



Liberté • Égalité • Fraternité

RÉPUBLIQUE FRANÇAISE

Ministère de l'économie et des finances

Ministère des affaires sociales et de la santé

Ministère du travail, de l'emploi, de la formation professionnelle et du dialogue social

**Direction de la recherche, des études,
de l'évaluation et des statistiques**

**Sous-direction de l'observation de la
solidarité**

Bureau de la jeunesse et de la famille

Dossier suivi par : Marie Acs

Tel : +33 (0) 1 40 56 81 78

Fax : +33 (0) 1 40 56 84 08

Mél : <mailto:marie.acs@sante.gouv.fr>

Paris, le 17 janvier 2013

DREES-BJF N° 03 / 2013

Chemin d'accès document :

2013_03_Appariement des enfants des tableaux 13 et 28
de l'EFL.docx

**Note à l'attention de Guillemette Buisson
Division enquêtes et études démographiques
Institut national de la statistique et des études économiques**

Objet : Appariement et apurement des enfants des tableaux 13 et 28 de l'EFL

Cette note présente les traitements qui ont été effectués pour rendre diffusables les informations du tableau 28 qui renseigne le mode de garde habituel et la scolarisation éventuelle des enfants de moins de 4 ans de la famille.

1 Définition du champ

La première étape consiste à définir le champ que nous souhaitons retenir pour l'utilisation des informations du tableau 28.

Le champ demandé dans le tableau 28 est celui des enfants de moins de 4 ans qui vivent dans le logement enquêté. Or le fait de vivre dans le logement peut être interprété de différentes manières par les répondants ; en particulier, on pourra trouver dans le tableau 28 des enfants qui sont dans le logement qu'une partie du temps (garde alternée par exemple). Ensuite, la limite d'âge n'a pas toujours été respectée et certains enfants ayant 4 ans, voire plus, ont tout de même une ligne leur correspondant dans le tableau 28. Enfin, le lien de l'enfant avec le répondant et son conjoint n'est pas précisé dans l'intitulé : un enfant d'une autre personne peut ainsi être cité dans le tableau 28.

Par ailleurs, le tableau 13 décrit l'ensemble des enfants du répondant ou de son conjoint vivant dans le logement y compris ceux n'y vivant qu'une partie du temps seulement. Dans le tableau 13, une variable permet de déterminer si l'enfant vit tout le temps dans le logement, la moitié du temps ou plus, ou moins de la moitié du temps.

Enfin, les enfants présents dans le tableau 13 ont fait l'objet d'un appariement avec les enfants recensés dans le logement. Or, le champ des enfants du recensement correspond théoriquement aux enfants présents dans le logement la plus grande partie de l'année.

Il existe ainsi trois champs différents pour étudier les enfants de moins de 4 ans :

Adresse postale : 14 avenue Duquesne 75350 PARIS 07 SP
Adresse visiteurs : 11 place des cinq martyrs du lycée Buffon - 75014 PARIS

- 1^{er} champ : les enfants déclarés dans le tableau 28.
- 2^{ème} champ : les enfants du tableau 13 nés en 2007 ou après. Ces enfants sont seulement ceux du répondant ou de son conjoint, et ne vivent pas forcément la majeure partie du temps dans le logement.
- 3^{ème} champ : les enfants de moins de 4 ans recensés dans le logement, c'est-à-dire vivant dans le logement la majeure partie du temps.

La question du tableau 28 concerne le mode de garde habituel des enfants. Aussi, il n'est pas pertinent de retenir les enfants qui ne passent qu'une faible quotité de temps dans le logement. D'ailleurs, la question concernant les modes de garde n'a été remplie que dans 40% des cas des enfants vivant moins de la moitié du temps dans le logement.

Enfin, le champ retenu pour l'apurement est celui des **enfants de moins de 4 ans vivant dans le logement la majeure partie du temps**. Par la suite, on se limitera aux moins de 3 ans pour la publication.

La construction du champ se fait en deux étapes : la première intègre les enfants vivant avec au moins un de leurs parents dans le logement, la seconde ajoute les enfants vivant sans aucun de leurs parents.

