\_mrrr2=10) et que le père habite avec la mère et ailleurs ne sait pas où, on enlève le ne sait pas (→res\_prrr2=00) (14 cas)

- 10 autres cas corrigés : le père vit avec la mère dans le lgt ou ailleurs et la mère est à « ne sait pas » : on met ailleurs (10) à la mère

**Fin du programme d'apurement** : à ce stade, les variables ont le nom et les modalités qu'elles auront dans le FPR.

L'étape suivante consiste à redresser de la non-réponse les variables concernant l'état vital, à savoir Q19viv\_m, Q19adec\_m, Q25viv\_p et q25adec\_p

## Redressement de la non-réponse à Q19viv m, Q19adec m, Q25viv p et q25adec p

### 1) Q19viv\_m et Q25viv\_p

après apurement et avant redressement de la non-réponse :

q19viv_m	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	216184	60.09	216184	60.09
2	132029	36.70	348213	96.79
Z	11557	3.21	359770	100.00

q25viv_p	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	160385	44.58	160385	44.58
2	182886	50.83	343271	95.41
Z	16499	4.59	359770	100.00
			Cumulative	Cumulative

On utilise des tranches d'âge quinquennal et la variable Dipl (dernier diplôme obtenu) de l'EAR et on impute par hot-deck



tri=tage!!dipl;

```
length tage $ 8.;
if 19<=agego<=25 then tage='19-25';else
if 26<=agego<=30 then tage='26-30';else
if 31<=agego<=35 then tage='31-35';else
if 36<=agego<=40 then tage='36-40';else
if 41<=agego<=45 then tage='41-45';else
if 46<=agego<=50 then tage='46-50';else
if 50<=agego<=55 then tage='51-55';else
if 56<=agego<=60 then tage='56-60';else
if 61<=agego<=65 then tage='61-65';else
if 66<=agego<=70 then tage='66-70';else
if 71<=agego<=75 then tage='71-75';else
if 76<=agego<=80 then tage='76-80';else
if 81<=agego<=86 then tage='81-86';else
if 86<=agego then tage='86 et '+';
```

#### DIPL

ZZ	Non-réponse
01	Pas de scolarité
02	Aucun diplôme mais scolarité jusqu'en école primaire ou au collège
03	Aucun diplôme mais scolarité au-delà du collège
11	CEP (certificat d'études primaires)
12	BEPC, brevet élémentaire, brevet des collèges
13	CAP, brevet de compagnon
14	BEP
15	Baccalauréat général, brevet supérieur
16	Baccalauréat technologique ou professionnel, brevet professionnel ou de technicien, BEA, BEC, BEI, BEH, capacité en droit
17	Diplôme de 1er cycle universitaire BTS, DUT, diplôme des professions sociales ou de la santé, d'infirmier(ère)
18	Diplôme de 2° ou 3° cycle universitaire (y compris médecine, pharmacie, dentaire), diplôme d'ingénieur, d'une grande école, doctorat, etc

#### Après imputations :

	q19viv_m	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	1	222821	61.93	222821	61.93
	2	136949	38.07	359770	100.00

	I_q19viv_m	Frequency	Percent	Cumulative Frequency	Cumulative Percent
identique à la brute	0	343511	95.48	343511	95.48
apurée	1	4702	1.31	348213	96.79
imputée	2	11557	3.21	359770	100.00

	q25viv_p	Frequency	Percent	Cumulative Frequency	Cumulative Percent
	1	167454	46.54	167454	46.54
	2	192316	53.46	359770	100.00



	I_q25viv_p	Frequency	Percent	Cumulative Frequency	Cumulative Percent
identique à la brute	0	337976	93.94	337976	93.94
apurée	1	5295	1.47	343271	95.41
imputée	2	16499	4.59	359770	100.00

## 2)- Q19adec\_m et q25adec\_p

On utilise des tranches d'âge quinquennal et la variable Diplc (dernier diplôme obtenu condensé) de l'EAR et on impute par hot-deck

**tri=tage!!diplc;**

```
length tage $ 8.;
if 19<=agego<=25 then tage='19-25';else
if 26<=agego<=30 then tage='26-30';else
if 31<=agego<=35 then tage='31-35';else
if 36<=agego<=40 then tage='36-40';else
if 41<=agego<=45 then tage='41-45';else
if 46<=agego<=50 then tage='46-50';else
if 50<=agego<=55 then tage='51-55';else
if 56<=agego<=60 then tage='56-60';else
if 61<=agego<=65 then tage='61-65';else
if 66<=agego<=70 then tage='66-70';else
if 71<=agego<=75 then tage='71-75';else
if 76<=agego<=80 then tage='76-80';else
if 81<=agego<=86 then tage='81-86';else
if 86<=agego then tage='86 et +';

length diplc $ 1.;
if dipl in ('01','02','03') then diplc='1';else
if dipl in ('11','12','13') then diplc='2';else
if dipl in ('14','15','16') then diplc='3';else
if dipl in ('17','18') then diplc='4';
else diplc='Z';
```



# Grands-parents

L'Insee a réalisé l'apurement de ces variables

## Apurement des variables Q27grp x, Q27nbpenf x, Q27agpenf x

### Q27grp\_x:

Variable brute après traitement des biffés et des raturés :

grp00	Frequency	Percent	Cumulative Frequency	Cumulative Percent
00	67872	18.87	67872	18.87
01	178867	49.72	246739	68.58
10	112852	31.37	359591	99.95
11	179	0.05	359770	100.00

on a 19% de non renseignés

Les principales corrections ont été effectuées à partir des choix suivants :

- si l'individu n'a pas d'enfants, il ne peut pas avoir de petits-enfants
- un individu qui n'a pas au moins 35 ans ne peut pas avoir de petits-enfants
- quand l'aîné des enfants a moins de 16 ans dans l'année, un individu ne peut pas avoir de petits-enfants
- s'il y a moins de 15 ans entre l'aîné des enfants et l'aîné des petits-enfants il y a une erreur et on considère qu'il n'y a pas de petits-enfants.
- lorsque l'âge de l'aîné des enfants est égal à 1 près à l'âge de l'aîné des petits-enfants, on suppose que l'enquêté a décrit dans cette partie ses enfants

Une soixantaine de relevés ont d'autre part été corrigés après vérification manuelle.

### Après apurement et avant redressement de la non-réponse :

grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	110457	30.70	110457	30.70
2	234556	65.20	345013	95.90
Z	14757	4.10	359770	100.00



## Apurement des questions sur les logements dans lesquels vivent les répondants (questions 29 à 36)

Ces apurements ont été réalisés par l'Ined.

### Création de variables synthétiques

Dans un premier temps les variables brutes issues de cases à cocher sont recodées afin de réduire le nombre de réponses possibles comme développé dans les principes généraux.

Afin de faciliter les choix de corrections, on crée ensuite des variables synthétiques avec des modalités renseignées pour l'ensemble des répondants, même les cas avec plusieurs cases cochées. Le principe est de coder à 9 pour la non-réponse, 88 quand il n'y a que des cases biffées ou de renseigner l'ensemble des numéros des modalités cochées. Par exemple : 5 quand seule la modalité 5 est cochée, 12 quand les modalités 1 et 2 sont cochées, 234 si les modalités 2 3 et 4 sont cochées

Par exemple, pour la variable Q29TPS\_X

I_TPS	Pourcentage
1	91.80
2	0.51
3	0.70
4	0.70
5	0.58
6	0.09
9	5.36
12	0.02
13	0.06
14	0.03
15	0.02
16	0.00
23	0.00
24	0.01
25	0.01
26	0.00
34	0.02
35	0.01
45	0.01
46	0.00
56	0.00
88	0.06
123	0.00
124	0.00
125	0.00
126	0.00
134	0.00
135	0.00
145	0.00
234	0.00
235	0.00
345	0.00
346	0.00
1234	0.00
12345	0.00
12356	0.00
123456	0.00



## Correction de la question 29

Quand il n'y a qu'une case cochée dans la variable synthétique, Q29TPS prend la valeur de cette variable.

Quand il y a plusieurs cases cochées dans variable synthétique créée précédemment on décide de coder Q29 tps de la façon suivante:

12 en 2  
13 en 3  
14 en 4  
15 en 5  
24 en 2  
25 en 2  
34 en 3  
35 en 3

L'objectif est de retenir la modalité la plus précise.

Il reste à reclasser les modalités 16, 23, 26, 45, 46, 56, les modalités indiquant plus de 2 cases cochées et la modalité 88 (seulement des cases biffées).

Pour la modalité 88 ce ne sont que des questionnaires rayés (car les cases sont biffées) par des personnes qui ne pensaient pas être concernées par cette partie du questionnaire. Ce sont donc des non-réponses.

Quand la variable synthétique vaut 88 ou 9 on code Q29tps= « Z » indiquant la non-réponse.

Pour les cas de réponses multiples résiduelles, on considère qu'elles se divisent en deux types :

-les cas où la personne dit vivre toute l'année dans ce logement, ie les cas où la réponse 1 a été cochée parmi d'autres. Dans ce cas on code Q29TPS=1.

-les cas où la personne dit vivre moins que toute l'année dans ce logement, ie les cas où la réponse 1 n'a pas été cochée. Dans ce cas on code Q29TPS=5 : une partie du temps en alternance car cette modalité exprime bien la confusion des multi-réponses

Q29tps	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1	333905	92.81	333905	92.81
2	1313	0.36	335218	93.18
3	2814	0.78	338032	93.96
4	2556	0.71	340588	94.67
5	2158	0.60	342746	95.27
6	341	0.09	343087	95.36
Z	16683	4.64	359770	100.00

## Correction de la question 30.

La variable synthétique I\_TPSM prend les modalités suivantes (1,2,3,9, 12, 13, 23, 123 et 88).

Quand il n'y a qu'une case cochée dans la variable synthétique, Q30TPSM prend la valeur de cette variable.

Pour la modalité 88 ce ne sont que des questionnaires rayés (car les cases sont biffées) par des personnes qui ne pensaient pas être concernées par cette partie du questionnaire. Ce sont donc des non-réponses.

Quand la variable synthétique vaut 88 ou 9 on code Q30TPSM= « Z » indiquant la non-réponse.

Pour les cas de réponses multiples on recode par la réponse qui indique la durée la plus longue.



Enfin si Q30TPSM est en non-réponse mais que Q29TPS=1 alors Q30TPSM est codée à 1

Q30tpsm	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1	338940	94.21	338940	94.21
2	3303	0.92	342243	95.13
3	321	0.09	342564	95.22
Z	17206	4.78	359770	100.00

## **Création de variables d'aide à la décision**

Variable NBREPLOG : nombre de réponse aux variables 32 à 36.

Dans cette variable, on compte une réponse dès que l'on a pas une non-réponse propre (par exemple q34 ne vaut pas « 000000 »).

Variable BLOCLOG

- 0 : aucune réponse dans le bloc logement (q32 à q36)
- 1 : au moins une réponse « claire » (au moins une cases cochée)
- 2 : au moins une réponse mais biffée ou raturée

## **Traitements des cases biffées pour les question 31 à 36**

Pour la question 31 et le bloc :

On a regardé les cas où certaines variables étaient bien cochées et d'autres biffées dans le bloc logement : s'il y a au moins un "oui" à Q31LGTSEC\_X et que d'autre(s) variable(s) sont cochées entre q32 et q36, on a transformé les biffées en cochées. Les cases biffées de la question 31 sont traitées différemment selon les réponses au bloc et aux questions 29 et 30.

Si l'enquête déclare clairement ne pas loger dans d'autre(s) logement(s), on transforme les cases biffées en non réponse puis en « non concernée » après correction de la variable 31.

## **Correction de la question 31**

Cette question admet plusieurs réponses en cas d'existence d'un autre logement. En revanche, on ne peut avoir coché : « je n'ai pas d'autre logement » avec une autre modalité, puisqu'elle indiquerait l'existence d'un autre logement. On décide de créer 6 variables. Une variable qui dit simplement si oui ou non il y a un autre logement (Q31lgt\_sec) puis 5 variables (Q31lgt\_sec\_fam, Q31lgt\_sec\_trav, Q31lgt\_sec\_et, q31lgt\_sec\_vac q31\_lgt\_sec\_aut) sur la (ou les) raison(s) de cet autre logement. Ces 5 variables sont mises en valeur manquante (pour « non concernée ») quand la variable Q31lgt\_sec indique qu'il n'y a pas d'autre logement.

1) si selon la variable synthétique I\_lgtsec une seule case est cochée et indique l'existence d'un autre logement (modalité 1 à 5) on code Q31lgt\_sec=1 et on renseigne la variable de la raison correspondante à 1 pour oui et les 4 autres à 2 pour non. Par exemple si I\_lgtsec=1 alors Q31lgt\_sec=1 Q31lgt\_sec\_fam=1 et Q31lgt\_sec\_trav=2, Q31lgt\_sec\_et=2, q31lgt\_sec\_vac=2 q31\_lgt\_sec\_aut=2.

2) si selon la variable synthétique I\_lgtsec une seule case est cochée et indique qu'il n'y a pas d'autre logement ou une non-réponse.



Si les questions 32, 34, 35, et 36 ne sont pas renseignées (bloclog=0) on code Q31lgt\_sec= 2 et les 5 variables sur la raison de cet autre logement sont mises en « non concernée » (valeur manquante).

Si au moins une question des questions 32, 34, 35, et 36 est renseignée (bloclog différent de 0)

S'il y a au moins 2 réponses sur 4 de renseignées : on code Q31lgt\_sec= 1 et les 5 variables sur la raison de cet autre logement sont mises en « non-réponse » (Z). Certains de ces cas ont été revus par l'Insee par la suite grâce à l'examen des images des bulletins EFL. Quand l'enquête indiquait ne pas avoir d'autres logements et que les seules variables renseignées étaient la localisation du logement dans la même commune que la résidence de l'enquête ou en l'absence de commune dans le même département on a recodé Q31lgt\_sec= 2 et les 5 variables sur la raison de cet autre logement sont mises en « non concernée » (valeur manquante). En effet, dans ces cas il est vraisemblable que l'enquête a indiqué la commune (ou le département) de sa seule résidence, d'autant que le retour aux images a montré que cette ligne était souvent rayée quoique saisie.

S'il n'y a qu'une réponse on code Q31lgt\_sec= 2 et les 5 variables sur la raison de cet autre logement sont mises en « non concernée » (valeur manquante).

## **Utilisation des précisions aux questions 32 et 36**

### **Précision à la question 32- « Nature » de l'autre logement**

- création d'une variable i\_natautlgtprec qui transforme la précision en une modalité de la question 32 sur la nature du logement
- création d'une variable i\_natautlgtprec b qui transforme la précision en une modalité de la question 36 sur la propriété du logement
- création d'une variable raisonautlog qui transforme la précision selon « l'utilisation » du logement : travail/vacances.

La création de trois variables différentes permet de récupérer les doubles informations. Par exemple, dans le cas d'un enquêteur ayant noté « dans la maison d'un ami », on récupère l'information que c'est une maison pour la question 32 et que c'est chez un ami pour la question 36.

### **La variable i\_natautlgtprec :**

Les précisions suivantes sont reclassées dans la modalité 1 de la question 32 : chambre en résidence universitaire, internat, foyer, caserne :

"CHAMBRE D INTERNAT"  
"BASE AERIENNE"  
"CHAMBRE D ETUDIANT"  
"KOT ETUDIANT"  
"MONASTERE"  
"PENSION"  
"CASERNE"  
"CANTONEMENT"  
"CASERNEMENT"  
"CHANTIER PETROLIER"  
"PLATEFORME PETROLIERE"  
"PETROLIER"  
"PLATE FORME"  
"MILITAIRE"  
"REGIMENT"  
"CENTRE EDUCATIF"  
"AUBERGE DE JEUNESSE"  
"CROUS"  
"CITE UNIVERSITAIRE"  
"LYCEE"  
"CAMP DE MANOEUVRE"  
"CAMP DE TRAVAILLEUR"  
"CAMP ONU"



"INTERNAT"  
"INTERNE"  
"INSTERNAT"  
"PENSIONNAT"  
"FOYER"  
"FOYERS"  
"COLLECTIVITE"  
"CENTRE DE SECOURS"  
"ETABLISSEMENT MILITAIRE"  
"SOUS MARIN"  
"INSTITUT"  
"CENTRE D HEBERGEMENT POUR ENFANT"  
"CENTRE D HEBERGEMENT POUR HANDICAPES"  
"CENTRE POUR HANDICAPES"  
"CENTRE REEDUCATION"  
"ETABLISSEMENT MEDICO SOCIAL"  
"ETABLISSEMENT DE SOIN"  
"CLINIQUE"  
"IME CENTRE"  
"PSYCHIATRIQUE"  
"HOSPITALISATION"  
"CENTRE ADAPTE AU HANDICAP"  
"CENTRE DE SOINS"  
"HANDICAPES"  
"HOSPITALIER"  
"CENTRE MEDICAL"  
"PARAMEDICAL"  
"AUTISTE CENTRE DE VIE"  
"HOSPITAL"  
"HOPITAL"

Les précisions suivantes sont reclassées dans la modalité 2 de la question 32 : mobil-home, caravane, habitation mobile :

"BATEAU"  
"BATEAUX"  
"YOURTE"  
"CAMPING CAR"  
"CAMPIGNE"  
"BUNGALOW"  
"BENGALOW"  
"TENTE"  
"NAVIRE"  
"NAVIRES"  
"VOILIER"  
"PENICHE"  
"FOURGONNETTE"  
"HABITATION MOBILE"  
"MOBIL HOM"  
"MOBIL HOMME"  
"MOBILE"  
"MOBILHOMME"  
"CARAVANNE"  
"CAMPING"  
"CAMPINGS"  
"CAMIONNETTE"  
"CAMIONNETTE"  
"CARAVANE"  
"POID LOURD"  
"POIDS LOURD"  
"MOBILHOME"



"MOBIL HOME"  
"ROUTIER"  
"VEHICULE"  
"ABRI DE JARDIN"  
"ALGECO"  
"CAMION"

Les précisions suivantes sont reclassées dans la modalité 3 de la question 32 : appartement ou maison :

"HLM"  
"PAVILLON"  
"PAVILLONS"  
"STUDIO"  
"STUDIOT"  
"STUDETTE"  
"VILLA"  
"APPART"  
"APPARTEMENT"  
"PRESBYTERE"  
"MAISON"  
"T1"  
"T2"  
"T3"  
"F1"  
"F2"  
"F3"  
"F4"  
"F 3"  
"F 4"  
"F5"  
"1 PIECE"  
"2 PIECES"  
"3 PIECES"  
"5 PIECES"  
"TYPE 3"  
"TYPE 4"  
"DE FONCTION"  
"FONCTION"  
"LOGEMENT DE TRAVAIL"  
"BARRAQUE"  
"3 CHAMBRES"  
"MEUBLE"  
"FERME"  
"FERMETTE"  
"CHATEAU"  
"CHAMBRE DE BONNE"  
"CHBRE DE BONNES"  
"CHAMBRE"  
"CHAMBRES"  
"ATELIER"  
"CABANE EN DUR"  
"LOGE GARDIEN"  
"GARDIEN"  
"APT DE TRAVAIL"  
"BOULANGERIE"  
"CASE"  
"IMMEUBLE"  
"COMMERCE"  
"MAISONNETTE"  
"MAISONNETTE"



"MAISOON DE CAMPAGNE"  
"MAGASIN"  
"DOMAINE"  
"GRANGE"  
"MANOIR"  
"OPAC"  
"MAISONS"  
"OPPAC"  
"MAS"  
"LCNAS GENDARMERIE"  
"LOGEMENT SOCIAL "  
"RESIDENCE"  
"RESIDENCES"  
"RESIDENSE"  
"VILLAGE"  
"CHALET"

Les précisions suivantes sont reclassées dans la modalité 4 de la question 32 : un autre type de logement :

"REFUGE EN MONTAGNE"  
"GROTTE"  
"GARAGE"  
"ABRI DE PECHE"  
"ABRIS DE PECHE"  
"CABANON"  
"CASOT"  
"LOCAL"

Les précisions suivantes sont reclassées dans une nouvelle modalité : modalité 5 de la question 32 : hotel, club de vacances chambre d'hôte :

"CLUB VACANCES"  
"CLUB MED"  
"CLUB DE VACANCE"  
"CLUB DE VACANCES"  
"CLUBS VACANCES"  
"HOTEL"  
"HOTELS"  
"CHAMBRE D HOTE"  
"CHAMBRES D HOTES"  
"VILLAGE VACANCES"  
"CENTRE DE VACANCES"  
"CENTRE VACANCES"  
"CENTRES VACANCES"  
"CENTRES DE VACANCES"  
"GUEST HOUSE"  
"NOVOTEL "  
"GITES"  
"GITE"

**La variable i\_natautlgtprecb :**

Les précisions suivantes alimentent une nouvelle modalité à la question 36 : le logement appartient à des amis :

"AMIS"  
"AMIES"  
"AMI"  
"AMICAL"  
"COPINE"  
"COPINES"



"COPAIN"  
"AMICAL"  
"AMIS"  
"AMICALEMENT"  
"POTE"  
"AMIE"  
"COPAINS"

Les précisions suivantes alimentent une nouvelle modalité à la question 36 : le logement appartient à des autres membres de la famille :

"GRAND MERE"  
"GRANDS PARENTS"  
"FAMILLE"  
"FAMILIALE"  
"FAMILLES"  
"FRERE"  
"FRERES"  
"SOEUR"  
"ONCLE"  
"TANTE"  
"BEAU FILS"  
"BEAU FRERE"  
"COUSIN"  
"COUSINE"  
"TATA"  
"COUSINES"  
"COUSINS"  
"SŒURS"  
"FAMILLIALE"  
"GENDRE"  
"GRAND-MERE"  
"GRANDPARENT"  
"GRANDS-PARENTS"  
"NEVEUX"  
"ONCLES"  
"PETIT ENFANT"  
"NIECE"  
"FAMILIAL"  
"FILLEULE"  
"MAMIE"  
"NEVEU"  
"PETIT FILS"  
"MARRAINE"

Les précisions suivantes sont reclassées dans la modalité 4 de la question 36 : le logement appartient aux parents ou beaux-parents :

"PARENTS"  
"BEAUX PARENTS"  
"PARENT"  
"PARENTAL"  
"PARENTALE"  
"BEAU PERE"  
"PAPA"  
"MAMAN"  
"PERE"  
"MERE"  
"BELLE MERE"

Les précisions suivantes sont reclassées dans la modalité 3 de la question 36 : le logement appartient aux enfants :



"FILS"  
"FILLE"  
"ENFANT"  
"FILLES"  
"ENFANTS"

Les précisions suivantes sont reclassées dans la modalité 1 de la question 36 : le logement appartient au répondant :

"MOI"  
"PROPRIETAIRE"  
"A MON PROPRIETAIRE"  
"AU PROPRIETAIRE"  
"BIEN PERSONNEL"  
"C EST MON LOGEMENT"  
"CELUI DU PROPRIETAIRE"  
"J EN SUIS LE PROPRIETAIRE"  
"LE MIEN"  
"LE MIENS"  
"LOCATION"  
"LOGEMENT QUE JE LOUE"  
"PROPRIETAIRES"  
"LOCATIF"  
"LOGEMENT LOUE"  
"QUE JE LOUE"  
"PROPRIETAIRE"  
"LOUER"  
"MA PROPRIETE"  
"MA RESIDENCE"  
"MON HABITATION"  
"MON LOGEMENT"  
"PERSONNEL"  
"PROPRIETE"  
"LOCATAIRE"  
"LOCATION"  
"A NOUS DEUX"  
"ACHAT EN COUPLE"  
"ACHETER A DEUX"  
"ACHETER EN COMMUN"  
"APPARTENANT AU COUPLE"  
"APPARTENANT A TOUS LES DEUX"  
"APPARTIENT AU COUPLE"  
"CONJOINTEMENT"  
"AUX 2 EPOUX"  
"LOGEMENT DU COUPLE"  
"COUPLE"  
"LE NOTRE"  
"A TOUS LES DEUX"  
"NOTRE LOGEMENT"  
"UN DE NOS LOGEMENTS"

Les précisions suivantes sont reclassées dans la modalité 2 de la question 36 : le logement appartient au conjoint du répondant :

"CONJOINT"  
"CONJOINTE"  
"EPOUX"  
"EPOUSE"  
"CONJOIN"  
"MARI"  
"A NOUS DEUX"  
"ACHAT EN COUPLE"



"ACHETER A DEUX"  
"ACHETER EN COMMUN"  
"APPARTENANT AU COUPLE"  
"APPARTENANT A TOUS LES DEUX"  
"APPARTIENT AU COUPLE"  
"CONJOINTEMENT"  
"AUX 2 EPOUX"  
"LOGEMENT DU COUPLE"  
"COUPLE"  
"LE NOTRE"  
"A TOUS LES DEUX"  
"NOTRE LOGEMENT"  
"UN DE NOS LOGEMENTS"

### **La variable raisonautlog**

indication vacances/ travail (logement lié à l'activité professionnelle ou aux vacances)

Lié aux vacances

"VACANCES"  
"TIME SHARE"  
"TIMESHARE"  
"VOYAGE"  
"VOYAGES"  
"CLUB MED"  
"GUEST HOUSE"  
"VACANCE"

Lié au travail

"TRAVAIL"  
"FONCTIONS"  
"PROFESSIONNEL"  
"CHANTIER"  
"CASERNE"  
"FONCTION"

Environ 300 précisions seulement n'ont pu être utilisées soit moins de 10% d'entre elles.

### **Correction de la question 32**

Création de la variable Q32natautlgt dans laquelle on intègre les précisions codées.

1) S'il n'y a pas de réponse, on distingue les non-réponses des « non concernées » avec la variable 31 : si q31lgt\_sec indique qu'il y a un autre logement (modalité 1) ou est en non réponse (modalité Z) la question 32 est en non réponse si la question 31 indique qu'il n'y a pas d'autre logement (modalité 2) la question 32 est mise en « non concernée » (valeur manquante).

2) S'il y a une seule case cochée qui correspond aux modalités 1, 2 ou 3, on garde la réponse de l'enquête.

3) Si deux cases sont cochées, dont la case 4 (autre type de logement) :  
- s'il n'y a pas de précision, on garde seulement la première case cochée  
- si la précision correspond à la première case cochée, on garde seulement la première case cochée.  
- si la précision correspond à une autre nature que la première case cochée, on code Q32natautlgt en 4 (autre).

4) Si seulement la case 4 est cochée (autre type de logement), on recode la variable selon la précision, en reclassant dans les modalités existantes et en créant de nouvelles modalités.



S'il n'y a pas de précision, on laisse la modalité 4 (autre, sans précision)

5) Si la variable `i_natautlgt` est codée en 88 (que des cases biffées) , on distingue les non-réponses des « non concernée » avec la variable 31.

6) Cas des autres réponses multiples : il s'agit de personnes qui ont plusieurs autres logements. Or on ne peut déterminer lequel des logements secondaires est le plus important. On décide de recoder `Q32natautlgt` en 4 (autre).

7) contrôle de cohérence finale avec la question 31 s'il n'y a pas d'autre logement la question 32 est mise en « non concernée ».

### **Correction de la question 34**

1) S'il n'y a pas de réponse, on distingue les non-réponses des « non concernées » avec la variable 31 : si `q31lgt_sec` indique qu'il a y un autre logement (modalité 1) ou est en non réponse (modalité Z) la question 32 est en non réponse si la question 31 indique qu'il n'y a pas d'autre logement (modalité 2) la question 32 est mise en « non concernée » (valeur manquante).

2) S'il y a une seule case cochée on garde la réponse de l'enquêté.

3) Si la variable synthétique `i_tpsautlgt` est codée en 88 (que des cases biffées) , on distingue les non-réponses des « non concernée » avec la variable 31.

5) Cas des réponses multiples : On recode tous les cas multiples en 5 (« Une partie du temps, en alternance »)

6) contrôle de cohérence finale avec la question 31 s'il n'y a pas d'autre logement la question 34 est mise en « non concernée ».

### **Correction de la question 35**

La question 35 accepte plusieurs réponses possibles. Cinq variables sont créées correspondant à chaque modalité. Elles peuvent toutes être renseignées.

La cohérence avec la question 31 est assurée :s'il n'y a pas d'autre logement d'après la question 31, les variables issues de la question 35 sont mises en « non concernée ».

### **Correction de la question 36**

La question 36 acceptant plusieurs réponses possibles on crée autant de variables que de réponses possibles (les 5 modalités d'origine + les 2 modalités créées à partir des précisions : amis/copains et autre membre de la famille).

Création de 7 variables :

`Q36propautlgt_ego` : un des autres logements appartient à l'enquêté.

`Q36propautlgt_ami` : un des autres logements appartient à un copain.

`Q36propautlgt_c` : un des autres logements appartient au conjoint de l'enquêté.

`Q36propautlgt_enf` : un des autres logements appartient à un enfant de l'enquêté.

`Q36propautlgt_par` : un des autres logements appartient au parent de l'enquêté.

`Q36propautlgt_aut_fam` : un des autres logements appartient à l'enquêté.

`Q36propautlgt_aut` : un des autres logements appartient à quelqu'un d'autres.

1) S'il n'y a pas de réponse, on distingue les non-réponses des « non concernées » avec la variable 31 : si `q31lgt_sec` indique qu'il a y un autre logement (modalité 1) ou est en non réponse (modalité Z) les 7 variables issues de la question 36 sont en non réponse. Si la question 31 indique qu'il n'y a pas



d'autre logement (modalité 2), les 7 variables de la question 36 sont mises en « non concernée » (valeur manquante).

2) S'il y a une seule case cochée qui correspond aux modalités 1 à 4, la variable correspondant à la modalité cochée est mise en « 1 » pour oui et les 6 autres en « 2 » pour non.

3) Si seulement la case 5 est cochée (autre), on code la variable correspondante à la précision à « 1 » pour oui et les 6 autres à « 2 » pour non. S'il n'y a pas de précision ou qu'elle n'est pas reconnue, on code Q36propautlgt\_aut = « 1 » et les 6 autres à « 2 ».

4) Si plusieurs cases sont cochées:

- si la case 5 est cochée et la précision correspond à une case cochée, on Q36propautlgt\_aut= « 2 » et les autres variables correspondant aux cases cochées à « 1 »

- sinon les variables correspondant aux cases cochées sont mises à « 1 » pour oui et les autres à « 2 ».

5) Si la variable i\_propautlgt est codée en 88 (que des cases biffées) , on distingue les non-réponses des « non concernée » avec la variable 31.

6) contrôle de cohérence finale avec la question 31 s'il n'y a pas d'autre logement la question 36 est mise en « non concernée ».



# Imputation aléatoire des données manquantes

## L'année de mise en couple (Q6annee c I)

Elle est imputée aléatoirement en fonction de l'âge, du dernier diplôme obtenu, des années éventuellement remplies de pacs et/ou de mariage et de l'année éventuellement remplie de l'ainé des enfants commun au couple actuel.

Programme d'imputation correspondant :

```
data rajout_imput ; set efldef.efl_individu_imput_nat
(keep= cabefl q7annee_pacs_i q8annee_mari_i );
proc sort; by cabefl;run;
data indiv0 ; merge efldef.efl_individu_v10
(keep= cabefl q3couple q4anai_c qlanai q6annee_c q7pacs q8mari
q6annee_c_x q0sexe dipl reg_code
q7annee_pacs q8annee_mari) rajout_imput; by cabefl;
/* if cabefl='1F00476612' then qlanai='1940';** valeur aberrante */
qlage=2010-qlanai;
length tage $ 8.;
if 18<=qlage<=21 then tage='18-21';else
if 22<=qlage<=25 then tage='22-25';else
if 26<=qlage<=30 then tage='26-30';else
if 31<=qlage<=35 then tage='31-35';else
if 36<=qlage<=40 then tage='36-40';else
if 41<=qlage<=45 then tage='41-45';else
if 46<=qlage<=50 then tage='46-50';else
if 51<=qlage<=55 then tage='51-55';else
if 56<=qlage<=60 then tage='56-60';else
if 61<=qlage<=65 then tage='61-65';else
if 66<=qlage<=70 then tage='66-70';else
if 71<=qlage<=75 then tage='71-75';else
if 76<=qlage then tage='76 et +';
length tage2 $ 8.;
if 18<=qlage<=21 then tage2='18-21';else
if 22<=qlage<=25 then tage2='22-25';else
if 26<=qlage<=30 then tage2='26-30';else
if 31<=qlage<=35 then tage2='31-35';else
if 36<=qlage<=40 then tage2='36-40';else
if 41<=qlage<=45 then tage2='41-45';else
if 46<=qlage<=50 then tage2='46-50';else
if 51<=qlage<=60 then tage2='51-60';else
if 61<=qlage then tage2='61 et +';
length diplc $ 1.;
if dipl in ('01','02','03') then diplc='1';else
if dipl in ('11','12','13') then diplc='2';else
if dipl in ('14','15','16') then diplc='3';else
if dipl in ('17','18') then diplc='4';
else diplc='Z';
eff=1;
if cabefl='1H00360501' then dipl='11';run;

proc freq data=indiv0; table q6annee_c tage2;run;
data enfant0;set efldef.efl_enfant_v9 (keep=cabefl e13_14anai e13_14part
e13_14cpart q6annee_c loca);
if e13_14part='1' and e13_14cpart='1';** enfants d'ego et du conjoint
actuel**;
if e13_14anai ne 9999;run;
```



```

data enfant; set enfant0;
if q6annee_c not in (.,9999) and e13_14anai-q6annee_c<0 then delete;run;
proc sort data=enfant; by cabefl e13_14anai; run;
data aine (rename=(e13_14anai=dnaine)); set enfant; by cabefl;
if first.cabefl;drop e13_14part e13_14cpart; run;
proc freq data=aine ;table dnaine;run;

*** q6annee_c **;
data indiv; merge indiv0 aine; by cabefl ;if q3couple in ('1','2');**242
273 inds **;
length annee_max 8. anneemax $12.;
if q7pacs='1' then
do;annee_max=q7annee_pacs_i;anneemax='q7annee_pacs';end;else
if q8mari='1' then
do;annee_max=q8annee_mari_i;anneemax='q8annee_mari';end;else
annee_max=.;
run;
proc freq data=indiv ;table q7annee_pacs_i q8annee_mari_i anneemax; run;
proc freq data=indiv ;table q7pacs q8mari q7annee_pacs q8annee_mari
q7annee_pacs_i q8annee_mari_i; run;

proc freq data=indiv ;table dnaine q6annee_c q7annee_pacs_i q8annee_mari_i
annee_max anneemax; run;
** 242 273 ind et 34941 Z à q6annee_c soit 14,4% **;

data indiv1; set indiv;
length attente 8. agegonaisaine 8.;
if anneemax in ('q7annee_pacs','q8annee_mari') and q6annee_c not in
(.,9999) then do;
if anneemax='q7annee_pacs' then attente=q7annee_pacs_i-q6annee_c;else
if anneemax='q8annee_mari' then attente=q8annee_mari_i-q6annee_c;end;else
attente=99;

if anneemax in ('q7annee_pacs','q8annee_mari') then do;
agegocontrat=annee_max-qlanai;
if dnaine ne . then do;
if dnaine>=annee_max then enfapres='oui'; else enfapres='non';end;
else if dnaine=. then enfapres=' ';
end;

if q6annee_c ne 9999 then agemiseencouple=q6annee_c-qlanai;

if dnaine ne . then agegonaisaine=dnaine-qlanai;
length tragegonaisaine $ 7.;
if 14<=agegonaisaine<=20 then tragegonaisaine='14-20';else
if 21<=agegonaisaine<=23 then tragegonaisaine='21-23';else
if 24<=agegonaisaine<=26 then tragegonaisaine='24-26';else
if 27<=agegonaisaine<=29 then tragegonaisaine='27-29';else
if 30<=agegonaisaine<=32 then tragegonaisaine='30-32';else
if 33<=agegonaisaine<=35 then tragegonaisaine='33-35';else
if 36<=agegonaisaine then tragegonaisaine='36 et +';
run;
proc freq data=indiv1 ;table attente dnaine enfapres agemiseencouple
agegocontrat; run;
proc freq data=indiv1 ;table tage tage2 agegonaisaine tragegonaisaine ;
run;

proc print data=indiv1 (where=(agegocontrat ne . and agegocontrat<14));
var cabefl q7annee_pacs_i q8annee_mari_i q8annee_mari qlanai q6annee_c
attente

```



```

dnaine agemiseencouple agegocontrat;run;

data pb; set indiv1; if attente ne . and attente<0;run;
proc print data=pb; var q7annee_pacs_i q8annee_mari_i q6annee_c ;run;
data indiv2; set indiv1;
q6annee_cc=q6annee_c;
if attente ne . and attente<0 then do;
q6annee_cc=annee_max;attente=0;;end;**110 cas**;
run;
proc freq data=indiv2 ;table q6annee_cc attente ; run;

proc sort data=indiv2; by cabefl;run;
proc freq data=indiv2;table q6annee_cc enfapres;run;
proc freq data=indiv2 (where=(q6annee_cc=9999));table q6annee_cc attente
enfapres;run;

** 1er cas : on a une date de pacs ou de mariage et pas
d'enfants **;
** 6 727 à imputer **;

data indiv21;set indiv2;**35 866 ind **;
if (q7annee_pacs_i ne . or q8annee_mari_i ne .) and enfapres=' ';
if anneemax='q7annee_pacs' and tage='26-30' and diplc='Z' then diplc='4';
run;
proc freq data=indiv21;table q6annee_cc attente anneemax tage diplc diplc
agemiseencouple agegocontrat;run; ** 24 273 non reponses pour q6couple_c**;
proc freq data=indiv21 (where=(attente ne 99));table q6annee_cc ;run;
data rep nonrep;set indiv21;
if attente=99 then output nonrep;
else output rep;
proc sort data=rep ; by anneemax tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table attente; run;

proc sort data=nonrep; by anneemax tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class anneemax tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class anneemax tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by anneemax tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by anneemax tage diplc;
alea=ranuni(14);
tri=anneemax!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc attente_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;

```



```

        else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename attente=attente_i;run;

data res0 ;merge nonrep rep3;
q6annee=annee_max-attente_i;
agemiseencouple=q6annee-qlanai;
if agemiseencouple <15 then attente_i=0;**284 **;
run;
proc freq data=res0;table attente_i agemiseencouple;run;

data res(drop=agemiseencouple q6annee); set res0;run;
proc freq data=res;table attente_i;run;
proc sort; by cabefl; run;

data indiv22; merge indiv21 (in=a) res(keep=cabefl attente_i) ; by
cabefl;if a;
if attente_i ne . then q6annee_c_i=annee_max-attente_i;else
if attente_i=. then q6annee_c_i=q6annee_c;
run;

proc freq data=indiv22; table attente_i dnaine q6annee_c_i ;run;**6 727
imputés **;

data verif; set res;
q6annee_c_i=annee_max-attente_i;
agemiseencouple=q6annee_c_i-qlanai;
proc freq data=verif; table attente_i q6annee_c_i agemiseencouple;run;

** 2ème cas : on a une date de pacs ou de mariage et enfants
nés après **;
** 17 546 à imputer **;

data indiv31;set indiv2;**126 994 ind **;
if (q7annee_pacs_i ne . or q8annee_mari_i ne .) and enfapres='oui';
if anneemax='q7annee_pacs' and tage='26-30' and diplc='Z' then diplc='4';
if tragegonaisaine='14-20' and anneemax='q8annee_mari' and tage='18-21' and
diplc in ('3','4') then diplc='3';
if tragegonaisaine='36 et +' and anneemax='q7annee_pacs' and tage='31-35'
and diplc in ('3','4') then diplc='3';
if tragegonaisaine='36 et +' and anneemax='q7annee_pacs' and tage='46-50'
and diplc in ('1','2') then diplc='1';
if tragegonaisaine='36 et +' and anneemax='q7annee_pacs' and tage='51-55'
and diplc in ('3','4') then diplc='3';
if tragegonaisaine='36 et +' and anneemax='q7annee_pacs' and tage='56-60'
and diplc in ('3','4') then diplc='3';
run;
proc freq data=indiv31;table q6annee_cc tragegonaisaine attente anneemax
tage diplc agemiseencouple agegocontrat;run; ** 17 546 non reponses pour
q6couple_c**;
proc freq data=indiv31 (where=(attente ne 99));table q6annee_cc ;run;
data rep nonrep;set indiv31;
if attente=99 then output nonrep;
else output rep;
proc sort data=rep ; by tragegonaisaine anneemax tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table attente; run;

proc sort data=nonrep; by tragegonaisaine anneemax tage diplc ;run;
*proc print data=nonrep (obs=10); run;

```



```

proc summary data=nonrep nway ; ** non-réponses** ;
class tragegonaisaine anneemax tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses** ;
class tragegonaisaine anneemax tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by tragegonaisaine anneemax tage diplc
;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tragegonaisaine anneemax tage diplc;
alea=ranuni(14);
tri=tragegonaisaine!!anneemax!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc attente_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename attente=attente_i;run;

data res0 ;merge nonrep rep3;
q6annee=annee_max-attente_i;
if dnaine ne . then pb=dnaine-q6annee;else pb=.;
agemiseencouple=q6annee-q1anai;
if agemiseencouple <15 then attente_i=0;**114 **;
run;
proc freq data=res0;table pb agemiseencouple;run;

data res (drop=agemiseencouple q6annee) ;set res0; run;

proc freq data=res;table attente_i;run;
proc sort; by cabefl; run;

data indiv32; merge indiv31 (in=a) res(keep=cabefl attente_i) ; by
cabefl;if a;
if attente_i ne . then q6annee_c_i=annee_max-attente_i;else
if attente_i=. then q6annee_c_i=q6annee_c;
run;

proc freq data=indiv32; table attente_i dnaine q6annee_c_i ;run;**17 546
imputés **;

/*****
3ème cas ceux qui n'ont pas de date de pacs ou de mariage mais une
dnaine + on a une date de pacs ou de mariage et enfants nés avant
46 096 individus et 7001 "9999" à q6couple_c soit 15%
*****/

data indiv41; set indiv2;
if ( (q7annee_pacs_i = . and q8annee_mari_i = .) or
((q7annee_pacs_i ne . or q8annee_mari_i ne .) and enfapres='non') )
and dnaine ne .;

```



```

if tragegonaisaine in ('14-20','21-23','24-26') and tage2 in ('46-50','56-
60','61 et +') and diplc in ('1','2') then do;
tragegonaisaine='14-20';tage2='46-50';diplc='1';end;
if tragegonaisaine in ('14-20','21-23','24-26') and tage2 in ('46-50','56-
60','61 et +') and diplc in ('3','4') then do;
tragegonaisaine='14-20';tage2='46-50';diplc='3';end;
if tragegonaisaine = '30-32' and tage2='61 et +' and diplc in ('2','3') then
diplc='2';
run;
proc freq data=indiv41;table q6annee_c tragegonaisaine tage2 diplc;run;
proc freq data=indiv41 (where=(q6annee_c=9999));table tage2 diplc;run;
data rep nonrep;set indiv41;
if q6annee_c=9999 then output nonrep;
else output rep;
proc sort data=rep ; by tragegonaisaine tage2 diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q6annee_c; run;

proc sort data=nonrep; by tragegonaisaine tage2 diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class tragegonaisaine tage2 diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class tragegonaisaine tage2 diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by tragegonaisaine tage2 diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tragegonaisaine tage2 diplc;
alea=ranuni(14);
tri=tragegonaisaine!!tage2!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q6annee_c_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q6annee_c=q6annee_c_i;run;

data res ;merge nonrep rep3;
if dnaine ne . then pb=dnaine-q6annee_c_i;else pb=.;**1837**;
if pb ne . and pb<0 then q6annee_c_i=dnaine-1;**1837 **;
agemiseencouple=q6annee_c_i-qlanai;
if agemiseencouple<15 then q6annee_c_i=dnaine-1;**260 **;
run;
Proc freq data=res;
tables pb agemiseencouple;run;
/*data verif; set res;
if dnaine ne . then pb=dnaine-q6annee_c_i;else pb=.;
agemiseencouple=q6annee_c_i-qlanai;
proc freq data=verif; table pb agemiseencouple ;run;*/

```



```

proc sort; by cabefl; run;

Proc freq data=res;
tables q6annee_c_i;run;

data indiv42; merge indiv41 (in=a) res(keep=cabefl q6annee_c_i) ; by
cabefl;if a;
if q6annee_c_i= . then q6annee_c_i=q6annee_c;run;

proc freq data=indiv42; table dnaine q6annee_c_i
agemiseencouple;run;**7001 imputés **;

** 4ème cas les restants : ni date de pacs, ni date mariage ,
ni enfant **;
** 33 417 inds et 3667 "9999" soit 11% **;

data indiv51 ;set indiv2;
if q7annee_pacs_i =. and q8annee_mari_i=. and dnaine=.;run;
proc freq data=indiv51;table q6annee_c ;run;
proc freq data=indiv51 (where=(q6annee_c=9999));table tage diplc;run;
data rep nonrep;set indiv51;
if q6annee_c=9999 then output nonrep;
else output rep;
proc sort data=rep ; by tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q6annee_c; run;

proc sort data=nonrep; by tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tage diplc;
alea=ranuni(14);
tri=tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q6annee_c_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q6annee_c=q6annee_c_i;run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

```



```

Proc freq data=res;
tables q6annee_c_i;run;

data indiv52;merge indiv51(in=a) res (keep=cabefl q6annee_c_i);by cabefl;if
a;
if q6annee_c_i=. then q6annee_c_i=q6annee_c;
agemiseencouple=q6annee_c_i-qlanai;
run;
Proc freq data=indiv52 ;
tables q6annee_c_i agemiseencouple;
run;

data indivi; set indiv22 (keep=cabefl q6annee_c_x q6annee_c q6annee_c_i)
indiv32 (keep=cabefl q6annee_c_x q6annee_c q6annee_c_i)
indiv42 (keep=cabefl q6annee_c_x q6annee_c q6annee_c_i)
indiv52 (keep=cabefl q6annee_c_x q6annee_c q6annee_c_i);
;

**indicateur : I_q6annee_c**;
length I_q6annee_c 8. q6anneeci $4.;
q6anneeci=put(q6annee_c_i,4.);
if q6annee_c='9999' then I_q6annee_c=2; else
if q6anneeci=q6annee_c_x then I_q6annee_c=0; else
I_q6annee_c=1;
run;

Proc freq data=indivi ;
tables q6annee_c_x q6annee_c I_q6annee_c q6annee_c_i;
run;

data test; set indivi;
if q6annee_c ne 9999 and q6annee_c ne q6annee_c_i;run;

data efldef.rajout_q6_bis;set indivi (keep=cabefl q6annee_c_i
I_q6annee_c);run;

proc freq data=efldef.rajout_q6_bis ;
table q6annee_c_i I_q6annee_c/missing;run;

** verif ***;
data verif; set indiv52 ;
**indicateur : I_q6annee_c**;
length I_q6annee_c 8. q6anneeci $4.;
q6anneeci=put(q6annee_c_i,4.);
if q6annee_c='9999' then I_q6annee_c=2; else
if q6anneeci=q6annee_c_x then I_q6annee_c=0; else
I_q6annee_c=1;
run;

data verif2;set verif;
if I_q6annee_c=2;
if dnaine ne . then pb=dnaine-q6annee_c_i;else pb=.;
agemiseencouple=q6annee_c_i-qlanai;
proc freq data=verif2 ; table pb agemiseencouple agegocontrat;run;

data verif2;set indivi;
if q6annee_c ne 9999 and q6annee_c ne q6annee_c_i;run;

```



## **L'année de naissance et le sexe des enfants (E13 14anai I et E13 14sexe I)**

Le sexe et l'année de naissance des enfants sont imputés ensemble par hotdeck.

1) Pour les enfants du couple actuel, les donneurs sont des enfants du couple actuel sans non-réponse pour le sexe et l'année de naissance et dont l'année de naissance est postérieure à la date de mise en couple.

C'est l'écart entre la date de mise en couple et la date naissance des enfants qui est imputé.

Les donneurs sont choisis dans les mêmes groupes que les receveurs. Tous les enfants d'un même individu sont imputés à partir des enfants d'un seul individu donneur.

Les groupes sont classés (4 % des enfants) par :  
tableau 13 ou 14  
tranche d'âge quinquennal  
diplôme sur 4 positions  
nombre d'enfants du couple du tableau 13 ou 14  
tranche d'âge quinquennal d'année de mise en couple

Pour les enfants restant (0,2 %), on réduit les critères :  
tableau 13 ou 14  
tranche d'âge décennal d'année de mise en couple  
nombre d'enfants du couple du tableau 13 ou 14

Pour les enfants restant (0,2 %), on réduit les critères :  
tableau 13 ou 14  
nombre d'enfants du couple du tableau 13 ou 14  
Ce sont généralement des enfants issus de familles très nombreuses.

2) Pour les enfants qui ne sont pas du couple actuel, les donneurs sont des enfants qui ne sont pas du couple actuel sans non-réponse pour le sexe et l'année de naissance et dont l'année de naissance est antérieure à la séparation ou au décès ou dont le parent n'a jamais vécu en couple.

C'est l'écart entre la date de séparation/décès et la date de naissance des enfants qui est imputé. Les parents qui n'ont jamais vécu en couple ont une date de séparation à 8888.

Les donneurs sont choisis dans les mêmes groupes que les receveurs. Tous les enfants d'un même individu sont imputés à partir des enfants d'un seul individu donneur.

Les groupes sont classés (4 % des enfants) par :  
tableau 13 ou 14  
type de séparation (décès, séparation, n'a jamais vécu en couple)  
tranche d'âge quinquennal  
diplôme sur 4 positions  
nombre d'enfants du couple du tableau 13 ou 14  
tranche d'âge quinquennal d'année de séparation

Pour les enfants restant (0,5 %), on réduit les critères :  
tableau 13 ou 14  
tranche d'âge décennal  
tranche d'âge décennal d'année de séparation  
nombre d'enfants du couple du tableau 13 ou 14

Pour les enfants restant (0,3 %), on réduit les critères :  
tableau 13 ou 14  
type de séparation (décès, séparation, n'a jamais vécu en couple)  
nombre d'enfants du couple du tableau 13 ou 14



Dans ce dernier cas, c'est l'âge auquel on est devenu parent qui est imputé. On en déduit ensuite l'année de naissance des enfants.

```
/***** cas 1 : les enfants sont du couple actuel, on utilise loca age
diplome nombre d'enfant année de mise en couple *****/

data individu0;set table.efl_individu_v9(keep=cabefl q3couple dipl cabefl
q6annee_c q10annee_s q10annee_d);

data couple;set imput.efl_individu_imput_nat(keep=cabefl q6annee_c_i
q10annee_s_i q10annee_d_i q10cause_r_i);

data individu;merge individu0 couple;by cabefl;

data enfant0;set table.efl_enfant_v9(keep=ident cabefl e13_14anai loca
e13_14part e13_14cpart e13_14part_x e13_14cpart_x qlage e13_14sexe
e13_14sexe_x);
if e13_14part='1' & e13_14cpart='1';

proc sort data=enfant0;by cabefl e13_14anai;

data cal;set enfant0;by cabefl;
if first.cabefl;
an_aine=e13_14anai;

data enfant2;merge enfant0 (in=a) individu cal(keep=cabefl an_aine);by
cabefl;
if a;
if q6annee_c_i ne 9999;
run;

data enfant3;set enfant2;by cabefl;
retain nb9l nb9h nbt1 nbth s9l s9h;
if first.cabefl then do;nb9l=0;nb9h=0;nbt1=0;nbth=0;s9l=0;s9h=0;end;
if loca='logemen' & e13_14anai=9999 then nb9l=nb9l+1;
if loca ne 'logemen' & e13_14anai=9999 then nb9h=nb9h+1;
if loca='logemen' & e13_14sexe='Z' then s9l=s9l+1;
if loca ne 'logemen' & e13_14sexe='Z' then s9h=s9h+1;
if loca='logemen' then nbt1=nbt1+1;
if loca ne 'logemen' then nbth=nbth+1;
if last.cabefl;

data enfant4;merge enfant2 enfant3(keep=cabefl nb9h nb9l nbt1 nbth s9l
s9h);by cabefl;
if loca='logemen' then do;
nb9=nb9l;
nbt=nbt1;
s9=s9l;
end;
if loca ne 'logemen' then do;
nb9=nb9h;
nbt=nbth;
s9=s9h;
end;
if dipl in ('01','02','03') then diplc='1';
if dipl in ('11','12','13') then diplc='2';
if dipl in ('14','15','16') then diplc='3';
if dipl in ('17','18') then diplc='4';
if dipl='ZZ' & qlage le 59 then diplc='3';
if dipl='ZZ' & qlage ge 60 then diplc='1';
```



```

if 18<=qlage<=29 then tage='18-29';
if 30<=qlage<=34 then tage='30-34';
if 35<=qlage<=39 then tage='35-39';
if 40<=qlage<=44 then tage='40-44';
if 45<=qlage<=49 then tage='45-49';
if 50<=qlage<=54 then tage='50-54';
if 55<=qlage<=59 then tage='55-59';
if 60<=qlage<=69 then tage='60-69';
if 70<=qlage<=79 then tage='70-79';
if 80<=qlage then tage='80-99';
if q3couple in ('1','2') then agem=q6annee_c_i-(2010-qlage);
if agem le 24 then agem2=1824;
if 25 le agem le 29 then agem2=2529;
if 30 le agem le 34 then agem2=3034;
if 35 le agem le 39 then agem2=3539;
if 40 le agem le 99 then agem2=4099;
if q6annee_c_i le 1949 then q6annee_c_i2=3049;
if 1950 le q6annee_c_i le 1959 then q6annee_c_i2=5059;
if 1960 le q6annee_c_i le 1969 then q6annee_c_i2=6059;
if 1970 le q6annee_c_i le 1979 then q6annee_c_i2=7079;
if 1980 le q6annee_c_i le 1984 then q6annee_c_i2=8084;
if 1985 le q6annee_c_i le 1989 then q6annee_c_i2=8589;
if 1990 le q6annee_c_i le 1991 then q6annee_c_i2=9091;
if 1992 le q6annee_c_i le 1993 then q6annee_c_i2=9293;
if 1994 le q6annee_c_i le 1999 then q6annee_c_i2=9499;
if 2000 le q6annee_c_i le 2004 then q6annee_c_i2=2004;
if 2005 le q6annee_c_i then q6annee_c_i2=2005;
if 18<=qlage<=39 then tage3='18-39';
if 40<=qlage<=49 then tage3='40-49';
if 50<=qlage<=59 then tage3='50-59';
if 60<=qlage then tage3='60-99';
if agem le 27 then agem3=1827;
if 28 le agem le 33 then agem3=2833;
if 34 le agem le 99 then agem3=3499;
if q6annee_c_i le 1949 then q6annee_c_i3=3049;
if 1950 le q6annee_c_i le 1959 then q6annee_c_i3=5059;
if 1960 le q6annee_c_i le 1969 then q6annee_c_i3=6059;
if 1970 le q6annee_c_i le 1979 then q6annee_c_i3=7079;
if 1980 le q6annee_c_i le 1989 then q6annee_c_i3=8089;
if 1990 le q6annee_c_i le 1999 then q6annee_c_i3=9099;
if 2000 le q6annee_c_i then q6annee_c_i3=2000;
if e13_14anai ne 9999 then agep=e13_14anai-(2010-qlage);
eff=1;run;

data rep nonrep;set enfant4;
if nb9>0 ! s9>0 then output nonrep;
if nb9=0 & s9=0 & (an_aine ge q6annee_c_i) then output rep;
run;
proc sort data=rep ; by loca tage diplc nbt q6annee_c_i2;
proc sort data=nonrep; by loca tage diplc nbt q6annee_c_i2;

proc summary data=nonrep nway ;class loca tage diplc nbt q6annee_c_i2;var
eff;output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ;class loca tage diplc nbt q6annee_c_i2;var
eff;output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by loca tage diplc nbt q6annee_c_i2;
if nbnonrep>nbrep;

```



```

proc sort data=nonrep;by loca tage diplc nbt q6annee_c_i2;

data nonrep2;merge nonrep verif(in=b keep=loca tage diplc nbt
q6annee_c_i2);by loca tage diplc nbt q6annee_c_i2;
if not b;
run;

proc summary data=nonrep2 nway ;class loca tage diplc nbt q6annee_c_i2;var
eff;output out=tabnonrep2 (drop=_type_ _freq_) sum=nbnonrep2;
proc print data=tabnonrep2; run;

data rep2;merge rep tabnonrep2;by loca tage diplc nbt q6annee_c_i2;
proc sort data=rep2; by cabefl nbt loca;run;

data rep3;set rep2;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!tage!!diplc!!nbt!!q6annee_c_i2;

data rep4;merge rep2 rep3(keep=cabefl nbt loca tri alea);by cabefl nbt
loca;
cabefl2=cabefl;
proc sort data=rep4; by tri alea;run;
data rep5;set rep4;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=nbnonrep2 then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;
rename qlage=qlage5;
rename e13_14sexe=e13_14sexe5;
run;

data res1 ;merge nonrep2 rep5(keep=e13_14anai5 q6annee_c_i5 qlage5 cabefl2
e13_14sexe5);
if q3couple in ('1','2') then e13_14anai_i=q6annee_c_i+(e13_14anai5-
q6annee_c_i5);
e13_14sexe_i=e13_14sexe5;
run;

/***** les enfants sont du couple actuel *****/

/**** cas 2 : les enfants sont du couple actuel, on utilise loca, nombre
d'enfant, année de mise en couple **/
proc sort data=nonrep;by loca tage diplc nbt q6annee_c_i2;
data nonrep2A;merge nonrep verif(in=b keep=loca tage diplc nbt
q6annee_c_i2);by loca tage diplc nbt q6annee_c_i2;
if b;
run;

proc summary data=nonrep2A nway ;class loca q6annee_c_i3 nbt;var eff;output
out=tabnonrep2A (drop=_type_ _freq_) sum=nbnonrep2A;
proc print data=tabnonrep2A; run;

proc summary data=rep nway ;class loca q6annee_c_i3 nbt;var eff;output
out=tabrepA(drop=_type_ _freq_) sum=nbrepA;
proc print data=tabrepA; run;

data verifA; merge tabrepA tabnonrep2A; by loca q6annee_c_i3 nbt;
if nbnonrep2A>nbrepA;

```



```

proc print;run;
proc sort data=nonrep2A;by loca q6annee_c_i3 nbt;

data nonrep2B;merge nonrep2A verifA(in=b keep= loca q6annee_c_i3 nbt);by
loca q6annee_c_i3 nbt;
if not b;

data nonrep2C;merge nonrep2A verifA(in=b keep= loca q6annee_c_i3 nbt);by
loca q6annee_c_i3 nbt;
if b;
run;

proc summary data=nonrep2B nway ;class loca q6annee_c_i3 nbt;var eff;output
out=tabnonrep2B (drop=_type_ _freq_) sum=nbnonrep2B;
proc print data=tabnonrep2B; run;
proc sort data=rep;by loca q6annee_c_i3 nbt;

data rep2B;merge rep tabnonrep2B;by loca q6annee_c_i3 nbt;
proc sort data=rep2B; by cabefl nbt loca;run;

data rep3B;set rep2B;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!q6annee_c_i3!!nbt;

data rep4B;merge rep2B rep3B(keep=cabefl nbt loca tri alea);by cabefl nbt
loca;
cabefl2=cabefl;
proc sort data=rep4B; by tri alea;run;

data rep5B;set rep4B;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=nbnonrep2B then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;
rename qlage=qlage5;
rename e13_14sexe=e13_14sexe5;
run;

data res1B ;merge nonrep2B rep5B(keep=e13_14anai5 q6annee_c_i5 qlage5
cabefl2 e13_14sexe5);
if q3couple in ('1','2') then e13_14anai_i=q6annee_c_i+(e13_14anai5-
q6annee_c_i5);
e13_14sexe_i=e13_14sexe5;
run;

/***** les enfants sont du couple actuel *****/

/**** cas 3 : les enfants sont du couple actuel, on utilise loca, nombre
d'enfant **/

proc summary data=nonrep2C nway ;class loca nbt;var eff;output
out=tabnonrep2C (drop=_type_ _freq_) sum=nbnonrep2C;
proc print data=tabnonrep2C; run;

proc summary data=rep nway ;class loca nbt;var eff;output
out=tabrepC(drop=_type_ _freq_) sum=nbrepC;
proc print data=tabrepC; run;

```



```

data verifC; merge tabrepC tabnonrep2C; by loca nbt;
if nbnonrep2C>nbrepC;
proc print;run;
proc sort data=nonrep2C;by loca nbt;
data nonrep2d;merge nonrep2C verifC(in=b keep= loca nbt);by loca nbt;
if not b;
run;

proc summary data=nonrep2d nway ;class loca nbt;var eff;output
out=tabnonrep2d (drop=_type_ _freq_) sum=nbnonrep2d;
proc print data=tabnonrep2d; run;
proc sort data=rep;by loca nbt;

data rep2d;merge rep tabnonrep2d;by loca nbt;
proc sort data=rep2d; by cabefl nbt loca;run;

data rep3d;set rep2d;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!nbt;

data rep4d;merge rep2d rep3d(keep=cabefl nbt loca tri alea);by cabefl nbt
loca;
cabefl2=cabefl;

proc sort data=rep4d; by tri alea;run;
data rep5d;set rep4d;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=nbnonrep2d then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;
rename qlage=qlage5;
rename e13_14sexe=e13_14sexe5;
run;

data res1d ;merge nonrep2d rep5d(keep=e13_14anai5 q6annee_c_i5 qlage5
cabefl2 e13_14sexe5);
if q3couple in ('1','2') then e13_14anai_i=q6annee_c_i+(e13_14anai5-
q6annee_c_i5);
e13_14sexe_i=e13_14sexe5;
run;

/***** les enfants ne sont pas du couple actuel *****/

/**** cas 4 : les enfants ne sont pas du couple actuel, on utilise loca,
type de séparation, age, diplome, nombre d'enfants, annee de séparation
décès **/

data Pindividu0;set table.efl_individu_v9(keep=cabefl q3couple dipl cabefl
q6annee_c q10annee_s q10annee_d);

data Pcouple;set imput.efl_individu_imput_nat(keep=cabefl q6annee_c_i
q10annee_s_i q10annee_d_i q10cause_r_i);

data Pindividu;merge Pindividu0 Pcouple;by cabefl;

```



```

data Penfant0;set table.ef1_enfant_v9(keep=ident cabefl e13_14anai loca
e13_14part e13_14cpart e13_14part_x e13_14cpart_x qlage e13_14sexe
e13_14sexe_x);
if e13_14part ne '1' ! e13_14cpart ne '1';
proc sort data=Penfant0;by cabefl e13_14anai;

data Pcal;set Penfant0;by cabefl;
if last.cabefl;
an_benj=e13_14anai;

data Penfant2;merge Penfant0 (in=a) Pindividu Pcal(keep=cabefl an_benj);by
cabefl;
if a;
if q3couple='4' ! (q3couple in ('1','2','3') & (.<q6annee_c_i<9999 !
.<q10annee_s_i<9999 ! .<q10annee_d_i<9999));
run;

data Penfant3;set Penfant2;by cabefl;
retain nb9l nb9h nbt1 nbth s9l s9h;
if first.cabefl then do;nb9l=0;nb9h=0;nbt1=0;nbth=0;s9l=0;s9h=0;end;
if loca='logemen' & e13_14anai=9999 then nb9l=nb9l+1;
if loca ne 'logemen' & e13_14anai=9999 then nb9h=nb9h+1;
if loca='logemen' & e13_14sexe='Z' then s9l=s9l+1;
if loca ne 'logemen' & e13_14sexe='Z' then s9h=s9h+1;
if loca='logemen' then nbt1=nbt1+1;
if loca ne 'logemen' then nbth=nbth+1;
if last.cabefl;

data Penfant4;merge Penfant2 Penfant3(keep=cabefl nb9l nb9h nbt1 nbth s9l
s9h);by cabefl;

if loca='logemen' then do;
nb9=nb9l;
nbt=nbt1;
s9=s9l;
end;

if loca ne 'logemen' then do;
nb9=nb9h;
nbt=nbth;
s9=s9h;
end;
if dipl in ('01','02','03') then diplc='1';
if dipl in ('11','12','13') then diplc='2';
if dipl in ('14','15','16') then diplc='3';
if dipl in ('17','18') then diplc='4';
if dipl='ZZ' & qlage le 59 then diplc='3';
if dipl='ZZ' & qlage ge 60 then diplc='1';
if 18<=qlage<=29 then tage='18-29';
if 30<=qlage<=34 then tage='30-34';
if 35<=qlage<=39 then tage='35-39';
if 40<=qlage<=44 then tage='40-44';
if 45<=qlage<=49 then tage='45-49';
if 50<=qlage<=54 then tage='50-54';
if 55<=qlage<=59 then tage='55-59';
if 60<=qlage<=64 then tage='60-64';
if 65<=qlage<=69 then tage='65-69';
if 70<=qlage<=74 then tage='70-74';
if 75<=qlage<=79 then tage='75-79';
if 80<=qlage<=84 then tage='80-84';
if 85<=qlage<=89 then tage='95-89';

```



```

if 90<=qlage<=94 then tage='90-94';
if 95<=qlage      then tage='95-99';
if .<q10annee_s_i<9999 then q6annee_c_i=q10annee_s_i;
if .<q10annee_d_i<9999 then q6annee_c_i=q10annee_d_i;
if q3couple='4'      then q6annee_c_i=8888;
agem=q6annee_c_i-(2010-qlage);
if      agem le 24 then agem2=1824;
if 25 le agem le 29 then agem2=2529;
if 30 le agem le 34 then agem2=3034;
if 35 le agem le 39 then agem2=3539;
if 40 le agem le 49 then agem2=4049;
if 50 le agem le 59 then agem2=5059;
if 60 le agem le 98 then agem2=6099;
if 99 le agem      then agem2=8888;
if      q6annee_c_i le 1949 then q6annee_c_i2=3049;
if 1950 le q6annee_c_i le 1959 then q6annee_c_i2=5059;
if 1960 le q6annee_c_i le 1969 then q6annee_c_i2=6059;
if 1970 le q6annee_c_i le 1974 then q6annee_c_i2=7074;
if 1975 le q6annee_c_i le 1979 then q6annee_c_i2=7579;
if 1980 le q6annee_c_i le 1984 then q6annee_c_i2=8084;
if 1985 le q6annee_c_i le 1989 then q6annee_c_i2=8589;
if 1990 le q6annee_c_i le 1994 then q6annee_c_i2=9094;
if 1995 le q6annee_c_i le 1999 then q6annee_c_i2=9599;
if 2000 le q6annee_c_i le 2004 then q6annee_c_i2=2004;
if 2005 le q6annee_c_i < 8888 then q6annee_c_i2=2005;
if 8888 le q6annee_c_i      then q6annee_c_i2=8888;
if q3couple in ('1','2') then type='coupl';
if q3couple =  '4'      then type='jamai';
if q10cause_r_i='1'    then type='autre';
if q10cause_r_i='2'    then type='autre';
if q3couple in ('1','2') then type2='coupl';
if q3couple =  '4'      then type2='jamai';
if q10cause_r_i='1'    then type2='separ';
if q10cause_r_i='2'    then type2='deces';
if 18<=qlage<=39 then tage3='18-39';
if 40<=qlage<=49 then tage3='40-49';
if 50<=qlage<=59 then tage3='50-59';
if 60<=qlage<=69 then tage3='60-69';
if 70<=qlage<=79 then tage3='70-79';
if 80<=qlage      then tage3='80-99';
if      agem le 27 then agem3=1827;
if 28 le agem le 33 then agem3=2833;
if 34 le agem le 99 then agem3=3499;
if      q6annee_c_i le 1949 then q6annee_c_i3=3049;
if 1950 le q6annee_c_i le 1959 then q6annee_c_i3=5059;
if 1960 le q6annee_c_i le 1969 then q6annee_c_i3=6059;
if 1970 le q6annee_c_i le 1979 then q6annee_c_i3=7079;
if 1980 le q6annee_c_i le 1989 then q6annee_c_i3=8089;
if 1990 le q6annee_c_i le 1999 then q6annee_c_i3=9099;
if 2000 le q6annee_c_i < 8888 then q6annee_c_i3=2000;
if 8888 le q6annee_c_i      then q6annee_c_i3=8888;
if e13_14anai ne 9999 then agep=e13_14anai-(2010-qlage);
eff=1;

data Prep Pnonrep;set Penfant4;
if nb9>0 ! s9>0 then output Pnonrep;
if (.<q6annee_c_i<9999 & nb9=0 & s9=0 & an_benj le sum(q6annee_c_i,1)) !
(.<q10annee_s_i<9999 & nb9=0 & s9=0 & an_benj le sum(q10annee_s_i,1)) !
(.<q10annee_d_i<9999 & nb9=0 & s9=0 & an_benj le sum(q10annee_d_i,1))
! (q3couple='4' & s9=0 & nb9=0) then output Prep;
run;

```



```

proc sort data=Prep ; by loca type tage diplc nbt q6annee_c_i2;
proc sort data=Pnonrep; by loca type tage diplc nbt q6annee_c_i2;

proc summary data=Pnonrep nway ;class loca type tage diplc nbt
q6annee_c_i2;var eff;output out=Ptabnonrep (drop=_type_ _freq_)
sum=Pnbnonrep;
proc print data=Ptabnonrep; run;

proc summary data=Prep nway ;class loca type tage diplc nbt
q6annee_c_i2;var eff;output out=Ptabrep(drop=_type_ _freq_) sum=Pnbrep;
proc print data=Ptabrep; run;

data Pverif; merge Ptabrep Ptabnonrep; by loca type tage diplc nbt
q6annee_c_i2;
if Pnbnonrep>Pnbrep;
run;
proc sort data=Pnonrep;by loca type tage diplc nbt q6annee_c_i2;
data Pnonrep2;merge Pnonrep Pverif(in=b keep=loca type tage diplc nbt
q6annee_c_i2);by loca type tage diplc nbt q6annee_c_i2;
if not b;
run;

proc summary data=Pnonrep2 nway ;class loca type tage diplc nbt
q6annee_c_i2;var eff;output out=Ptabnonrep2 (drop=_type_ _freq_)
sum=Pnbnonrep2;
proc print data=Ptabnonrep2;sum Pnbnonrep2;run;

data Prep2;merge Prep Ptabnonrep2;by loca type tage diplc nbt q6annee_c_i2;
proc sort data=Prep2; by cabefl nbt loca;run;

data Prep3;set Prep2;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!type!!tage!!diplc!!nbt!!q6annee_c_i2;
run;

data Prep4;merge Prep2 Prep3(keep=cabefl tri alea nbt loca);by cabefl nbt
loca;
cabefl2=cabefl;
proc sort data=Prep4; by tri alea;run;

data Prep5;set Prep4;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=Pnbnonrep2 then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;
rename qlage=qlage5;
rename e13_14sexe=e13_14sexe5;
run;

data Pres1 ;merge Pnonrep2 Prep5(keep=e13_14anai5 q6annee_c_i5 qlage5
cabefl2 e13_14sexe5);
e13_14anai_i=q6annee_c_i+(e13_14anai5-q6annee_c_i5);
e13_14sexe_i=e13_14sexe5;
run;

/***** les enfants ne sont pas du couple actuel *****/

```



```

/**** cas 5 : les enfants ne sont pas du couple actuel, on utilise loca,
age, nombre d'enfants annee de séparation décès **/

proc sort data=Pnonrep;by loca type tage diplc nbt q6annee_c_i2;
data Pnonrep2A;merge Pnonrep Pverif(in=b keep=loca type tage diplc nbt
q6annee_c_i2);by loca type tage diplc nbt q6annee_c_i2;
if b;
run;

proc summary data=Pnonrep2A nway ;class loca tage3 q6annee_c_i3 nbt;var
eff;output out=Ptabnonrep2A (drop=_type_ _freq_) sum=Pnbnonrep2A;
proc print data=Ptabnonrep2A; run;

proc summary data=Prep nway ;class loca tage3 q6annee_c_i3 nbt;var
eff;output out=PtabrepA(drop=_type_ _freq_) sum=PnbrepA;
proc print data=PtabrepA; run;

data PverifA; merge PtabrepA Ptabnonrep2A; by loca tage3 q6annee_c_i3 nbt;
if Pnbnonrep2A>PnbrepA;
proc print;run;
proc sort data=Pnonrep2A;by loca tage3 q6annee_c_i3 nbt;

data Pnonrep2B;merge Pnonrep2A PverifA(in=b keep= loca tage3 q6annee_c_i3
nbt);by loca tage3 q6annee_c_i3 nbt;
if not b;
run;
proc freq;tables loca tage3 q6annee_c_i3 nbt;run;

data Pnonrep2C;merge Pnonrep2A PverifA(in=b keep= loca tage3 q6annee_c_i3
nbt);by loca tage3 q6annee_c_i3 nbt;
if b;
run;

proc summary data=Pnonrep2B nway;class loca tage3 q6annee_c_i3 nbt;var
eff;output out=Ptabnonrep2B (drop=_type_ _freq_) sum=Pnbnonrep2B;
proc print data=Ptabnonrep2B;sum Pnbnonrep2B; run;

proc sort data=Prep;by loca tage3 q6annee_c_i3 nbt;
data Prep2B;merge Prep Ptabnonrep2B;by loca tage3 q6annee_c_i3 nbt;
proc sort data=Prep2B; by cabefl nbt loca;run;

data Prep3B;set Prep2B;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!tage3!!q6annee_c_i3!!nbt;

data Prep4B;merge Prep2B Prep3B(keep=cabefl nbt loca tri alea);by cabefl
nbt loca;
cabefl2=cabefl;

proc sort data=Prep4B; by tri alea;run;

data Prep5B;set Prep4B;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=Pnbnonrep2B then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;
rename qlage=qlage5;
rename e13_14sexe=e13_14sexe5;

```



```

run;

data Pres1B ;merge Pnonrep2B Prep5B(keep=e13_14anai5 q6annee_c_i5 qlage5
cabefl2 e13_14sexe5);
e13_14anai_i=q6annee_c_i+(e13_14anai5-q6annee_c_i5);
e13_14sexe_i=e13_14sexe5;
run;

/***** les enfants ne sont pas du couple actuel *****/

/**** cas 6 : les enfants ne sont pas du couple actuel, on utilise loca,
nombre d'enfants **/

proc summary data=Pnonrep2c nway ;class loca type2 nbt;var eff;output
out=Ptabnonrep2c (drop=_type_ _freq_) sum=Pnbnonrep2c;
proc print data=Ptabnonrep2c; run;

proc summary data=Prep nway ;class loca type2 nbt;var eff;output
out=Ptabrepc(drop=_type_ _freq_) sum=Pnbrepc;
proc print data=PtabrepA; run;

data Pverifc; merge Ptabrepc Ptabnonrep2c; by loca type2 nbt;
if Pnbnonrep2c>Pnbrepc;
proc print;run;

proc sort data=Pnonrep2c;by loca type2 nbt;

data Pnonrep2d;merge Pnonrep2c Pverifc(in=b keep= loca type2 nbt);by loca
type2 nbt;
if not b;
run;

proc summary data=Pnonrep2d nway;class loca type2 nbt;var eff;output
out=Ptabnonrep2d (drop=_type_ _freq_) sum=Pnbnonrep2d;
proc print data=Ptabnonrep2d;sum Pnbnonrep2d; run;

proc sort data=Prep;by loca type2 nbt;

data Prep2d;merge Prep Ptabnonrep2d;by loca type2 nbt;

proc sort data=Prep2d; by cabefl nbt loca;run;

data Prep3d;set Prep2d;by cabefl nbt loca;
if first.cabefl ! first.nbt ! first.loca;
alea=ranuni(14);
tri=loca!!type2!!nbt;

data Prep4d;merge Prep2d Prep3d(keep=cabefl nbt loca tri alea);by cabefl
nbt loca;
cabefl2=cabefl;

proc sort data=Prep4d; by tri alea;run;

data Prep5d;set Prep4d;by tri;
retain cpt;
if first.tri then cpt=0;
cpt=cpt+1;
if cpt<=Pnbnonrep2d then output;
rename e13_14anai=e13_14anai5;
rename q6annee_c_i=q6annee_c_i5;

```



```

rename qlage=qlage5;
rename agep=agep5;
rename e13_14sexe=e13_14sexe5;
run;

data Pres1d ;merge Pnonrep2d Prep5d(keep=e13_14anai5 q6annee_c_i5 qlage5
agep5 cabefl2 loca type2 e13_14sexe5);
e13_14anai_i=2010-qlage+agep5;
e13_14sexe_i=e13_14sexe5;
run;

proc sort data=res1 ;by ident;
proc sort data=res1b;by ident;
proc sort data=res1d;by ident;
proc sort data=Pres1 ;by ident;
proc sort data=Pres1b;by ident;
proc sort data=Pres1d;by ident;
proc sort data=enfant4;by ident;
proc sort data=Penfant4;by ident;

data result/*imput.efl_enfant_imput_nat(keep=ident e13_14anai_i
e13_14sexe_i)*/;merge
                                enfant4(in=a keep=ident e13_14anai qlage
loca cabefl E13_14PART E13_14CPART E13_14PART_x E13_14CPART_x
q6annee_c_i tage3 nbt nb9 agep e13_14sexe s9 q3couple
                                q6annee_c q10annee_s q10annee_s_i
q10annee_d q10annee_d_i)
                                res1 (in=b keep=ident e13_14anai_i cabefl2
e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
e13_14sexe_i e13_14sexe5)
                                res1b (in=c keep=ident e13_14anai_i cabefl2
e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
e13_14sexe_i e13_14sexe5)
                                res1d (in=d keep=ident e13_14anai_i cabefl2
e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
e13_14sexe_i e13_14sexe5)

                                Penfant4(in=a keep=ident e13_14anai qlage
loca cabefl E13_14PART E13_14CPART E13_14PART_x E13_14CPART_x
q6annee_c_i tage3 nbt nb9 agep an_benj e13_14sexe s9 q3couple
                                q6annee_c q10annee_s q10annee_s_i
q10annee_d q10annee_d_i)
                                Pres1 (in=b keep=ident e13_14anai_i
cabefl2 e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
type2 e13_14sexe_i e13_14sexe5)
                                Pres1b (in=c keep=ident e13_14anai_i
cabefl2 e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
type2 e13_14sexe_i e13_14sexe5)
                                Pres1d (in=d keep=ident e13_14anai_i
cabefl2 e13_14anai5 q6annee_c_i5 qlage5 tage3 nbt agem3 tage3 q6annee_c_i3
type2 agep5 e13_14sexe_i e13_14sexe5)
;by ident;
if a;

if e13_14anai ne 9999 then e13_14anai_i=e13_14anai;
if e13_14sexe ne 'Z' then e13_14sexe_i=e13_14sexe;
if e13_14anai_i>2011 then e13_14anai_i=2010;
if ident='1F00717987_1 ' then e13_14anai_i=1966;
if ident='1F00717987_2 ' then e13_14anai_i=1967;
if ident='1F00717987_3 ' then e13_14anai_i=1968;
if ident='1F00717987_4 ' then e13_14anai_i=1970;

```



```

if ident='1F00717987_5 ' then e13_14anai_i=1971;
if ident='1F00717987_6 ' then e13_14anai_i=1972;
if ident='1F00717987_7 ' then e13_14anai_i=1974;
if ident='1F00717987_8 ' then e13_14anai_i=1975;
if ident='1F00717987_9 ' then e13_14anai_i=1976;
if ident='1F00717987_10' then e13_14anai_i=1978;
if ident='1F00717987_11' then e13_14anai_i=1980;
if ident='1F00717987_12' then e13_14anai_i=1981;
if ident='1F00717987_13' then e13_14anai_i=1982;
if ident='1F00717987_14' then e13_14anai_i=1985;
if ident='1F00717987_15' then e13_14anai_i=1990;
if ident='1F00717987_16' then e13_14anai_i=1992;
if ident='1F00472005_3 ' then e13_14anai_i=1959;
if ident='1F00472005_4 ' then e13_14anai_i=1961;
if ident='1F00472005_5 ' then e13_14anai_i=1964;
if ident='1F00472005_6 ' then e13_14anai_i=1965;
if ident='1F00472005_7 ' then e13_14anai_i=1966;
if ident='1F00472005_8 ' then e13_14anai_i=1968;
if ident='1F00472005_9 ' then e13_14anai_i=1969;
if ident='1F00472005_10' then e13_14anai_i=1970;
if ident='1F00472005_11' then e13_14anai_i=1971;
if ident='1F00472005_12' then e13_14anai_i=1972;
if ident='1F00472005_13' then e13_14anai_i=1973;
if ident='1F00472005_14' then e13_14anai_i=1975;
if ident='1F00472005_15' then e13_14anai_i=1980;
if ident='1F00591318_1 ' then e13_14anai_i=1954;
if ident='1F00591318_2 ' then e13_14anai_i=1956;
if ident='1F00591318_3 ' then e13_14anai_i=1957;
if ident='1F00591318_4 ' then e13_14anai_i=1960;
if ident='1F00591318_5 ' then e13_14anai_i=1961;
if ident='1F00591318_6 ' then e13_14anai_i=1962;
if ident='1F00591318_7 ' then e13_14anai_i=1964;
if ident='1F00591318_8 ' then e13_14anai_i=1967;
if ident='1F00591318_9 ' then e13_14anai_i=1968;
if ident='1F00591318_10' then e13_14anai_i=1970;
if ident='1F00591318_11' then e13_14anai_i=1973;
if ident='1F00591318_12' then e13_14anai_i=1976;
if e13_14anai_i in (.,9999) then do; /*** 137 cas **/
if qlage ge 40 then e13_14anai_i=2010-qlage+40-int(20*ranuni(14));
if 30 le qlage le 40 then e13_14anai_i=2010-qlage+30-int(10*ranuni(14));
if 25 le qlage le 29 then e13_14anai_i=2010-qlage+25-int(5*ranuni(14));
if 18 le qlage le 24 then e13_14anai_i=2010;
end;
if e13_14sexe_i=' ' then do; /*** 147 cas ***/
x=int(2*ranuni(14));
if x=0 then e13_14sexe_i='1';
if x=1 then e13_14sexe_i='2';
end;
if e13_14anai_i ne 9999 then agep_i=e13_14anai_i-(2010-qlage);
if -10 le q6annee_c_i-e13_14anai_i le -1 & e13_14anai=9999 & e13_14cpart ne
'1' then do; /*** 305 cas **/
e13_14anai_i=e13_14anai_i-10;
agep_i=agep_i-10;
end;
if -20 le q6annee_c_i-e13_14anai_i le -11 & e13_14anai=9999 & e13_14cpart
ne '1' then do; /*** 49 cas **/
e13_14anai_i=e13_14anai_i-20;
agep_i=agep_i-20;
end;
if -40 le q6annee_c_i-e13_14anai_i le -21 & e13_14anai=9999 & e13_14cpart
ne '1' then do; /*** 7 cas **/