Les enfants nés en 2007 ou après vivant la majeure partie du temps dans le logement avec au moins un de leurs parents sont déclarés dans le tableau 13. Or le champ du tableau 13 étant plus étendu que le champ des enfants vivant la majeure partie du temps dans le logement on ne conservera que les enfants du tableau 13 appariés avec l'Enquête Annuelle du Recensement (EAR), et les enfants présents dans le tableau 13 qui ne sont pas appariés avec l'EAR mais déclarant être présents dans le logement tout le temps.

Les enfants de moins de 4 ans vivant la majeure partie du temps sans aucun de leurs parents (enfants sous tutelle par exemple) sont également dans le champ retenu. Ces enfants sont présents à la fois dans le tableau 28 et l'EAR mais ils sont absents du tableau 13 puisqu'ils ne vivent pas avec leurs parents.

Au final, on compte dans ce champ 45 251 enfants.

2 Appariement des enfants

2.1 Appariement des enfants du tableau 28 avec le tableau 13

La sélection des enfants du tableau 13 se fait selon les conditions précitées : on conserve uniquement les enfants de moins de 4 ans appariés avec l'EAR, et ceux qui ne sont pas appariés avec l'EAR mais qui déclarent vivre tout le temps dans le logement.

Étape 1 : l'appariement entre les enfants du tableau 13 et ceux du tableau 28 se fait dans le même questionnaire EFL et sur le prénom. On apparie directement 86 % des enfants du tableau 13 (voir annexe pour les effectifs concernés par chacune des étapes).

Il arrive qu'il y ait des petites erreurs sur les prénoms. On réalise donc le même appariement sur le prénom et dans un même questionnaire EFL en acceptant cette fois-ci une erreur sur deux caractères maximum (fonction *complev* de SAS) entre le prénom figurant dans le tableau 13 et celui du tableau 28.

Étape 2 : des cas simples où deux enfants sont déclarés sur la même ligne du tableau 28 ont été identifiés. On a créé de nouvelles lignes dans le tableau 28 pour les enfants concernés. Et l'appariement entre les enfants du tableau 13 et ceux du tableau 28 qui ont été scindés est réalisé sur les prénoms dans le même questionnaire.

Étape 3 : certains enfants ont plusieurs prénoms ou un prénom et un nom de famille pour un tableau mais seulement un prénom dans l'autre. Pour traiter ces cas, on compare les chaînes de caractères présentes dans les variables prénoms des tableaux une à une. Lorsque deux chaînes de caractères sont identiques à deux caractères près, on valide l'appariement entre l'enfant du tableau 13 et celui du tableau 28. La comparaison est réalisée grâce à la fonction *scan* de SAS.

Étape 4 : les prénoms ne correspondent pas pour les enfants restant. Dans ce cas, on compare les nombres d'enfants non appariés des tableaux 13 et 28. Lorsqu'il y a le même nombre d'enfants restant dans un

même questionnaire, on les apparie directement sans tenir compte du prénom. La majorité des cas correspond à des prénoms à blanc (62 %).

S'il n'y a qu'un seul enfant dans chacun des tableaux 13 et 28, l'appariement ne pose pas de problème. Lorsqu'il y a deux enfants dans les deux tableaux, pour décider quels enfants doivent être appariés ensemble, on applique la règle suivante : l'enfant le plus vieux des deux (information donnée par le tableau 13 via la date de naissance) s'apparie avec l'enfant fréquentant le plus souvent l'école (information présente dans le tableau 28 via la variable scolarité). Cette étape permet d'apparier environ 2 % des enfants du tableau 13.

Étape 5 : on élargit l'appariement aux enfants du même logement, et non plus seulement du même questionnaire. En effet, certains enfants sont déclarés dans le tableau 13 d'un questionnaire du logement et dans le tableau 28 d'un autre questionnaire du même logement. L'appariement se fait dans le logement sur le prénom en acceptant une différence sur maximum deux caractères.

Étape 6 : parmi les enfants du tableau 13 qui n'ont pas encore été appariés avec le tableau 28, un grand nombre se trouvent dans des questionnaires pour lesquels aucun enfant n'a été déclaré dans le tableau 28. Ces enfants sont conservés dans le champ des enfants de moins de 4 ans vivant la majeure partie du temps dans le logement, mais ils n'auront pas les informations concernant les variables du tableau 28. Ils seront traités comme des non-réponses aux variables mode de garde et scolarité, et représente 8 % du champ final.

Étape 7 : il reste des enfants dans les deux tableaux mais pas en nombre identique et pour lesquels le prénom du tableau 13 ne correspond pas à celui du tableau 28. Ces cas, peu nombreux, sont traités à la main lorsque c'est possible de la façon suivante :

Si plusieurs enfants sont déclarés dans une même ligne du tableau 28 et que ce dédoublement n'a pas été repéré à l'étape 2, on crée de nouvelles lignes du tableau 28 permettant d'apparier les enfants du tableau 13.

Si les prénoms des enfants sont vides dans au moins un des tableaux (appariements précédents impossibles), et que l'on repère grâce à l'EAR qu'un enfant de plus de 4 ans est probablement déclaré dans le tableau 28 en plus d'un enfant qui a vraiment moins de 4 ans (ce qui fausse les comptes entre les tableaux puisque deux enfants sont comptés dans le tableau 28 et un seul dans le tableau 13), on conserve dans le tableau 28 la ligne correspondant au plus jeune des deux enfants suivant la règle de scolarité, c'est-à-dire la plus faible durée passée à l'école.

Étape 8 : enfin, les derniers enfants du tableau 13 n'ayant pas pu être appariés sont conservés dans le champ, et les informations sur le mode de garde et la scolarité ne seront pas renseignées pour ces enfants.

2.2 Appariement des enfants du tableau 28 avec l'EAR

Le champ que nous souhaitons obtenir est le champ des enfants de moins de 4 ans vivant la majeure partie du temps dans le logement. Cette seconde partie permet d'apparier les enfants de moins de 4 ans vivant la majeure partie du temps dans le logement sans aucun de leurs parents, par exemple les enfants sous tutelle. Pour ces cas, nous n'avons pas le prénom des enfants dans l'EAR pour affiner l'appariement.

Étape 1 : on compare les nombres d'enfants présents dans le logement à partir de l'EAR et à partir du ou des tableaux 28 du logement. Lorsqu'il y a le même nombre d'enfants dans un même logement, on les apparie. S'il n'y a qu'un seul enfant dans l'EAR et le tableau 28, l'appariement ne pose pas de problème. Lorsqu'il y a deux enfants dans les deux sources, pour décider quels enfants s'apparient ensemble, on applique la règle suivante : l'enfant le plus vieux des deux (information donnée par l'EAR via la date de naissance) s'apparie avec l'enfant fréquentant le plus souvent l'école (information présente dans le tableau 28 via la variable scolarité).

Étape 2 : on repère les enfants du tableau 28 qui sont hors champ. Pour ces cas, il ne reste plus d'enfant de moins de 4 ans dans l'EAR à apparier mais des enfants sont présents dans le tableau 28. En effet, soit les enfants du tableau 28 ont en fait plus de 4 ans, soit les enfants de l'EAR ont déjà été appariés avec le tableau 28 dans un autre questionnaire EFL, soit il n'y a aucun enfant dans le logement dans l'EAR (ce peut être des enfants présents qu'une partie de l'année dans le logement et donc non recensés).

Ces enfants ne seront pas dans le champ des enfants de moins de 4 ans vivant la majeure partie du temps dans le logement.

Étape 3 : pour les cas restants, il y a des enfants dans l'EAR et le tableau 28 mais pas en nombre identique. Ces cas, peu nombreux, sont traités à la main.

Lorsqu'un enfant présent dans le logement apparaît dans plusieurs questionnaires différents de ce logement, en général les lignes des tableaux 28 correspondant à cet enfant sont identiques. On conserve alors une seule ligne pour cet enfant, peu importe le questionnaire duquel elle provient.

Si les lignes du tableau 28 sont différentes, on choisit la ligne la mieux remplie, c'est-à-dire celle qui a le moins de non-réponse, de rature ou de biffage.

Pour le reste des enfants, le choix n'est pas possible, aussi ils ne seront pas appariés.