```



```

e13_14anai_i=e13_14anai_i-34;
agep_i=agep_i-34;
end;
if .<agep_i le 4 & e13_14anai=9999 then do; /***** 11 cas ***/
e13_14anai_i=e13_14anai_i+12;
agep_i=agep_i+12;
end;
if 5 le agep_i le 9 & e13_14anai=9999 then do; /*** 57 cas **/
e13_14anai_i=e13_14anai_i+10;
agep_i=agep_i+10;
end;
if 10 le agep_i le 14 & e13_14anai=9999 then do; /*** 480 cas **/
e13_14anai_i=e13_14anai_i+5;
agep_i=agep_i+5;
end;
if 15 le agep_i le 16 & e13_14anai=9999 then do; /**** 747 cas ***/
e13_14anai_i=e13_14anai_i+3;
agep_i=agep_i+3;
end;
if 51 le agep_i le 55 & e13_14anai=9999 then do; /*** 110 cas **/
e13_14anai_i=e13_14anai_i-6;
agep_i=agep_i-6;
end;
if 56 le agep_i le 90 & e13_14anai=9999 then do; /*** 106 cas ***/
e13_14anai_i=e13_14anai_i-13;
agep_i=agep_i-13;
end;
if e13_14anai ne 9999 then ageenf=2010-e13_14anai;
if e13_14anai_i ne 9999 then ageenf_i=2010-e13_14anai_i;
if .< ageenf le 09 then ageenf2='0009';
if 10 le ageenf le 14 then ageenf2='1014';
if 15 le ageenf le 17 then ageenf2='1517';
if 18 le ageenf le 20 then ageenf2='1820';
if 21 le ageenf le 24 then ageenf2='2124';
if 25 le ageenf le 29 then ageenf2='2529';
if 30 le ageenf le 34 then ageenf2='3034';
if 35 le ageenf le 39 then ageenf2='3539';
if 40 le ageenf le 49 then ageenf2='4049';
if 50 le ageenf le 59 then ageenf2='5059';
if 60 le ageenf le 69 then ageenf2='6069';
if 70 le ageenf then ageenf2='7079';
if e13_14anai = 9999 then ageenf2='ZZZZ';
if .< ageenf_i le 09 then ageenf_i2='0009';
if 10 le ageenf_i le 14 then ageenf_i2='1014';
if 15 le ageenf_i le 17 then ageenf_i2='1517';
if 18 le ageenf_i le 20 then ageenf_i2='1820';
if 21 le ageenf_i le 24 then ageenf_i2='2124';
if 25 le ageenf_i le 29 then ageenf_i2='2529';
if 30 le ageenf_i le 34 then ageenf_i2='3034';
if 35 le ageenf_i le 39 then ageenf_i2='3539';
if 40 le ageenf_i le 49 then ageenf_i2='4049';
if 50 le ageenf_i le 54 then ageenf_i2='5054';
if 55 le ageenf_i le 59 then ageenf_i2='5559';
if 60 le ageenf_i le 69 then ageenf_i2='6069';
if 70 le ageenf_i then ageenf_i2='7079';
if 18 le qlage le 19 then qlage2='1819';
if 20 le qlage le 29 then qlage2='2029';
if 30 le qlage le 39 then qlage2='3039';
if 40 le qlage le 49 then qlage2='4049';
if 50 le qlage le 59 then qlage2='5059';
if 60 le qlage le 69 then qlage2='6069';

```



```

if 70 le qlage le 79 then qlage2='7079';
if 80 le qlage le 89 then qlage2='8089';
if 90 le qlage le 99 then qlage2='9099';
if .< agep le 17 then agep2='0017';
if 18 le agep le 24 then agep2='1824';
if 25 le agep le 29 then agep2='2529';
if 30 le agep le 34 then agep2='3034';
if 35 le agep le 39 then agep2='3539';
if 40 le agep le 44 then agep2='4044';
if 45 le agep then agep2='4599';
if .< agep_i le 17 then agep_i2='0017';
if 18 le agep_i le 24 then agep_i2='1824';
if 25 le agep_i le 29 then agep_i2='2529';
if 30 le agep_i le 34 then agep_i2='3034';
if 35 le agep_i le 39 then agep_i2='3539';
if 40 le agep_i le 44 then agep_i2='4044';
if 45 le agep_i then agep_i2='4599';

/** verification **/
if e13_14part='1' & e13_14cpart = '1' & e13_14anai=9999 & e13_14anai_i ne
9999 & q6annee_c_i-e13_14anai_i>0 then erreur='oui';
if e13_14part='1' & e13_14cpart = '1' & e13_14anai=9999 & e13_14anai_i ne
9999 & q6annee_c_i-e13_14anai_i le 0 then erreur='non';
if e13_14part='1' & e13_14cpart = '1' & e13_14anai_i ne 9999 & q6annee_c_i-
e13_14anai_i>0 then erreur2='oui';
if e13_14part='1' & e13_14cpart = '1' & e13_14anai_i ne 9999 & q6annee_c_i-
e13_14anai_i le 0 then erreur2='non';

if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai=9999 & e13_14anai_i
ne 9999 & q10annee_s not in (.,9999) & q10annee_s_i-e13_14anai_i<0 then
erreur3='oui';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai=9999 & e13_14anai_i
ne 9999 & q10annee_s not in (.,9999) & q10annee_s_i-e13_14anai_i ge 0 then
erreur3='non';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai ne 9999 &
e13_14anai_i ne 9999 & q10annee_s not in (.,9999) & q10annee_s_i-
e13_14anai_i<0 then erreur5='oui';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai ne 9999 &
e13_14anai_i ne 9999 & q10annee_s not in (.,9999) & q10annee_s_i-
e13_14anai_i ge 0 then erreur5='non';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai=9999 & e13_14anai_i
ne 9999 & q10annee_d not in (.,9999) & q10annee_d_i-e13_14anai_i<0 then
erreur7='oui';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai=9999 & e13_14anai_i
ne 9999 & q10annee_d not in (.,9999) & q10annee_d_i-e13_14anai_i ge 0 then
erreur7='non';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai ne 9999 &
e13_14anai_i ne 9999 & q10annee_d not in (.,9999) & q10annee_d_i-
e13_14anai_i<0 then erreur9='oui';
if e13_14part ne '1' ! e13_14cpart ne '1' & e13_14anai ne 9999 &
e13_14anai_i ne 9999 & q10annee_d not in (.,9999) & q10annee_d_i-
e13_14anai_i ge 0 then erreur9='non';
run;

/** verification **/
proc freq data=result;where loca ne 'logemen';tables ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 18 le qlage le 29;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 30 le qlage le 39;tables
ageenf2*ageenf_i2;run;

```



```

proc freq data=result;where loca ne 'logemen' & 40 le qlage le 49;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 50 le qlage le 59;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 60 le qlage le 69;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 70 le qlage le 79;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 80 le qlage le 89;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' & 90 le qlage le 99;tables
ageenf2*ageenf_i2;run;
proc freq data=result;where loca ne 'logemen' ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 18 le qlage le 29 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 30 le qlage le 39 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 40 le qlage le 49 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 50 le qlage le 59 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 60 le qlage le 69 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 70 le qlage le 79 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 80 le qlage le 89 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 90 le qlage le 99 ;tables
ageenf2*agep_i2/missing;run;
proc freq data=result;where loca ne 'logemen' & 60 le qlage le 69;tables
agep2;run;
proc freq data=result;where loca ne 'logemen' & 60 le qlage le 69;tables
agep_i2/missing;run;

proc print data=pres1B;where loca ne 'logemen' & 30 le qlage le 39 &
(e13_14part ne '1' ! e13_14cpart ne '1');
var loca ident e13_14part e13_14cpart e13_14anai e13_14anai_i e13_14sexe
e13_14sexe_i qlage q6annee_c_i q6annee_c_i5 e13_14anai5 e13_14sexe5
qlage5 nbt cabefl2;run;

```



**Temps passé dans le logement et autre(s) logement(s) éventuel(s) des enfants (E13tps1gt I, E13aut\_lgt I, E13lgt aut par I, E13lgt etude I, E13lgt autre I)**

**Dans un premier temps, on impute l'ensemble de ces variables pour les enfants ne vivant pas avec leur deux parents.**

En effet il ne peut y avoir de non-réponse qu'a l'ensemble de ces variables après apurement :

E13tps1gt

E13aut\_lgt

	Fréquence			Total
	Pourcentage			
	Pctage en ligne		Z	
	1	2		
	Pctage en col.			
1	0	225126	0	225126
	0.00	82.68	0.00	82.68
	0.00	100.00	0.00	
	0.00	100.00	0.00	
2	14301	0	0	14301
	5.25	0.00	0.00	5.25
	100.00	0.00	0.00	
	31.70	0.00	0.00	
3	18587	0	0	18587
	6.83	0.00	0.00	6.83
	100.00	0.00	0.00	
	41.20	0.00	0.00	
4	12227	0	0	12227
	4.49	0.00	0.00	4.49
	100.00	0.00	0.00	
	27.10	0.00	0.00	
Z	0	0	2054	2054
	0.00	0.00	0.75	0.75
	0.00	0.00	100.00	
	0.00	0.00	100.00	
Total	45115	225126	2054	272295
	16.57	82.68	0.75	100.00

Quand, parmi les mineurs d'une fratrie d'enfants d'unions précédentes un des enfants mineurs est répondant on affecte à l'ensemble des enfants mineurs non répondants de la fratrie la réponse d'un des répondants.

Dans les autres cas toute la fratrie d'enfants mineurs d'unions précédentes est non-répondante, on impute aléatoirement l'ainé selon son âge et le type de rupture puis on affecte à l'ensemble des enfants mineurs non répondants de la fratrie la réponse imputée à l'ainé.

Les majeurs sont imputées aléatoirement selon l'âge, le sexe et la cause de la rupture parmi l'ensemble des majeurs ne vivant pas avec leur deux parents.

Il reste ensuite à répartir uniquement la variable E13tps1gt codée en « 4 » c'est-à-dire pas tout le temps mais quotité non précisée entre les modalités « 2 » : la moitié du temps ou plus et « 3 » moins de la moitié du temps. Le principe est le même que précédemment.

La nature de logement peut ne pas être précisée même si l'on sait qu'il y a un autre logement. Elle est ensuite imputée pour ces enfants parmi les enfants ne vivant pas avec leur deux parents et ayant plusieurs logement en fonction de des réponses des autres membres de la fratrie, de l'âge, du sexe et de la cause de la rupture sur un principe similaire.



Dans un second temps, on répartit pour les enfants vivant pas avec leur deux parents les temps passé dans le logement différents de tout le temps sans précision (modalité '4') entre la moitié du temps ou plus (modalité '2') ou moins de la moitié du temps (modalité '3') selon l'âge et la nature de l'autre logement..

Programme d'imputation correspondant :

```

/*début imputation : l'imputation ne se fait que sur des enfants qui ne
vivent pas avec leur deux parents
on crée une table avec l'ensemble de ces enfants*/

data enfpar_sep3;
set enf; /*table EFL_enfant après imputation de l'année de naissance et du
sexe*/
where partcpart in ("12", "21", "1 ");
drop partcpart;
proc sort; by ident;
run;
proc freq data=enfpar_sep3 (where=(e13tpsigt="Z" ));
tables cabefl / noprint out=nonrep;
run;
/* on impute aléatoirement le temps passé dans le logement selon la cause
de la rupture, le sexe et l'âge les cas des majeurs,
des situations où il n'y a qu'un enfant et qu'il est non répondant*/
/* pour cela on compte le nombre de non réponses*/
data nonrep2;
set nonrep;
nbnr=count;
keep cabefl nbnr;
run;
proc freq data=enfpar_sep3 (where=(loca="logemen" ));
tables cabefl / noprint out=nbenf;
run;
/* pour cela on compte le nombre d'enfants*/
data nbenf2;
set nbenf;
nbenf=count;
keep cabefl nbenf;
run;
proc sort data=enfpar_sep3;
by cabefl;
run;

/*on apparie ces comptages*/
data enfpar_sep4;
merge enfpar_sep3 nonrep2 nbenf2;
by cabefl;
run;
proc freq data=enfpar_sep4;
tables nbnr*nbenf;
run;
data enfpar_seplog;
set enfpar_sep4;
where loca="logemen" ;
run;
proc freq data=enfpar_seplog;
tables e13tpsigt;
run;

/*quand il y a moins de non réponse que d'enfants de parents séparées on
donne une des réponses aléatoirement

```



```

à l'ensemble des cas de non-réponses*/
data pasttnonrep;
set enfpar_seplog;
where (nbnr=1 and nbenf>1) or (nbnr=2 and nbenf>2) or (nbnr=3 and nbenf>3)
or (nbnr=4 and nbenf>4)
or (nbnr=5 and nbenf>5) or (nbnr=6 and nbenf>6) and E13_14anai_I<1993;
if e13tpslgt="Z" then delete;
e13tpslgt_i=e13tpslgt;
e13aut_lgt_i=e13aut_lgt;
e13lgt_aut_par_i=e13lgt_aut_par;
e13lgt_autre_i=e13lgt_autre;
e13lgt_etude_i=e13lgt_etude;
keep cabefl e13tpslgt_i e13aut_lgt_i e13lgt_aut_par_i e13lgt_autre_i
e13lgt_etude_i;
run;
proc sort data=pasttnonrep nodupkeys;
by cabefl e13tpslgt_i;
run;
proc freq data=pasttnonrep;
tables e13tpslgt_i;
run;
proc sort data=enfpar_seplog;
by cabefl ;
run;
data enfpar_seplog2;
merge enfpar_seplog pasttnonrep;
by cabefl;
run;
proc sort data=enfpar_seplog2;
by q10cause_r_i e13_14sexe_i e13_14anai_I ;
run;

data enfpar_seplog3;
set enfpar_seplog2;
e13tpslgt2=lag(e13tpslgt);
e13aut_lgt2=lag(e13aut_lgt);
e13lgt_aut_par2=lag(e13lgt_aut_par);
e13lgt_autre2=lag(e13lgt_autre);
e13lgt_etude2=lag(e13lgt_etude);
e13tpslgt3=lag(e13tpslgt2);
e13aut_lgt3=lag(e13aut_lgt2);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);

if E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1) then do;
if e13tpslgt='Z' and e13tpslgt2 ne "Z" then do;
e13tpslgt_i2=e13tpslgt2;
e13aut_lgt_i2=e13aut_lgt2;
e13lgt_aut_par_i2=e13lgt_aut_par2;
e13lgt_autre_i2=e13lgt_autre2;
e13lgt_etude_i2=e13lgt_etude2;
end;

if e13tpslgt='Z' and e13tpslgt2 = "Z" then do;
e13tpslgt_i2=e13tpslgt3;
e13aut_lgt_i2=e13aut_lgt3;
e13lgt_aut_par_i2=e13lgt_aut_par3;
e13lgt_autre_i2=e13lgt_autre3;
e13lgt_etude_i2=e13lgt_etude3;
end;

```



```

if e13tpslgt ne 'Z' then do;
e13tpslgt_i2=e13tpslgt;
e13aut_lgt_i2=e13aut_lgt;
e13lgt_aut_par_i2=e13lgt_aut_par;
e13lgt_autre_i2=e13lgt_autre;
e13lgt_etude_i2=e13lgt_etude;
end;
end;

else if e13tpslgt="Z" and E13tpslgt_i ne " " then do;
e13tpslgt_i2=e13tpslgt_i;
e13aut_lgt_i2=e13aut_lgt_i;
e13lgt_aut_par_i2=e13lgt_aut_par_i;
e13lgt_autre_i2=e13lgt_autre_i;
e13lgt_etude_i2=e13lgt_etude_i;
end;

else do;
e13tpslgt_i2=e13tpslgt;
e13aut_lgt_i2=e13aut_lgt;
e13lgt_aut_par_i2=e13lgt_aut_par;
e13lgt_autre_i2=e13lgt_autre;
e13lgt_etude_i2=e13lgt_etude;
end;
drop e13tpslgt2 e13aut_lgt2 e13lgt_aut_par2 e13lgt_autre2 e13lgt_etude2
e13tpslgt3 e13aut_lgt3 e13lgt_aut_par3 e13lgt_autre3 e13lgt_etude3;
run;

data enfpar_seplog4;
set enfpar_seplog3;
if q10cause_r_i="2" and e13tpslgt_i2='Z' then do;
e13tpslgt_i2="1";
e13aut_lgt_i2="2";
e13lgt_aut_par_i2=" ";
e13lgt_autre_i2=" ";
e13lgt_etude_i2=" ";
end;
run;

proc freq data=enfpar_seplog4;
tables e13tpslgt_i2 e13tpslgt_i2*nbnr
e13tpslgt_i2*q10cause_r_i/*e13_14sexe_i*e13tpslgt_i*q10cause_r_i*/;
run;
proc freq data=enfpar_seplog4 (where=(e13tpslgt_i2="Z"));
tables nbnr*nbenf q10cause_r_i/*e13_14sexe_i*e13tpslgt_i*q10cause_r_i*/
e13_14anai_i;
run;

/*pour les mineurs restants toutes la fratrie qui n'est pas du couple
actuel est non répondante on impute l'ainé aléatoirement
par cause de la rupture,sexe et âge et on donne à toute la fratrie la même
chose*/
proc sort data=enfpar_seplog4;
by cabefl e13_14anai_i;
run;
proc sort data=enfpar_seplog4 nodupkeys nodup out=aine;
by cabefl;
run;
proc sort data=aine;
by q10cause_r_i e13_14sexe_i e13_14anai_I ;

```



```

run;
data aine2;
set aine;
e13tpslgt2=lag(e13tpslgt);
e13aut_lgt2=lag(e13aut_lgt);
e13lgt_aut_par2=lag(e13lgt_aut_par);
e13lgt_autre2=lag(e13lgt_autre);
e13lgt_etude2=lag(e13lgt_etude);
e13tpslgt3=lag(e13tpslgt2);
e13aut_lgt3=lag(e13aut_lgt2);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);

if e13tpslgt_i2='Z' and e13tpslgt2 ne "Z" then do;
e13tpslgt_i3=e13tpslgt2;
e13aut_lgt_i3=e13aut_lgt2;
e13lgt_aut_par_i3=e13lgt_aut_par2;
e13lgt_autre_i3=e13lgt_autre2;
e13lgt_etude_i3=e13lgt_etude2;
end;

if e13tpslgt_i2='Z' and e13tpslgt2 = "Z" then do;
e13tpslgt_i3=e13tpslgt3;
e13aut_lgt_i3=e13aut_lgt3;
e13lgt_aut_par_i3=e13lgt_aut_par3;
e13lgt_autre_i3=e13lgt_autre3;
e13lgt_etude_i3=e13lgt_etude3;
end;

if e13tpslgt_i2 ne 'Z' then do;
e13tpslgt_i3=e13tpslgt_i2;
e13aut_lgt_i3=e13aut_lgt_i2;
e13lgt_aut_par_i3=e13lgt_aut_par_i2;
e13lgt_autre_i3=e13lgt_autre_i2;
e13lgt_etude_i3=e13lgt_etude_i2;
end;

drop e13tpslgt2 e13aut_lgt2 e13lgt_aut_par2 e13lgt_autre2 e13lgt_etude2
e13tpslgt3 e13aut_lgt3 e13lgt_aut_par3 e13lgt_autre3 e13lgt_etude3;
run;

proc freq data=aine2;
tables e13tpslgt_i2 * e13tpslgt_i3;
run;
data ainefi;
set aine2;
keep cabefl e13tpslgt_i3 e13aut_lgt_i3 e13lgt_aut_par_i3 e13lgt_autre_i3
e13lgt_etude_i3;
proc sort;
by cabefl;
run;
data enfpar_seplog5;
merge enfpar_seplog4 ainefi;
by cabefl;
run;

data enfpar_seplog6;
set enfpar_seplog5;
if e13tpslgt_i2 ne 'Z' then do;
e13tpslgt_i4=e13tpslgt_i2;

```



```

e13aut_lgt_i4=e13aut_lgt_i2;
e13lgt_aut_par_i4=e13lgt_aut_par_i2;
e13lgt_autre_i4=e13lgt_autre_i2;
e13lgt_etude_i4=e13lgt_etude_i2;
end;
if e13tpslgt_i2 = 'Z' then do;
e13tpslgt_i4=e13tpslgt_i3;
e13aut_lgt_i4=e13aut_lgt_i3;
e13lgt_aut_par_i4=e13lgt_aut_par_i3;
e13lgt_autre_i4=e13lgt_autre_i3;
e13lgt_etude_i4=e13lgt_etude_i3;
end;
if e13_14anai_i<1993 then mineur=2;
else mineur=1;
run;
proc freq data=enfpar_seplog6;
tables e13tpslgt_i4*e13tpslgt e13tpslgt_i4*mineur;
run;
proc freq data=enfpar_seplog6 (where=(e13tpslgt ne "Z"));
tables e13tpslgt;
run;

/* toute la non-réponse à e13tpslgt est imputé mais avec des 4*/
/*On répartit les 4 entre 2 et 3*/

data psttletps;
set enfpar_seplog6;
where e13tpslgt_i4 in ("2", "3", "4");
run;
proc sort data=psttletps;
by e13_14sexe_i e13_14anai_I ;
run;
data psttletps2;
set psttletps;
e13tpslgt2=lag(e13tpslgt);
e13tpslgt3=lag(e13tpslgt2);
if (E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1)) and e13tpslgt_i4='4' and
e13tpslgt2 in ("2", "3") then
e13tpslgt_i5=e13tpslgt2;
else if (E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1)) and e13tpslgt_i4='4'
and e13tpslgt2 not in ("2", "3")
and e13tpslgt3 in ("2", "3") then e13tpslgt_i5=e13tpslgt3;
else if (E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1)) and e13tpslgt_i4 ne
'4' then e13tpslgt_i5=e13tpslgt_i4;
else e13tpslgt_i5=e13tpslgt_i4;
drop e13tpslgt2 e13tpslgt3;
run;
proc freq data=psttletps2;
tables e13tpslgt_i4 e13tpslgt_i4*e13tpslgt_i5 e13tpslgt_i5*mineur;
run;
proc freq data=psttletps2 (where=(e13tpslgt_i5="4"));
tables nbnr*nbenf e13_14anai_i;
run;
data psttletps3;
set psttletps2;
if e13tpslgt_i5="4" and mineur=2 then e13tpslgt_i5="3";
run;

proc sort data=psttletps3;
by cabefl descending e13_14anai_i;
run;

```



```

proc sort data=psttletps3 nodupkeys nodup out=cad;
by cabefl;
run;
proc sort data=cad;
by e13_14sexe_i e13_14anai_I ;
run;
data cad2;
set cad;
e13tpslgt2=lag(e13tpslgt);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i5='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i6=e13tpslgt2;
else if e13tpslgt_i5='4' and e13tpslgt2 not in ("2", "3")
and e13tpslgt3 in ("2", "3") then e13tpslgt_i6=e13tpslgt3;
else e13tpslgt_i6=e13tpslgt_i5;
drop e13tpslgt2 e13tpslgt3;
proc sort;
by cabefl;
run;
proc freq data=cad2;
tables e13tpslgt_i6;
run;
data psttletps4;
merge psttletps3 cad2;
by cabefl;
run;
data psttletps5;
set psttletps4;
if e13tpslgt_i5='4' and e13tpslgt_i6 ne "4" then e13tpslgt_i6=e13tpslgt_i6;
else e13tpslgt_i6=e13tpslgt_i5;
run;
proc freq data=psttletps5;
tables e13tpslgt_i6 e13tpslgt_i6*e13tpslgt_i5 e13tpslgt_i6*mineur;
run;
proc freq data=psttletps5(where=(e13tpslgt_i6="4"));
tables e13_14anai_i;
run;
proc sort data=psttletps5;
by cabefl e13_14anai_i;
run;
proc sort data=psttletps5 nodupkeys nodup out=ain;
by cabefl;
run;
proc sort data=ain;
by e13_14sexe_i e13_14anai_I ;
run;
data ain2;
set ain;
e13tpslgt2=lag(e13tpslgt);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i6='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i7=e13tpslgt2;
else if e13tpslgt_i6='4' and e13tpslgt2 not in ("2", "3")
and e13tpslgt3 in ("2", "3") then e13tpslgt_i7=e13tpslgt3;
else e13tpslgt_i7=e13tpslgt_i6;
drop e13tpslgt2 e13tpslgt3;
proc sort;
by cabefl;
run;
proc freq data=ain2;
tables e13tpslgt_i7;

```



```

run;
data psttletps6;
merge psttletps5 ain2;
by cabefl;
run;
data psttletps7;
set psttletps6;
if e13tpslgt_i6='4' and e13tpslgt_i7 ne "4" then e13tpslgt_i7=e13tpslgt_i7;
else e13tpslgt_i7=e13tpslgt_i6;
run;
proc freq data=psttletps7;
tables e13tpslgt_i7 e13tpslgt_i7*e13tpslgt_i6 e13tpslgt_i7*mineur;
run;
proc sort data=psttletps7;
by e13_14anai_i;
run;
data psttletps8;
set psttletps7;
e13tpslgt2=lag(e13tpslgt_i7);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i7='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i8=e13tpslgt2;
else if e13tpslgt_i7='4' and e13tpslgt2 not in ("2", "3")
and e13tpslgt3 in ("2", "3") then e13tpslgt_i8=e13tpslgt3;
else e13tpslgt_i8=e13tpslgt_i7;
drop e13tpslgt2 e13tpslgt3;
proc sort;
by cabefl;
run;
proc freq data=psttletps8;
tables e13tpslgt_i8 e13tpslgt_i8*e13tpslgt_i6 e13tpslgt_i8*mineur;
run;
proc sort data=psttletps8 ;
by descending e13_14anai_i;
run;
data psttletps9;
set psttletps8;
e13tpslgt2=lag(e13tpslgt_i8);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i8='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i9=e13tpslgt2;
else if e13tpslgt_i8='4' and e13tpslgt2 not in ("2", "3")
and e13tpslgt3 in ("2", "3") then e13tpslgt_i9=e13tpslgt3;
else e13tpslgt_i9=e13tpslgt_i8;
drop e13tpslgt2 e13tpslgt3;
proc sort;
by cabefl;
run;
proc freq data=psttletps9;
tables e13tpslgt_i9 e13tpslgt_i9*e13tpslgt_i6 e13tpslgt_i9*mineur;
run;

/*imputation de la non réponse sur la raison de l'autre logement
restante*/;
proc sort data=psttletps9;
by e13tpslgt_i9 e13lgt_aut_par_i4 e13_14sexe_i e13_14anai_i;
run;
data psttletps10;
set psttletps9;
e13lgt_aut_par2=lag(e13lgt_aut_par_i4);

```



```

e13lgt_autre2=lag(e13lgt_autre_i4);
e13lgt_etude2=lag(e13lgt_etude_i4);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1) then do;
if e13lgt_aut_par_i4='Z' and e13lgt_etude_i4='Z' and e13lgt_autre_i4='Z'
and e13lgt_aut_par2 not in ( "Z", " ") then do;
e13lgt_aut_par_i5=e13lgt_aut_par2;
e13lgt_autre_i5=e13lgt_autre2;
e13lgt_etude_i5=e13lgt_etude2;
end;
else if e13lgt_aut_par_i4='Z' and e13lgt_etude_i4='Z' and
e13lgt_autre_i4='Z' and e13lgt_aut_par2 in ( "Z", " ")
and e13lgt_aut_par3 not in ( "Z", " ") then do;
e13lgt_aut_par_i5=e13lgt_aut_par3;
e13lgt_autre_i5=e13lgt_autre3;
e13lgt_etude_i5=e13lgt_etude3;
end;
else if e13lgt_aut_par_i4='2' and e13lgt_etude_i4='Z' and
e13lgt_autre_i4='Z' and e13lgt_autre2 not in ( "Z", " ") then do;
e13lgt_aut_par_i5="2";
e13lgt_autre_i5=e13lgt_autre2;
e13lgt_etude_i5=e13lgt_etude2;
end;
else if e13lgt_aut_par_i4='2' and e13lgt_etude_i4='Z' and
e13lgt_autre_i4='Z' and e13lgt_autre2 in ( "Z", " ")
and e13lgt_autre3 not in ( "Z", " ") then do;
e13lgt_aut_par_i5="2";
e13lgt_autre_i5=e13lgt_autre3;
e13lgt_etude_i5=e13lgt_etude3;
end;
else do;
e13lgt_aut_par_i5=e13lgt_aut_par_i4;
e13lgt_autre_i5=e13lgt_autre_i4;
e13lgt_etude_i5=e13lgt_etude_i4;
end;
end;
else do;
e13lgt_aut_par_i5=e13lgt_aut_par_i4;
e13lgt_autre_i5=e13lgt_autre_i4;
e13lgt_etude_i5=e13lgt_etude_i4;
end;
end;
drop e13lgt_aut_par2 e13lgt_autre2 e13lgt_etude2 e13lgt_aut_par3
e13lgt_autre3 e13lgt_etude3;
run;
proc freq data=psttletps10;
tables
e13tpslgt_i9
e13aut_lgt_i4
e13lgt_aut_par_i4
e13lgt_autre_i4
e13lgt_etude_i4
e13lgt_aut_par_i5
e13lgt_autre_i5
e13lgt_etude_i5
e13lgt_aut_par_i5*mineur
e13lgt_autre_i5*mineur
e13lgt_etude_i5*mineur;
run;
proc sort data=psttletps9;

```



```

by descending cabefl ;
run;
proc sort data=psttletps10;
by e13lgt_aut_par_i5 e13_14anai_i;
run;
data psttletps12;
set psttletps10;
e13lgt_aut_par2=lag(e13lgt_aut_par_i5);
e13lgt_autre2=lag(e13lgt_autre_i5);
e13lgt_etude2=lag(e13lgt_etude_i5);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if E13_14anai_I<1993 or (nbnr=nbenf and nbnr=1) then do;
if e13lgt_aut_par_i5 = 'Z' and e13lgt_etude_i5='Z' and e13lgt_autre_i5='Z'
and e13lgt_aut_par2 not in ( "Z", " ") then do;
e13lgt_aut_par_i6=e13lgt_aut_par2;
e13lgt_autre_i6=e13lgt_autre2;
e13lgt_etude_i6=e13lgt_etude2;
end;
else if e13lgt_aut_par_i5 = '2' and e13lgt_etude_i5='Z' and
e13lgt_autre_i5='Z' and e13lgt_autre2 not in ( "Z", " ") then do;
e13lgt_aut_par_i6='2';
e13lgt_autre_i6=e13lgt_autre2;
e13lgt_etude_i6=e13lgt_etude2;
end;
else if e13lgt_aut_par_i5='Z' and e13lgt_etude_i5='Z' and
e13lgt_autre_i5='Z' and e13lgt_aut_par2 in ( "Z", " ")
and e13lgt_aut_par3 not in ( "Z", " ") then do;
e13lgt_aut_par_i6=e13lgt_aut_par3;
e13lgt_autre_i6=e13lgt_autre3;
e13lgt_etude_i6=e13lgt_etude3;
end;
else if e13lgt_aut_par_i5='2' and e13lgt_etude_i5='Z' and
e13lgt_autre_i5='Z' and e13lgt_autre2 in ( "Z", " ")
and e13lgt_autre3 not in ( "Z", " ") then do;
e13lgt_aut_par_i6=2;
e13lgt_autre_i6=e13lgt_autre3;
e13lgt_etude_i6=e13lgt_etude3;
end;
else do;
e13lgt_aut_par_i6=e13lgt_aut_par_i5;
e13lgt_autre_i6=e13lgt_autre_i5;
e13lgt_etude_i6=e13lgt_etude_i5;
end;
end;
else do;
e13lgt_aut_par_i6=e13lgt_aut_par_i5;
e13lgt_autre_i6=e13lgt_autre_i5;
e13lgt_etude_i6=e13lgt_etude_i5;
end;
end;
drop e13tps1gt2 e13aut_lgt2 e13lgt_aut_par2 e13lgt_autre2 e13lgt_etude2
e13tps1gt3 e13aut_lgt3 e13lgt_aut_par3 e13lgt_autre3 e13lgt_etude3;
run;
proc freq data=psttletps12 (where =(e13_14anai_i<1971));
tables
e13lgt_aut_par_i6
e13lgt_autre_i6
e13lgt_etude_i6;
run;
proc freq data=psttletps12 (where=(mineur=1 and e13lgt_aut_par_i6='2'));

```



```

tables
e13lgt_autre_i6*
e13lgt_etude_i6;
run;
/* environ 1% des plus de 35 ans n'ont pas la raison autre: on met à autre
raisonnablement*/
data psttletps13;
set psttletps12;
if e13lgt_aut_par_i6 in ("Z", "2") and e13lgt_etude_i6='Z' and
e13lgt_autre_i6='Z' and e13_14anai_i<1976 then do;
e13lgt_aut_par_i6="2";
e13lgt_autre_i6="1";
e13lgt_etude_i6="2";
end;
if e13lgt_aut_par_i6 in ("1") and e13lgt_etude_i6='Z' and
e13lgt_autre_i6='Z' then do;
e13lgt_aut_par_i6="1";
e13lgt_autre_i6="2";
e13lgt_etude_i6="2";
end;
proc sort;
by cabefl;
run;
proc freq data=psttletps13;
tables
e13tpslgt_i9
e13aut_lgt_i4
e13lgt_aut_par_i5
e13lgt_autre_i5
e13lgt_etude_i5
e13lgt_aut_par_i6
e13lgt_autre_i6
e13lgt_etude_i6
e13lgt_aut_par_i6*mineur
e13lgt_autre_i6*mineur
e13lgt_etude_i6*mineur;
run;
proc freq data=psttletps13 (where =(e13lgt_etude_i6="Z"));
tables E13_14anai_i;
run;
data min;
set psttletps13 (where=(mineur=1));
proc sort nodupkeys nodup out=min2;
by CABEFL;
run;
proc freq data=min2;
tables
e13lgt_aut_par_i6
e13lgt_autre_i6
e13lgt_etude_i6;
run;

proc sort data=min2;
by e13_14anai_i;
run;
data min3;
set min2;
e13lgt_aut_par2=lag(e13lgt_aut_par_i6);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
if e13lgt_aut_par_i6 = 'Z' and e13lgt_aut_par2 not in ( "Z", " ") then
e13lgt_aut_par_i7=e13lgt_aut_par2;

```



```

else if e13lgt_aut_par_i6='Z' and e13lgt_aut_par2 in ( "Z", " ") and
e13lgt_aut_par3 not in ( "Z", " ") then
e13lgt_aut_par_i7=e13lgt_aut_par3;
else e13lgt_aut_par_i7=e13lgt_aut_par_i6;
*drop e13lgt_aut_par2 e13lgt_aut_par3 ;
run;
proc freq data=min3;
tables
e13lgt_aut_par_i7;
run;
/* il n'en reste que 10 non imputée on les met comme les 90% de oui*/
data min4;
set min3;
if e13lgt_aut_par_i7="Z" then e13lgt_aut_par_i7="1";
keep cabefl e13lgt_aut_par_i7;
proc sort; by cabefl;
run;
data maj;
set psttletps13 (where=(mineur=2));
proc sort nodupkeys nodup out=maj2;
by CABEFL;
run;
proc freq data=maj2;
tables
e13lgt_aut_par_i6
e13lgt_autre_i6
e13lgt_etude_i6;
run;
proc sort data=maj2;
by e13_14anai_i;
run;
data maj3;
set maj2;
e13lgt_aut_par2=lag(e13lgt_aut_par_i6);
e13lgt_aut_par3=lag(e13lgt_aut_par2);
e13lgt_autre2=lag(e13lgt_autre_i6);
e13lgt_etude2=lag(e13lgt_etude_i6);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if e13lgt_aut_par_i6 = 'Z' and e13lgt_aut_par2 not in ( "Z", " ") then do;
e13lgt_aut_par_i7=e13lgt_aut_par2;
e13lgt_autre_i7=e13lgt_autre2;
e13lgt_etude_i7=e13lgt_etude2;
end;
else if e13lgt_aut_par_i6='Z' and e13lgt_aut_par2 in ( "Z", " ") and
e13lgt_aut_par3 not in ( "Z", " ") then do;
e13lgt_aut_par_i7=e13lgt_aut_par3;
e13lgt_autre_i7=e13lgt_autre3;
e13lgt_etude_i7=e13lgt_etude3;
end;
else do;
e13lgt_aut_par_i7=e13lgt_aut_par_i6;
e13lgt_autre_i7=e13lgt_autre_i6;
e13lgt_etude_i7=e13lgt_etude_i6;
end;
*drop e13lgt_aut_par2 e13lgt_aut_par3 ;
run;
proc freq data=maj3;
tables
e13lgt_aut_par_i7 e13lgt_autre_i7 e13lgt_etude_i7
e13lgt_etude_i7*e13_14anai_i;

```



```

run;
proc sort data=maj3;
by descending e13_14anai_i;
run;
data maj4;
set maj3;
e13lgt_aut_par2=lag(e13lgt_aut_par_i7);
e13lgt_aut_par3=lag(e13lgt_aut_par);
e13lgt_autre2=lag(e13lgt_autre_i7);
e13lgt_etude2=lag(e13lgt_etude_i7);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if e13lgt_aut_par_i7 = 'Z' and e13lgt_aut_par2 not in ( "Z", " ") then do;
e13lgt_aut_par_i8=e13lgt_aut_par2;
e13lgt_autre_i8=e13lgt_autre2;
e13lgt_etude_i8=e13lgt_etude2;
end;
else if e13lgt_aut_par_i7='Z' and e13lgt_aut_par2 in ( "Z", " ") and
e13lgt_aut_par3 not in ( "Z", " ") then do;
e13lgt_aut_par_i8=e13lgt_aut_par3;
e13lgt_autre_i8=e13lgt_autre3;
e13lgt_etude_i8=e13lgt_etude3;
end;
else do;
e13lgt_aut_par_i8=e13lgt_aut_par_i7;
e13lgt_autre_i8=e13lgt_autre_i7;
e13lgt_etude_i8=e13lgt_etude_i7;
end;
*drop e13lgt_aut_par2 e13lgt_aut_par3 ;
run;
/* il n'en reste que 2 non imputée né en 1990 et 1989*/
data maj5;
set maj4;
if e13lgt_aut_par_i8="Z" then do;
e13lgt_aut_par_i8="1";
e13lgt_autre_i8="2";
e13lgt_etude_i8="2";
end;
keep cabefl e13lgt_aut_par_i8 e13lgt_autre_i8 e13lgt_etude_i8;
proc sort; by cabefl;
run;
proc freq data=maj4;
tables
e13lgt_aut_par_i8 e13lgt_autre_i8 e13lgt_etude_i8*e13_14anai_i;
run;
data psttletps14;
merge psttletps13 min4 maj5;
by cabefl;
if e13lgt_aut_par_i6="Z" and e13lgt_aut_par_i7 ne " " then
e13lgt_aut_par_i6=e13lgt_aut_par_i7;
if e13lgt_aut_par_i6="Z" and e13lgt_aut_par_i8 ne " " then
e13lgt_aut_par_i6=e13lgt_aut_par_i8;
if e13lgt_autre_i6="Z" and e13lgt_autre_i8 ne " " then
e13lgt_autre_i6=e13lgt_autre_i8;
if e13lgt_etude_i6="Z" and e13lgt_etude_i8 ne " " then
e13lgt_etude_i6=e13lgt_etude_i8;
run;
data psttletps15;
set psttletps14;
if e13lgt_aut_par_i6="1" and e13lgt_etude_i6="Z" then do;
e13lgt_autre_i6="2" ;