Étape 4 : Les enfants de moins de 4 ans vivant dans le logement la majeure partie du temps et qui n'ont pas été appariés avec le tableau 28, ne sont pas conservés dans le champ final.

2.3 Âge des enfants

La seule variable donnant une information sur l'âge des enfants dans le tableau 13 est l'année de naissance, or les enfants nés en 2007 peuvent avoir plus de 4 ans s'ils sont nés en début d'année (si leur anniversaire arrive avant le jour de l'interrogation). Pour pallier ce problème, on considérera l'âge des enfants au 31/12/2010. Le champ de la diffusion sera celui des enfants du logement ayant moins de 4 ans au 31/12/2010 (nés en 2007, 2008, 2009, 2010, 2011).

Ensuite, certains enfants n'ont pas la même année de naissance entre le tableau 13 et l'EAR. Lorsque ces enfants sont présents dans le tableau 28, on peut supposer qu'ils ont effectivement moins de 4 ans puisque deux informations se recoupent (une des leurs années de naissance est supérieure ou égale à 2007 et présence dans le tableau 28).

Cas 1 : l'année de naissance dans le tableau 13 est initialement antérieure à 2007, et la date de naissance de l'EAR est postérieure à 2007, et ces enfants s'apparient avec le tableau 28. On peut supposer qu'ils ont bien moins de 4 ans puisqu'ils sont dans le tableau 28. La date de naissance du tableau 13 a été corrigée par celle de l'EAR, et ces enfants sont donc conservés dans le champ des enfants de moins de 4 ans vivant la majeure partie du temps dans le logement. On dénombre 122 enfants dans ce cas.

Cas 2 : l'année de naissance dans le tableau 13 est antérieure à 2007, et la date de naissance de l'EAR est postérieure à 2007, mais aucun de ces enfants n'est apparié avec le tableau 28 (qui est vide pour ces questionnaires). Le problème viendrait ici de la date de naissance de l'EAR qui doit être erronée puisque ces enfants ne sont pas dans le tableau 28 ils doivent avoir plus de 4 ans. Ces enfants ne sont pas dans le champ des enfants de moins de 4 ans vivant la majeure partie du temps dans le logement car leur année de naissance dans le tableau 13 est antérieure à 2007. On dénombre 149 enfants dans ce cas.

Cas 3 : l'année de naissance dans le tableau 13 est postérieure à 2007, mais l'année de naissance de l'EAR est antérieure à 2007, et ces enfants s'apparient avec le tableau 28. On peut supposer qu'ils ont en réalité moins de 4 ans puisqu'ils sont dans le tableau 28. Le problème viendrait ici d'une erreur dans l'EAR. Ces enfants sont conservés dans le champ des enfants de moins de 4 ans du logement (étant donné que le champ a été filtré à partir de l'année de naissance provenant du tableau 13), mais il y aura une incohérence entre leur date de naissance dans l'EAR et dans l'EFL. On dénombre 421 enfants dans ce cas.

Les enfants présents dans le tableau 13, déclarant vivre tout le temps dans le logement qui ne sont appariés ni avec l'EAR, ni avec le tableau 28. Se pose alors la question de savoir si ces enfants sont vraiment dans le logement ? Il y a peut-être ici une erreur sur la variable *tpsigt* qui renseigne la quotité de temps passé par l'enfant dans le logement ? Faute d'information supplémentaire, ces enfants sont dans le champ des enfants de moins de 4 ans vivant dans le logement et ils seront non-répondants pour les variables issues du tableau 28.

L'année de naissance des enfants présents dans le champ final correspond à leur année de naissance dans le tableau 13, si cette dernière est inconnue (non-enseignée ou non-appariée) alors l'année de naissance est celle provenant du recensement.

2.4 Bilan de l'appariement

Dans le champ des enfants de moins de 4 ans au 31/12/2010 vivant dans le logement, il y a 45 251 enfants. Parmi eux, 44 946 proviennent du T13 (99,3 %), et 305 proviennent des cas où il y a des enfants dans le recensement et dans le T28 (0,7 %) mais pas dans le T13.