```



```

e13lgt_etude_i6="2";
end;
drop e13lgt_autre_i8 e13lgt_etude_i8 e13lgt_aut_par_i7 e13lgt_aut_par_i8;
run;
proc freq data=psttletps15;
tables
e13tps1gt_i9
e13aut_lgt_i4
e13lgt_aut_par_i6
e13lgt_autre_i6*e13lgt_aut_par_i6
e13lgt_etude_i6*e13lgt_aut_par_i6
e13lgt_aut_par_i6*mineur
e13lgt_autre_i6*mineur
e13lgt_etude_i6*mineur;
run;
proc freq data=psttletps15 (where=(e13lgt_etude_i6="Z"));
tables e13_14anai_i;
run;
/* il reste 707 mineurs qui ne voient pas leur autre parent à répartir
entre étude et autre parmi les enfant
du même age qui ne voit pas leur parent*/
data pasautreparent;set psttletps15;
where e13lgt_aut_par_i6="2";
run;
proc sort data=pasautreparent nodupkeys nodup out=autre;
by cabefl;
run;
proc sort data=autre;
by e13_14sexe_i e13_14anai_i;
run;
data autre2;
set autre;
e13lgt_autre2=lag(e13lgt_autre_i6);
e13lgt_etude2=lag(e13lgt_etude_i6);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if e13lgt_etude_i6 ="Z" and e13lgt_etude2 not in ("Z", " ") then do;
e13lgt_etude_i7=e13lgt_etude2;
e13lgt_autre_i7=e13lgt_autre2;
end;
else if e13lgt_etude_i6 ="Z" and e13lgt_etude2 in ("Z", " ") and
e13lgt_etude3 not in ("Z", " ") then do;
e13lgt_etude_i7=e13lgt_etude3;
e13lgt_autre_i7=e13lgt_autre3;
end;
else do;
e13lgt_etude_i7=e13lgt_etude_i6;
e13lgt_autre_i7=e13lgt_autre_i6;
end;
run;
proc freq data=autre2 ;
tables e13lgt_etude_i7 e13lgt_autre_i7;
run;
proc sort data=autre2;
by e13_14anai_i /*e13_14sexe_i*/;
run;
data autre3;
set autre2;
e13lgt_autre2=lag(e13lgt_autre_i7);
e13lgt_etude2=lag(e13lgt_etude_i7);
e13lgt_autre3=lag(e13lgt_autre2);

```



```

e13lgt_etude3=lag(e13lgt_etude2);
if e13lgt_etude_i7 ="Z" and e13lgt_etude2 not in ("Z", " ") then do;
e13lgt_etude_i8=e13lgt_etude2;
e13lgt_autre_i8=e13lgt_autre2;
end;
else if e13lgt_etude_i7 ="Z" and e13lgt_etude2 in ("Z", " ") and
e13lgt_etude3 not in ("Z", " ") then do;
e13lgt_etude_i8=e13lgt_etude3;
e13lgt_autre_i8=e13lgt_autre3;
end;
else do;
e13lgt_etude_i8=e13lgt_etude_i7;
e13lgt_autre_i8=e13lgt_autre_i7;
end;
keep cabefl e13lgt_etude_i8 e13lgt_autre_i8;
proc sort;
by cabefl;
run;

data psttletps16;
merge psttletps15 autre3;
by cabefl;
if e13lgt_autre_i6="Z" and e13lgt_autre_i8 ne " " then
e13lgt_autre_i6=e13lgt_autre_i8;
if e13lgt_etude_i6="Z" and e13lgt_etude_i8 ne " " then
e13lgt_etude_i6=e13lgt_etude_i8;
drop e13lgt_etude_i8 e13lgt_autre_i8;
proc sort;
by e13lgt_aut_par_i6 e13tps1gt_i9 e13_14anai_i;
run;
proc freq data=psttletps16;
tables
e13tps1gt_i9
e13aut_lgt_i4
e13lgt_aut_par_i6
e13lgt_autre_i6*e13lgt_aut_par_i6
e13lgt_etude_i6*e13lgt_aut_par_i6
e13lgt_aut_par_i6*mineur
e13lgt_autre_i6*mineur
e13lgt_etude_i6*mineur;
run;
proc freq data=psttletps16 (where =(e13lgt_aut_par_i6="2"));
tables
mineur
e13lgt_autre_i6*
e13lgt_etude_i6;
run;
data psttletps17;
set psttletps16;
e13lgt_autre2=lag(e13lgt_autre_i6);
e13lgt_etude2=lag(e13lgt_etude_i6);
e13lgt_autre3=lag(e13lgt_autre2);
e13lgt_etude3=lag(e13lgt_etude2);
if e13lgt_etude_i6 ="Z" and e13lgt_etude2 not in ("Z", " ") then do;
e13lgt_etude_i8=e13lgt_etude2;
e13lgt_autre_i8=e13lgt_autre2;
end;
else if e13lgt_etude_i6 ="Z" and e13lgt_etude2 in ("Z", " ") and
e13lgt_etude3 not in ("Z", " ") then do;
e13lgt_etude_i8=e13lgt_etude3;
e13lgt_autre_i8=e13lgt_autre3;

```



```

end;
else if e13lgt_aut_par_i6="2" and e13lgt_etude_i6 ="2" and e13lgt_autre_i6
="2"
and (e13lgt_etude2 ne "2" or e13lgt_autre2 ne "2") then do;
e13lgt_etude_i8=e13lgt_etude2;
e13lgt_autre_i8=e13lgt_autre2;
end;
else if e13lgt_aut_par_i6="2" and e13lgt_etude_i6 ="2" and e13lgt_autre_i6
="2"
and (e13lgt_etude2 = "2" and e13lgt_autre2 = "2") then do;
e13lgt_etude_i8=e13lgt_etude3;
e13lgt_autre_i8=e13lgt_autre3;
end;
else do;
e13lgt_etude_i8=e13lgt_etude_i6;
e13lgt_autre_i8=e13lgt_autre_i6;
end;
proc freq data=psttletps17;
tables
e13tps1gt_i9
e13aut_lgt_i4
e13lgt_aut_par_i6
e13lgt_autre_i8*e13lgt_aut_par_i6
e13lgt_etude_i8*e13lgt_aut_par_i6
e13lgt_aut_par_i6*mineur
e13lgt_autre_i8*mineur
e13lgt_etude_i8*mineur;
run;
proc freq data=psttletps17 (where =(e13lgt_aut_par_i6="2"));
tables
mineur e13lgt_autre_i8 e13lgt_etude_i8
e13lgt_autre_i8*
e13lgt_etude_i8;
run;
/*il ne reste que 3 cas pb que l'on met en autre cas on se trompe moins en
la mettant dans une cat ouverte*/
data psttletps18;
set psttletps17;
if e13lgt_autre_i8="Z" then e13lgt_autre_i8="1";
if e13lgt_etude_i8="Z" then e13lgt_etude_i8="2";
if e13lgt_etude_i8="2" and e13lgt_autre_i8="2" and e13lgt_aut_par_i6="2"
then e13lgt_autre_i8="1";
keep ident e13lgt_etude_i8 e13lgt_autre_i8 e13lgt_aut_par_i6 e13tps1gt_i9;
proc sort;
by ident;
run;
proc sort data=enfpar_seplog6 ;
by ident;
run;
data enfpar_seplog7;
merge enfpar_seplog6 psttletps18;
by ident;
if e13tps1gt_i9 ne " " then e13tps1gt_i4=e13tps1gt_i9;
if e13lgt_aut_par_i6 ne " " then e13lgt_aut_par_i4=e13lgt_aut_par_i6;
if e13lgt_autre_i8 ne " " then e13lgt_autre_i4=e13lgt_autre_i8;
if e13lgt_etude_i8 ne " " then e13lgt_etude_i4=e13lgt_etude_i8;
keep ident e13tps1gt_i4 e13aut_lgt_i4 e13lgt_aut_par_i4 e13lgt_autre_i4
e13lgt_etude_i4;
run;
proc freq data=enfpar_seplog7;
tables

```



```

e13tpslgt_i4
e13aut_lgt_i4
e13lgt_aut_par_i4
e13lgt_autre_i4
e13lgt_etude_i4;
run;
proc sort data=traj.enf_apur;
by ident;
run;

/*l'imputation ne se fait que sur des enfnats qui vivent avec leur deux
parents
on crée une table avec l'ensemble de ces enfants*/

data enf2par;
set enf3b;
where partcpart in ("11");
drop partcpart;
proc sort; by ident;
run;

proc freq data=enf2par;
tables
e13tpslgt e13aut_lgt e13lgt_aut_par e13lgt_autre e13lgt_etude;
run;

/*que 4 "Z" pour e13tpslgt et e13aut_lgt les enfants sont mineurs ont les
impute chez leur parents tout le temps*/

data enf2par2;
set enf2par;
if e13tpslgt="Z" then do;
e13tpslgt_il="1";
e13aut_lgt_il ="2";
e13lgt_etude_il=" ";
e13lgt_aut_par_il=" ";
e13lgt_autre_il=" ";
end;
else do;
e13tpslgt_il=e13tpslgt;
e13aut_lgt_il =e13aut_lgt;
e13lgt_aut_par_il=e13lgt_aut_par;
e13lgt_etude_il=e13lgt_etude;
e13lgt_autre_il=e13lgt_autre;
end;
run;
proc freq data=enf2par2 (where=(e13tpslgt="4" and e13lgt_etude="Z"));
tables e13tpslgt_x e13autlgt_x e13_14anai_x cabbi_enf;
run;
proc freq data=enf2par2;
tables
e13tpslgt_il e13aut_lgt_il e13lgt_aut_par_il e13lgt_autre_il
e13lgt_etude_il e13tpslgt_il*e13lgt_etude_il e13tpslgt_il*e13lgt_autre_il;
run;

/* on répartie les e13tpslgt="4" entre 2 et 3 selon l'age et la nature de
l'autre logement*/
data psttletps_par2;
set enf2par2;
where e13tpslgt_il in ("2", "3", "4");
run;

```



```

proc freq data=psttletps_par2;
tables
e13tpslgt_i1;
run;
proc sort data=psttletps_par2;
by e13_14anai_I e13lgt_etude_i1 e13lgt_autre_i1 ;
run;
data psttletps2_par2;
set psttletps_par2;
e13tpslgt2=lag(e13tpslgt_i1);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i1='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i2=e13tpslgt2;
else if e13tpslgt_i1='4' and e13tpslgt2 not in ("2", "3") and e13tpslgt3
in ("2", "3") then e13tpslgt_i2=e13tpslgt3;
else e13tpslgt_i2=e13tpslgt_i1;
drop e13tpslgt2 e13tpslgt3;
run;
proc freq data=psttletps2_par2;
tables e13tpslgt_i1*e13tpslgt_i2;
run;
proc sort data=psttletps2_par2;
by descending e13_14anai_I ;
run;
data psttletps3_par2;
set psttletps2_par2;
e13tpslgt2=lag(e13tpslgt_i2);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i2='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i3=e13tpslgt2;
else if e13tpslgt_i2='4' and e13tpslgt2 not in ("2", "3") and e13tpslgt3
in ("2", "3") then e13tpslgt_i3=e13tpslgt3;
else e13tpslgt_i3=e13tpslgt_i2;
drop e13tpslgt2 e13tpslgt3;
if e13_14anai_i<1951 then clage=1;
else if 1951<=e13_14anai_i<1981 then clage=2;
else if 1981<=e13_14anai_i<1993 then clage=3;
else if 1993<=e13_14anai_i<1998 then clage=4;
else clage=5;
run;
proc sort data=psttletps3_par2;
by ident;
run;
proc sort data=psttletps3_par2;
by clage;
run;
data psttletps4_par2;
set psttletps3_par2;
e13tpslgt2=lag(e13tpslgt_i3);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i3='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i4=e13tpslgt2;
else if e13tpslgt_i3='4' and e13tpslgt2 not in ("2", "3") and e13tpslgt3
in ("2", "3") then e13tpslgt_i4=e13tpslgt3;
else e13tpslgt_i4=e13tpslgt_i3;
drop e13tpslgt2 e13tpslgt3 clage;
run;
proc freq data=psttletps4_par2;
tables e13tpslgt_i1*e13tpslgt_i4 e13_14anai_i*e13tpslgt_i4
e13_14anai_i*e13tpslgt e13_14anai*e13tpslgt e13_14anai_i*e13lgt_etude
e13_14anai*e13lgt_etude;

```



```

run;
proc sort data=psttletps4_par2;
by e13_14anai_i;
run;
data psttletps5_par2;
set psttletps4_par2;
e13tpslgt2=lag(e13tpslgt_i4);
e13tpslgt3=lag(e13tpslgt2);
if e13tpslgt_i4='4' and e13tpslgt2 in ("2", "3") then
e13tpslgt_i5=e13tpslgt2;
else if e13tpslgt_i4='4' and e13tpslgt2 not in ("2", "3") and e13tpslgt3
in ("2", "3") then e13tpslgt_i5=e13tpslgt3;
else e13tpslgt_i5=e13tpslgt_i4;
drop e13tpslgt2 e13tpslgt3 clage;
run;
proc freq data=psttletps5_par2;
tables e13tpslgt_i1*e13tpslgt_i5 e13_14anai_i*e13tpslgt_i5
e13_14anai_i*e13tpslgt e13_14anai*e13tpslgt e13_14anai_i*e13lgt_etude
e13_14anai*e13lgt_etude;
run;
data psttletpsfi_par2;
set psttletps5_par2;
if e13tpslgt_i5="4" then e13tpslgt_i5="3" /*7 cas né dans les années 60
majo en 3*/;
keep ident e13tpslgt_i5;proc sort; by ident;
run;
data enf3par2;
merge enf2par2 psttletpsfi_par2;
by ident;
if e13tpslgt_i5=" " then e13tpslgt_i5=e13tpslgt_i1;
if e13_14anai_i<1975 and e13lgt_etude_i1="Z" then do;
e13lgt_etude_i1="2";
e13lgt_autre_i1="1";
end;
if e13_14anai_i>2001 and e13lgt_etude_i1="Z" then do;
e13lgt_etude_i1="2";
e13lgt_autre_i1="1";
end;
run;
proc freq data=enf3par2;
tables e13tpslgt_i5 e13_14anai_i*e13tpslgt_i5 e13_14anai_i*e13lgt_etude
e13_14anai*e13lgt_etude e13lgt_etude_i1 e13lgt_autre_i1;
run;
proc sort data=enf3par2;
by e13_14anai_i e13tpslgt_i5;
run;
data enf4par2;
set enf3par2;
e13lgt_etude2=lag(e13lgt_etude_i1);
e13lgt_etude3=lag(e13lgt_etude2);
e13lgt_autre2=lag(e13lgt_autre_i1);
e13lgt_autre3=lag(e13lgt_autre2);
if e13lgt_etude_i1="Z" and e13lgt_etude2 not in (" ", "Z") then do;
e13lgt_etude_i2=e13lgt_etude2;
e13lgt_autre_i2=e13lgt_autre2;
end;
else if e13lgt_etude_i1="Z" and e13lgt_etude2 in (" ", "Z") and
e13lgt_etude3 not in (" ", "Z") then do;
e13lgt_etude_i2=e13lgt_etude3;
e13lgt_autre_i2=e13lgt_autre3;
end;

```



```

else do;
e13lgt_etude_i2=e13lgt_etude_i1;
e13lgt_autre_i2=e13lgt_autre_i1;
end;
if e13_14anai_i<1951 then clage=1;
else if 1951<=e13_14anai_i<1981 then clage=2;
else if 1981<=e13_14anai_i<1993 then clage=3;
else if 1993<=e13_14anai_i<1998 then clage=4;
else clage=5;
drop e13lgt_etude2 e13lgt_etude3 e13lgt_autre2 e13lgt_autre3;
run;
proc freq data=enf4par2;
tables e13lgt_etude_i2 e13lgt_autre_i2;
run;
proc sort data=enf4par2;
by ident;
run;
proc sort data=enf4par2;
by clage e13tpslgt_i5;
run;

data enf5par2;
set enf4par2;
e13lgt_etude2=lag(e13lgt_etude_i2);
e13lgt_etude3=lag(e13lgt_etude2);
e13lgt_autre2=lag(e13lgt_autre_i2);
e13lgt_autre3=lag(e13lgt_autre2);
if e13lgt_etude_i2="Z" and e13lgt_etude2 not in (" ", "Z") then do;
e13lgt_etude_i3=e13lgt_etude2;
e13lgt_autre_i3=e13lgt_autre2;
end;
else if e13lgt_etude_i2="Z" and e13lgt_etude2 in (" ", "Z") and
e13lgt_etude3 not in (" ", "Z") then do;
e13lgt_etude_i3=e13lgt_etude3;
e13lgt_autre_i3=e13lgt_autre3;
end;
else do;
e13lgt_etude_i3=e13lgt_etude_i2;
e13lgt_autre_i3=e13lgt_autre_i2;
end;
drop e13lgt_etude2 e13lgt_etude3 e13lgt_autre2 e13lgt_autre3 clage;
run;
proc freq data=enf5par2;
tables e13lgt_etude_i3 e13lgt_autre_i3;
run;
proc freq data=enf5par2;
tables e13lgt_etude_i3*e13tpslgt_i5 e13lgt_autre_i3*e13tpslgt_i5;
run;
proc freq data=enf5par2(where=(e13lgt_etude_i3="Z"));
tables e13_14anai_i;
run;
data pb1;
set enf5par2;
where e13lgt_aut_par_i1="2" and e13lgt_autre_i3="2" and
e13lgt_etude_i3="2";
run;
data enf6par2;
set enf5par2;
if e13lgt_etude_i3="Z" then do;
e13lgt_etude_i3="2";
e13lgt_autre_i3="1";

```



```

end;
if e13tpslgt_i5="1" then do;
e13lgt_etude_i3=" ";
e13lgt_autre_i3=" ";
end;
if e13tpslgt_i5 ne "1" and e13lgt_aut_par_i1="2" and e13lgt_autre_i3="2"
and e13lgt_etude_i3="2" then do;
e13lgt_aut_par_i1="2";
e13lgt_autre_i3="2";
e13lgt_etude_i3="1";
end;
keep ident e13tpslgt_i5 e13aut_lgt_i1 e13lgt_aut_par_i1 e13lgt_etude_i3
e13lgt_autre_i3;
proc sort;
by ident;
run;
/*on récupère les données imputées*/
data traj.enf_apur_imput;
merge traj.enf_apur /*base des données apurées sur tps passé et autre
logement*/ enfpar_seplog7 enf6par2;
by ident;
if e13tpslgt in ("Z", "4") and e13tpslgt_i4 ne " " then
e13tpslgt_i=e13tpslgt_i4;
else if e13tpslgt in ("Z", "4") and e13tpslgt_i5 ne " " then
e13tpslgt_i=e13tpslgt_i5;
else e13tpslgt_i=e13tpslgt;
if e13aut_lgt in ("Z") and e13aut_lgt_i4 ne " " then
e13aut_lgt_i=e13aut_lgt_i4;
else if e13aut_lgt in ("Z") and e13aut_lgt_i1 ne " " then
e13aut_lgt_i=e13aut_lgt_i1;
else e13aut_lgt_i=e13aut_lgt;
if e13lgt_aut_par in ("Z") and e13lgt_aut_par_i4 ne " " then
e13lgt_aut_par_i=e13lgt_aut_par_i4;
else if e13lgt_aut_par in ("Z") and e13lgt_aut_par_i1 ne " " then
e13lgt_aut_par_i=e13lgt_aut_par_i1;
else e13lgt_aut_par_i=e13lgt_aut_par;
if e13lgt_autre in ("Z") and e13lgt_autre_i4 ne " " then
e13lgt_autre_i=e13lgt_autre_i4;
else if e13lgt_autre in ("Z") and e13lgt_autre_i3 ne " " then
e13lgt_autre_i=e13lgt_autre_i3;
else e13lgt_autre_i=e13lgt_autre;
if e13lgt_etude in ("Z") and e13lgt_etude_i4 ne " " then
e13lgt_etude_i=e13lgt_etude_i4;
else if e13lgt_etude in ("Z") and e13lgt_etude_i3 ne " " then
e13lgt_etude_i=e13lgt_etude_i3;
else e13lgt_etude_i=e13lgt_etude;
drop e13tpslgt_i4 e13tpslgt_i5
e13aut_lgt_i4 e13aut_lgt_i1
e13lgt_aut_par_i4 e13lgt_aut_par_i1
e13lgt_autre_i4 e13lgt_autre_i3
e13lgt_etude_i4 e13lgt_etude_i3;
run;
data traj.enf_imput;
set traj.enf_apur_imput;
if e13aut_lgt_i="2" then do;
e13lgt_aut_par_i=" ";
e13lgt_etude_i=" ";
e13lgt_autre_i=" ";
end;
drop e13lgt_aut_par e13aut_lgt e13tpslgt e13lgt_autre e13lgt_etude ;
run;

```



## Lieu de naissance des parents (Q15 et Q21)

Programme d'imputation correspondant :

```
data indic (keep=cabefl paym payp); set efl111.efl_fc11v5 (keep= cabefl
cod_pay_m cod_pay_p);
length paym $3. payp $3.;
paym=substr(cod_pay_m,1,3);
payp=substr(cod_pay_p,1,3);
proc freq data=indic;table paym payp;;run;

data ear; set efljuill.ear_individu_v7 (keep=cabefl pnaic lnaia);
if cabefl ne ' ';
** pays de naissance ;
proc sort data=ear; by cabefl; run;
proc freq data=ear; table pnaic lnaia; run;

data nouvays; set pays.rajout_efl_pays; run;
data ind ; merge efljuill.efl_individu_v9 (keep= cabefl q15lnai q15reg
q21lnai q21reg reg_code qlanai dipl reg_code )
nouvays;by cabefl;
qlage=2010-qlanai;run;
data indiv0; merge ind (in=a) ear indic; by cabefl;
if a;
/*length tage $ 8.;
if 18<=qlage<=35 then tage='18-35';else
if 36<=qlage<=50 then tage='36-50';else
if 51<=qlage<=65 then tage='51-65';else
if 66<=qlage then tage='66 et +';*/
length tage $ 8.;
if 18<=qlage<=30 then tage='18-30';else
if 31<=qlage<=40 then tage='31-40';else
if 41<=qlage<=50 then tage='41-50';else
if 51<=qlage<=60 then tage='51-60';else
if 61<=qlage<=70 then tage='61-70';else
if 71<=qlage then tage='71 et +';
if cabefl='1H00360501' then dipl='11';
length diplc $ 1.;
if dipl in ('01','02','03') then diplc='1';else
if dipl in ('11','12','13') then diplc='2';else
if dipl in ('14','15','16') then diplc='3';else
if dipl in ('17','18') then diplc='4';
else diplc='Z';
eff=1;
run;
proc freq data=indiv0;table q15lnai q21lnai ; run;
** si le pays de naissance est renseigné pour l'un des 2, le père ou la
mère ***;
** on met la même chose pour l'autre ***;
** si ego est né à l'étranger on met parents nés à l'étranger ***;
data indiv; set indiv0;
if q15lnai='Z' and q21lnai ne 'Z' then q15lnai=q21lnai;**8105 **;
if q21lnai='Z' and q15lnai ne 'Z' then q21lnai=q15lnai;**16884 ;
if lnaia='2' then do;
if q15lnai='Z' then q15lnai='7';
if q21lnai='Z' then q21lnai='7';
end;
if tage='71 et +' and diplc='Z' and reg_code='54' and q15lnai='Z' then do;
```



```

q15lnai='2';q21lnai='2';q15reg='54';q21reg='54';end;
run;
proc freq data=indiv0; table lnaia q15lnai*q21lnai;run;**35085 q15 à Z et
43864 q21 à Z ;
proc freq data=indiv; table q15lnai*q21lnai;run; **23317 q15 et q21 à Z;

proc freq data=indiv(where=(q15lnai ne 'Z')) ;table q15lnai; run;
proc freq data=indiv(where=(q21lnai ne 'Z')) ;table q21lnai; run;

*** q15lnai q21lnai : 23317 à recoder *****;

proc freq data=indiv; table reg_code tage diplc;run;
proc freq data=indiv(where=(q15lnai ne 'Z' and q21lnai ne 'Z')) ;table
q15lnai q21lnai; run;

data rep nonrep;set indiv;
if q15lnai='Z' and q21lnai='Z' then output nonrep;
else output rep;
proc sort data=rep ; by lnaia reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q15lnai q21lnai; run;

proc sort data=nonrep; by lnaia reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class lnaia reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class lnaia reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by lnaia reg_code tage diplc ;
if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by lnaia reg_code tage diplc;
alea=ranuni(14);
tri=lnaia !!reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q15lnai_i q21lnai_i cpt nbnonrep tri);set
rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q15lnai=q15lnai_i q21lnai=q21lnai_i;run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q15lnai_i q21lnai_i;run;

```



```

data indiva;merge indiv(in=a) res (keep=cabefl q15lnai_i q21lnai_i);by
cabefl;if a;
if q15lnai_i=' ' then q15lnai_i=q15lnai;
if q21lnai_i=' ' then q21lnai_i=q21lnai;
run;

Proc freq data=indiva (where=(q15lnai ne 'Z' and q21lnai ne 'Z'));
tables q15lnai*q21lnai ;run;

Proc freq data=indiva ;
tables q15lnai_i*q21lnai_i ;run;

Proc freq data=indiva ;
tables q15lnai_i q21lnai_i;
run;

data catcouple0; set indiv(where=(q15lnai ne 'Z' and q21lnai ne 'Z')) ;
length catcouple $ 9.;
if (q15lnai = '7' and q21lnai = '7') then catcouple='2etranger'; else
if (q15lnai ne '7' and q21lnai = '7') or (q15lnai='7' and q21lnai ne '7')
then catcouple='mixte'; else
catcouple='2français';run;
proc freq data=catcouple0; table catcouple; run;
data catcouplei; set indiva ;
length catcouple $ 9.;
if (q15lnai_i = '7' and q21lnai_i = '7') then catcouple='2etranger'; else
if (q15lnai_i ne '7' and q21lnai_i = '7') or (q15lnai_i='7' and q21lnai_i ne
'7') then catcouple='mixte'; else
catcouple='2français';run;
proc freq data=catcouplei; table catcouple; run;

***** Pays de naissance
*****;
proc freq data=indiva (where=(q15pay not in (' ','ZZ')));table q15pay;run;
proc freq data=indiva (where=(q21pay not in (' ','ZZ')));table q21pay;run;

data indiv2a;set indiva;
q15payc=q15pay;
q21payc=q21pay;
q15regc=q15reg;
q21regc=q21reg;
** si parent né à l'étranger et pays non renseigné et ego né à l'étranger
on met le même pays ;
if q15pay='ZZ' then do;
if q15lnai_i ne '7' then q15payc=' ';else
if q15lnai_i='7' and pnaic ne '99' then q15payc=pnaic;
end;
if q15lnai_i='7' and q15reg='ZZ' then q15regc='99';
if q21pay='ZZ' then do;
if q21lnai_i ne '7' and q21pay='ZZ' then q21payc=' ';else
if q21lnai_i='7' and pnaic ne '99' then q21payc=pnaic;
end;
if q15reg='ZZ' then do;
if q15lnai_i in ('2','3') then q15regc=reg_code;
if q15lnai_i='5' then q15regc='D0';
if q15lnai_i='6' then q15regc='CO';
if q15lnai_i='7' then q15regc='99';
end;
if q21reg='ZZ' then do;
if q21lnai_i in ('2','3') then q21regc=reg_code;

```



```

if q21lnai_i='5' then q21regc='D0';
if q21lnai_i='6' then q21regc='CO';
if q21lnai_i='7' then q21regc='99';
end;
run;

proc freq data=indiv2a; table q15pay q15payc q15lnai_i q15reg q15regc
/missing;run;

proc freq data=indiv2a; table q15payc q21payc;run;

** si 2 nés à l'étrangers et 1 non-réponse, on met la même chose pour les 2
**;
data indiv2;set indiv2a;
if q15payc='ZZ' and q21payc not in (' ','ZZ') then q15payc=q21payc;
if q21payc='ZZ' and q15payc not in (' ','ZZ') then q21payc=q15payc;
proc freq data=indiv2; table q15payc q21payc q15regc q21regc/missing;run;
data verif;set indiv2;
q15=q15payc; if q15payc not in (' ','ZZ') then q15='OK';
q21=q21payc; if q21payc not in (' ','ZZ') then q21='OK';
proc freq data=verif; table q15*q21 /missing;run;

** q15payc- q21payc : 3390 valeurs à imputer, 1566 non réponses pour les 2
***;
** on impute ensemble les 1566 : strate = reg_code*tage*diplc***;
proc freq data=indiv2 (where=(q15payc not in (' ','ZZ') and q21payc not in
(' ','ZZ')));
table q15payc q21payc; run;

data rep nonrep;set indiv2 (where=(q15payc ne ' ' and q21payc ne ' '));
if (q15payc='ZZ' and q21payc='ZZ') then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

```



```

data rep3 (keep=tage diplc q15pay_i q21pay_i cpt nbnonrep tri);set rep2;by
tri;
retain cpt;
if first.tri then cpt=1;
           else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q15payc=q15pay_i q21payc=q21pay_i;run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q15pay_i q21pay_i;run;

data indiv3;merge indiv2(in=a) res (keep=cabefl q15pay_i q21pay_i);by
cabefl;if a;
if q15pay_i=' ' then q15pay_i=q15payc;
if q21pay_i=' ' then q21pay_i=q21payc;
run;
Proc freq data=indiv3 (where=(q15pay_i ne ' ' and q21pay_i ne ' '));
tables q15pay_i q21pay_i ;
run;

** on impute les 740 ZZ restant de q15payc**
data indivq15; set indiv2; if q15payc ne ' ' and q21payc = ' ' ;run;**10665
obs **;
data rep nonrep;set indivq15 ;
if q15payc='ZZ' then output nonrep;
           else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q15payc ; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
  proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
  proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q15pay_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
           else cpt=cpt+1;

```



```

if cpt<=nbnonrep then output;
rename q15payc=q15pay_i ;run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q15pay_i ;run;

data indiv4;merge indiv3(in=a) res (keep=cabefl q15pay_i );by cabefl;if a;
if q15pay_i=' ' then q15pay_i=q15payc;
run;
Proc freq data=indiv4;
tables q15pay_i q21pay_i /missing;
run;

** même chose pour q21payc **;
** on impute les 1084 ZZ restant de q21payc**;
data indivq21; set indiv2; if q21payc ne ' ' and q15payc = ' ' ;run;**14370
**;
data rep nonrep;set indivq21 ;
if q21payc='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q21payc ; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q21pay_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q21payc=q21pay_i ;run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

```



```

Proc freq data=res;
tables q21pay_i ;run;

data indiv5;merge indiv4(in=a) res (keep=cabefl q21pay_i );by cabefl;if a;
if q21pay_i=' ' then q21pay_i=q21payc;
run;
Proc freq data=indiv5;
tables q21pay_i q21payc /missing;
run;
Proc freq data=indiv5;
tables q15pay_i q15regc q21pay_i q21regc/missing;
run;

***** REGIONS *****;
***** q15reg et q21reg ***;
proc freq data=indiva (where=(q15reg ne 'ZZ'));table q15reg;run;
proc freq data=indiva (where=(q21reg ne 'ZZ'));table q21reg;run;

data indiv6;set indiv5;
if q15pay_i ne ' ' then q15regc='99' ;
if q21pay_i ne ' ' then q21regc='99' ;

if q15regc='ZZ' and q21regc not in ('CO','DO','99','ZZ') then
q15regc=q21regc;
if q21regc='ZZ' and q15regc not in ('CO','DO','99','ZZ') then
q21regc=q15regc;
Proc freq data=indiv6;
tables q15regc q21regc /missing;
run;

data indiv1521; set indiv6;
if q15regc not in ('99','CO','DO') and q21regc not in ('99','CO','DO');run;
Proc freq data=indiv1521;
tables q15regc q21regc /missing;
run;

*** on impute q15reg et q21reg en même temps **;
data rep nonrep;set indiv1521 ;
if q15regc='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q15regc q21regc; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;

```



```

if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q15reg_i q21reg_i cpt nbnonrep tri);set rep2;by
tri;
retain cpt;
if first.tri then cpt=1;
           else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q15regc=q15reg_i q21regc=q21reg_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q15reg_i q21reg_i;run;

data indiv7;merge indiv6(in=a) res (keep=cabefl q15reg_i q21reg_i);by
cabefl;
if a;
if q15reg_i=' ' then q15reg_i=q15regc;
if q21reg_i=' ' then q21reg_i=q21regc;
run;

Proc freq data=indiv7;
tables q15reg_i q21reg_i/missing;
run;

data verif; set indiv7;
q15=q15reg_i; if q15reg_i not in ('ZZ','99','CO','DO') then q15='OK';
q21=q21reg_i; if q21reg_i not in ('ZZ','99','CO','DO') then q21='OK';
proc freq data=verif;table q15*q21;run;

** il reste 421 Z à imputer pour Q15 et 276 pour Q21 **;
*** on impute q15reg en france metro pour être cohérent avec q15lnai_i **;

data indiv15; set indiv7;
if q15regc not in ('99','CO','DO') and q21regc in ('99','CO','DO');run;
Proc freq data=indiv15;
tables q15regc q21regc /missing;run;

data rep nonrep;set indiv15 ;
if q15regc='ZZ' then output nonrep;
           else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q15regc ; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;

```



```

proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses** ;
class reg_code tage diplc ;
var eff ;
output out=tabrep(drop=_type_ _freq_) sum=nbrep ;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;
proc print data=verif;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc ;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q15reg_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q15regc=q15reg_i ;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q15reg_i ;run;

data indiv8;merge indiv7(in=a) res (keep=cabefl q15reg_i );by cabefl;if a;
if q15reg_i=' ' then q15reg_i=q15regc;
run;

Proc freq data=indiv8;
tables q15reg_i q21reg_i/missing;
run;

data verif; set indiv8;
q15=q15reg_i; if q15reg_i not in ('99','CO','DO') then q15='OK';
q21=q21reg_i; if q21reg_i not in ('ZZ','99','CO','DO') then q21='OK';
proc freq data=verif;table q15*q21;run;

** il reste 276 Z à imputer pour Q21 **;
*** on impute q21reg **;
data indiv21; set indiv6;
if q21regc not in ('99','CO','DO') and q15regc in ('99','CO','DO');run;
Proc freq data=indiv21;
tables q21regc /missing;run;

data rep nonrep;set indiv21 ;
if q21regc='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q21regc ; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

```



```

proc summary data=nonrep nway ; ** non-réponses** ;
class reg_code tage diplc ;
var eff ;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep ;
proc print data=tabnonrep ; run ;

proc summary data=rep nway ; ** réponses** ;
class reg_code tage diplc ;
var eff ;
output out=tabrep(drop=_type_ _freq_) sum=nbrep ;
proc print data=tabrep ; run ;

data verif ; merge tabrep tabnonrep ; by reg_code tage diplc ;
if nbnonrep>nbrep ; run ;
proc print data=verif ; run ;
data rep2 ; merge rep tabnonrep ; by reg_code tage diplc ;
alea=ranuni(14) ;
tri=reg_code!!tage!!diplc ;
proc sort data=rep2 ; by tri alea ; run ;
*proc print data=rep2 (obs=10) ; run ;

data rep3 (keep=tage diplc q21reg_i cpt nbnonrep tri) ; set rep2 ; by tri ;
retain cpt ;
if first.tri then cpt=1 ;
else cpt=cpt+1 ;
if cpt<=nbnonrep then output ;
rename q21regc=q21reg_i ;
run ;

data res ; merge nonrep rep3 ; run ;
proc sort ; by cabefl ; run ;

Proc freq data=res ;
tables q21reg_i ; run ;

data indiv9 ; merge indiv8(in=a) res (keep=cabefl q21reg_i ) ; by cabefl ; if a ;
if q21reg_i=' ' then q21reg_i=q21regc ; run ;

Proc freq data=indiv9 ;
tables q21reg_i q21regc/missing ;
run ;

data efldef.rajout_q15q21 ; set indiv9 ;
keep cabefl q15pay q15payd q15lnai_i q15pay_i q15reg_i q21pay q21payd
q21lnai_i q21pay_i q21reg_i ; run ;
proc freq data=efldef.rajout_q15q21 ;
table q15pay q15payd q15lnai_i q15pay_i q15reg_i q21pay q21payd q21lnai_i
q21pay_i q21reg_i ; run ;

** verifs ** ;

data ind1 ; set indiv0 ;
if q15pay ne 'ZZ' ; run ;
proc freq data=ind1 ; table q15lnai q15pay ; run ;
data ind1 ; set indiv0 ;
if q21pay ne 'ZZ' ; run ;
proc freq data=ind1 ; table q21lnai q21pay ; run ;
data indi ; set efldef.rajout_q15q21 ;
if q15pay_i=' ' then q15pay_i='FR' ;
if q21pay_i=' ' then q21pay_i='FR' ; run ;

```



```
proc freq data=indi; table q15pay_i*q21pay_i/missing;run;
```

## Situation vitale des parents (Q19 et Q25)

Programme d'imputation correspondant :

```
** redressement des viv et adec par hot-deck ****;  
  
** viv ****;  
  
proc freq data=indiv ;  
table dipl diplc; run;  
  
proc sort data=indiv; by tage;  
proc freq data=indiv ;by tage;  
table q19vivm*diplc;weight poids; run;  
  
** mère - q19viv_m ****;  
data rep nonrep;set indiv;  
if q19vivm='Z' then output nonrep;  
    else output rep;  
proc sort data=rep ; by tage dipl ;  
proc print data=rep (obs=10); run;  
proc freq data=rep ;table q19vivm; run;  
  
proc sort data=nonrep; by tage dipl ;run;  
proc print data=nonrep (obs=10); run;  
  
proc summary data=nonrep nway ; ** non-réponses**;  
class tage dipl ;  
var eff;  
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;  
/*proc print data=tabnonrep; run;*/  
  
proc summary data=rep nway ; ** réponses**;  
class tage dipl ;  
var eff;  
output out=tabrep(drop=_type_ _freq_) sum=nbrep;  
/*proc print data=tabrep; run;*/  
  
data verif; merge tabrep tabnonrep; by tage dipl ;  
if nbnonrep>nbrep;run;  
  
data rep2;merge rep tabnonrep;by tage dipl;  
alea=ranuni(14);  
tri=tage!!dipl;  
proc sort data=rep2; by tri alea;run;  
proc print data=rep2 (obs=10) ; run;  
  
data rep3 (keep=tage dipl q19viv_m cpt nbnonrep tri);set rep2;by tri;  
retain cpt;  
if first.tri then cpt=1;  
    else cpt=cpt+1;  
if cpt<=nbnonrep then output;  
rename q19vivm=q19viv_m;  
run;  
  
/*proc print data=rep3 ; run;*/
```



```

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q19vivm q19viv_m;
run;
proc sort data=indiv; by cabefl;run;
Proc freq data=indiv;
tables q19vivm q19viv_m_x ;
run;
data indivm1;merge indiv(in=a) res (keep=cabefl q19viv_m);by cabefl;
if a;
if q19viv_m=' ' then q19viv_m=q19vivm;

length I_q19viv_m 8.;
if q19viv_m_x='01' then q19vivmx='2'; else
if q19viv_m_x='10' then q19vivmx='1';

if q19vivm=q19vivmx then I_q19viv_m=0; else
if q19vivm='Z' then I_q19viv_m=2; else
I_q19viv_m=1;
run;

;
Proc freq data=indivm1;
tables q19vivmx q19viv_m I_q19viv_m q19vivm;
run;

** père *****;
data rep nonrep;set indiv;
if q25vivp='Z' then output nonrep;
else output rep;
proc sort data=rep ; by tage dipl ;run;
proc print data=rep (obs=10); run;
proc freq data=rep ;table q25vivp; run;

proc sort data=nonrep; by tage dipl ;run;
proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class tage dipl ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
/*proc print data=tabnonrep; run;*/

proc summary data=rep nway ; ** réponses**;
class tage dipl ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
/*proc print data=tabrep; run;*/

data verif; merge tabrep tabnonrep; by tage dipl ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tage dipl;
alea=ranuni(14);
tri=tage!!dipl;
proc sort data=rep2; by tri alea;run;
proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage dipl q25viv_p cpt nbnonrep tri);set rep2;by tri;

```



```

retain cpt;
if first.tri then cpt=1;
                else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q25vivp=q25viv_p;
run;

/*proc print data=rep3 ; run;*/
data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q25vivp q25viv_p;
run;
proc sort data=indiv; by cabefl;run;
data indivp1;merge indiv(in=a) res (keep=cabefl q25viv_p);by cabefl;
if a;
if q25viv_p=' ' then q25viv_p=q25vivp;

length I_q25viv_p 8.;
if q25viv_p_x='01' then q25vivpx='2'; else
if q25viv_p_x='10' then q25vivpx='1';

if q25viv_p=q25vivpx then I_q25viv_p=0; else
if q25vivp='Z' then I_q25viv_p=2; else
I_q25viv_p=1;

;
Proc freq data=indivp1;
tables q25viv_p_x q25viv_p I_q25viv_p q25vivp;
run;

data rajoutviv; merge indivm1 (keep= cabefl q19viv_m I_q19viv_m q19vivm)
                    indivp1 (keep= cabefl q25viv_p I_q25viv_p q25vivp) ;
by cabefl;run;
Proc freq data=rajoutviv (drop=cabefl);
tables _all_;
run;
data indivr1 (drop=q19adec_m q25adec_p); merge  indiv
                                                rajoutviv;by cabefl;

q19adecm=q19adec_m;
if q19viv_m='1' then q19adecm=' ';
q25adecp=q25adec_p;
if q25viv_p='1' then q25adecp=' ';
if cabefl = '1H00390228' then diplc='2';
if cabefl = '1F00407353' then diplc='2';
if cabefl = '1H00411475' then diplc='2';
run;

** mère - q19adec_m *****;
data t2m; set indivr1 (drop=tage);
if q19viv_m='2';
length tage $ 8.;
if 19<=agego<=25 then tage='19-25';else
if 26<=agego<=30 then tage='26-30';else
if 31<=agego<=35 then tage='31-35';else
if 36<=agego<=40 then tage='36-40';else
if 41<=agego<=45 then tage='41-45';else
if 46<=agego<=50 then tage='46-50';else
if 50<=agego<=55 then tage='51-55';else

```



```

if 56<=agego<=60 then tage='56-60';else
if 61<=agego<=65 then tage='61-65';else
if 66<=agego<=70 then tage='66-70';else
if 71<=agego<=75 then tage='71-75';else
if 76<=agego<=80 then tage='76-80';else
if 81<=agego<=86 then tage='81-86';else
if 86<=agego      then tage='86 et +';

run;

Proc freq data=t2m;
tables q19adecm tage dipl;
run;

data rep nonrep;set t2m;
if q19adecm='9999' then output nonrep;
                    else output rep;
proc sort data=rep ; by tage diplc ;
proc print data=rep (obs=10); run;
proc freq data=rep ;table tage; run;

proc sort data=nonrep; by tage diplc ;run;
proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
/*proc print data=tabnonrep; run;*/

proc summary data=rep nway ; ** réponses**;
class tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
/*proc print data=tabrep; run;*/

data verif; merge tabrep tabnonrep; by tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tage diplc;
alea=ranuni(14);
tri=tage!!diplc;
proc sort data=rep2; by tri alea;run;
proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q19adec_m cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
                    else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q19adecm=q19adec_m;
run;

/*proc print data=rep3 ; run;*/
data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q19adecm q19adec_m;
run;

```



```

proc sort data=indivr1; by cabefl;run;
data indivm2;merge indivr1(in=a) res (keep=cabefl q19adec_m);by cabefl;
if a;
if q19adec_m=' ' then q19adec_m=q19adecm;
q19adecmn=sum(q19adec_m);
if q19adecmn ne . then agegodecm=q19adecmn-q1anai;else agegodecm=.;
run;
proc freq data=indivm2; table q19viv_m q19adecmn agegodecm;run;

** on a des date deces mere < datenaissance ego ==> on met moyenne de la
classe **;
data pbm; set indivm2;
if agegodecm ne . and agegodecm<0 ;run;**81 obs**;
proc freq data=pbm; table tage*diplc;run;
data rep2;set rep;
q19adecmn=sum(q19adecm);
agegodecm=q19adecmn-q1anai;run;
proc means data=rep2 nway;
class tage diplc;
var q19adecmn ;
weight poid;
output out=tmoym mean=moym;run;
proc sort data=indivm2;by tage diplc;run;
data indivm3; merge indivm2 tmoym (keep=tage diplc moym); by tage diplc;
if agegodecm ne . and agegodecm<0 then q19adec_m=moym;

** indicateur **;
data avant;set efl11.vivres_avant_mars14 (keep=cabefl q19adec_m );rename
q19adec_m=q19adecmavant;run;
proc sort data=indivm3;by cabefl;run;
data indivm3 ; merge indivm30 avant; by cabefl;
length I_q19adec_m 8.;
if q19adecmavant=q19adec_m_x then I_q19adec_m=0; else
if q19adecmavant='9999' then I_q19adec_m=2; else
I_q19adec_m=1;
run;
proc freq data=indivm3; table q19adec_m I_q19adec_m;run;

**verif **;
data pbm2; set indivm3;
q19adecmn=sum(q19adec_m);
agegodecm=q19adecmn-q1anai;
if agegodecm ne . and agegodecm<0 ;
run;

proc sort data=indivm3; by cabefl; run;
Proc freq data=indivm3;* (where=(q19vivm='2'));
tables q19viv_m q19adec_m I_q19adec_m q19adecm;
run;

** père - q25adec_p *****;
data t2p; set indivr1;
if q25viv_p='2';run;

data rep nonrep;set t2p;
if q25adecp='9999' then output nonrep;
else output rep;
proc sort data=rep ; by tage diplc ;
proc print data=rep (obs=10); run;
proc freq data=rep ;table q25adecp; run;

```



```

proc sort data=nonrep; by tage diplc ;run;
proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by tage diplc;
alea=ranuni(14);
tri=tage!!diplc;
proc sort data=rep2; by tri alea;run;
proc print data=rep2 (obs=10) ; run;

data rep3 (keep= q25adec_p cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
           else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q25adecp=q25adec_p;
run;

/*proc print data=rep3 ; run;*/
data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q25adecp q25adec_p;
run;
proc sort data=indivrl; by cabefl;run;
data indivp2;merge indivrl(in=a) res (keep=cabefl q25adec_p);by cabefl;
if a;
if q25adec_p=' ' then q25adec_p=q25adecp;
q25adecpn=sum(q25adec_p);
if q25adecpn not in (.,9999) then agegodecp=q25adecpn-q1anai;else
agegodecp=.;run;

Proc freq data=indivp2;* (where=(q25viv_p='2'));
tables agegodecp q25vivp q25adecp q25adec_p q25adecp;
run;

** on a des date deces pere < datenaissance ego ==> on met moyenne de la
classe **;
** et 17 non redressés ==> on met moyenne de la classe **;
data pbp; set indivp2;
if (agegodecp ne . and agegodecp<-1) or q25adecpn=9999 ;run;**120 +17
obs**;

proc freq data=pbp; table tage*diplc;run;

```



```

data rep2;set rep;
q25adecpn=sum(q25adecp);
agegodecm=q25adecpn-qlanai;run;
proc means data=rep2 nway;
class tage diplc;
var q25adecpn ;
weight poids;
output out=tmoyp mean=moyp;run;
proc sort data=indivp2;by tage diplc;run;
data indivp30; merge indivp2 tmoyp (keep=tage diplc moyp); by tage diplc;
if (agegodecp ne . and agegodecp<-1) or q25adecpn=9999 then
q25adec_p=moyp;

data avant;set efl11.vivres_avant_mars14 (keep=cabefl q25adec_p );rename
q25adec_p=q25adecpavant;run;
proc sort data=indivp30;by cabefl;run;
data indivp3; merge indivp30 avant; by cabefl;
length I_q25adec_p 8.;
if q25adecpavant=q25adec_p_x then I_q25adec_p=0; else
if q25adecpavant='9999' then I_q25adec_p=2; else
I_q25adec_p=1;
run;
proc freq data=indivp3; table q25adecpavant I_q25adec_p;run;
/*length I_q25adec_p 8.;
if q25adec_p=q25adec_p_x then I_q25adec_p=0; else
if q25adecp='9999' then I_q25adec_p=2; else
I_q25adec_p=1;
run*/
**verif **;
data pbp2; set indivp3;
q25adecpn=sum(q25adec_p);
agegodecp=q25adecpn-qlanai;
if agegodecp ne . and agegodecp<-1 ;
run;
Proc freq data=indivp3;* (where=(q19vivm='2'));
tables q25viv_p q25adec_p I_q25adec_p q25adecp;
run;
proc sort data=indivp3; by cabefl;run;
proc sort data=indivm3; by cabefl;run;
data redress_viv_adec;
merge rajoutviv (keep=cabefl q19viv_m I_q19viv_m q25viv_p I_q25viv_p)
indivm3 (keep=cabefl q19adec_m I_q19adec_m)
indivp3 (keep=cabefl q25adec_p I_q25adec_p)
efl11.vivres_avant_mars14 (keep= cabefl Q20res_lgt_M Q20res_aut_M
Q20comr_M Q20depr_M Q20domr_M
Q26res_avecm_P Q26res_lgt_P Q26res_aut_P Q26comr_P
Q26depr_P Q26domr_P);
by cabefl;
if q19viv_M='2' then do;Q20res_lgt_M=' '; Q20res_aut_M=' ';
Q20comr_M=' '; Q20depr_M=' ';Q20domr_M='
';Q26res_avecm_P=' ';end;

if q25viv_P='2' then do;Q26res_avecm_P=' '; Q26res_lgt_P=' ';
Q26res_aut_P=' ';
Q26comr_P=' '; Q26depr_P=' ';Q26domr_P=' ';end;

run;

proc freq data=redress_viv_adec (drop=cabefl Q20comr_M Q26comr_P);
table _all_;
run;

```



```

*** mise à jour des qresid *****;
data individu;set efljuill.efl_individu_v9; keep cabefl q20resid q26resid
q20reg q26reg;run;
proc sort data=individu;by cabefl;run;

data maj; merge individu redress_viv_adec; by cabefl;
if q19viv_m='2' then do;q20resid=' ';q20reg=' ';Q20comr_M=' ';Q20depr_M='
';Q20domr_M=' ';;end;

if q19viv_m='1' and q20res_aut_m='ZZ' then do;
q20resid='Z';q20reg='ZZ';Q20comr_M='ZZZZZ';Q20depr_M='ZZ';Q20domr_M='Z';end
;

if q19viv_m='1' and q20res_aut_m='00' then do;q20resid=' ';q20reg=' ';
Q20comr_M=' ';Q20depr_M=' ';Q20domr_M=' ';end;

if q25viv_p='2' then do;q26resid=' ';q26reg=' ';
;Q26comr_P=' '; Q26depr_P=' '; Q26domr_P=' ';end;

if q25viv_p='1' and q26res_aut_p='ZZ' then do;q26resid='Z';q26reg='ZZ';
Q26comr_P='ZZZZZ'; Q26depr_P='ZZ'; Q26domr_P='Z';
end;
if q25viv_p='1' and q26res_aut_p='00' then do;q26resid=' ';q26reg=' ';
Q26comr_P=' '; Q26depr_P=' '; Q26domr_P=' ';end;
run;

data verif; set maj;
if q20comr_m not in (' ','ZZZZZ') then q20comr_m='oui';
if q26comr_p not in (' ','ZZZZZ') then q26comr_p='oui';
run;
proc freq data=verif;
table q20res_aut_m q20comr_m Q20depr_m Q20domr_m q26res_aut_p q26comr_p
Q26depr_p Q26domr_p;run;

data efl111.redress_viv_adec_avril14; set maj;
run;
proc freq data=efl111.redress_viv_adec_avril14 (drop=cabefl Q20comr_M
Q26comr_P);
table _all_;
run;

```

## Lieu de résidence des parents (Q20 et Q26).

Programme d'imputation correspondant :

```

*** on calcule des régions de résidences initiales à partir des
départements ****;
*** pour fabriquer les indicateurs i_q20reg et i_q26reg ****;
data indic0; set efl111.efl_fc11v5 (keep= cabefl res_m res_p depr_m
depr_p);run;
** calcul de reg_m et reg_p à partir de depr_m et depr_p pour calculer les
indicateurs **;
** de q20reg et q26reg ****;
proc freq data=indic0;table depr_m depr_p;run;

data depr_m; set indic0; keep cabefl depr_m;if depr_m ne ' ';run;
proc sort data=depr_m;by depr_m;
data deprm (keep=cabefl reg_m depr_m); merge depr_m (in=a) efl111.regdept
(rename=(dep=depr_m)); by depr_m;if a;

```



```

if depr_m='97' then region='D0';
if depr_m='98' then region='CO';
if depr_m='99' then region='99';
if depr_m='20' then region='94';
rename region=reg_m;run;
proc freq data=deprm;table depr_m reg_m;run;** reg_m : région de résidence
de la mère, variable brute initiale **;

data depr_p; set indic0; keep cabefl depr_p;if depr_p ne ' ';run;
proc sort data=depr_p;by depr_p;
data deprp (keep=cabefl reg_p depr_p); merge depr_p (in=a)
efll1.regdept(rename=(dep=depr_p)); by depr_p;if a;
if depr_p='97' then region='D0';
if depr_p='98' then region='CO';
if depr_p='99' then region='99';
if depr_p='20' then region='94';
rename region=reg_p;run;
proc freq data=deprp;table depr_p reg_p;run;** reg_p : région de résidence
du père, variable brute initiale **;

proc sort data=deprm;by cabefl;
proc sort data=deprp;by cabefl;

data indic; merge indic0 deprm (keep=cabefl reg_m) deprp (keep=cabefl
reg_p); by cabefl;run;

** on fait des classes d'âge pour ego **;

/*data ind ; set efljuill.efl_individu_v9
(keep= cabefl q19viv_m q20res_lgt_m q20res_aut_m q20reg q20resid
q25viv_p q26res_avecm_p q26res_lgt_p q26res_aut_p q26reg q26resid qlanai
dipl reg_code poids_ind);
qlage=2010-qlanai;run;*/
data nouvvar;set efl111.redress_viv_adec_mars14 (keep= cabefl q20res_lgt_m
q20res_aut_m q20reg q20resid
q25viv_p q26res_avecm_p q26res_lgt_p q26res_aut_p q26reg q26resid);
data ind ; set efljuill.efl_individu_v9
(keep= cabefl q19viv_m q25viv_p qlanai dipl reg_code poids_ind);
qlage=2010-qlanai;run;
data indiv; merge ind (in=a) nouvvar indic; by cabefl;
if a;
length tage $ 8.;
if 18<=qlage<=25 then tage='18-25';else
if 26<=qlage<=30 then tage='26-30';else
if 31<=qlage<=35 then tage='31-35';else
if 36<=qlage<=40 then tage='36-40';else
if 41<=qlage<=45 then tage='41-45';else
if 46<=qlage<=50 then tage='46-50';else
if 51<=qlage<=55 then tage='51-55';else
if 56<=qlage<=60 then tage='56-60';else
if 61<=qlage<=65 then tage='61-65';else
if 66<=qlage then tage='66 et +';
length diplc $ 1.;
if dipl in ('01','02','03') then diplc='1';else
if dipl in ('11','12','13') then diplc='2';else
if dipl in ('14','15','16') then diplc='3';else
if dipl in ('17','18') then diplc='4';
else diplc='Z';
poids=poids_ind/1000;
eff=1;

```



```

if cabefl='1H00360501' then dipl='11';run;** après retour au questionnaire,
apurement car ressort toujours dans les imputations;
proc freq data=indiv; table dipl; run;

*** mère - q20res_lgt_m q20res_aut_m q20reg q20resid
*****;

*** q20res_lgt_m : hotdeck strates reg_code*tage*dipl **;
data indivm1; set indiv ;if q19viv_m='1';
if reg_code='94' and tage='66 et +' and diplc in ('1','2') then diplc='2';
run;** mères en vie **;
proc freq data=indivm1; table q20res_lgt_m*q20res_aut_m;run;

proc freq data=indivm1; table reg_code tage diplc;run;
proc freq data=indivm1 ;table q20res_lgt_m; run;
proc freq data=indivm1 (where=(q20res_lgt_m ne 'Z')) ;table q20res_lgt_m;
run;
data rep nonrep;set indivm1;
if q20res_lgt_m='Z' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q20res_lgt_m; run;

proc sort data=nonrep; by reg_code tage dipc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q20res_lgt_m_i cpt nbnonrep tri);set rep2;by
tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q20res_lgt_m=q20res_lgt_m_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

```



```

Proc freq data=res;
tables q20res_lgt_m_i;
run;

data indivmla;merge indivml(in=a) res (keep=cabefl q20res_lgt_m_i);by
cabefl;
if a;
if q20res_lgt_m_i=' ' then q20res_lgt_m_i=q20res_lgt_m;
**indicateur : I_q20res_lgt_m**;
resm1=substr(res_m,1,1);
resm2=substr(res_m,2,2);
length I_q20res_lgt_m 8.;
if q20res_lgt_m='Z' then I_q20res_lgt_m=2; else
if q20res_lgt_m_i=resm1 then I_q20res_lgt_m=0; else
I_q20res_lgt_m=1;
run;

;
Proc freq data=indivmla ;
tables res_m resm1 I_q20res_lgt_m q20res_lgt_m q20res_lgt_m_i;
run;

*** q20res_aut_m ***;
** si la mère est vivante et vit ailleurs ne sait pas où :
q20res_aut_m='01' ;
** on l'impute aussi, donc seulement 2 possibilités vit ailleurs ou ne vit
pas ailleurs;
** on crée une nouvelle variable à imputer q20resautm à 3 modalités **;
data indivmla; set indivmla;
q20resautm=q20res_aut_m;
if q20res_aut_m='01' then q20resautm='10';
if q20res_lgt_m_i='1' and reg_code='21' and tage='66 et +' and diplc in
('2','3') then diplc='2';
if q20res_lgt_m_i='1' and reg_code='25' and tage='56-60' and diplc in
('3','4') then diplc='3';
if q20res_lgt_m_i='1' and reg_code='52' and tage='61-65' and diplc in
('3','4') then diplc='3';
if q20res_lgt_m_i='1' and reg_code='53' and tage='66 et +' and diplc in
('1','2') then diplc='1';
if q20res_lgt_m_i='1' and reg_code='53' and tage='66 et +' and diplc in
('3','4') then diplc='3';
run;

proc freq data=indivmla ;table q20res_aut_m q20resautm; run;

** imputation par hot-deck q20res_lgt_m_i*reg_code*tage*dipl***;

data rep nonrep;set indivmla;
if q20resautm='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by q20res_lgt_m_i reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q20res_aut_m; run;

proc sort data=nonrep; by q20res_lgt_m_i reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class q20res_lgt_m_i reg_code tage diplc ;

```



```

var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class q20res_lgt_m_i reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by q20res_lgt_m_i reg_code tage diplc
;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by q20res_lgt_m_i reg_code tage diplc;
alea=ranuni(14);
tri=q20res_lgt_m_i!!reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage diplc q20res_aut_m_i cpt nbnonrep tri);set rep2;by
tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q20resautm=q20res_aut_m_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
Proc freq data=res;
tables q20res_aut_m_i;
run;

data indivmlb;merge indivmla(in=a) res (keep=cabefl q20res_aut_m_i);by
cabefl;
if a;
if q20res_aut_m_i=' ' then q20res_aut_m_i=q20resautm;
**indicateur : I_q20res_aut_m**;
resm2=substr(res_m,2,2);
length I_q20res_aut_m 8.;
if q20resautm='ZZ' then I_q20res_aut_m=2; else
if q20res_aut_m_i=resm2 then I_q20res_aut_m=0; else
I_q20res_aut_m=1;
** si ne vit pas ailleurs on met q20reg et qresid à blanc **;
if q20res_aut_m_i='00' then do; q20reg=' ';q20resid=' ';end;
run;
Proc freq data=indivmlb ;
tables I_q20res_aut_m q20res_aut_m_i ;
run;

data rajoutml; set indivmlb;
keep cabefl I_q20res_lgt_m q20res_lgt_m_i I_q20res_aut_m
q20res_aut_m_i;run;

*** on impute d'abord q20resid **;
data indivm2; set indivmlb ;
if q20resid='9' then q20resid='Z';** on impute les Z et les 9 en même temps
**;

```



```

if q20res_aut_m_i='10';run;** 195300 individus doivent avoir un
q20resid/q20reg rempli **;
proc freq data=indivm2 ;table q20resid; run;

data rep nonrep;set indivm2;
if q20resid='Z' then output nonrep;
    else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q20resid; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage reg_code q20resid_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
    else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q20resid=q20resid_i;
run;
data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
data indivm2a;merge indivm2(in=a) res (keep=cabefl q20resid_i);by cabefl;
if a;
if q20resid_i=' ' then q20resid_i=q20resid;run;
data indivm2a; set indivm2a;
** indicateur **;
length I_q20resid 8.;
if q20resid='Z' then I_q20resid=2; else
I_q20resid=1;
run;

*** q20reg **;
data indivm2b; set indivm2a;
** on met en cohérence q20resid avec q20reg **;
q20regc=q20reg;
if q20resid_i='7' then q20regc='99';
if q20resid_i in ('1','2','3','8') then q20regc=reg_code;
if q20resid_i='5' then q20regc='D0';

```



```

if q20resid_i='6' then q20regc='CO';
reg=q20regc;
if reg not in ('99','ZZ') then reg='RE';
proc freq data=indivm2a ;table q20reg ; run;**19498 ZZ**;
proc freq data=indivm2b ;table q20regc q20regc*q20resid_i reg
*q20resid_i/missing; run;**4805 ZZ**;

** les ZZ restant ont un Q20resid_i = 4 (4805)**;
** on peut imputer les 4805 **;
data indivm3a; set indivm2b;if Q20resid_i = '4';run;** mères résidant hors
reg_code **;
proc freq data=indivm3a ;table reg_code*q20regc; run;
proc freq data=indivm3a ;table q20regc; run;

data rep nonrep;set indivm3a ;
if q20regc='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q20reg; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc ;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage reg_code q20reg_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q20regc=q20reg_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
Proc freq data=res;
tables q20reg_i;
run;

** verif *****;

```



```

data verif;set res; if q20reg_i=reg_code; run;* 0 obs **;
*****;

data indivm4a;merge indivm2b(in=a) res (keep=cabefl q20reg_i);by cabefl;
if a;
if q20reg_i=' ' then q20reg_i=q20regc;
length i_q20reg 8.;
if q20reg='ZZ' then I_q20reg=2; else
if q20reg_i=reg_m then I_q20reg=0; else
I_q20reg=1;
run;
Proc freq data=indivm4a ;
tables q20reg I_q20reg q20reg_i;
run;
data rajoutm2; set indivm4a;
keep cabefl q20reg_i I_q20reg q20resid_i I_q20resid;run;
data rajoutm; merge rajoutm1 rajoutm2; by cabefl;
proc freq data=rajoutm (drop=cabefl); table _all_;run;

data individu; merge indiv (keep=cabefl poids_ind q20res_lgt_m
q20res_aut_m q20reg q20resid)
rajoutm; by cabefl; run;

proc freq data=individu ;
table q20res_lgt_m q20res_lgt_m_i I_q20res_lgt_m
q20res_aut_m q20res_aut_m_i I_q20res_aut_m
q20reg q20reg_i i_q20reg
q20resid q20resid_i i_q20resid
;run;

*** même chose pour le père ***;

*** q26res_avec_m_p ***;
proc freq data=indiv (where=(q25viv_p='1' and q19viv_m='1')); table
reg_code tage diplc;run;
data indivpla; set indiv ;if q25viv_p='1' and q19viv_m='1';
if reg_code='43' and tage='66 et +' and diplc in('1','2') then diplc='2';
if reg_code='72' and tage='66 et +' and diplc in('1','2') then diplc='2';
if reg_code='83' and tage='61-65' and diplc in('1','2') then diplc='2';
;run;
proc freq data=indivpla ;table q26res_avec_m_p; run;
proc freq data=indivpla (where=(q26res_avec_m_p ne 'Z')) ;table
q26res_avec_m_p; run;
data rep nonrep;set indivpla;
if q26res_avec_m_p='Z' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q26res_avec_m_p; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;
proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc ;
var eff;

```



```

output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;
data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage dipl q26res_avecm_p_i cpt nbnonrep tri);set rep2;by
tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26res_avecm_p=q26res_avecm_p_i;
run;
data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
Proc freq data=res;
tables q26res_avecm_p_i;
run;

/*Proc freq data=indivpla ;
tables res_p;run;*/

data indivplb;merge indivpla(in=a) res (keep=cabefl q26res_avecm_p_i)
individum (keep=cabefl q20res_lgt_m_i q20res_aut_m_i q20reg_i
q20resid_i);by cabefl;
if a;
if q26res_avecm_p_i=' ' then q26res_avecm_p_i=q26res_avecm_p;
**indicateur : I_q26res_avecm_p**;
resp1=substr(res_p,1,1);
resp2=substr(res_p,2,1);
resp3=substr(res_p,3,2);
length I_q26res_avecm_p 8.;
if q26res_avecm_p='Z' then I_q26res_avecm_p=2; else
if q26res_avecm_p_i=resp1 then I_q26res_avecm_p=0; else
I_q26res_avecm_p=1;
run;
Proc freq data=indivplb ;
tables resp1 I_q26res_avecm_p q26res_avecm_p q26res_avecm_p_i q26reg
q26resid;
run;

data rajoutp1; set indivplb; keep cabefl I_q26res_avecm_p
q26res_avecm_p_i;run;

*** q26res_lgt_p **;
** verifs avant apurement de q26res_lgt_p **;
data verif0; set indivpla;if q26res_avecm_p='1';
proc freq data=verif0 ;table q20res_lgt_m*q26res_lgt_p;run;
/*
Q20res_lgt_M
Q26res_lgt_P

Frequency,
Percent ,

```



```

Row Pct  ,
Col Pct  ,0          ,1          , Total
ffffffff^ffffffff^ffffffff^
0          , 79949 , 353 , 80302
          , 82.25 , 0.36 , 82.61
          , 99.56 , 0.44 ,
          , 100.00 , 2.05 ,
ffffffff^ffffffff^ffffffff^
1          , 0 , 16902 , 16902
          , 0.00 , 17.39 , 17.39
          , 0.00 , 100.00 ,
          , 0.00 , 97.95 ,
ffffffff^ffffffff^ffffffff^
Total      79949  17255  97204
          82.25  17.75  100.00*/

```

```
/* même comportement ==> on apure */
```

```

data indivp2a ;merge indiv indivplb;by cabefl;if q25viv_p='1';
q26res_lgt_pc=q26res_lgt_p;
if q26res_avecm_p='Z' and q26res_avecm_p_i='1' and q26res_lgt_p='Z' then
q26res_lgt_pc=q20res_lgt_m_i;
*if q26res_avecm_p='Z' and q26res_avecm_p_i='1' and q26res_aut_p='ZZ' then
q26res_aut_p='00';
if reg_code='43' and tage='66 et +' and diplc in ('3','4') then diplc='3';
if reg_code='53' and tage='66 et +' and diplc in ('1','2') then diplc='2';
if reg_code='83' and tage='51-55' and diplc='Z' then diplc='2';
if reg_code='94' and tage='66 et +' and diplc in ('1','2') then diplc='2';
run;
Proc freq data=indivp2a ;tables q26res_avecm_p_i q26res_lgt_p
q26res_lgt_pc ;run;
proc freq data=indivp2a ;table q26res_lgt_pc; run;
** q26res_avecm_p_i ne ' ' **;
data indivp2a1;set indivp2a (where=(q26res_avecm_p_i ne ' ')) ;run;
proc freq data=indivp2a1 ;table q26res_lgt_pc; run;
data rep nonrep;set indivp2a1 ;
if q26res_lgt_pc='Z' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc ; run;
proc sort data=nonrep; by reg_code tage diplc ;run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc ;
alea=ranuni(14);
tri=reg_code!!tage!!diplc ;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

```



```

data rep3 (keep= reg_code tage diplc q26res_lgt_p_i cpt nbnonrep tri);set
rep2;by tri;
retain cpt;
if first.tri then cpt=1;
                else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26res_lgt_pc=q26res_lgt_p_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
Proc freq data=res;
tables q26res_lgt_p_i;run;
Proc freq data=indivp2a1 ;tables res_p;run;

data indivp2b1;merge indivp2a1(in=a) res (keep=cabefl q26res_lgt_p_i);by
cabefl;
if a;
if q26res_lgt_p_i=' ' then q26res_lgt_p_i=q26res_lgt_pc;
  **indicateur : I_q26res_lgt_p**;
resp1=substr(res_p,1,1);
resp2=substr(res_p,2,1);
resp3=substr(res_p,3,2);
length I_q26res_lgt_p 8.;
if q26res_lgt_pc='Z' then I_q26res_lgt_p=2; else
if q26res_lgt_p_i=resp2 then I_q26res_lgt_p=0; else
I_q26res_lgt_p=1;
run;

Proc freq data=indivp2b1 ;
tables q26res_lgt_pc I_q26res_lgt_p q26res_lgt_p_i;
run;

** q26res_avecm_p_i ne ' ' **;
data indivp2a2;set indivp2a (where=(q26res_avecm_p_i = ' ')) ;
if reg_code='21' and tage in ('56-60','61-65','66 et +') and diplc in
('1','2') then do;
tage='56-60'; diplc='1';end;
if reg_code in ('23','24','42','53','82','83') and tage in ('61-65','66 et
+') and diplc in ('1','2') then do;
tage='61-65'; diplc='1';end;
if reg_code in ('41','83') and tage = '26-30' and diplc in ('1','2') then
diplc='1';
if reg_code='94' and tage = '66 et +' and diplc in ('3','4') then
diplc='3';
run;
run;
proc freq data=indivp2a2 ;table q26res_lgt_pc; run;
data rep nonrep;set indivp2a2 ;
if q26res_lgt_pc='Z' then output nonrep;
                else output rep;
proc sort data=rep ; by reg_code tage diplc ; run;
proc sort data=nonrep; by reg_code tage diplc ;run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

```



```

proc summary data=rep nway ; ** réponses** ;
class reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc ;
alea=ranuni(14);
tri=reg_code!!tage!!diplc ;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep= reg_code tage diplc q26res_lgt_p_i cpt nbnonrep tri);set
rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26res_lgt_pc=q26res_lgt_p_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;
Proc freq data=res;
tables q26res_lgt_p_i;run;
Proc freq data=indivp2a2 ;tables res_p;run;

data indivp2b2;merge indivp2a2(in=a) res (keep=cabefl q26res_lgt_p_i);by
cabefl;
if a;
if q26res_lgt_p_i=' ' then q26res_lgt_p_i=q26res_lgt_pc;
**indicateur : I_q26res_lgt_p** ;
resp1=substr(res_p,1,1);
resp2=substr(res_p,2,1);
resp3=substr(res_p,3,2);
length I_q26res_lgt_p 8.;
if q26res_lgt_pc='Z' then I_q26res_lgt_p=2; else
if q26res_lgt_p_i=resp2 then I_q26res_lgt_p=0; else
I_q26res_lgt_p=1;
run;

Proc freq data=indivp2b2 ;
tables q26res_lgt_pc I_q26res_lgt_p q26res_lgt_p_i;
run;

data indivp2b; set indivp2b1 indivp2b2;run;
proc sort data=indivp2b; by cabefl;run;
Proc freq data=indivp2b ;
tables q26res_lgt_pc I_q26res_lgt_p q26res_lgt_p_i;
run;

*** q26res_aut_p ** ;
data indivp3a; set indivp2b;
q26resautp=q26res_aut_p;
if q26res_aut_p='01' then q26resautp='10';
if tage in ('61-65','66 et +') then tage='61-65';

```



```

if reg_code in ('52','53','91') and tage in ('56-60','61-65') then
tage='56-60';
if reg_code in ('11','22') and tage in ('51-65','56-60','61-65') then
tage='56-60';
if reg_code in ('23','26','83') and tage in ('51-65','56-60','61-65') and
diplc in ('1','2') then do;
tage='56-60'; diplc='1';end;
if reg_code in ('23','26','83') and tage in ('51-65','56-60','61-65') and
diplc in ('3','4') then do;
tage='56-60'; diplc='3';end;
if reg_code='22' and tage='36-40' and diplc in ('1','2') then diplc='1';
if reg_code='22' and tage='36-40' and diplc in ('3','4') then diplc='3';
if reg_code='94' and tage='51-55' and diplc in ('1','2') then diplc='1';
run;

proc freq data=indivp3a ;table q26res_aut_p q26resautp; run;
proc freq data=indivp3a (where=(q26resautp ne 'ZZ')) ;table q26resautp;
run;

data rep nonrep;set indivp3a;
if q26resautp='ZZ' then output nonrep;
else output rep;
proc sort data=rep ; by q26res_lgt_p_i reg_code tage diplc ;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q26res_aut_p; run;

proc sort data=nonrep; by q26res_lgt_p_i reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses** ;
class q26res_lgt_p_i reg_code tage diplc ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses** ;
class q26res_lgt_p_i reg_code tage diplc ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by q26res_lgt_p_i reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by q26res_lgt_p_i reg_code tage diplc;
alea=ranuni(14);
tri=q26res_lgt_p_i!!reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage dipl q26res_aut_p_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26resautp=q26res_aut_p_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

```



```

Proc freq data=res;
tables q26res_aut_p_i;
run;

data indivp3b;merge indivp3a(in=a) res (keep=cabefl q26res_aut_p_i);by
cabefl;
if a;
if q26res_aut_p_i=' ' then q26res_aut_p_i=q26resautp;
**indicateur : I_q26res_aut_p**;
resp3=substr(res_p,3,2);
length I_q26res_aut_p 8.;
if q26res_aut_p='ZZ' then I_q26res_aut_p=2; else
if q26res_aut_p_i=resp3 then I_q26res_aut_p=0; else
I_q26res_aut_p=1;
if q26res_aut_p_i='00' then do; q26reg=' ';q26resid=' ';end;
run;

Proc freq data=indivp3b ;
tables res_p resp3 I_q26res_aut_p q26resautp q26res_aut_p_i q26reg
q26resid;
run;

data rajoutp2; set indivp3b;
keep cabefl I_q26res_lgt_p q26res_lgt_p_i I_q26res_aut_p
q26res_aut_p_i;run;

*** q26resid - q26reg **;
data indivp4a;merge indivp3b (in=a) individu (keep=cabefl q20res_aut_m_i
q20reg_i q20resid_i) ;by cabefl;
if a;
run;
proc freq data=indivp4a; table q26reg q26resid;run;

data indivp4b; set indivp4a ;if q26res_aut_p_i='10';
if tage='66 et +' and reg_code='94' then reg_code='93';
if reg_code='22' and tage='66 et +' and diplc in ('3','4') then diplc='3';
if reg_code='42' and tage='61-65' and diplc in ('1','2') then diplc='2';
if reg_code='42' and tage='66 et +' and diplc in ('1','2','3') then
diplc='3';
if reg_code='52' and tage='66 et +' and diplc in ('1','2') then diplc='2';
if reg_code='94' and tage='61-65' and diplc in ('3','4') then diplc='3';
if reg_code='82' and tage='36-40' and diplc='Z' then diplc='3';
run;
proc freq data=indivp4b ;table tage reg_code q26res_aut_p q26res_aut_p_i
q26reg q26resid; run;

*** on impute d'abord q26resid **;
data indivp4; set indivp4b ;
if q26resid='9' then q26resid='Z';** on impute les Z et les 9 en même temps
**;
if q26res_aut_p_i='10';run;** 64 730 individus doivent avoir un
q20resid/q20reg rempli **;

proc freq data=indivp4 ;table q26resid; run;
proc freq data=indivp4 ;table reg_code tage diplc; run;

data rep nonrep;set indivp4;
if q26resid='Z' then output nonrep;
else output rep;
proc sort data=rep ; by reg_code tage diplc;;

```



```

*proc print data=rep (obs=10); run;
proc freq data=rep ;table q26resid; run;

proc sort data=nonrep; by reg_code tage diplc; ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc; ;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc; ;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc; ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;
data rep3 (keep=tage reg_code q26resid_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26resid=q26resid_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

data indivp4c;merge indivp4(in=a) res (keep=cabefl q26resid_i);by cabefl;
if a;
if q26resid_i=' ' then q26resid_i=q26resid;
** indicateur **;
length I_q26resid 8.;
if q26resid='Z' then I_q26resid=2; else
I_q26resid=1;
run;

*** q26reg **;
data indivp4d; set indivp4c;
** on met en cohérence q26resid avec q26reg ***;
q26regc=q26reg;
if q26resid_i='7' then q26regc='99';
if q26resid_i in ('1','2','3','8') then q26regc=reg_code;
if q26resid_i='5' then q26regc='D0';
if q26resid_i='6' then q26regc='C0';
reg=q26regc;
if reg not in ('99','ZZ') then reg='RE';
proc freq data=indivp4c ;table q26reg ; run;**12 357 ZZ**;
proc freq data=indivp4d ;table q26regc q26regc*q26resid_i
reg*q26resid_i/missing; run;**3509 ZZ**;

** les ZZ restant ont un Q26resid_i = 4 (3509)***;

```



```

** on peut imputer les 3509 **;
data indivp5a; set indivp4d;if Q26resid_i = '4';run;** pères résidant hors
reg_code **;
proc freq data=indivp5a ;table reg_code*q26regc reg_code*reg; run;

proc freq data=indivp4d ;table q26regc; run;

data rep nonrep;set indivp5a ;
if q26regc='ZZ' then output nonrep;
    else output rep;
proc sort data=rep ; by reg_code tage diplc;
*proc print data=rep (obs=10); run;
proc freq data=rep ;table q26reg; run;

proc sort data=nonrep; by reg_code tage diplc ;run;
*proc print data=nonrep (obs=10); run;

proc summary data=nonrep nway ; ** non-réponses**;
class reg_code tage diplc;
var eff;
output out=tabnonrep (drop=_type_ _freq_) sum=nbnonrep;
    proc print data=tabnonrep; run;

proc summary data=rep nway ; ** réponses**;
class reg_code tage diplc;
var eff;
output out=tabrep(drop=_type_ _freq_) sum=nbrep;
    proc print data=tabrep; run;

data verif; merge tabrep tabnonrep; by reg_code tage diplc ;
if nbnonrep>nbrep;run;

data rep2;merge rep tabnonrep;by reg_code tage diplc;
alea=ranuni(14);
tri=reg_code!!tage!!diplc;
proc sort data=rep2; by tri alea;run;
*proc print data=rep2 (obs=10) ; run;

data rep3 (keep=tage reg_code q26reg_i cpt nbnonrep tri);set rep2;by tri;
retain cpt;
if first.tri then cpt=1;
    else cpt=cpt+1;
if cpt<=nbnonrep then output;
rename q26regc=q26reg_i;
run;

data res ;merge nonrep rep3;run;
proc sort; by cabefl; run;

Proc freq data=res;
tables q26reg_i;
run;

** verif ****;
data verif;set res; if q26reg_i=reg_code; run;* 0 obs **;
****;

data indivp5b;merge indivp4d(in=a) res (keep=cabefl q26reg_i);by cabefl;
if a;
if q26reg_i=' ' then q26reg_i=q26regc;
length i_q26reg 8.;

```



```

if q26reg='ZZ' then I_q26reg=2; else
if q26reg_i=reg_m then I_q26reg=0; else
I_q26reg=1;
run;

proc freq data=indivp5b (where=(q26reg ne 'ZZ'));table q26regc; run;
Proc freq data=indivp5b ;
tables q26reg I_q26reg q26reg_i;
run;

data rajoutp3; set indivp5b;
keep cabefl q26reg_i I_q26reg q26resid_i I_q26resid;run;
Proc freq data=rajoutp3 (drop=cabefl) ;
tables _all_;
run;

data rajoutp; merge rajoutp1 rajoutp2 rajoutp3; by cabefl;run;
proc freq data=rajoutp (drop=cabefl); table _all_;run;

data individup; merge indiv (keep=cabefl poids_ind q26res_avec_m_p
q26res_lgt_p q26res_aut_p q26reg q26resid)
rajoutp; by cabefl; run;
proc freq data=individup (drop=cabefl poids_ind);
table q26res_avec_m_p q26res_avec_m_p_i I_q26res_avec_m_p
q26res_lgt_p q26res_lgt_p_i I_q26res_lgt_p
q26res_aut_p q26res_aut_p_i I_q26res_aut_p
q26reg q26reg_i I_q26reg
q26resid q26resid_i I_q26resid
;run;

data rajout_q20q26; merge individum individup; by cabefl; run;

data efldef.rajout_q20q26;set rajout_q20q26;run;

proc freq data=efldef.rajout_q20q26;
table
q20res_lgt_m q20res_lgt_m_i I_q20res_lgt_m
q20res_aut_m q20res_aut_m_i I_q20res_aut_m
q20reg q20reg_i i_q20reg
q20resid q20resid_i i_q20resid
q26res_avec_m_p q26res_avec_m_p_i I_q26res_avec_m_p
q26res_lgt_p q26res_lgt_p_i I_q26res_lgt_p
q26res_aut_p q26res_aut_p_i I_q26res_aut_p
q26reg q26reg_i I_q26reg
q26resid q26resid_i I_q26resid/missing
;run;

** envoi nathalie **;
data efldef.rajout_q15q21q20q26;merge
efldef.rajout_q15q21
efldef.rajout_q20q26
(keep= cabefl
q20res_lgt_m_i I_q20res_lgt_m
q20res_aut_m_i I_q20res_aut_m
q20reg_i i_q20reg
q20resid_i i_q20resid
q26res_avec_m_p_i I_q26res_avec_m_p
q26res_lgt_p_i I_q26res_lgt_p
q26res_aut_p_i I_q26res_aut_p
q26reg_i I_q26reg