Il y a 3 578 cas de non-réponse sur les variables mode de garde et scolarisation, due à un défaut d'appariement : enfants présents dans le T13 et EAR pour lesquels l'appariement avec le T28 n'a pas réussi, soit 7,9 % du champ.

Il y a 542 enfants du champ final qui ne sont pas appariés avec le recensement.

Le champ final compte 45 251 enfants dont 92,1 % sont appariés avec le tableau 28 (cf. Annexe : Tableau récapitulatif de l'appariement).

Tableau 1 : Bilan appariement

	Nombre d'enfants	%
Total dans le champ	45 251	100,0%
appariement	41 673	92,1%
non appariement	3 578	7,9%
appariement direct	40 095	88,6%

Champ : Enfants de moins de 4 ans vivant dans le logement la majeure partie du temps.

3 Apurement des variables

L'objectif des travaux d'apurement est de traiter la non-réponse, et de prendre une décision en cas de réponse multiple. En effet, alors que l'intitulé précisait bien qu'un seul mode de garde habituel devait être rempli, certains questionnaires comptent plusieurs cases cochées à cette question.

Tableau 2 : Fréquence de la non-réponse ou des réponses multiples

Scolarisation		
non réponse	6 328	14,0%
<i>due à l'appariement</i>	3 576	7,9%
<i>à la question</i>	3 134	6,9%
réponse simple	38 425	84,9%
réponse multiple	116	0,3%
Total	45 251	
Mode de garde	nombre d'enfant	%
total de non réponse	4 826	10,7%
<i>due à l'appariement</i>	3 576	7,9%
<i>à la question</i>	1 250	2,8%
réponse simple	35 645	78,8%
réponse multiple	4 780	10,6%
Total	45 251	

Champ : Enfants de moins de 4 ans vivant dans le logement la majeure partie du temps.

Seuls 76 enfants sont dans le cas où les deux variables (mode de garde et scolarisation) ne sont pas renseignées. Lorsque le répondant n'a pas répondu à la question sur le mode de garde habituel, dans 94 % des cas il a répondu à la question sur la scolarisation de l'enfant. De même, lorsque l'enquêté n'a pas répondu à la question sur la scolarisation de l'enfant, dans 98 % des cas il a répondu à la question sur le mode de garde de l'enfant.

On traite en premier lieu la variable sur la scolarisation, puis on utilisera cette dernière comme information afin d'apurer la variable concernant le mode de garde.

3.1 Variable Scolarisation

La non-réponse à la question de scolarisation est décroissante avec l'âge de l'enfant. Les plus jeunes nés en 2010 ou 2011 ont les taux de non-réponse les plus élevés à cette question (10 % contre 2 % pour les enfants nés en 2007). On peut supposer que lorsque l'enfant est très jeune il ne va pas à l'école et le répondant ne s'est pas senti concerné par cette question.

Les résultats provenant de l'EFL sont comparés à ceux de l'enquête mode de garde de 2007, et aux taux de scolarisations de la Depp. L'âge des enfants dans les données provenant de la Depp correspond à l'année de naissance (année civile), par exemple pour la rentrée 2010, les enfants nés en 2008 ont 2 ans.

Les réponses multiples sont traitées de la manière suivante : on conserve la situation de scolarisation la plus élevée parmi celles déclarées. Cela concerne 0,3 % des enfants.

Les résultats sont pondérés par la variable de pondération poids_men_fam_enf.

Pour les enfants de moins de 2 ans :

Les enfants de 0 ou 1 an (nés en 2009, 2010 et 2011) ne peuvent pas être inscrits dans un établissement scolaire. Ainsi, ces enfants seront considérés comme n'étant pas scolarisés.

Tableau 3 : Scolarisation des enfants de 0 et 1 an

Les 0 et 1 an (en %)	Mode de garde 2007	Depp	EFL brut	EFL redressé
% d'enfants scolarisés hors non réponse				
<i>journée</i>	1,1	0	0,6	0
<i>matin</i>	0,9	-	0,5	0
	0,2	-	0,1	0
% d'enfants non scolarisés hors non réponse				
Total	98,9	100	81,7	100
% de non réponse	100	100	100	100
	0	0	17,6	0

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête dans les écoles publiques et privées de l'enseignement préélémentaire et élémentaire, Depp, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de 0 et 1 an (nés en 2009 ou après).