```



```

        q26resid_i          I_q26resid);
; by cabefl;run;

data verif; merge efldef.res_pere_mere_apur
(rename=(q20res_lgt_m=q20reslgtm  q20res_aut_m=q20resautm q20reg=q20rega
q20resid=q20resida
q26res_avec_m_p=q26resavec_m_p q26res_lgt_p=q26reslgt_p q26res_aut_p=q26resaut_p
q26reg=q26rega q26resid=q26resida))
efldef.rajout_q20q26; by cabefl;run;

data compar; set verif;
if q26resid ne q26resida;run;
proc contents data=efldef.rajout_q15q21q20q26;run;

```

## Imputation aléatoire de la non-réponse des variables Q27grp I, Q27nbpenf I

### - Q27grp I

On utilise une variable de tranches d'âge, le nombre d'enfants, le sexe et une variable touslgt qui vaut 1 si tous les enfants d'ego habite le logement et 0 sinon, et on impute par hot-deck

```

tri=trage!!touslgt!!nbe !!q0sexe;

length trage $ 8.;
if 19<=agego<=32 then trage='19-32';else
if 33<=agego<=54 then trage='33-54';else
if 55<=agego<=59 then trage='55-59';else
if 60<=agego<=69 then trage='60-69';else
if 70<=agego then trage='70 et +';

length nbe $ 8.;
if q11nb_enf=0 then nbe='0';else
if q11nb_enf=1 then nbe='1';else
if q11nb_enf=2 then nbe='2';else
if q11nb_enf>=3 then nbe='3';

```

les 14757 non-réponses sont ainsi imputées :

Q27grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	6067	41.11	6067	41.11
2	8690	58.89	14757	100.00

Finalement après imputations :

Q27grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	116524	32.39	116524	32.39
2	243246	67.61	359770	100.00



	I_q27grp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
identique à la brute	0	287524	79.92	287524	79.92
apurée	1	57489	15.98	345013	95.90
imputée	2	14757	4.10	359770	100.00

- **Q27nbpenf I** :

On a 8615 non-réponses

On utilise une variable de tranches d'âge, le nombre d'enfants, et une variable enflgt à 3 modalités (« aucun » si aucun enfant n'habite le lgt, « tous », s'ils sont tous dans le logement et « part » sinon), et on impute par hot-deck

```
tri=trage!!enflgt!!nbe;
```

```
length trage $ 8.;
if 19<=agego<=32 then trage='19-32';else
if 33<=agego<=54 then trage='33-54';else
if 55<=agego<=59 then trage='55-59';else
if 60<=agego then trage='60 et +';
```

```
length nbe $ 8.;
if q11nb_enf=0 then nbe='0';else
if q11nb_enf=1 then nbe='1';else
if q11nb_enf=2 then nbe='2';else
if q11nb_enf>=3 then nbe='3';
```

Résultat de l'imputation :

q27nbpenf	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	2046	23.75	2046	23.75
2	2130	24.72	4176	48.47
3	1257	14.59	5433	63.06
4	984	11.42	6417	74.49
5	639	7.42	7056	81.90
6	432	5.01	7488	86.92
7	316	3.67	7804	90.59
8	204	2.37	8008	92.95
9	168	1.95	8176	94.90
10	91	1.06	8267	95.96
11	79	0.92	8346	96.88
12	64	0.74	8410	97.62
13	56	0.65	8466	98.27
14	28	0.33	8494	98.60
15	38	0.44	8532	99.04
16	19	0.22	8551	99.26
17	8	0.09	8559	99.35
18	12	0.14	8571	99.49
19	5	0.06	8576	99.55
20	7	0.08	8583	99.63
21	6	0.07	8589	99.70
22	5	0.06	8594	99.76
23	2	0.02	8596	99.78
24	3	0.03	8599	99.81



25	5	0.06	8604	99.87
26	1	0.01	8605	99.88
27	1	0.01	8606	99.90
28	2	0.02	8608	99.92
30	1	0.01	8609	99.93
32	2	0.02	8611	99.95
35	1	0.01	8612	99.97
39	2	0.02	8614	99.99
40	1	0.01	8615	100.00

	I_q27nbpenf	Frequency	Percent	Cumulative Frequency	Cumulative Percent
identique à la brute	0	108338	30.11	108338	30.11
apurée	1	242817	67.49	351155	97.61
imputée	2	8615	2.39	359770	100.00

## **Les autres données imputées aléatoirement ( I )**

Dans la table individu, le mois de naissance du répondant (Q1mnai\_i), les nombres de frères et sœurs du répondant (Q2), l'année et le mois de naissance du conjoint du répondant (Q4mnai\_C\_I et Q4anai\_C\_I), le mois de mise en couple (Q6mois\_C\_I), l'année et le mois de pacs (Q7annee\_pacs\_I, Q7mois\_Pacs\_I), l'année et le mois de mariage (Q8annee\_mari\_I, Q8mois\_mari\_I), le temps passé par le conjoint dans le logement (Q9tps\_logr\_I), La cause de la rupture et son année (Q10cause\_R\_I, Q10annee\_D\_I, Q10annee\_S\_I), la catégorie sociale et le statut de l'emploi de la mère et du père du répondant (Q17cs\_m\_i, Q18statut\_m\_i et Q23cs\_p\_i et Q24statut\_p\_i), ainsi que les données sur les autres résidences du répondant (Q29tps\_i, Q30tpsm\_i, Q31lgt\_sec\_i, Q31lgt\_sec\_fam\_i, Q31lgt\_sec\_trav\_i, Q31lgt\_sec\_et\_i, Q31lgt\_sec\_vac\_i, Q31lgt\_sec\_aut\_i, Q32natautlgt\_i, Q34tpsautlgt\_i, Q35persautlgt\_ego\_i, Q35persautlgt\_C\_i, Q35persautlgt\_fam\_i, Q35persautlgt\_hfam\_i, Q35persautlgt\_pers\_i, Q36propautlgt\_ego\_i, Q36propautlgt\_ami\_i, Q36propautlgt\_C\_i, Q36propautlgt\_enf\_i, Q36propautlgt\_par\_i, Q36propautlgt\_autfam\_i, Q36propautlgt\_aut\_i) sont issues du programme développé par l'INED contenu dans le document ci-dessous :

[imputation INED.doc](#)

## **La variable E13 14VIVANT I**

Elle est produite à partir du programme développée par l'INED en utilisant l'année de naissance et le sexe imputés par l'INSEE.

Cette imputation conduisait à considérer un trop grand nombre d'enfants ayant quitté le domicile de leur mère avant 1 an pour une autre raison que le décès. Les enfants nés avant 1993, partis avant 1 an du domicile de leur mère, et imputés par le programme en vivant ont ensuite été re-classés en décédés.



# Annexe



**NOTE**

*A l'attention des membres du groupe d'exploitation de  
l'enquête Famille et Logement*

Dossier suivi par :  
Blanpain Nathalie  
Tél. : 01 41 17 65 49  
Mél : [DG75-F170@insee.fr](mailto:DG75-F170@insee.fr)

Paris, le 11 juin 2012  
N° 1290/DG75-F170/

**Objet : Repérage des questionnaires suite, vide, doublon, jumeau, hors champ et appariement de l'enquête Famille et logements avec l'EAR**

Le premier objet de cette note est de présenter les différentes étapes avant l'appariement de l'EFL :

- repérage des questionnaires « principaux » et « suite »
- trois cas questionnaires « principaux » et « suite » du tableau28
- repérage des questionnaires quasiment vides
- repérage des questionnaires mineurs
- repérage des doublons et jumeaux

Le deuxième objet est la méthode d'appariement. Elle est présentée en point 6 et 7.

Le point 8 présente l'apurement de la variable SEXE\_X.

Le point 9 présente le repérage des questionnaires hors champ.

**DESTINATAIRES** : Nathalie Blanpain, Aude Lapinte, Liliane Lincot, Chantal Rebiscoul, Xavier Niel, Pascal Breuil, Marie Volhuer, Thibaut De Saint Pol, Solveig Vanovermeir, Emmanuelle Crenner, Jean-Baptiste Rudant, Chantal Dutilleul, Pauline Domingo, Estelle Bailly, Laurent Toulemon, Julie Chapon, Christophe Imbert, Loïc Trabut, David Lessault, Pascal ardilly, Ali hachid



## 1. Repérage des questionnaires « principaux » et « suite »

Les enquêtés devaient remplir un second questionnaire pour les tableaux enfants 13 ou 14 lorsqu'il y avait plus de 6 enfants dans le logement ou plus de 7 enfants hors logement.

Il s'agit de créer trois variables :

- **PRINCIPAL=1** pour le questionnaire ayant le tableau 13 ou 14 entièrement renseigné
- **SUITE=1** pour le ou les questionnaires détaillant les enfants qui n'ont pas pu être inscrits sur le questionnaire principal
- **CABEFL\_PRINCIPAL=CABEFL** du questionnaire principal

1.1. Dans un premier temps, on repère les questionnaires liés en cherchant deux ou trois questionnaires pour une même personne (même logement, même sexe, même date de naissance).

Si un des questionnaires a la dernière ligne du tableau 13 renseignée et un nombre d'enfants d'ego total ou dans le logement différent de 6 ou

Si un des questionnaires a la dernière ligne du tableau 14 renseignée et un nombre d'enfants d'ego total ou hors du logement différent de 7,

→ il s'agit du questionnaire principal et le second questionnaire est le questionnaire suite

1.2. Dans un deuxième temps, on repère les questionnaires quasiment vides avec au moins un enfant renseigné dans le tableau 13 ou 14.

1.2.a) Si un seul des questionnaires du logement a la dernière ligne du tableau 13 renseignée et un nombre d'enfants d'ego total ou dans le logement différent de 6 ou

Si un seul des questionnaires du logement a la dernière ligne du tableau 14 renseignée et un nombre d'enfants d'ego total ou hors du logement différent de 7,

→ il s'agit du questionnaire principal et le questionnaire quasiment vide est le questionnaire suite

1.2.b) Si deux des questionnaires du logement ont la dernière ligne du tableau 13 renseignée et un nombre d'enfants d'ego total ou dans le logement différent de 6 ou

Si deux des questionnaires du logement ont la dernière ligne du tableau 14 renseignée et un nombre d'enfants d'ego total ou hors du logement différent de 7,

→ un des questionnaires du logement ayant la dernière ligne du tableau 13 ou 14 renseignée est le questionnaire principal (celui avec un nombre d'enfants d'ego



renseigné). L'autre ainsi que le questionnaire quasiment vide sont des questionnaires suites. Il s'agit généralement d'individus ayant plus de 14 enfants hors logement.

1.3 Dans un troisième temps, on sélectionne à nouveau les questionnaires quasiment vides avec au moins un enfant renseigné dans le tableau 13 ou 14

1.3.a) Si un seul des questionnaires du district a la dernière ligne du tableau 13 renseignée et un nombre d'enfants d'ego total ou dans le logement différent de 6 ou

Si un seul des questionnaires du district a la dernière ligne du tableau 14 renseignée et un nombre d'enfants d'ego total ou hors du logement différent de 7,

→ il s'agit du questionnaire principal et le questionnaire quasiment vide est le questionnaire suite

1.3.b) Si plusieurs des questionnaires du district ont la dernière ligne du tableau 13 renseignée et un nombre d'enfants d'ego total ou dans le logement différent de 6 ou

Si plusieurs des questionnaires du district ont la dernière ligne du tableau 14 renseignée et un nombre d'enfants d'ego total ou hors du logement différent de 7,

→ traitement manuel pour déterminer les questionnaires principaux et suite

principal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366518	99.85	366518	99.85
1	557	0.15	367075	100.00

suite	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366514	99.85	366514	99.85
1	561	0.15	367075	100.00



## 2. Trois cas questionnaires « principaux » et « suite » du tableau28

Il s'agit de créer deux variables :

- PRINCIPAL28=1 pour le questionnaire principal ayant le tableau 28 renseigné
- SUITE28=1 pour le ou les questionnaires détaillant des enfants supplémentaires au tableau28

### bulletins « suite » tableau 28

```

17196__0001_0570011F00433746  2 17 12 1980 . 1 . .      2
17196__0001_0570011F00436637  2 17 12 1980 1 . . .      1

69383__0202_0060071F00631569  2 11 07 1976 . 1 . .      2
69383__0202_0060071F00631612  2 11 07 1976 1 . . .      1

16245__0004_0620011F00775550  2 09 12 1972 1 . . .      2
16245__0004_0620011F00433547  2 09 12 1972 . 1 . .      1
  
```

principal28	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	367072	100.00	367072	100.00
1	3	0.00	367075	100.00

suite28	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	367072	100.00	367072	100.00
1	3	0.00	367075	100.00

## 3. Repérage des questionnaires quasiment vides

Il s'agit de créer une variable VIDE=1

- si on considère que le questionnaire n'a pas suffisamment de réponses pour être exploité.

On crée un compteur qui comptabilise 1 si

- l'année et le mois d'ego vide
- la variable couple vide
- le sexe et l'année du conjoint l'année de mise en couple, l'année de mariage, la cause de rupture, l'année de séparation, l'année de décès vides, et vit ou a vécu en couple
- les variables enfant vides
- les variables mère vides
- les variables père vides
- la variable grand-parent vide
- les variables logement vides



compteur	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	264469	72.27	264469	72.27
1	66521	18.18	330990	90.45
2	18058	4.93	349048	95.38
3	7299	1.99	356347	97.38
4	4197	1.15	360544	98.52
5	2162	0.59	362706	99.11
6	964	0.26	363670	99.38
7	1351	0.37	365021	99.75
8	930	0.25	365951	100.00

Frequency Missing = 1124

On supprime les questionnaires dont le compteur vaut 7 ou 8, et ceux dont le compteur vaut 6 et dont la variable couple, le bloc conjoint sont vides, et le nombre d'enfants est vide ou supérieur à 6

**Pour compteur=8**, tous les blocs sont vides → VIDE=1

**Pour compteur=7**, un seul bloc renseigné (le plus fréquent l'année d'ego) → VIDE=1

**Pour compteur=6**, VIDE=1 sauf si la variable couple renseigné ou le bloc conjoint renseigné ou le nombre d'enfants vaut 1,2,3,4,5,6

Exemple : les blocs renseignés sont :

l'année d'ego + bloc couple → on garde car info centrale

l'année d'ego + bloc conjoint → on garde car info centrale

l'année d'ego + bloc logement → VIDE=1

l'année d'ego + bloc enfant → VIDE=1 si NB\_ENF à blanc ou supérieur à 6 (il peut s'agir d'un questionnaire suite dont le questionnaire principal est difficile à trouver ou n'est pas présent). 74 questionnaires sont dans ce cas.

→ on garde si NB\_ENF vaut 1,2,3,4,5,6 (il ne s'agit pas de questionnaires suite, car les questionnaires suite ont en général un nombre d'enfant à blanc ou supérieur à 6)

**Pour compteur=5**, les blocs renseignés les plus fréquents sont :

L'année d'ego+couple+bloc conjoint → on garde

L'année d'ego+bloc mère+bloc père → on garde

L'année d'ego+bloc conjoint+bloc enfant → on garde

vide	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	364423	99.28	364423	99.28
1	2652	0.72	367075	100.00

On supprime donc 0,72% des questionnaires soit 2652 bulletins

```
Data test10;set test9;
```

```
if suite ne 1 & principal ne 1 & suite28 ne 1 & principal28 ne 1
then do;
```

```
/* suppression des questionnaires très peu renseignés */
compteur=sum((mnai_x=' ' & anai_x=' '), (couple not in
('1000', '0100', '0010', '0001'))),
```



```

(anai_c=' ' & sexe_c not in ('10','01') & annee_c=' ' &
annee_mari=' ' & cause_r not in ('01','10') & annee_s=' ' &
annee_d=' ' & couple ne '0001'),
(enfan not in ('10','01') & nb_enf=' ' &
prenom_131=' ' & prenom_141='
' & anai_131=' ' & anai_141=' '),
(statut_m not in ('1000','0100','0010','0001') & viv_m not in
('10','01') & nat_m not in ('10','01')),
(statut_p not in ('1000','0100','0010','0001') & viv_p not in
('10','01') & nat_p not in ('10','01')),
(grp not in ('01','10')),
(tps not in ('100000','010000','001000','000100','000010','000001')
& tpsm not in ('100','010','001') &
lgtsec not in
('100000','010000','001000','000100','000010','000001') & natautlgt
not in ('1000','0100','0010','0001'))
);

infoprincipal=sum((couple not in ('1000','0100','0010','0001')),
(anai_c=' ' & sexe_c not in ('10','01') & annee_c=' ' &
annee_mari=' ' & cause_r not in ('10','01') & annee_s=' ' &
annee_d=' ' & couple ne '0001'),
(nb_enf not in ('01','02','03','04','05','06'))
);

if compteur in (7,8) ! (compteur=6 & infoprincipal=3) then vide=1;
end;

infoprincipal=sum((couple not in
('1000','0100','0010','0001')), (anai_c=' ' & sexe_c not in
('10','01') & cause_r not in ('10','01') & couple ne '0001'));

if compteur in (7,8) ! (compteur=6 & infoprincipal=2) then vide=1;
end;

```

## 4. Repérage des questionnaires mineurs

**Il s'agit de créer une variable MINEUR=1 pour les personnes nées en 1993 ou après**

On supprime les questionnaires des mineurs au 1<sup>er</sup> janvier 2011 (nés en 1993 ou après)

mineur	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366669	99.89	366669	99.89
1	406	0.11	367075	100.00



## 5. Repérage des doublons et jumeaux

Il s'agit de créer une variable :

**DOUBLON=1** s'il s'agit de la même personne qui a rempli deux questionnaires ou plus ou d'une personne qui a rempli un questionnaire pour elle-même et son conjoint

**DOUBLONPRINCIPAL=1** pour le questionnaire doublon qui a le plus de réponses renseignées

**JUMEAU=1** s'il s'agit de deux personnes nées le même jour (la date de naissance est renseignée) dans le même logement (ces deux personnes ne sont pas nécessairement de la même fratrie)

On considère que le questionnaire est un doublon si la date de naissance du conjoint ou la date de mise en couple ou la date de mariage ou le prénom du premier enfant du tableau 13 ou le prénom du premier enfant du tableau 14 sont renseignés et identiques pour une même personne (même logement, même sexe, même date de naissance) .

Pour déterminer le doublon principal, on retient celui ayant le plus de variables renseignées. En cas d'égalité, on retient le premier.

Les jumeaux correspondent à tous les autres cas (deux personnes dans le même logement ayant le même sexe et la même date de naissance, mais n'étant pas considérées comme DOUBLON). On ne peut toutefois pas exclure qu'il s'agit d'un doublon, sans conjoint et sans enfant.

doublon	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366983	99.97	366983	99.97
1	92	0.03	367075	100.00

doublonprincipal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	367029	99.99	367029	99.99
1	46	0.01	367075	100.00

jumeau	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366654	99.89	366654	99.89
1	421	0.11	367075	100.00

```
/** repérage des doublons **/  
data test12;merge test10 test11(keep=apparind nbreind2 jnai_c1  
jnai_c2 mnai_c1 mnai_c2 anai_c1 anai_c2 mois_c1 mois_c2 annee_c1  
annee_c2 mois_mari1 mois_mari2  
annee_mari1  
annee_mari2 prenom_1311 prenom_1312 prenom_1411 prenom_1412  
compteurdoublon1 compteurdoublon2 cabefl1 cabefl2);by apparind;  
  
if jnai_c1 ne ' ' & mnai_c1 ne ' ' & anai_c1 ne ' ' then do;
```



```

if jnai_c1=jnai_c2 & mnai_c1=mnai_c2 & anai_c1=anai_c2 then
doublon=1;
end;

if mois_c1 ne ' ' & annee_c1 ne ' ' then do;
if mois_c1=mois_c2 & annee_c1=annee_c2 then doublon=1;
end;

if mois_mari1 ne ' ' & annee_mari1 ne ' ' then do;
if mois_mari1=mois_mari2 & annee_mari1=annee_mari2 then doublon=1;
end;

if prenom_1311 ne ' ' then
do;
if prenom_1311=prenom_1312 then doublon=1;
end;

if prenom_1411 ne ' ' then
do;
if prenom_1411=prenom_1412 then doublon=1;
end;

if doublon=1 then do;
if compteurdoublon1 ge compteurdoublon2 & cabefl=cabefl1 then
DOUBLONPRINCIPAL=1;
if compteurdoublon2>compteurdoublon1 & cabefl=cabefl2 then
DOUBLONPRINCIPAL=1;
end;

if nbreind2 in (2,3) & doublon=. then jumeau=1;

```

## 6. Appariement

On supprime tout d'abord les questionnaires suite (tableau 13,14 ou 28), les questionnaires doublons non principaux et les questionnaires quasiment vides et les jumeaux, soit 363 007 individus. Les jumeaux sont traités spécifiquement.

```

if suite=1 ! (doublon=1 & doublonprincipal ne 1) ! vide=1 !
suite28=1 ! jumeau=1 then delete;

```

**6.1 On réalise tout d'abord un appariement direct entre l'EFL et le fichier d'appariement (même logement, même sexe, même année de naissance et année de naissance ; la date de naissance est renseignée)**

326 798 individus de l'EFL s'apparient directement avec 1 seul individu du fichier d'appariement.

61 individus de l'EFL s'apparient avec 2 individus du fichier d'appariement

1 individu de l'EFL s'apparient avec 3 individus ou plus du fichier d'appariement

36 147 ne s'apparient pas

On calcule une variable de score de qualité d'appariement entre l'EFL et le fichier d'appariement (à partir des variables vie en couple, est marié, mois et année de naissance du conjoint, sexe du conjoint, a des enfants dans le logement, nombre d'enfants de la famille, année de naissance des enfants).



Pour les 62 individus s'appariant avec 2 individus ou plus, on retient le BI ayant le score le plus élevé. En cas d'égalité, on en retient un aléatoirement (il semble s'agir de quelques cas de doublons dans l'Ear).

Programme score :

```

positif&no=0;
negatif&no=0;

if substr(couple,1,1)='1' & (couple_r&no = '1' ! moco_r&no in
('21','22')) then positif&no=positif&no+1;
if substr(couple,1,1)='1' & couple_r&no = '2' then
negatif&no=negatif&no-1;

if (substr(couple,3,1)='1' ! substr(couple,4,1)='1') & couple_r&no
='1' then negatif&no=negatif&no-1;
if (substr(couple,3,1)='1' ! substr(couple,4,1)='1' ! (cause_r in
('10','01') & couple='0000')) & couple_r&no = '2' then
positif&no=positif&no+1;

if substr(couple,2,1)='1' & moco_r&no in ('21','22') then
negatif&no=negatif&no-1;

if mari='10' & matr_r&no='2' then positif&no=positif&no+1;
if substr(couple,3,1) ne '1' & substr(couple,4,1) ne '1' & cause_r
ne '01' & mari='10' & matr_r&no in ('1','3','4') then
negatif&no=negatif&no-1;

if mari='01' & matr_r&no='2' then negatif&no=negatif&no-1;
if mari='01' & matr_r&no in ('1','3','4') then
positif&no=positif&no+1;

if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & jnai_c=jnai_conjoint&no &
mnai_c=mnai_conjoint&no & anai_c=anai_conjoint&no then
positif&no=positif&no+10;

if (jnai_c ne ' ' & mnai_c ne ' ' & anai_c ne ' ') & (
(jnai_c=hj1 & mnai_c=hml & anai_c=hal ) ! (jnai_c=hj2 & mnai_c=hm2 &
anai_c=ha2) )
then positif&no=positif&no+5;

if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & jnai_c=jnai_conjoint&no &
mnai_c=mnai_conjoint&no &
anai_c ne anai_conjoint&no then positif&no=positif&no+2;
if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & jnai_c=jnai_conjoint&no & mnai_c ne
mnai_conjoint&no &
anai_c=anai_conjoint&no then positif&no=positif&no+2;
if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & jnai_c ne jnai_conjoint&no &
mnai_c=mnai_conjoint&no &
anai_c=anai_conjoint&no then positif&no=positif&no+2;

if (jnai_x ne jnai_c ! mnai_x ne mnai_c ! anai_x ne anai_c) &
couple='1000' then do;
if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & (mnai_c ne mnai_conjoint&no & anai_c ne
anai_conjoint&no) then negatif&no=negatif&no-1;

```



```

if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & (mnai_c = mnai_conjoint&no & anai_c ne
anai_conjoint&no) then negatif&no=negatif&no-1;
if mnai_c ne ' ' & mnai_conjoint&no ne . & anai_c ne ' ' &
anai_conjoint&no ne . & (mnai_c ne mnai_conjoint&no & anai_c =
anai_conjoint&no) then negatif&no=negatif&no-1;
end;

if (mnai_c=' ' ! mnai_conjoint&no=.) & anai_c ne ' ' &
anai_conjoint&no ne . & anai_c=anai_conjoint&no then
positif&no=positif&no+2;

if (jnai_x ne jnai_c ! mnai_x ne mnai_c ! anai_x ne anai_c) &
couple='1000' then do;
if (mnai_c=' ' ! mnai_conjoint&no=.) & anai_c ne ' ' &
anai_conjoint&no ne . & anai_c ne anai_conjoint&no & couple='1000'
then negatif&no=negatif&no-1;
end;

if mnai_c ne ' ' & mnai_conjoint&no ne . & (anai_c=' ' !
anai_conjoint&no=.) & mnai_c=mnai_conjoint&no then
positif&no=positif&no+2;

if (jnai_x ne jnai_c ! mnai_x ne mnai_c ! anai_x ne anai_c) &
couple='1000' then do;
if mnai_c ne ' ' & mnai_conjoint&no ne . & (anai_c=' ' !
anai_conjoint&no=.) & mnai_c ne mnai_conjoint&no & couple='1000'
then negatif&no=negatif&no-1;
end;

if sexe_c='1' & sexe_conjoint&no='1' then positif&no=positif&no+1;
if sexe_c='2' & sexe_conjoint&no='2' then positif&no=positif&no+1;

if sexe_c='1' & sexe_conjoint&no='2' then negatif&no=negatif&no-1;
if sexe_c='2' & sexe_conjoint&no='1' then negatif&no=negatif&no-1;

if moco_r&no in ('22','23') & (nb_enf_lgt>'00' ! nb_enf_lgt_c>'00' !
prenom_131 ne ' ') then positif&no=positif&no+1;
if moco_r&no in ('11','12','21','31','32') & (nb_enf_lgt>'00' !
nb_enf_lgt_c>'00' ! prenom_131 ne ' ') then negatif&no=negatif&no-
1;

if moco_r&no in ('22','23') & (prenom_131 = ' ' & sexe_131='00')
then negatif&no=negatif&no-1;
if moco_r&no in ('11','12','21','31','32') & (prenom_131 = ' ' &
sexe_131='00') then positif&no=positif&no+1;

if moco_r&no in ('22','23') & nenf_r&no=1 & ( ((prenom_131 ne '
' ! sexe_131 ne '00') & (prenom_132=' ' & sexe_132='00')) !
nbsreenfdans=1) then positif&no=positif&no+1;
if moco_r&no in ('22','23') & nenf_r&no=2 & ( ((prenom_132 ne '
' ! sexe_132 ne '00') & (prenom_133=' ' & sexe_133='00')) !
nbsreenfdans=2) then positif&no=positif&no+1;
if moco_r&no in ('22','23') & nenf_r&no=3 & ( ((prenom_133 ne '
' ! sexe_133 ne '00') & (prenom_134=' ' & sexe_134='00')) !
nbsreenfdans=3) then positif&no=positif&no+1;
if moco_r&no in ('22','23') & nenf_r&no=4 & ( ((prenom_134 ne '
' ! sexe_134 ne '00') & (prenom_135=' ' & sexe_135='00')) !
nbsreenfdans=4) then positif&no=positif&no+1;

```



```

if moco_r&no in ('22','23') & nenf_r&no=5 & (((prenom_135 ne '
' ! sexe_135 ne '00') & (prenom_136=' ' & sexe_136='00')) !
nbreenfdans=5) then positif&no=positif&no+1;
if moco_r&no in ('22','23') & nenf_r&no ge 6 & (((prenom_136 ne '
' ! sexe_136 ne '00')
) !
nbreenfdans>5) then positif&no=positif&no+1;

if moco_r&no in ('22','23') & nenf_r&no=1 & (((prenom_131= ' '
& sexe_131 = '00') ! (prenom_132 ne ' ' ! sexe_132 ne '00')) &
nbreenfdans ne 1)
then negatif&no=negatif&no-1;
if moco_r&no in ('22','23') & nenf_r&no=2 & (((prenom_132= ' '
& sexe_132 = '00') ! (prenom_133 ne ' ' ! sexe_133 ne '00')) &
nbreenfdans ne 2)
then negatif&no=negatif&no-1;
if moco_r&no in ('22','23') & nenf_r&no=3 & (((prenom_133= ' '
& sexe_133 = '00') ! (prenom_134 ne ' ' ! sexe_134 ne '00')) &
nbreenfdans ne 3)
then negatif&no=negatif&no-1;
if moco_r&no in ('22','23') & nenf_r&no=4 & (((prenom_134= ' '
& sexe_134 = '00') ! (prenom_135 ne ' ' ! sexe_135 ne '00')) &
nbreenfdans ne 4)
then negatif&no=negatif&no-1;
if moco_r&no in ('22','23') & nenf_r&no=5 & (((prenom_135= ' '
& sexe_135 = '00') ! (prenom_136 ne ' ' ! sexe_136 ne '00')) &
nbreenfdans ne 5)
then negatif&no=negatif&no-1;
if moco_r&no in ('22','23') & nenf_r&no ge 6 & (((prenom_136= ' '
& sexe_136 = '00')
)
& nbreenfdans not in (6,7,8,9,10,11,12,13,14,15,16,17))
then negatif&no=negatif&no-1;

if anai_131 ne ' ' & (anai_131=a1&no ! anai_131=a2&no !
anai_131=a3&no ! anai_131=a4&no) then positif&no=positif&no+1;
if anai_132 ne ' ' & (anai_132=a1&no ! anai_132=a2&no !
anai_132=a3&no ! anai_132=a4&no) then positif&no=positif&no+1;
if anai_133 ne ' ' & (anai_133=a1&no ! anai_133=a2&no !
anai_133=a3&no ! anai_133=a4&no) then positif&no=positif&no+1;
if anai_134 ne ' ' & (anai_134=a1&no ! anai_134=a2&no !
anai_134=a3&no ! anai_134=a4&no) then positif&no=positif&no+1;
sc&no=sum(positif&no,negatif&no);

```

#### The FREQ Procedure

sc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-3	41	0.01	41	0.01
-2	229	0.07	270	0.08
-1	1004	0.31	1274	0.39
0	5825	1.78	7099	2.17
1	6008	1.84	13107	4.01
2	88026	26.93	101133	30.94
3	6637	2.03	107770	32.97
4	10524	3.22	118294	36.19
5	5586	1.71	123880	37.90
6	2660	0.81	126540	38.71
7	1718	0.53	128258	39.24
8	1411	0.43	129669	39.67
9	751	0.23	130420	39.90
10	2018	0.62	132438	40.52



11	9865	3.02	142303	43.54
12	12194	3.73	154497	47.27
13	84578	25.88	239075	73.14
14	24979	7.64	264054	80.79
15	35144	10.75	299198	91.54
16	19986	6.11	319184	97.65
17	5834	1.78	325018	99.44
18	1501	0.46	326519	99.90
19	160	0.05	326679	99.94
20	116	0.04	326795	99.98
21	51	0.02	326846	100.00
22	11	0.00	326857	100.00
23	3	0.00	326860	100.00

Les scores négatifs semblent correspondre le plus souvent à des erreurs dans l'AMF de l'Ear et non à des cas d'erreur dans l'appariement.

**6.2 On réalise un appariement entre l'EFL et le fichier d'appariement en supprimant une ou plusieurs variables d'appariement (la date de naissance est renseignée) :**

appariement	Frequency	Percent	Cumulative Frequency	Cumulative Percent
01fa -direct	326860	90.04	326860	90.04
02fa -sans jour	2648	0.73	329508	90.77
03fa -sans mois	2025	0.56	331533	91.33
04fa -sans année+score pas mauvais	2955	0.81	334488	92.14
05fa -sans sexe	1165	0.32	335653	92.46
06fa -sans logement	4801	1.32	340454	93.79
07fa -sans adresse	5078	1.40	345532	95.19
08fa -sans fil	103	0.03	345635	95.21
09fa -sans district	998	0.27	346633	95.49
10fa -sans cil	127	0.03	346760	95.52
11fa -sans logement adresse	1075	0.30	347835	95.82
12fa -sans fil jour	1	0.00	347836	95.82
13fa -sans fil année+score pas mauvais	1	0.00	347837	95.82
15fa -sans jour année+score pas mauvais	539	0.15	348376	95.97
16fa -sans commune	234	0.06	348610	96.03
17fa -sans departement	82	0.02	348692	96.06
18fa -sans district adresse+score très bon	79	0.02	348771	96.08
19fa -sans district logement+score très bon	17	0.00	348788	96.08
20fa -sans departement adresse+score très bon	7	0.00	348795	96.08
21fa -sans cil district	2	0.00	348797	96.09
22fa -sans cil logement	7	0.00	348804	96.09
74 non apparie	14203	3.91	363007	100.00

Si le questionnaire EFL s'apparie avec plusieurs bulletins (ce qui est très rare), on retient celui ayant le score le plus élevé. En cas de score identique, on en retient un aléatoirement.

Les scores d'appariement ont été étudiés. Si le score d'appariement est très faible, on ne retient que les questionnaires ayant un très bon score (+score très bon ci dessus). Ce score est variable selon le type de famille :

```
if (moco_r in ('11','12','31','32') & sc>1) ! (moco_r=' ' & sc>2) !
(moco_r='40' & sc>3) ! (moco_r='23' & sc>4) !
(moco_r='21' & sc>12) ! (moco_r='22' & sc>13)
```

Si le score d'appariement est un peu faible, on élimine les questionnaires ayant un très mauvais score (+score pas mauvais ci dessus). Ce score est variable selon le type de famille :



```

if (moco_r='23' & sc>-2) ! (moco_r in ('31','32') & sc>-1) ! (moco_r
in (' ', '11', '12', '40') & sc>0) ! (moco_r='21' & sc>3) !
(moco_r='22' & sc>6)

```

6.3 On réalise un appariement direct entre l'EFL et l'Ear, puis en supprimant une ou plusieurs variables d'appariement (la date de naissance est renseignée) :

L'appariement **avec l'Ear** correspond à des questionnaires dont la case EFL n'a pas été cochée. En gras, ci dessous :

appariement	Frequency	Percent	Cumulative Frequency	Cumulative Percent
01fa -direct	326860	90.04	326860	90.04
02fa -sans jour	2648	0.73	329508	90.77
03fa -sans mois	2025	0.56	331533	91.33
04fa -sans année+score pas mauvais	2955	0.81	334488	92.14
05fa -sans sexe	1165	0.32	335653	92.46
06fa -sans logement	4801	1.32	340454	93.79
07fa -sans adresse	5078	1.40	345532	95.19
08fa -sans fil	103	0.03	345635	95.21
09fa -sans district	998	0.27	346633	95.49
10fa -sans cil	127	0.03	346760	95.52
11fa -sans logement adresse	1075	0.30	347835	95.82
12fa -sans fil jour	1	0.00	347836	95.82
13fa -sans fil année+score pas mauvais	1	0.00	347837	95.82
15fa -sans jour année+score pas mauvais	539	0.15	348376	95.97
16fa -sans commune	234	0.06	348610	96.03
17fa -sans departement	82	0.02	348692	96.06
18fa -sans district adresse+score très bon	79	0.02	348771	96.08
19fa -sans district logement+score très bon	17	0.00	348788	96.08
20fa -sans departement adresse+score très bon	7	0.00	348795	96.08
21fa -sans cil district	2	0.00	348797	96.09
22fa -sans cil logement	7	0.00	348804	96.09
<b>30ear-direct</b>	<b>2758</b>	<b>0.76</b>	<b>351562</b>	<b>96.85</b>
<b>31ear-sans logement</b>	<b>3582</b>	<b>0.99</b>	<b>355144</b>	<b>97.83</b>
<b>32ear-sans adresse</b>	<b>511</b>	<b>0.14</b>	<b>355655</b>	<b>97.97</b>
<b>33ear-sans logement adresse</b>	<b>2228</b>	<b>0.61</b>	<b>357883</b>	<b>98.59</b>
<b>34ear-sans district</b>	<b>13</b>	<b>0.00</b>	<b>357896</b>	<b>98.59</b>
<b>35ear-sans commune+score très bon</b>	<b>6</b>	<b>0.00</b>	<b>357902</b>	<b>98.59</b>
<b>36ear-sans logement sexe+score très bon</b>	<b>44</b>	<b>0.01</b>	<b>357946</b>	<b>98.61</b>
<b>37ear-sans adresse sexe+score très bon</b>	<b>4</b>	<b>0.00</b>	<b>357950</b>	<b>98.61</b>
<b>38ear-sans logement adresse sexe+score très bon</b>	<b>34</b>	<b>0.01</b>	<b>357984</b>	<b>98.62</b>
<b>39ear-sans logement année+score pas mauvais</b>	<b>144</b>	<b>0.04</b>	<b>358128</b>	<b>98.66</b>
<b>40ear-sans adresse année+score très bon</b>	<b>11</b>	<b>0.00</b>	<b>358139</b>	<b>98.66</b>
<b>41ear-sans logementadresse année+ score très bon</b>	<b>72</b>	<b>0.02</b>	<b>358211</b>	<b>98.68</b>
<b>42ear-sans logement mois+score pas mauvais</b>	<b>87</b>	<b>0.02</b>	<b>358298</b>	<b>98.70</b>
<b>43ear-sans adresse mois+score pas mauvais</b>	<b>16</b>	<b>0.00</b>	<b>358314</b>	<b>98.71</b>
<b>44ear-sans logement adresse mois+score très bon</b>	<b>41</b>	<b>0.01</b>	<b>358355</b>	<b>98.72</b>
<b>45ear-sans logement jour+score pas mauvais</b>	<b>132</b>	<b>0.04</b>	<b>358487</b>	<b>98.75</b>
<b>46ear-sans adresse jour+score très bon</b>	<b>13</b>	<b>0.00</b>	<b>358500</b>	<b>98.76</b>
<b>47ear-sans logement adresse jour+score très bon</b>	<b>69</b>	<b>0.02</b>	<b>358569</b>	<b>98.78</b>
74 non apparie	4438	1.22	363007	100.00

Si le questionnaire EFL s'apparie avec plusieurs bulletins (ce qui est très rare), on retient celui ayant le score le plus élevé. En cas de score identique, on en retient un aléatoirement.



Les scores d'appariement ont été étudiés. Si le score d'appariement est très faible, on ne retient que les questionnaires ayant un très bon score (+score très bon ci dessus). Ce score est variable selon le type de famille

```
if (moco_r in ('11','12','31','32') & sc>1) ! (moco_r=' ' & sc>2) !
(moco_r='40' & sc>3) ! (moco_r='23' & sc>4) !
(moco_r='21' & sc>12) ! (moco_r='22' & sc>13)
```

Si le score d'appariement est un peu faible, on élimine les questionnaires ayant un très mauvais score (+score pas mauvais ci dessus). Ce score est variable selon le type de famille

```
if (moco_r='23' & sc>-2) ! (moco_r in ('31','32') & sc>-1) ! (moco_r
in (' ','11','12','40') & sc>0) ! (moco_r='21' & sc>3) !
(moco_r='22' & sc>6)
```

#### 6.4 On réalise un appariement entre l'EFL et le fichier d'appariement, en supprimant une ou plusieurs variables d'appariement :

On apparie **avec le fichier d'appariement** sans la date de naissance puis sans la date de naissance et le sexe (si la date de naissance est à blanc dans le fichier d'appariement ou dans l'EFL) :

appariement	Frequency	Percent
01fa -direct	326860	90.04
02fa -sans jour	2648	0.73
03fa -sans mois	2025	0.56
04fa -sans année+score pas mauvais	2955	0.81
05fa -sans sexe	1165	0.32
06fa -sans logement	4801	1.32
07fa -sans adresse	5078	1.40
08fa -sans fil	103	0.03
09fa -sans district	998	0.27
10fa -sans cil	127	0.03
11fa -sans logement adresse	1075	0.30
12fa -sans fil jour	1	0.00
13fa -sans fil année+score pas mauvais	1	0.00
15fa -sans jour année+score pas mauvais	539	0.15
16fa -sans commune	234	0.06
17fa -sans departement	82	0.02
18fa -sans district adresse+score très bon	79	0.02
19fa -sans district logement+score très bon	17	0.00
20fa -sans departement adresse+score très bon	7	0.00
21fa -sans cil district	2	0.00
22fa -sans cil logement	7	0.00
30ear-direct	2758	0.76
31ear-sans logement	3582	0.99
32ear-sans adresse	511	0.14
33ear-sans logement adresse	2228	0.61
34ear-sans district	13	0.00
35ear-sans commune+score très bon	6	0.00
36ear-sans logement sexe+score très bon	44	0.01
37ear-sans adresse sexe+score très bon	4	0.00
38ear-sans logement adresse sexe+score très bon	34	0.01
39ear-sans logement année+score pas mauvais	144	0.04



40ear-sans adresse année+score très bon	11	0.00
41ear-sans logementadresse année+ score très bon	72	0.02
42ear-sans logement mois+score pas mauvais	87	0.02
43ear-sans adresse mois+score pas mauvais	16	0.00
44ear-sans logement adresse mois+score très bon	41	0.01
45ear-sans logement jour+score pas mauvais	132	0.04
46ear-sans adresse jour+score très bon	13	0.00
47ear-sans logement adresse jour+score très bon	69	0.02
<b>51fa -sans date naissance,date naissance blanc EFL+score pas mauvais</b>	<b>468</b>	<b>0.13</b>
<b>52fa -sans sexe date naissances,date naissance blanc EFL+score très bon</b>	<b>20</b>	<b>0.01</b>
<b>53fa -sans date naissance,date naissance blanc fa+score tres bon</b>	<b>133</b>	<b>0.04</b>
<b>54fa -sans sexe date naissances,date naissance blanc fa+score très bon</b>	<b>19</b>	<b>0.01</b>
74 non apparie	3798	1.05

6.5 On réalise un appariement entre l'EFL et l'Ear, en supprimant une ou plusieurs variables d'appariement :

On apparie **avec l'Ear** sans la date de naissance puis sans la date de naissance et le sexe (si la date de naissance est à blanc dans l'EFL ou dans l'Ear) :

appariement	Frequency	Percent
01fa -direct	326860	90.04
02fa -sans jour	2648	0.73
03fa -sans mois	2025	0.56
04fa -sans année+score pas mauvais	2955	0.81
05fa -sans sexe	1165	0.32
06fa -sans logement	4801	1.32
07fa -sans adresse	5078	1.40
08fa -sans fil	103	0.03
09fa -sans district	998	0.27
10fa -sans cil	127	0.03
11fa -sans logement adresse	1075	0.30
12fa -sans fil jour	1	0.00
13fa -sans fil année+score pas mauvais	1	0.00
15fa -sans jour année+score pas mauvais	539	0.15
16fa -sans commune	234	0.06
17fa -sans departement	82	0.02
18fa -sans district adresse+score très bon	79	0.02
19fa -sans district logement+score très bon	17	0.00
20fa -sans departement adresse+score très bon	7	0.00
21fa -sans cil district	2	0.00
22fa -sans cil logement	7	0.00
30ear-direct	2758	0.76
31ear-sans logement	3582	0.99
32ear-sans adresse	511	0.14
33ear-sans logement adresse	2228	0.61
34ear-sans district	13	0.00
35ear-sans commune+score très bon	6	0.00
36ear-sans logement sexe+score très bon	44	0.01
37ear-sans adresse sexe+score très bon	4	0.00
38ear-sans logement adresse sexe+score très bon	34	0.01
39ear-sans logement année+score pas mauvais	144	0.04
40ear-sans adresse année+score très bon	11	0.00
41ear-sans logementadresse année+ score très bon	72	0.02
42ear-sans logement mois+score pas mauvais	87	0.02
43ear-sans adresse mois+score pas mauvais	16	0.00
44ear-sans logement adresse mois+score très bon	41	0.01



45ear-sans logement jour+score pas mauvais	132	0.04
46ear-sans adresse jour+score très bon	13	0.00
47ear-sans logement adresse jour+score très bon	69	0.02
51fa -sans date naissance,date naissance blanc EFL+score pas mauvais	468	0.13
52fa -sans sexe date naissance,date naissance blanc EFL+score très bon	20	0.01
53fa -sans date naissance,date naissance blanc fa+score tres bon	133	0.04
54fa -sans sexe date naissances,date naissance blanc fa+score très bon	19	0.01
<b>61ear-sans date naissance,date naissance blanc EFL+score très bon</b>	<b>18</b>	<b>0.00</b>
<b>62ear-sans sexe date naissance,date naissance blanc EFL+score très bon</b>	<b>1</b>	<b>0.00</b>
74 non apparie	3779	1.04

6.6 On réalise un appariement entre l'EFL et le fichier d'appariement, en supprimant une ou plusieurs variables d'appariement :

On apparie **avec le fichier d'appariement** sans la date de naissance  
(la date de naissance est renseignée dans le fichier d'appariement et dans l'EFL)

Le score d'appariement est très faible, on ne retient que les questionnaires ayant un très bon score.

```
if (moco_r in ('11','12','31','32') & sc>1) ! (moco_r=' ' & sc>2) !
(moco_r='40' & sc>3) ! (moco_r='23' & sc>4) !
(moco_r='21' & sc>12) ! (moco_r='22' & sc>13)
```

appariement	Frequency	Percent
01fa -direct	326860	90.04
02fa -sans jour	2648	0.73
03fa -sans mois	2025	0.56
04fa -sans année+score pas mauvais	2955	0.81
05fa -sans sexe	1165	0.32
06fa -sans logement	4801	1.32
07fa -sans adresse	5078	1.40
08fa -sans fil	103	0.03
09fa -sans district	998	0.27
10fa -sans cil	127	0.03
11fa -sans logement adresse	1075	0.30
12fa -sans fil jour	1	0.00
13fa -sans fil année+score pas mauvais	1	0.00
15fa -sans jour année+score pas mauvais	539	0.15
16fa -sans commune	234	0.06
17fa -sans departement	82	0.02
18fa -sans district adresse+score très bon	79	0.02
19fa -sans district logement+score très bon	17	0.00
20fa -sans departement adresse+score très bon	7	0.00
21fa -sans cil district	2	0.00
22fa -sans cil logement	7	0.00
30ear-direct	2758	0.76
31ear-sans logement	3582	0.99
32ear-sans adresse	511	0.14
33ear-sans logement adresse	2228	0.61
34ear-sans district	13	0.00
35ear-sans commune+score très bon	6	0.00
36ear-sans logement sexe+score très bon	44	0.01
37ear-sans adresse sexe+score très bon	4	0.00
38ear-sans logement adresse sexe+score très bon	34	0.01
39ear-sans logement année+score pas mauvais	144	0.04
40ear-sans adresse année+score très bon	11	0.00
41ear-sans logementadresse année+ score très bon	72	0.02



42ear-sans logement mois+score pas mauvais	87	0.02
43ear-sans adresse mois+score pas mauvais	16	0.00
44ear-sans logement adresse mois+score très bon	41	0.01
45ear-sans logement jour+score pas mauvais	132	0.04
46ear-sans adresse jour+score très bon	13	0.00
47ear-sans logement adresse jour+score très bon	69	0.02
51fa -sans date naissance,date naissance blanc EFL+score pas mauvais	468	0.13
52fa -sans sexe date naissance,date naissance blanc EFL+score très bon	20	0.01
53fa -sans date naissance,date naissance blanc fa+score tres bon	133	0.04
54fa -sans sexe date naissance,date naissance blanc fa+score très bon	19	0.01
61ear-sans date naissance,date naissance blanc EFL+score très bon	18	0.00
62ear-sans sexe date naissance,date naissance blanc EFL+score très bon	1	0.00
<b>71fa -sans date naissance, date naissance renseignée+ score très bon</b>	<b>189</b>	<b>0.05</b>
74 non apparie	3590	0.99

6.7 On réalise un appariement entre l'EFL et l'Ear, en supprimant une ou plusieurs variables d'appariement :

On apparie avec l'Ear sans la date de naissance,  
(la date de naissance est renseignée dans le fichier d'appariement et dans l'Ear) :

Le score d'appariement est très faible, on ne retient que les questionnaires ayant un très bon score.

```
if (moco_r in ('11','12','31','32') & sc>1) ! (moco_r=' ' & sc>2) !
(moco_r='40' & sc>3) ! (moco_r='23' & sc>4) !
(moco_r='21' & sc>12) ! (moco_r='22' & sc>13)
```

appariement	Frequency	Percent
01fa -direct	326860	90.04
02fa -sans jour	2648	0.73
03fa -sans mois	2025	0.56
04fa -sans année+score pas mauvais	2955	0.81
05fa -sans sexe	1165	0.32
06fa -sans logement	4801	1.32
07fa -sans adresse	5078	1.40
08fa -sans fil	103	0.03
09fa -sans district	998	0.27
10fa -sans cil	127	0.03
11fa -sans logement adresse	1075	0.30
12fa -sans fil jour	1	0.00
13fa -sans fil année+score pas mauvais	1	0.00
15fa -sans jour année+score pas mauvais	539	0.15
16fa -sans commune	234	0.06
17fa -sans departement	82	0.02
18fa -sans district adresse+score très bon	79	0.02
19fa -sans district logement+score très bon	17	0.00
20fa -sans departement adresse+score très bon	7	0.00
21fa -sans cil district	2	0.00
22fa -sans cil logement	7	0.00
30ear-direct	2758	0.76
31ear-sans logement	3582	0.99
32ear-sans adresse	511	0.14
33ear-sans logement adresse	2228	0.61
34ear-sans district	13	0.00
35ear-sans commune+score très bon	6	0.00
36ear-sans logement sexe+score très bon	44	0.01



37ear-sans adresse sexe+score très bon	4	0.00
38ear-sans logement adresse sexe+score très bon	34	0.01
39ear-sans logement année+score pas mauvais	144	0.04
40ear-sans adresse année+score très bon	11	0.00
41ear-sans logementadresse année+ score très bon	72	0.02
42ear-sans logement mois+score pas mauvais	87	0.02
43ear-sans adresse mois+score pas mauvais	16	0.00
44ear-sans logement adresse mois+score très bon	41	0.01
45ear-sans logement jour+score pas mauvais	132	0.04
46ear-sans adresse jour+score très bon	13	0.00
47ear-sans logement adresse jour+score très bon	69	0.02
51fa -sans date naissance,date naissance blanc EFL+score pas mauvais	468	0.13
52fa -sans sexe date naissnce,date naissance blanc EFL+score très bon	20	0.01
53fa -sans date naissance,date naissance blanc fa+score tres bon	133	0.04
54fa -sans sexe date naissnce,date naissance blanc fa+score très bon	19	0.01
61ear-sans date naissance,date naissance blanc EFL+score très bon	18	0.00
62ear-sans sexe date naissnce,date naissance blanc EFL+score très bon	1	0.00
71fa -sans date naissance, date naissance renseignée+ score très bon	189	0.05
<b>73ear-sans date naissance, date naissance renseignée + score très bon</b>	<b>16</b>	<b>0.00</b>
74 non apparie	3574	0.98

On a 3574 questionnaires non appariés.

sc	Frequency	Cumulative	
		Percent	Frequency
-4	1	0.00	1
-3	47	0.01	48
-2	254	0.07	302
-1	1170	0.33	1472
0	6617	1.84	8089
1	7804	2.17	15893
2	98014	27.27	113907
3	7701	2.14	121608
4	11752	3.27	133360
5	6226	1.73	139586
6	2994	0.83	142580
7	1979	0.55	144559
8	1730	0.48	146289
9	857	0.24	147146
10	2258	0.63	149404
11	10691	2.97	160095
12	13321	3.71	173416
13	91222	25.38	264638
14	27020	7.52	291658
15	37906	10.55	329564
16	21511	5.98	351075
17	6342	1.76	357417
18	1624	0.45	359041
19	183	0.05	359224
20	136	0.04	359360
21	56	0.02	359416
22	13	0.00	359429
23	4	0.00	359433

### 6.8 appariement jumeau

On réapplique la même méthode que précédemment.



Si des « jumeaux » s'apparient à un seul BI, on apparie avec le jumeau dont le score est le plus élevé ou aléatoirement en cas d'égalité.

The FREQ Procedure

appariement	Frequency	Percent
01fa -direct	343	81.47
02fa -sans jour	1	0.24
05fa -sans sexe	1	0.24
06fa -sans logement	3	0.71
07fa -sans adresse	4	0.95
09fa -sans district	1	0.24
11fa -sans logement adresse	7	1.66
30ear-direct	1	0.24
33ear-sans logement adresse	1	0.24
51fa -sans date naissance,date naissance blanc EFL+score pas mauvais	1	0.24
61ear-sans date naissance,date naissance blanc EFL+score très bon	1	0.24
62ear-sans sexe date naissnce,date naissance blanc EFL+score très bon	1	0.24
71fa -sans date naissance, date naissance renseignée+ score très bon	7	1.66
73ear-sans date naissance, date naissance renseignée + score très bon	2	0.48
74 non apparie	47	11.16

Cumulative appariement	Cumulative	
	Frequency	Percent
—		
01fa -direct	343	81.47
02fa -sans jour	344	81.71
05fa -sans sexe	345	81.95
06fa -sans logement	348	82.66
07fa -sans adresse	352	83.61
09fa -sans district	353	83.85
11fa -sans logement adresse	360	85.51
30ear-direct	361	85.75
33ear-sans logement adresse	362	85.99
51fa -sans date naissance,date naissance blanc EFL+score pas mauvais	363	86.22
61ear-sans date naissance,date naissance blanc EFL+score très bon	364	86.46
62ear-sans sexe date naissnce,date naissance blanc EFL+score très bon	365	86.70
71fa -sans date naissance, date naissance renseignée+ score très bon	372	88.36
73ear-sans date naissance, date naissance renseignée + score très bon	374	88.84
74 non apparie	421	100.00



## The FREQ Procedure

score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-1	3	0.80	3	0.80
0	14	3.74	17	4.55
1	16	4.28	33	8.82
2	321	85.83	354	94.65
3	2	0.53	356	95.19
4	2	0.53	358	95.72
5	1	0.27	359	95.99
6	3	0.80	362	96.79
10	2	0.53	364	97.33
12	1	0.27	365	97.59
13	3	0.80	368	98.40
14	1	0.27	369	98.66
15	2	0.53	371	99.20
16	2	0.53	373	99.73
17	1	0.27	374	100.00

Au final, on apparie 99,0 % des questionnaires EFL.

Le taux d'appariement des jumeaux est moins bon que celui de l'ensemble de l'échantillon. Une partie des « jumeaux » sont probablement des doublons.

Dans les logements où un questionnaire jumeau est apparié et l'autre non,

- le questionnaire non apparié est classé dans doublon et non dans jumeau.
- le questionnaire apparié est classé dans doublonprincipal et non dans jumeau

jumeau	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366706	99.90	366706	99.90
1	369	0.10	367075	100.00

doublon	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	366937	99.96	366937	99.96
1	138	0.04	367075	100.00

doublonprincipal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	367006	99.98	367006	99.98
1	69	0.02	367075	100.00

### 6.9 Vérifications

a) On étudie si les personnes dans un même logement EFL (même département commune district adresse logement et aucune variable à blanc) s'apparient avec des personnes dans un même logement Ear.



1,35 % sont dans un même logement EFL et s'apparient à des BI dans deux logements Ear différent.

91 % de ces personnes ont été appariés sans le logement ou sans l'adresse. Pour ces personnes il ne s'agit donc pas d'erreur d'appariement, car il y a une erreur dans l'identifiant du logement de l'EFL (deux voisins sont codés à tort par l'enquêteur dans le même logement)

7 % de ces personnes ont été appariés sans le district. Pour ces personnes il ne s'agit donc pas d'erreur d'appariement, car il y a une erreur dans l'identifiant du district de l'EFL (deux personnes de la même commune sont codées à tort par l'enquêteur dans le même district).

**Pour la quasi-totalité de ces cas, il ne s'agit pas d'erreur d'appariement, il faut donc faire confiance à l'identifiant logement de l'EAR et non de l'EFL.**

b) Seul 0,02% sont dans un même logement et s'apparient dans deux logements EFL différent

**Pour la quasi-totalité des cas, il s'agit là aussi d'erreur de remplissage d'une variable département, commune ou district du cadre de l'EFL, il faut donc faire confiance à l'identifiant logement de l'EAR et non de l'EFL.**

## 7. Appariement complémentaire

La remarque de Laurent Toulemon lors de la réunion du groupe d'exploitation du 12 avril 2012 permet d'apparier 442 questionnaires supplémentaires. Il y avait 4 IRIS complètement non appariés (Ils ne figuraient pas dans le fichier d'appariement, de plus le numéro d'IRIS était erroné). Un appariement sur l'EAR au niveau de la commune, du sexe et de la date de naissance permet de récupérer ces cas et quelques autres.

Parmi les questionnaires qui étaient non appariés, on effectue un appariement avec l'EAR, sans le numéro d'iris, sans le rang d'adresse, et sans le logement.

Dans les 4 iris qui étaient complètement non appariés, il n'y a pas de conditions sur le score. 335 questionnaires sont ainsi appariés :

score	Frequency	Cumulative		
		Percent	Frequency	Percent
0	10	2.99	10	2.99
1	86	25.67	96	28.66
2	157	46.87	253	75.52
3	6	1.79	259	77.31
4	1	0.30	260	77.61
5	1	0.30	261	77.91
11	2	0.60	263	78.51
12	4	1.19	267	79.70
13	33	9.85	300	89.55
14	30	8.96	330	98.51
15	2	0.60	332	99.10
16	2	0.60	334	99.70
17	1	0.30	335	100.00

Dans les autres iris, on ne retient que les questionnaires ayant un très bon score. 107 questionnaires sont appariés :



The FREQ Procedure

score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	60	56.07	60	56.07
5	1	0.93	61	57.01
13	17	15.89	78	72.90
14	26	24.30	104	97.20
15	2	1.87	106	99.07
18	1	0.93	107	100.00

Au final, il reste 3160 questionnaires non appariés, soit 0,9 %.



## 8. Correction de la variable sexe\_x

Dans 1236 cas, le sexe de l'EFL est différent du sexe de l'EAR :

SEXE_X	sexe_rec			
		1	2	Total
1	Frequency	2474	120868	123664
	Percent	0.67	32.93	33.69
	Row Pct	2.00	97.74	0.26
	Col Pct	34.00	99.25	0.14
2	Frequency	4802	914	243411
	Percent	1.31	0.25	66.31
	Row Pct	1.97	0.38	97.65
	Col Pct	66.00	0.75	99.86
Total	7276	121782	238017	367075
	1.98	33.18	64.84	100.00

Parmi eux, la proportion qui se déclare en couple homosexuel est forte (35 %) :

The FREQ Procedure

SEXE_C	Frequency	Percent	Cumulative Frequency	Cumulative Percent
00	267	21.60	267	21.60
01	436	35.28	703	56.88
02	1	0.08	704	56.96
10	514	41.59	1218	98.54
11	4	0.32	1222	98.87
12	4	0.32	1226	99.19
13	5	0.40	1231	99.60
20	1	0.08	1232	99.68
21	1	0.08	1233	99.76
31	2	0.16	1235	99.92
33	1	0.08	1236	100.00

Lorsqu'il y a une incohérence de sexe, et que la personne se déclare en couple homosexuel, et qu'il n'y a pas une autre personne homosexuelle dans le logement, on garde la variable sexe de l'EAR. Dans tous les autres cas (couple hétérosexuel et célibataire), on garde la variable sexe de l'EFL.

```
if erreur='oui' & sexe_c in ('01','02','11','21','31','33') & nbhomo
ne 2 then sexe_x_r=sexe_rec;
```



Table of sexe\_x\_r by SEXE\_X

sexe_x_r		SEXE_X		
Frequency				
Percent				
Row Pct				
Col Pct	1	2		Total
1	123597	375		123972
	33.67	0.10		33.77
	99.70	0.30		
	99.95	0.15		
2	67	243036		243103
	0.02	66.21		66.23
	0.03	99.97		
	0.05	99.85		
Total	123664	243411		367075
	33.69	66.31		100.00

## 9. Repérage des hors champ

Il s'agit de créer une variable :

- HORSCHAMP=1 pour les questionnaires hors champ,

On supprime les personnes qui ne sont pas en ménage ordinaire.

On supprime également les hommes qui ont répondu à un questionnaire femme (et inversement)

On supprime les hommes appartenant à un iris femme, qui sont dans un logement où des femmes ont répondu (et inversement)

### 9.1 on supprime les questionnaires qui ne sont pas en ménage ordinaire

80 personnes ne vivent pas en ménage ordinaire.

CPOP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
01	359706	99.98	359706	99.98
11	2	0.00	359708	99.98
14	2	0.00	359710	99.98
21	76	0.02	359786	100.00

Frequency Missing = 7289

horschamp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-----------	-----------	---------	----------------------	--------------------



1	80	100.00	80	100.00
---	----	--------	----	--------

9.2 On supprime les hommes qui ont répondu à un questionnaire femme (et inversement)

Pour les hommes qui ont répondu à un questionnaire femme, certaines questions ne sont pas adaptées (tableau 13 : Êtes vous sa mère, Son père est-il votre conjoint actuel; tableau 14 : êtes vous sa mère). De plus, on ne sait pas si la personne a répondu pour elle-même ou son conjoint.

horschamp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	522	100.00	522	100.00

Frequency Missing = 366553

9.3 On supprime les hommes appartenant à un iris femme, qui sont dans un logement où des femmes ont répondu (et inversement)

13 personnes s'ajoutent aux hors champ

horschamp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	535	100.00	535	100.00

Frequency Missing = 366540



**NOTE**

Dossier suivi par :  
Aude Lapinte  
Tél. : 01 41 17 55 03  
Mél : [-DG75-F170](mailto:-DG75-F170)

Paris, le  
N° / DG75-F170 /

**Objet : Les enfants et les structures familiales: comparaison entre l'EFL et le Recensement**

L'EFL, si elle est associée à l'enquête annuelle du recensement (EAR) 2011, a ses questions propres pour identifier les enfants de la personne interrogée. Les liens de filiation sont a priori plus solides dans l'EFL que dans l'EAR, où, notamment pour les ménages complexes, les enfants ne sont pas toujours attribués aux « vrais » parents et plus précis puisqu'on est en mesure de distinguer les beaux-enfants des enfants. Plus largement, l'EFL a bénéficié de traitements spécifiques visant à assurer un meilleur degré de qualité des informations.

L'objectif de cette note est de comparer les estimations EFL et EAR des structures familiales et des nombres d'enfants y vivant. Cette comparaison est effectuée d'abord au niveau macro sans distinguer entre familles recomposées et familles « traditionnelles » (l'EAR ne permettant pas cette distinction). Pour mieux comprendre les différences, une comparaison des deux sources au niveau micro est ensuite réalisée.

Les comparaisons entre les sources sont faites à concepts comparables.

Les travaux de comparaison portent sur les enfants ayant moins de 18 ans (nés en 1993 ou après). Pour ces derniers, qui ne sont jamais directement enquêtés, les seules informations disponibles dans l'EFL proviennent du questionnaire d'un de leur (beau-)parent. Exclure les enfants de 18 ans ou plus permet de s'appuyer sur des informations homogènes quel que soit l'âge de l'enfant<sup>7</sup>.

Néanmoins, pour la construction de la typologie des familles, tous les enfants, sans limite d'âge, sont pris en compte.

---

<sup>7</sup> Pour les enfants majeurs, certains ont pu également répondre à l'enquête, ce qui poserait la question de la prise en compte de leurs réponses.



## I. Principes généraux du recensement, traitements réalisés dans le cadre de l'EFL et choix complémentaires retenus dans cette note

Les travaux menés dans cette note s'intéressent aux « enfants d'une famille » (au sens de l'analyse ménage famille). Un enfant, au sens du recensement, correspond à la définition suivante<sup>8</sup> : « personne vivant au sein du même ménage (au sens du recensement) que son (ses) parent(s) avec le(s)quel(s) elle forme une famille, quel que soit son âge, si elle est célibataire et n'a pas de conjoint ou d'enfant vivant dans le ménage ».

Il s'agit donc d'identifier clairement, au sein de chaque famille, qui peut être effectivement considéré comme un enfant et parmi eux, lesquels y vivent principalement.

### I.1 Rappel sur les traitements réalisés dans le cadre de l'EFL

Les données de l'EFL ont pour un certain nombre d'entre elles été apurées en s'appuyant sur les informations issues de l'EAR. Pour l'apurement le principe de cohérence interne à l'EFL a été privilégié et lorsque les informations EFL n'ont pas suffi à lever les incertitudes, la cohérence avec le RP a été retenue comme principe de décision (afin de ne pas générer de divergences entre EFL et RP en l'absence de bonnes raisons de le faire). Notamment, lorsque les compteurs d'enfants présents dans le logement<sup>9</sup> étaient supérieurs au nombre de lignes d'enfants déclarés dans le tableau T13, les individus identifiés comme enfants dans l'EAR ont été intégrés pour compléter les lacunes du tableau. Cette « intégration » a été faite aussi dans le cas de non-réponse aux données sur les enfants (compteurs + lignes du tableau) dans le cas où les enfants identifiés dans l'EAR vivent dans des ménages simples (s'ils sont dans des ménages complexes, on ne peut les relier de façon certaine à un adulte en particulier) (pour plus de détails voir note sur l'apurement).

### I.2 Analyse ménages - familles (AMF)

L'analyse ménages - familles (AMF) désigne, dans l'EAR, l'ensemble des opérations permettant d'identifier, parmi les personnes qui partagent habituellement un même logement (ménage), celles qui constituent un noyau familial. Ces traitements sont réalisés automatiquement grâce à des algorithmes mais, dans les cas complexes, ils requièrent une intervention humaine.

La détermination des liens familiaux existant entre les personnes d'un ménage utilise, dans un premier temps, des variables simples : l'âge, le sexe, la situation matrimoniale légale des personnes de ce ménage. L'analyse de ces informations permet le plus souvent de déterminer la structure familiale du ménage. Dans un second temps, lorsque le ménage est complexe ou lorsque ces premières variables s'avèrent incohérentes entre elles, on utilise les liens de parenté décrits dans la feuille de logement du recensement. La première phase de traitement est réalisée de façon automatique grâce à des algorithmes. La seconde nécessite une intervention humaine, le « codage manuel ».

L'AMF dans le recensement repose sur un certain nombre de choix, notamment :

- Pour des raisons historiques (continuité des séries), deux personnes de même sexe ne peuvent constituer un couple. Ce choix évite de considérer à tort des colocataires de même sexe comme un couple.
- Si dans un logement, vivent deux adultes avec enfants qui ne se déclarent pas tous les deux en couple mais sont tous les deux mariés, l'AMF les remet en couple et les enfants du logement vivent dans une famille « couple avec enfants ». En revanche, si ces deux adultes (dont au moins un ne se déclare pas en couple) ne

<sup>8</sup> Cette définition est utilisée pour les analyses « ménages-famille » des recensements (analyses usuelles pour les recensements, voir notamment *Recommandations de la conférence des statisticiens européens pour les recensements de la population et des logements de 2010*, Commission économique des Nations Unies pour l'Europe, préparé en collaboration avec Eurostat) visant à étudier la composition familiale au sein d'un ménage, et plus particulièrement les structures familiales susceptibles d'accueillir des enfants (couple sans ou avec enfant, famille monoparentale). Dans cette approche une personne appartient au plus à une seule famille.

<sup>9</sup> e11nb\_enf et e12nb\_enf\_c



sont pas mariés, le recours au codage manuel, même si les éléments sur lesquels il s'appuie ne sont pas équivoques, ne conduit pas toujours à les considérer comme un couple. On est alors en présence d'un parent de famille monoparentale et d'un adulte hors famille. Les enfants vivent en famille monoparentale bien qu'étant, dans un certain nombre de cas, sous le même toit que leurs deux parents.

- Un individu est considéré comme un enfant s'il vit avec au moins l'un de ses parents et s'il est célibataire, pas en couple avec quelqu'un du ménage et s'il n'est lui-même pas parent.

### I.3 Choix effectués dans l'EFL

Conformément aux choix retenus dans le recensement lui-même qui visent à éviter les doubles comptes, chaque personne est recensée (et prise en compte) dans son logement principal. Ainsi, il faut s'assurer qu'un enfant est compté une unique fois, même s'il partage son temps entre plusieurs résidences.

C'est pourquoi, seuls les enfants apparaissant dans le tableau T13 de l'EFL c'est-à-dire vivant (même une partie du temps seulement) avec l'un de leur (beau-)parent peuvent être pris en compte. Les informations sur le temps de présence dans le logement permettent en plus de filtrer les enfants à retenir pour ne pas les compter plusieurs fois.

Si la définition des enfants retenue dans l'EFL et la construction de la typologie des familles s'appuie fortement sur la logique RP, on note néanmoins une petite différence entre l'EAR et l'EFL. Elle concerne la manière de définir qui vit dans le logement. Le RP s'appuie sur des règles précises : sont comptées dans le logement, les personnes qui y passent la plus grande partie de leur temps, avec des traitements spécifiques pour les conjoints absents pour raison professionnelle, les étudiants majeurs, les élèves,... Ainsi, lorsque l'un des parents ne vit pas la semaine dans le logement de son conjoint et de ses enfants pour raisons professionnelles, il est, dans le RP, rattaché au logement de sa famille qui est alors une famille de type « couple avec enfants » et non monoparentale. Dans l'EFL, c'est l'appréciation de la personne qui répond qui compte, pour le conjoint - comme pour les enfants<sup>10</sup>. Dans le cas du conjoint qui ne vit dans le logement que le week-end après sa semaine de travail, le répondant EFL pourra estimer, au vu de sa situation au quotidien, que son conjoint vit dans un autre logement.

### I.4 Qui est considéré comme enfant dans l'EFL ?

Les enfants nés début 2011 après la date du début de collecte ne sont pas présents dans l'EAR, contrairement à l'EFL. En tout, 59 000 enfants nés en 2011 apparaissent dans l'EFL, contre 39 000 dans l'EAR. Les 20 000 de plus dans l'EFL correspondent à ceux nés entre la date de référence du recensement et la fin de l'enquête, aux erreurs de déclaration dans l'une des deux sources près. Par la suite, les enfants nés en 2011 (apparaissant dans l'EFL ou dans l'EAR) ne seront pas pris en compte (ou traités à part) pour les comparaisons pour raisonner à champ constant. On ne peut, en effet, faire mieux pour être à champ constant, ne disposant pas dans l'EFL du mois de naissance des enfants.

Les enfants retenus pour cette comparaison (une fois l'apurement effectué) et la mise en place d'une typologie de famille (et en général pour les études sur les enfants au sein des familles) sont ceux figurant dans le tableau T13 qui soit vivent tout le temps dans le logement (e13tps1gt=1), soit ont rempli un bulletin individuel dans l'EAR<sup>11</sup>. Les enfants supplémentaires que la deuxième partie de la définition conduit à prendre en compte ne passent pas tout leur temps dans le logement<sup>12</sup>:

- soit parce qu'ils passent la moitié du temps ou plus dans le logement<sup>13</sup> : il est logique de les conserver s'ils passent plus de la moitié du temps dans le logement. S'ils en passent exactement la moitié alors on suppose que les règles du recensement

<sup>10</sup> pour les enfants, ce décalage n'est pas un problème puisque la prise en compte dans le champ de ceux qui ont un bulletin individuel dans le recensement réintègre ceux qui n'ont pas e13tps1gt=1.

<sup>11</sup> e13tps1gt=1 ou cabbi\_e^=".

<sup>12</sup> Ils ont alors e13tps1gt différent de 1.

<sup>13</sup> e13tps1gt=2.



(présence dans le logement le jour de référence du RP) ont bien été appliquées, ce qui amène à retenir les enfants qui ont un BI<sup>14</sup>.

- soit parce que l'information sur le temps de présence n'est pas remplie,
- soit encore, dans des cas assez rares (0,6% de l'ensemble des enfants retenus), parce qu'ils ne résident qu'une petite partie du temps dans le logement<sup>15</sup> : ils peuvent, par exemple, passer l'essentiel de leur temps en internat, auquel cas, leur résidence principale dans le RP est celle de leur parent chez qui ils reviennent régulièrement, ou partager leur temps entre plus de deux logements sans qu'il y en ait un où ils vivent au moins la moitié du temps

Dans toutes ces situations, le fait qu'ils aient un bulletin individuel rempli dans l'EAR conduit à considérer qu'ils sont rattachés à ce logement.

Finalement, on retient 13,7 millions d' « enfants » mineurs parmi 18,1 millions sans limite d'âge ; 9 sur 10 (91%) résident tout le temps dans le logement<sup>16</sup>, au sens EFL.

A ce premier filtre doit être rajouté aussi celui qui porte sur les caractéristiques familiales de l'individu : est-il célibataire ? N'est-il pas en couple avec une personne du logement ? N'a-t-il pas d'enfant dans le logement ?

Dans l'EAR, est enfant l'individu qui répond à MOCO=11 (enfant d'un couple) ou 12 (enfant d'une famille monoparentale).

**Tableau 1 : Répartition des enfants mineurs dans l'EAR par mode de cohabitation:**

Mode de cohabitation (MOCO)	Ensemble des enfants mineurs		Hors ceux nés début 2011	
	Effectifs	Part (en %)	Effectifs	Part (en %)
11 : enfant d'un couple	11 080 000	81,4	11 050 000	81,4
12 : enfant d'une famille monoparentale	2 540 000	18,6	2 530 000	18,6
Ensemble	13 620 000	100,0	13 580 000	100,0

Champ : ensemble des individus mineurs enfants dans l'EAR, c'est-à-dire des mineurs tels que MOCO=11 ou 12. Individus du sous échantillon de l'EAR des logements avec au moins un répondant EFL.

L'EFL ne contient pas, en revanche, d'information systématique permettant de savoir si un enfant est célibataire, sans conjoint et sans enfant, conditions exigées pour être considérés comme un enfant, au sens du recensement. Le recours aux informations de l'EAR, lorsqu'elles existent, permettent d'affiner le champ des enfants à retenir.

Il s'agit de définir quels individus sont considérés comme étant enfant dans l'EFL. On peut s'entendre sur plusieurs définitions, plus ou moins proches du concept RP :

- une définition « large » (définition 1), où tous les enfants sont pris en compte dès lors qu'ils sont inscrits dans le tableau T13 des enfants du répondant ou du conjoint de celui-ci, s'ils ont un BI ou s'ils vivent tout le temps<sup>17</sup> dans le logement.
- Une définition plus en adéquation avec la définition RP (définition 2), où les enfants précédemment retenus doivent, s'ils apparaissent dans l'EAR être identifiés comme étant enfant dans le logement (d'un couple ou d'une personne seule)<sup>18</sup>. Rajouter cette contrainte permet notamment de ne pas considérer dans le champ des enfants des individus vivant avec leur(s) parent(s) mais étant eux-mêmes parents ou en couple avec une personne du logement :

<sup>14</sup> En effet, si ceux qui passent plus de la moitié du temps dans le logement doivent bien être conservés, la situation est plus ambiguë pour ceux qui passent exactement la moitié du temps dans le logement. S'ils passent l'autre moitié de leur temps dans un autre logement, les conserver conduirait à les compter deux fois. Or, l'essentiel des enfants mineurs qui ne vivent pas tout le temps dans le logement vit chez leur autre parent (78% de ceux pour lesquels on a une réponse à la question « vit-il aussi régulièrement chez son autre parent ? »). Prendre le risque de les compter deux fois, reviendrait à surestimer les enfants en familles monoparentales et recomposées au profit des enfants en famille traditionnelle.

<sup>15</sup> e13tpsigt=3.

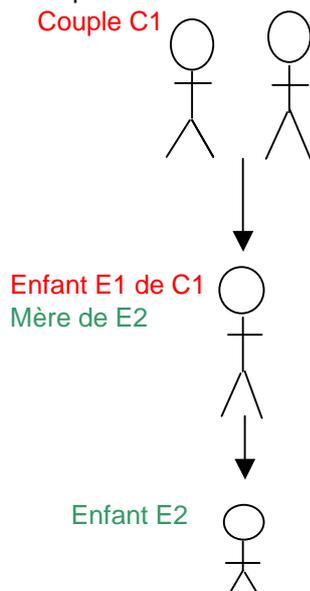
<sup>16</sup> e13tpsigt=1.

<sup>17</sup> e13tpsigt=1.

<sup>18</sup> MOCO=11 ou 12.



Exemple :



Dans une telle configuration, le ménage est composé de deux familles (au sens de l'AMF) : l'une formée du couple C1, l'autre formée de E1 et de son enfant E2. L'individu E1 ne peut être rattaché qu'à une seule famille : celle qu'il forme avec son enfant. Ses parents sont alors considérés comme un « couple sans enfant ».

Cette deuxième définition permet donc de prendre en compte la spécificité des familles multi générationnelles et de ne pas compter comme « enfants » des descendants ayant eux-mêmes construit leur propre famille, qui sont alors avant tout des parents.

Néanmoins, elle est sans doute un peu restrictive puisqu'elle cumule des contraintes des deux enquêtes. Un individu n'est compté dans le champ des enfants que s'il est à la fois déclaré comme l'enfant ou le bel-enfant d'un répondant EFL et si, par ailleurs côté EAR, il est codé comme enfant d'un couple ou d'une famille monoparentale<sup>19</sup>. Il suffit qu'une seule de ces enquêtes l'oublie ou le code mal pour qu'il soit exclu du champ des enfants.

En ce qui concerne les enfants du tableau T13 de l'EFL qui n'ont pas de BI dans l'EAR mais qui vivent tout le temps dans le logement, on les conserve, en supposant qu'il s'agit d'un oubli côté EAR. On n'applique pas de « contrôle » sur le fait qu'ils ne soient ni parent ni en couple avec quelqu'un du logement, car on ne dispose pas de l'information.

- Compte tenu de la proportion non négligeable d'enfants mineurs (environ 1%, soit 125 000) du tableau T13 (i.e. passant au moins une petite partie du temps dans le logement) codés comme vivant hors famille dans un ménage de plusieurs personnes côté EAR<sup>20</sup>, et au vu des éléments de la liste A<sup>21</sup> les concernant, on intègre dans la sélection précédente les individus s'ils sont célibataires<sup>22</sup>. On estime, en effet, que le codage RP a des chances relativement élevées d'être faux. On note notamment que parmi ces enfants, 57% vivraient en famille recomposée, contre 10% en moyenne (de ceux qui ont un BI). Ce constat laisse penser que la population des enfants EFL considérés comme personnes hors famille dans l'EAR provient de familles plus compliquées à appréhender. Cette approche correspond à la définition 3.

<sup>19</sup> MOCO=11 ou 12.

<sup>20</sup> MOCO=31

<sup>21</sup> tableau des habitants du logement, dans la feuille logement de l'EAR.

<sup>22</sup> célibataire dans l'EAR si apparié. Sinon, la plupart de ceux appariés étant célibataire, on conserve aussi les individus pour lesquels on n'a pas l'information du statut matrimonial.



Si l'on étudie plus précisément les individus mineurs de l'EAR considérés comme hors famille et célibataires, près des deux tiers sont mentionnés dans l'EFL en tant qu'enfant de quelqu'un du logement. Une partie de ceux n'apparaissant pas dans l'EFL vivent aussi peut-être avec un seul de leur parent qui ne correspond pas au sexe enquêté dans leur zone de collecte. Enfin, une partie - minoritaire - d'entre eux, correspondent vraisemblablement à des individus du logement sans lien de parenté avec les adultes qui y vivent (enfants en famille d'accueil, enfants vivant avec un autre membre de la famille, ...).

Plus la définition d'enfant est stricte, moins les effectifs d'enfants en famille recomposée<sup>23</sup> sont élevés : en effet, puisqu'on a moins d'enfant au sein des familles, les « risques » que l'un d'entre eux ne soient pas enfant des deux adultes du couple baissent. De plus, les individus qui sont exclus des enfants par la définition stricte appartiennent vraisemblablement plus souvent à des structures familiales plus compliquées au sein desquelles les familles recomposées peuvent être surreprésentées.

La définition 2 est la plus restrictive, on note ainsi une part légèrement plus faible d'enfants en famille recomposée (de l'ordre de 0,4 points) et un effectif global d'enfants de l'ordre d'1% plus bas qu'avec les définitions 1 et 3.