Règles retenues pour les enfants de moins de 2 ans :

- 1) En cas de non-réponse de l'enquêté, ou de non-appariement, l'enfant n'est pas scolarisé.

Pour les enfants âgées de 2 ans :

Plusieurs scénarios ont été envisagés pour traiter les cas de non-réponse à la question scolarisation des enfants de 2 ans :

- 1- Soit on conserve la non-réponse (colonne **EFL brut** dans le tableau ci-dessous). Cette variable EFL brute présente des taux de scolarisation plus élevés que ceux de la Depp (17 % contre 14 %), et surtout près de 13 % de non-réponse.
- 2- Soit on considère que les enquêtés n'ont pas répondu car ils ne sont pas concernés par cette question (uniquement pour les cas où l'enquêté n'a pas répondu, on laisse les non-appariés en non-réponse), ce qui signifie que ces enfants ne sont pas scolarisés (colonne **EFL redressé point 2**). La variable « EFL redressée Étape 1 » fournit des résultats se rapprochant de ceux de la Depp (16 % d'enfants de deux ans scolarisés contre 14 %) mais reste toutefois assez élevée, et présente un taux de non-réponse de 6 %. Si on conserve cette variable en l'état, la tranche des 2 ans conservera une part de non-réponse.

L'inconvénient de ces deux méthodes est qu'on ne traite pas tous les cas de non-réponse, et qu'il restera dans le fichier final des données incomplètes.

- 3- Pour traiter la non-réponse résiduelle, on réalise un modèle polytomique à trois modalités calculant pour chaque enfant la probabilité d'être scolarisé à temps complet, à temps partiel ou non scolarisé. Ce modèle est calculé à partir des variables disponibles : région, type de famille, mois de naissance, nombre d'enfant de moins de 10 ans du ménage, un des parents est homme ou femme au foyer. L'existence d'un homme ou d'une femme au foyer n'apparaît pas significativement déterminante et n'est pas retenue dans la modélisation finale.
- Afin de décider quelle modalité imputer, on tire dans une loi uniforme entre 0 et 1 un seuil d'acceptation. Si ce seuil est dans l'intervalle de probabilité de la modalité 1 alors on accepte celle-ci, s'il est dans l'intervalle cumulé de probabilité de la modalité 2 alors on accepte la 2, etc. On affecte alors la valeur de la modalité choisie à la variable scolarisation (colonne **EFL redressé point 3**). Dans 77 % des cas, cette technique permet d'obtenir la bonne modalité de la variable scolarisation (chiffre obtenu à partir des répondants).

Tableau 4 : Scolarisation des enfants de 2 ans

Les 2 ans (en %)	Mode de garde 2007	Depp	EFL brut	EFL redressé point 2	EFL redressé point 3
% d'enfants scolarisés hors non réponse	17,1	13,6	17,2	16,0	16,1
<i>journée</i>	5,3	-	7,2	6,7	6,7
<i>matin</i>	11,8	-	10,0	9,3	9,4
% d'enfants non scolarisés hors non réponse	82,9	86,4	82,8	84,0	83,9
Total	100	100	100	100	100
% de non réponse	0	0	13,1	6,2	0

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête dans les écoles publiques et privées de l'enseignement préélémentaire et élémentaire, Depp, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de 2 ans (nés en 2008).

En résumé, pour les deux ans, la règle suivante est appliquée :

- 1) En cas de non-réponse de l'enquêté, l'enfant n'est pas scolarisé.
- 2) En cas de non-appariement (l'enfant n'a pas de tableau 28), on impute une valeur déterminée à l'aide du modèle construit à partir des répondants.