Au final, les résultats obtenus avec les trois définitions ne diffèrent pas beaucoup les uns des autres. Notamment parce que peu d'individus de moins de 18 ans vivant avec au moins un de leurs parents sont en couple avec quelqu'un de leur logement ou sont eux-mêmes parents.

La sélection 3 est celle finalement conservée pour les études et intégrée dans le fichier de production et de recherche de l'enquête Famille et logements. C'est aussi cette définition qui est désormais considérée dans tout ce qui suit. On dénombre ainsi 13 690 000 enfants de moins de 18 ans.

On identifie alors les enfants par la variable ENF=1, qu'on peut ventiler par type de famille selon la variable FAMILLE.

Cette définition peut aussi être utilisée pour dénombrer le nombre de familles par la table *enfant*, en ne conservant qu'une observation enfant par bulletin EFL. En effet, dans les ménages « simples » avec une seule famille, la famille est portée par l'unique bulletin correspondant au parent répondant ; dans un ménage complexe avec plusieurs familles sans lien de parenté, chaque bulletin correspond à une famille ; enfin, dans les familles multi-générationnelles, se pose la question des générations « pivots », à la fois parent (et donc ayant rempli un bulletin si le parent est majeur) et enfant de l'un de ses parents qui a rempli un bulletin. Mais, le risque d'être compté deux fois dans deux types de familles différentes n'est, en fait, pas présent dans la mesure où ces « enfants-parents » ne sont pas considérés comme enfants puisque la variable mode de cohabitation, qui intervient dans la définition, correspond à « adulte avec enfant »<sup>24</sup>. Les familles comprenant un couple formé de deux conjoints de même sexe bénéficient d'un traitement spécifique (détail donné infra)<sup>25</sup>.

## I.5 Construction d'une typologie de famille

Les individus identifiés comme enfant dans l'EFL sont alors classés selon une typologie des familles déclinée en trois modalités :

- famille monoparentale, lorsqu'il n'y a qu'un seul parent
- couple avec enfants, qui se distingue en :
  - famille recomposée lorsqu'au moins un des enfants de la famille n'est pas l'enfant des deux membres du couple et
  - famille « traditionnelle »<sup>26</sup>, dans le cas contraire.

<sup>23</sup> La typologie du type de famille est développée plus bas, dans la partie II.2/.

<sup>24</sup> et qu'on ne conserve que les individus du tableau 13 qui sont enfants (MOCO=11 ou 12), voire personne hors famille (MOCO=31) célibataire dans l'EAR (ou qui n'ont pas d'informations EAR s'ils sont présents tout le temps).

<sup>25</sup> Dans le cas de couple de même sexe, il peut cependant y avoir un problème lors du passage de la table enfant à la table famille si au moins un enfant est rattaché en « parent principal » à chacun des deux adultes du couple, tous deux répondants à l'EFL. En pratique, ce problème concerne 11 ménages.

<sup>26</sup> On reprend ici le terme « traditionnelle » pour qualifier les familles qui ne sont ni recomposée ni monoparentale, utilisé dans plusieurs publications de l'Insee.



Comme dans l'EAR, les familles sont constituées au sein des ménages : chaque individu appartenant à un seul ménage (principalement) et donc à une seule famille.

La construction de cette typologie nécessite de répondre à deux questions :

1- l'individu répondant, qui est le parent ou le beau-parent des enfants mentionnés, a-t-il un conjoint dans le logement ?

- Si oui<sup>27</sup>, alors on est dans le cadre d'une famille de type couple avec enfants.
- Si non<sup>28</sup> (y compris, donc, s'il y a un conjoint non cohabitant), quelle qu'en soit la raison, alors on est en présence d'une famille monoparentale.

Dans le cas d'un adulte en couple avec un conjoint qui ne vit pas dans le logement, la famille est considérée comme monoparentale. On peut considérer, en effet, que les problématiques du quotidien de l'adulte du logement sont proches de celles d'un parent de famille monoparentale. L'enfant, quant à lui, vit, comme un enfant de famille monoparentale, l'essentiel de son temps avec un seul de ses parents.

Néanmoins, on sait préciser si le conjoint hors logement du parent avec lequel vivent ces enfants est le deuxième parent ou une autre personne.

2- Dans le cas de couple avec enfants, quels sont les liens de parenté qui existent entre les deux adultes et les enfants ?

Si l'un des enfants, au moins, n'est pas l'enfant des deux membres du couple<sup>29</sup>, alors on est en présence d'une famille recomposée.

Un enfant vivant dans le même logement que ses deux parents qui ne se considèrent plus en couple vit, selon cette logique, dans une famille monoparentale (avec le parent ayant répondu à l'EFL). La façon dont l'EFL est construite ne permet pas de prendre le point de vue de l'enfant, ce qui conduirait alors à le considérer en famille traditionnelle (ou recomposée si présence de demi-frères ou demi-sœurs) car résidant avec son père et sa mère dans un même logement. En effet, on sait la nature du lien qui existe entre la femme répondante EFL (respectivement l'homme) et l'enfant (enfant ou bel-enfant) et si le père (respectivement la mère) est le conjoint actuel de la personne répondante mais lorsque l'homme (respectivement la femme) n'est pas (plus) le conjoint actuel de la mère, on n'a aucune information sur la nature de son lien avec l'enfant. La mère n'étant pas en couple, la réponse à la question sur le conjoint actuel est vide.

Concernant les enfants vivant avec deux personnes de même sexe, on adopte la même logique : si les deux adultes se disent parent de l'enfant alors on est dans le cas soit d'une famille recomposée, si par ailleurs un des autres enfants du logement n'est l'enfant que d'un des membres du couple, soit d'une famille « traditionnelle » sinon. Néanmoins, on les classe ici, pour la comparaison avec l'EAR, en famille monoparentale, afin d'être sur des concepts équivalents.

## II. Comparaisons macroéconomiques des nombres de mineurs vivant en ménages ordinaires, d'enfants et de familles entre l'EFL et l'EAR.

La comparaison au niveau macro conduit à mettre en regard des effectifs d'enfants par types de famille définis à partir de l'EFL d'une part et du recensement d'autre part. On compare des effectifs globaux obtenus à partir de données parfois imputées.

### Pondérations et approches macro et micro :

Le décompte des enfants dans l'EFL nécessite d'utiliser un système de pondération.

Deux poids existent, conduisant à deux approches différentes des enfants :

- Le poids adulte répondant permet d'appréhender le nombre d'enfants des pères d'une part et celui des mères d'autre part,
- Le poids enfant décompte au sein des ménages les enfants au sens du recensement, c'est-à-dire dans leur résidence principale. Cette approche consiste à donner

<sup>27</sup> Q3couple='1'

<sup>28</sup> Q3couple^='1'

<sup>29</sup> E13\_14PART='2' ou E13\_14CPART='2'



un poids aux enfants appariés avec le recensement et à ceux (minoritaires) sans bulletin individuel dans le recensement considéré dans l'EFL comme « vivant tout le temps dans le logement ». Ainsi, les enfants ne sont comptés qu'une seule fois, dans le logement où ils passent le plus de temps.

C'est ce deuxième poids qu'on utilise ici.

Il est identique au poids ménage qui est obtenu par un partage des poids individus.

Pour certains types de famille, le poids est nul. C'est par exemple le cas des familles monoparentales vivant avec un isolé interrogé dans l'EFL tandis que le parent ne l'est pas (car du sexe non enquêté dans sa zone de collecte) ou encore lorsque cohabitent dans le même logement plusieurs familles qui ne sont pas toutes interrogées dans l'EFL.

L'approche macro « récupère » ces enfants via le jeu des pondérations. En effet, la pondération est construite à partir d'un calage sur marge réalisé de façon simultanée sur les caractéristiques des ménages, familles et enfants. Les enfants/familles/ménages sans poids sont compensés par le poids affecté aux enfants/familles/ménages de mêmes caractéristiques.

Le calage se fait sur les caractéristiques EAR des individus, ménages, familles, enfants, des répondants EFL par rapport à l'ensemble des répondants à l'EAR. Par exemple, pour le nombre d'enfants de l'individu, on regarde son nombre d'enfants dans l'EAR et non pas son nombre d'enfants dans l'EFL. Ce calage ne conduit donc pas à ce que le nombre d'enfants et les structures familiales de l'EFL correspondent « par construction » à ceux de l'EAR.

Dans l'approche macro la pondération permet donc de compenser les enfants qui n'apparaissent pas dans l'EFL, parce qu'ils vivent dans un ménage où leur (unique) parent n'a pas répondu à l'EFL, ce qui n'est pas le cas avec une comparaison micro (différence de réponse ou non pour un individu donné entre ses variables de l'EAR et celles de l'EFL). Néanmoins, l'approche macro ne permet pas d'explicitier les incohérences au sein d'un ménage entre EFL et EAR.

## II.1 comparaison de la répartition des enfants par type de famille entre les deux sources

Malgré le petit décalage conceptuel évoqué plus haut (les célibataires géographiques sont rattachés au logement où ils vivent les week-end ou les vacances dans l'EAR et pas toujours dans l'EFL où c'est l'appréciation du répondant qui prime) qui conduit à surestimer dans l'EFL par rapport à l'EAR les couples ne vivant pas ensemble et donc les familles monoparentales, les enfants en famille monoparentale restent plus nombreux dans l'EAR.

On obtient alors les répartitions suivantes :

**Tableau 2 : Répartition des enfants mineurs (effectifs pondérés) par type de famille**

		Ensemble	Hors nés en 2011
EFL	Couple avec enfants	11 240 000	11 180 000
	Traditionnelle	9 770 000	9 720 000
	Recomposée	1 470 000	1 460 000
	Monoparentale	2 450 000	2 450 000
	<b>Ensemble</b>	<b>13 690 000</b>	<b>13 630 000</b>
EAR 2011	Couple avec enfants	11 080 000	11 040 000
échantillon de ménages répondants	Monoparentale	2 540 000	2 540 000
	<b>Ensemble</b>	<b>13 620 000</b>	<b>13 580 000</b>

### dans l'EFL et dans l'EAR

Champ : ensemble des enfants mineurs de l'EFL d'une part, de l'EAR d'autre part.

Chiffres EAR : individus du sous échantillon de l'EAR des logements avec au moins un répondant EFL.

Le nombre d'enfants apparaissant dans l'EFL est légèrement supérieur à celui des enfants dans l'EAR.



Par ailleurs, la part d'enfants mineurs nés avant 2011 en famille monoparentale s'élève, d'après l'EFL, à 18,0 % contre 18,7 % dans l'EAR.

## II.2 Décompte des familles

Le passage à la notion de famille à partir de la table enfant de l'EFL est ensuite assez immédiat :

on conserve une observation par répondant qui déclare au moins un enfant de moins de 18 ans vivant principalement dans le logement. Pour les familles composées d'un couple de même sexe où les deux adultes sont chacun parent « principal<sup>30</sup> » d'au moins un enfant, il est nécessaire de faire un traitement spécifique pour ne conserver qu'une observation pour la famille.

Côté EAR, on conserve une observation par famille avec enfant de moins de 18 ans<sup>31</sup>.

**Tableau 3 : Répartition des familles avec au moins un enfant mineur dans l'EFL et dans l'EAR**

	Famille avec au moins un enfant mineur		Famille avec au moins un enfant mineur, hors ceux nés en 2011	
	EFL	EAR (sous échantillon)	EFL	EAR (sous échantillon)
couple + enfants	6 190 000 79,8	6 100 000 79,1	6 170 000 79,7	6 090 000 79,1
<i>traditionnelle</i>	5 470 000 70,5		5 450 000 70,4	
<i>recomposée</i>	720 000 9,3		720 000 9,3	
monoparentale	1 570 000 20,2	1 610 000 20,9	1 570 000 20,3	1 610 000 20,9
Ensemble	7 760 000 100,0	7 720 000 100,0	7 740 000 100,0	7 700 000 100,0

Champ : Familles avec au moins un enfant mineur (né après 1992).

Chiffres EAR : individus du sous échantillon de l'EAR des logements avec au moins un répondant EFL.

On dénombre 7 740 000 familles avec au moins un enfant mineur né avant 2011 à partir de l'EFL contre 7 700 000 avec l'EAR. Comme au niveau enfants, la part des familles monoparentales est plus élevée dans l'EAR que dans l'EFL (+0,6 points).

## III. Comparaison micro-économique du nombre d'enfants selon les situations familiales dans l'EFL et l'EAR

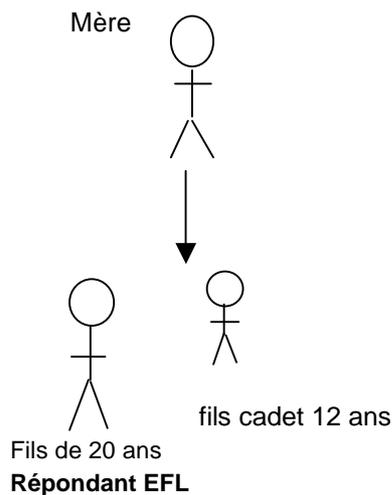
Au niveau micro, on compare les réponses apportées à l'EAR et à l'EFL pour chaque individu. Les comparaisons sont rendues plus complexes que les comparaisons macro du fait de l'échantillonnage de l'EFL. En effet, selon les zones, soit les hommes soit les femmes adultes répondent à l'EFL. Dans un ménage comptant au moins un adulte de chaque sexe qui ne sont pas en couple avec une personne du logement, par exemple une mère et son fils majeur, si les seuls enfants du ménage sont ceux de l'adulte n'ayant pas à répondre à l'EFL, il n'y a aucun enfant déclaré dans l'EFL.

Exemple :

<sup>30</sup> pour les couples de même sexe où les deux membres du couple répondent à l'EFL, on affecte aux enfants le cabefl d'un parent dit « principal » ainsi que le cabefl\_c de l'autre (beau-)parent. Lorsqu'un seul des adultes se dit être le père ou la mère, c'est celui-ci qui est associé principalement à l'enfant. Lorsque les deux adultes se déclarent mère ou père (cas plus rares), le parent principal est défini de façon arbitraire (on affecte l'enfant à l'un des parents).

<sup>31</sup> une famille EAR est alors identifiée par un unique identifiant cabfl !!numf.





Dans une zone de collecte où ce sont les hommes qui sont interrogés, le fils majeur sera l'unique répondant EFL de cette famille. Son frère n'apparaît donc pas dans la table enfant, alors que côté EAR, nous avons deux enfants vivant en famille monoparentale, dont un mineur.

Au niveau macro, cet effet est compensé par la pondération. Mais ce n'est pas le cas au niveau micro. Cet effet n'affecte que les familles monoparentales et les enfants vivant avec un couple de même sexe, et ne concerne pas ceux des couples de sexe différent. Ces familles se voient attribuer un poids nul (voir note sur la pondération n°137 calcul des pondérations de l'enquête EFL2011 et encadré plus haut).

Pour éviter d'alourdir la comparaison avec des individus qu'on ne retrouve pas au niveau macro, on exclut de la comparaison micro les données qui ont un poids nul.

### III.1 Écart sur le champ des enfants

Le nombre d'enfants mineurs au sens EAR est proche du nombre d'enfants mineurs répertoriés dans l'EFL.

On compte environ 900 observations d'enfants mineurs de plus dans l'EAR que dans l'EFL, si on se restreint à ceux ayant un poids non nuls.

On note, tout d'abord, un très faible écart lié à la sélection des enfants sur l'âge. Pour chacune des sources, environ 300 enfants mineurs ont 18 ans ou plus dans l'autre source.

Les variables années de naissance **anai** et **e13\_14anai** sur les deux sources ne sont pas toujours cohérentes et selon qu'on prend l'une ou l'autre en compte, on élimine ou on ne conserve pas exactement les mêmes individus.

Les âges pris en compte dans l'EFL et dans l'EAR sont les âges redressés, voire imputés. Si l'âge brut de l'enfant n'est pas renseigné dans l'EFL, mais qu'un appariement a néanmoins été rendu possible avec un enfant apparaissant côté EAR, alors l'âge e13\_14anai dans l'EFL est rempli par celui renseigné dans l'EAR dans les fichiers de diffusion (cf. document sur l'apurement).

#### Tableau 4 : Répartition des enfants EAR et EFL selon qu'ils sont ou non mineurs dans chacune des sources



		Age EAR					Total
		Mineurs		Total des mineurs	Majeurs	Non présents	
		enfant EAR (moco =11 ou 12)	non enfant EAR				
Age EFL	Mineurs	186 624	1 799	188 423	253	806	<b>189 482</b>
	Majeurs	334	27	361			
	Non présents	1 611		1 622			
	Total	<b>188 569</b>	<b>1 826</b>	<b>190 395</b>			

Champ : enfants au sens EFL ou EAR nés entre 1993 et 2010 compris, vivant principalement dans le logement, de poids non nul (Poids\_men\_fam\_enf=0 ou .).

Note : la définition d'enfants retenue est la définition 3.

L'âge EFL retenu est l'âge brut corrigé lorsque celui-ci n'est pas rempli dans l'EFL par l'âge EAR.

L'âge EAR retenu est l'âge imputé.

Par construction, lorsque l'année de naissance EFL est manquante, c'est qu'on ne retrouve pas l'enfant dans l'EAR.

Dans le cas contraire, l'année de naissance EAR aurait été récupérée pour combler la valeur manquante dans l'EFL.

En grisé : les individus hors champ : ni enfants mineurs EFL ni enfants majeurs EAR.

Environ 300 enfants mineurs dans chacune des sources ont une année de naissance antérieure à 1993 dans l'autre source.

Les 1 611 enfants mineurs EAR qu'on ne retrouve pas dans l'EFL (soit parce qu'ils ont été codés à tort comme enfant d'un ou des adultes présents dans le logement, soit parce qu'ils ont été omis dans l'EFL) « compensent » partiellement les 1799 mineurs de l'EAR codés comme individu hors famille (moco=31) et les 806 mineurs EFL qui n'ont pas de BI dans l'EAR.

En tout, on compte donc 1 945 enfants mineurs EAR hors champ dans l'EFL (soit non présents, soit majeurs) et 2 858 enfants mineurs EFL hors champ dans l'EAR (soit car MOCO='31', soit majeurs, soit non présents).

**Tableau 5 : Répartition des enfants mineurs non communs par type de famille**

		Effectifs	Part (en %)	Effectifs pondérés
Inclus dans champ EFL seulement	Couple avec	2 403	84,1	160 000
	Traditionnelle	1 230	43,0	80 000
	Recomposée	1 173	41,0	80 000
	Monoparentale	455	15,9	50 000
	<b>Ensemble</b>	<b>2 858</b>	<b>100,0</b>	<b>210 000</b>
Inclus dans champ EAR seulement	Couple avec enfant	1 507	77,5	100 000
	Monoparentale	438	22,5	60 000
	<b>Ensemble</b>	<b>1 945</b>	<b>100,0</b>	<b>160 000</b>

Champ : Enfants au sens EFL ou EAR nés entre 1993 et 2010, vivant principalement dans le logement, de poids non nul.

Par le jeu des pondérations, les observations d'enfants EFL absentes du champ EAR, moins nombreuses que les observations d'enfants EAR absentes du champ EFL, pèsent davantage sur les effectifs pondérés globaux.

**Tableau 6 : Écarts du nombre d'enfants par famille entre l'EFL et l'EAR imputables à la différence de champ**

	Écarts EFL - EAR	Écarts liés au champ	Écarts "résiduels"
Couple avec enfant	140 000	60 000	80 000
Famille monoparentale	- 90 000	-10 000	-80 000
<b>Ensemble</b>	<b>50 000</b>	<b>50 000</b>	<b>0</b>

Champ : enfants au sens EFL ou EAR nés entre 1993 et 2010, vivant principalement dans le logement.

Les écarts de champ des enfants retenus expliquent une partie des différences de répartition des enfants entre « couple avec enfant(s) » et famille monoparentale. En effet, les enfants « présents dans l'EAR et absents de l'EFL » sont plus souvent en famille monoparentale que



les enfants « présents dans l'EFL - absents de l'EAR » (23 % contre 16 % en pourcentages non pondérés et 38 % contre 24 % en pondérés).

Si on conservait la configuration familiale de l'EAR pour les enfants mineurs communs aux deux sources en rajoutant les enfants mineurs EFL non communs, on aurait 18,6% d'enfants en famille monoparentale. Sur les 0,7 points de différence observés sur données macro entre la part des enfants en famille monoparentale dans l'EFL et dans l'EAR, 0,1 point est ainsi dû à ces écarts de champ. Les 0,6 points restants sont dus à un classement différent dans l'EAR et l'EFL d'enfants présents simultanément dans les deux bases.

Ces écarts non expliqués par les différences de champ sont appelés dans le tableau écarts « résiduels ».

### **III.2 Qui sont les enfants EFL non considérés comme tels dans l'EAR ? En particulier, qui sont les enfants codés « isolés hors famille » dans l'EAR ?**

1799 individus « isolés hors famille » dans l'EAR sont considérés comme des enfants dans l'EFL, d'après la définition retenue. Ils vivent pour l'essentiel (plus des trois quarts) dans un logement considéré comme complexe au vu des informations collectées dans les bulletins individuels et dans lequel le traitement manuel du prestataire n'a pas été pris en compte. Cela peut, par exemple, être le cas si l'âge d'un enfant a été mal saisi (on trouve 1954 au lieu de 1994) : l'incohérence apparente qui ressort alors sur la différence d'âge entre les parents et l'enfant conduit le processus RP à rejeter le traitement du prestataire. L'année de naissance, fautive, de l'enfant est maintenue, mais le mode de cohabitation des personnes de la famille devient alors « personne hors famille ». Le traitement manuel du prestataire est rejeté alors qu'il aurait conduit à une variable correcte.

Pour ces 1400 cas, leur intégration dans le champ des enfants côté EFL est, sans conteste, validée.

On note que les deux tiers des enfants concernés par l'abandon du traitement du prestataire sont, côté EFL, des enfants de familles recomposées.

**Leur non prise en compte dans le champ des enfants dans l'EAR conduit à minorer la part des enfants vivant avec des couples au profit de ceux vivant en familles monoparentales.**

Les 400 autres individus mineurs « isolés hors famille » sont directement codés par le prestataire suite à l'examen des liens. Ils se répartissent dans environ 300 ménages. Dans les deux tiers des ces ménages, il n'y a aucun individu parent<sup>32</sup> et la majorité d'entre eux sont constitués uniquement d'individus « isolés hors famille ». Si dans certains cas (minoritaires), il s'agit bien d'une cohabitation de personnes hors famille (LIEN='colocataire') ou de situations plus difficiles à saisir (tous les liens de la liste A sont vides et les individus déclarent n'être pas en couple dans les BI), dans la majorité des cas, les liens sont clairement renseignés et indiquent une famille « classique » ('EPOUSE' + 'FILS' / 'FILLE') et les informations des BI sont aussi sans ambiguïté<sup>33</sup>. Il pourrait s'agir d'erreur du prestataire lors du codage des liens de ces personnes.

Un certain nombre d'enfants EFL classés comme hors famille dans l'EAR vivent avec d'autres personnes en famille et notamment au moins un individu parent.

Ces enfants sont dans des situations diverses, notamment :

- Les liens « belle-fille », « beau-fils » ou « bel enfant » sont parfois compris comme conjoint d'un des enfants, alors qu'il semble, au vu de ce qu'on trouve dans l'EFL, qu'il s'agisse bien des enfants du conjoint.
- Quelques-uns d'entre eux sont des enfants placés, en famille d'accueil.

On note à nouveau que le plus grand nombre de cas d'incohérences se trouvent sur les enfants en famille recomposée. **Leur non prise en compte dans le champ des enfants dans l'EAR conduit à nouveau à minorer la part des enfants vivant avec des couples au profit de ceux vivant en familles monoparentales.**

<sup>32</sup> pas de MOCO=22 ou 23.

<sup>33</sup> COUPLE\_X=1 pour les deux adultes les plus âgés, années de naissance entre les adultes en couple et avec les enfants supposés cohérentes.



**Au global, la réintégration de ces individus « isolés hors famille » en « enfants », lorsqu'ils apparaissent dans le champ des enfants de l'EFL, semble tout à fait adéquate.**

### III.3 Comparaison de la répartition des enfants présents dans les deux sources selon la typologie de famille entre les deux sources

En plus des écarts sur le champ des enfants signalés ci-dessus, on note aussi des incohérences entre les deux sources sur le classement des enfants mineurs présents de part et d'autre. Ainsi, des enfants sont codés comme « enfant de famille monoparentale » d'après une source alors qu'ils sont dans une famille traditionnelle ou recomposée d'après l'autre. On évalue ainsi à 265 le nombre de mineurs enfants de famille monoparentale dans l'EFL et enfants de couple dans l'EAR. Inversement, 647 observations correspondent à des enfants de familles monoparentales dans l'EAR et à des enfants de couple dans l'EFL.

**Tableau 7 : Répartition des individus mineurs enfants au sens EFL et EAR selon les typologies de famille**

		Mineurs EAR		
		Moco=11	Moco=12	Ensemble
Mineurs EFL	Traditionnelle	144 379	503	144 882
	Recomposée	20 666	144	20 810
	Monoparentale	265	20 667	20 932
	Ensemble	165 310	21 314	186 624

Champ : enfants au sens EFL et EAR, nés entre 1993 et 2010 compris.  
Hors observations de poids nuls.

Sur le champ des enfants communs aux deux sources, la part sur données pondérées des enfants mineurs en famille monoparentale est alors de 18,4 % dans l'EAR contre 17,9 % dans l'EFL, soit un écart de 0,5 points (contre 0,7 point sur l'ensemble des données).

#### a/ Enfants vivant en famille monoparentale dans l'EFL et avec un couple parental dans l'EAR

Parmi les 265 observations classées en enfant d'un couple<sup>34</sup> (MOCO=11) dans l'EAR et famille= « monoparentale » dans l'EFL, le tiers vit avec un parent en couple non cohabitant. Si cette non cohabitation se fait pour raison professionnelle, le conjoint qui vit ailleurs la semaine, est néanmoins rattaché au logement dans le recensement. Il ne s'agit pas d'une incohérence entre les sources, mais d'une différence de concept. Pour les observations restantes, difficile de trancher d'où vient le problème ; probablement un peu des deux sources.

#### b/ Enfants vivant en famille monoparentale dans l'EAR et avec un couple parental dans l'EFL

Parmi les 647 observations classées en enfant d'une famille monoparentale dans l'EAR et famille= « recomposée » ou « traditionnelle » dans l'EFL, on en compte 309 (48 %) qui vivent dans un logement où le sexe de l'un des membres du couple a été mal rempli dans l'EAR : contrairement à ce qui est indiqué dans l'EFL et en désaccord avec ce qu'on lit dans les informations de la liste A de la feuille logement du recensement, les deux membres du couple sont de même sexe dans leur bulletin individuel du recensement. Dans l'EAR pour la variable mode de cohabitation (MOCO) utilisée pour les analyses ménages famille un couple est composé d'un homme et d'une femme. Deux individus d'un même sexe ne peuvent donc pas former un couple au sens de MOCO, les deux adultes ont été recodés comme ne formant pas un couple et les enfants ont été rattachés à un seul des deux adultes pour former une famille monoparentale<sup>35</sup>. **La moitié (46 % sur données pondérées) de la sur-représentation des enfants en famille monoparentale dans l'EAR par rapport à l'EFL**

<sup>34</sup> MOCO=11.

<sup>35</sup> on a donc MOCO=31 pour l'un des adultes et MOCO=23 pour l'autre.



**vient ainsi d'erreur de saisie ou de remplissage sur le sexe d'un des membres du couple.**

Parmi les 338 autres enfants classés en famille monoparentale à l'EAR et comme vivant avec un couple parental dans l'EFL, 97 vivent dans un logement avec une seule famille sans personne hors famille au sens de l'EAR, c'est à dire une forme simple de ménage avec une famille.

Ces 97 observations d'enfants sont répartis dans 66 logements.

Pour la majorité d'entre eux, le (beau-)parent se déclare en couple dans l'EFL comme dans le BI, **mais il est la seule personne en couple avec un BI dans le logement** : soit le conjoint a oublié de remplir un BI, soit le conjoint est rattaché à un autre logement même si le répondant indique qu'il s'agit d'une situation de cohabitation, la notion pouvant dans des cas compliqués susciter différentes interprétations. Deux cas de figure sont alors possibles :

- soit le conjoint apparaît dans la liste A, cas le plus fréquent, et on peut imaginer que celui-ci a oublié de remplir son BI.
- Soit le conjoint n'apparaît pas dans la liste A. Difficile de trancher puisque les informations sont contradictoires.

Cependant, dans 70 % des cas, le répondant à l'EFL a indiqué (avant apurement / redressement) que le conjoint vivait « toute l'année (ou presque) » dans le logement (q9tps\_logr\_x='100000').

Pour les autres, minoritaires dans cette catégorie, le classement en famille monoparentale à l'EAR vient d'une erreur d'appréciation du lien par le prestataire entre les membres du couple : les deux parents déclarent bien vivre en couple dans leur bulletin individuel<sup>36</sup> et des liens de conjoint (ami / concubin / concubine / conjoint / épouse / époux / mari / pacsé / union libre) apparaissent dans la liste A des feuilles logement du recensement.

On dénombre aussi 227 observations d'enfants vivant dans un ménage composé d'une seule famille et d'une (ou plusieurs) personne(s) isolée(s) hors famille.

Parmi ceux-ci, environ un quart vit dans un logement dans lequel aucune ou une seule personne se déclare en couple sur leur BI. La procédure de codage du RP considère qu'il n'y a pas de couple dans le logement et que les enfants qui y vivent sont en famille monoparentale.

Pour les trois autres quarts, dans la plupart des cas, on a bien le même nombre de personnes dans la liste A que de bulletins individuels renseignés et les liens de parenté entre les individus sont généralement correctement remplis. On observe ainsi un nombre élevé de logement dans lequel on trouve un « mari », « époux(se) », une « femme », un « pacsé », ... Pour ces cas, les indicateurs de complexité du ménage indiquent que la variable mode de cohabitation de l'EAR a été construite en s'appuyant sur la lecture « manuelle » des liens de parenté de la liste A. Ce codage manuel considère alors, à tort, un lien « ami » au lieu de conjoint conduisant à considérer une famille monoparentale et un ou des isolés.

Enfin 13 enfants, répartis dans 9 logements, vivent dans un ménage avec deux familles, au sens de l'EAR. Dans la quasi-totalité de ces cas, il s'agit de familles multi générationnelles où un des grands « enfants » des adultes les plus âgés est lui-même parent et ne vit pas en couple : on a donc un couple avec ou sans enfant (les grands-parents dont plusieurs enfants peuvent vivre encore avec eux sans être eux-mêmes parents ou en couple) et une famille monoparentale (souvent la fille du couple précédent avec son ou ses enfants) ou une famille traditionnelle.

---

<sup>36</sup> COUPLE=COUPLE\_X=1.



#### IV. Part des enfants en familles recomposées et en familles traditionnelles dans l'EAR

Combien de familles recomposées et de familles traditionnelles dans l'EAR ? Ce calcul n'est pas possible dans l'EAR. On estime ici ces effectifs (tableau 8) en appliquant les taux obtenus avec l'EFL sur le total des familles de type « couple avec enfants ».

**Tableau 8 : Estimation de la répartition par type de famille des enfants EAR**

	EAR (sous échantillon)	
couple + enfants	6 100 000	79,1
<i>traditionnelle</i>	5 390 000	69,9
<i>recomposée</i>	710 000	9,2
monoparentale	1 610 000	20,9
Ensemble	7 720 000	100,0

Champ : enfants EAR mineurs nés avant 2011.



# Correction des poids Zus dans EFL

Travail réalisé par le Secrétariat Général du Comité Interministériel à la Ville

## *1. Correction des poids individuels*

Comme convenu lors de la dernière réunion, nous avons corrigé dans un 1er temps les poids individuels (table efl\_individu). Contrairement à ce qui était indiqué dans notre mail précédent, il nous a semblé peu judicieux de recalculer directement sur les chiffres populations en Zus et hors Zus du RP 2006. L'écart dans le temps, ainsi que l'impact éventuel de la rénovation urbaine, nous a conduit à préférer une approche s'assurant :

- que les populations résidant respectivement en Zus et hors Zus estimées par EFL soient conservées
- que le rapport homme/femme, aussi bien en Zus que hors Zus, corresponde à celui observé dans le RP 2006 (sur le champ des unités urbaines avec Zus)

Pour information, voici ci-dessous quelques éléments de comparaison avant/après correction des poids, sur le champ des adultes résidant dans des unités urbaines avec Zus. Les données grisées repèrent les données qui sont identiques, de par la méthode de correction des poids.

Notre population demeure un peu plus âgée que ce que le recensement nous permettait d'attendre. Nous allons voir en interne dans quelle mesure cela peut s'expliquer (rénovation urbaine, etc.).



	Unités urbaines avec Zus			En Zus			Hors Zus dans une UU avec Zus		
	EFL anciens poids	EFL nouveaux poids	RP 2006*	EFL anciens poids	EFL nouveaux poids	RP 2006*	EFL anciens poids	EFL nouveaux poids	RP 2006*
Hommes	12 803 712	12 509 904	11 765 557	1 096 964	1 250 412	1 349 860	11 706 748	11 259 492	10 415 698
Femmes	14 068 542	14 362 350	13 510 490	1 606 553	1 453 105	1 568 674	12 461 988	12 909 245	11 941 817
Ensemble	26 872 254	26 872 254	25 276 048	2 703 517	2 703 517	2 918 534	24 168 736	24 168 736	22 357 514
% Hommes	47,6	46,6	46,5	40,6	46,3	46,3	48,4	46,6	46,6
% Femmes	52,4	53,4	53,5	59,4	53,7	53,7	51,6	53,4	53,4
18_24	13,0	13,0	13,3	14,2	14,3	16,1	12,9	12,9	13,0
25_34	17,6	17,6	18,3	20,2	20,3	20,5	17,3	17,3	18,0
35_44	17,4	17,4	17,9	18,1	18,3	18,7	17,3	17,3	17,9
45_54	17,1	17,0	17,3	16,3	16,2	16,9	17,1	17,1	17,3
55_64	15,5	15,5	14,0	14,7	14,7	12,7	15,6	15,6	14,2
65epI	19,4	19,5	19,1	16,5	16,2	15,1	19,8	19,9	19,7
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Hommes									
18_24	13,4	13,4	13,8	15,2	15,2	16,8	13,2	13,2	13,4
25_34	18,4	18,5	19,4	21,3	21,3	21,2	18,1	18,1	19,1
35_44	18,2	18,3	18,7	19,6	19,6	19,3	18,1	18,1	18,6
45_54	17,6	17,5	17,5	15,8	15,8	16,6	17,7	17,7	17,6
55_64	15,5	15,5	14,3	14,5	14,5	13,3	15,6	15,6	14,4
65epI	16,9	16,9	16,4	13,6	13,6	12,8	17,3	17,3	16,8
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Femmes									
18_24	12,7	12,7	12,9	13,5	13,5	15,5	12,6	12,6	12,6
25_34	16,9	16,9	17,4	19,4	19,4	20,0	16,6	16,6	17,0
35_44	16,6	16,6	17,3	17,1	17,1	18,1	16,6	16,6	17,2
45_54	16,6	16,6	17,1	16,7	16,7	17,2	16,6	16,6	17,1
55_64	15,5	15,5	13,8	14,9	14,9	12,2	15,6	15,6	14,0
65epI	21,7	21,8	21,5	18,5	18,5	17,0	22,1	22,1	22,1
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

\* : 18 ans et plus résidant en métropole dans des ménages ordinaires

Nous avons en revanche rencontré deux problèmes.

Les nouveaux poids (npoids\_ind) ont la même propriété que les anciens, à savoir : tous les individus d'un même ménage de la table efl\_individu reçoivent le même poids (puisque tous les individus d'un même ménage de efl\_individu ont le même sexe, donc le même coefficient correctif). J'ai cependant 3 ménages de efl\_individu (1108490447, 1117141940, 1117214662) où des individus de sexe différent ont été interrogés et qui, après correction des poids, ne vérifient donc pas la condition précédente. C'est un problème très marginal, qui n'est pas gênant à ce stade, mais peut l'être éventuellement à la suivante (pour la correction des poids poids\_men\_fam\_enf). La solution la plus simple peut être de les éliminer. Comment gérez-vous ces ménages ? Sinon, j'ai pensé éventuellement à une solution alternative qui serait de remonter au niveau ménage pour effectuer le calage sur des totaux hommes et femmes, et de redistribuer ensuite les poids individuels.

Le second souci vient de la variable POSP (position professionnelle). Pour repérer les éventuelles déformations de l'échantillon Zus, nous avons effectivement comparé EFL en Zus au recensement sur différentes dimensions : âge, sexe, CS. Pour la CS, François Lebeau nous a cependant conseillé de travailler à partir de la variable POSP et nous a transmis les marges correspondantes. Mais nous trouvons un résultat surprenant sur cette



variable. Ci dessous la répartition des individus selon cette variable dans les deux sources. Pour EFL, les poids utilisés sont les poids individuels standard fournis dans la base. On a presque l'impression d'une inversion entre 1D et 1J d'une part, et 1H et 1I d'autre part.

Position professionnelle	EFL	RP 2006
1A Position professionnelle non précisée	5,2	5,5
1B Manœuvre, ouvrier spécialisé	4,4	4,7
1C Ouvrier qualifié ou hautement qualifié, technicien d'atelier	4,4	4,6
1D Employé (de bureau, de commerce, de la restauration, de maison...)	3,6	12,0
1E Agent de catégorie C ou D de la fonction publique	2,8	3,1
1F Agent de maîtrise administrative ou commerciale, VRP	2,6	2,7
1G Technicien (non cadre)	3,3	3,5
1H Agent de catégorie B de la fonction publique	7,3	2,8
1I Ingénieur, cadre d'entreprise	2,9	6,6
1J Agent de catégorie A de la fonction publique	12,0	3,1
21 Indépendant	3,0	2,4
22 Employeur	2,5	2,5
23 Aide familial	0,1	0,1
ZZ – sans objet	46,0	46,3
Total	100,0	100,0

Champ : adultes résidant dans des unités urbaines avec Zus

En revanche, la répartition dans EFL de la variable d'origine POSP\_X sur le même champ semble tout à fait correcte :

POSP_X	Fréquence	Pourcentage
1 Manœuvre, ouvrier spécialisé	1 207 998	4,5
2 Ouvrier qualifié ou hautement qualifié, technicien d'atelier	1 231 189	4,6
3 Technicien (non cadre)	977 731	3,6
4 Agent de catégorie B de la fonction publique	751 401	2,8
5 Agent de maîtrise administrative ou commerciale, VRP	701 975	2,6
6 Agent de catégorie A de la fonction publique	889 827	3,3
7 Ingénieur, cadre d'entreprise	2 069 225	7,7
8 Agent de catégorie C ou D de la fonction publique	778 608	2,9
9 Employé (par exemple : de bureau, de commerce, de la restauration, de maison)	3 289 775	12,2
Valeurs manquantes (sans objet)	14 974 526	55,7
Ensemble	26 872 254	100,0

Source : EFL

## 2. Correction des poids ménage-famille-enfant

Nous avons étudié la note qui nous a été envoyée ainsi que le programme de partage des poids, et nous avons bien compris la suggestion d'appliquer une simple règle de 3 entre les poids `poids_men_fam_enf` et `poids_ind` pour établir les poids `men_fam_enf` corrigés. En effet, la correction des poids individuels tels que nous l'appliquons (et notamment le fait que les poids individuels soient corrigés de façon uniforme au sein d'un même ménage), et la méthode du partage des poids pour déterminer les poids `men_fam_enf` tels qu'il est mis en œuvre dans le programme, impliquent que l'on devrait effectivement avoir des nouveaux poids ménage vérifiant:

$$npoids\_men\_fam\_enf=(npoids\_ind/poids\_ind)*poids\_men\_fam$$

Cependant lorsque nous lançons le programme en question (qui tourne sans problème), ce n'est pas le cas (à l'exception des ménages avec les structures familiales suivantes : `sfm=11`,



12, 30 (sfm sans enfant)). Nous souhaiterions juste nous assurer auprès de vous des raisons de cet écart avant de finaliser nos poids. Laurent Toulemon nous a indiqué que cela vient de la dernière étape de calage des poids ménage, réalisée après partage des poids. N'est-il pas cependant surprenant que les structures familiales 11, 12 et 30 ne soient pas du tout impactées ? (la note envoyée ne détaille pas les variables de calage utilisées au niveau ménage) Lors de la dernière réunion, vous aviez par ailleurs évoqué l'existence de deux programmes de partage des poids, l'un simple, l'autre plus complexe. La différence ne viendrait-elle pas également éventuellement de là ?

Signé :