Pour les enfants âgées de 3 ans :

D'après la Depp, la quasi-totalité des enfants de 3 ans sont scolarisés. On observe près de 15 % de non-réponse pour cette tranche d'âge. Plusieurs scénarios ont été détaillés :

- 1- On conserve la variable brute avec la non-réponse (**EFL brut**).
 - 2- Quasiment tous les enfants de 3 ans sont scolarisés, aussi lorsqu'il n'y a pas de réponse on considère que l'enfant va à l'école toute la journée. Cette correction donne des taux de scolarisation identiques à ceux de la Depp. Mais il est probable qu'elle surestime la part des enfants scolarisés la journée entière. (**EFL redressé point 2**)
 - 3- On réalise un modèle polytomique calculant pour chaque enfant la probabilité d'être scolarisé à temps complet, à temps partiel ou non scolarisé. La modélisation retient les mêmes variables que pour la scolarisation des enfants âgés deux ans, et l'existence d'un homme ou d'une femme au foyer est ici significativement déterminante et est retenue dans la modélisation finale.
- Afin de décider quelle modalité imputer, on tire dans une loi uniforme entre 0 et 1 un seuil d'acceptation. Si ce seuil est dans l'intervalle de probabilité de la modalité 1 alors on accepte celle-ci, s'il est dans l'intervalle de probabilité de la modalité 2 alors on accepte la 2, etc. On affecte alors la valeur de la modalité choisie à la variable scolarisation (colonne **EFL redressé**

point 3). Dans 63 % des cas, cette technique permet d'obtenir la modalité de la variable scolarisation déclarée (chiffre obtenu à partir des répondants).

Tableau 5 : Scolarisation des enfants de 3 ans

Les 3 ans (en %)	Mode de garde 2007	Depp	EFL brut	EFL redressé point 2	EFL redressé point 3
scolarisé	95	98	97,8	98,1	97,9
<i>journalée</i>	70	-	73,5	77,5	73,1
<i>matin</i>	25	-	24,3	20,7	24,8
non scolarisé	5	2	2,2	1,9	2,1
non réponse	0	0	15,0	0	0

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête dans les écoles publiques et privées de l'enseignement préélémentaire et élémentaire, Depp, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de 3 ans (nés en 2007).

Choix retenu pour les 3 ans :

- 1) En cas de non-réponse de l'enquêté, ou de non-appariement, on impute une valeur déterminée à l'aide du modèle construit à partir des répondants.

Bilan de l'apurement de la variable scolarisation :

Après redressement il n'y a plus de non-réponse pour la variable scolarisation.

Seules les proportions d'enfants scolarisés à 2 ans présentent un écart de plus de 2 points entre les taux donnés par l'EFL et par la Depp.

On compare les résultats issus de l'EFL et de la Depp des taux de scolarisation des enfants de 2 ans au niveau régional :

Tableau 6 : Comparaison par région des taux de scolarisation des enfants de 2 ans

	Sources	
	Depp	EFL redressé
Île-de-France	3,8 %	5,3%
Champagne-Ardenne	16,5 %	20,1%
Picardie	12,0 %	10,7%
Haute-Normandie	9,6 %	11,2%
Centre	6,9 %	9,8%
Basse-Normandie	16,7 %	20,8%
Bourgogne	16,8 %	15,4%
Nord-Pas-de-Calais	41,3 %	48,9%
Lorraine	13,8 %	15,0%
Alsace	5,4 %	6,4%
Franche-Comté	13,9 %	14,3%
Pays de la Loire	19,0 %	23,0%
Bretagne	37,0 %	43,1%
Poitou-Charentes	11,6 %	14,7%
Aquitaine	11,7 %	12,9%
Midi-Pyrénées	14,5 %	18,4%
Limousin	14,6 %	8,9%
Rhône-Alpes	12,0 %	10,8%
Auvergne	19,5 %	18,7%
Languedoc-Roussillon	12,7 %	14,5%

Provence-Alpes-Côte d'Azur	7,5 %	8,8%
Corse	11,2 %	11,2%
TOTAL	13,6 %	16,1%

Sources : Enquête dans les écoles publiques et privées de l'enseignement préélémentaire et élémentaire, Depp.

Enquête Famille et Logement 2011, Insee.

Champ : Enfants de 2 ans (nés en 2008).

Au niveau national, l'EFL surestime de 2,5 points le taux de scolarisation des 2 ans. Cet écart pourrait s'expliquer par l'effet des rentrées scolaires différées après septembre. D'après la Depp, ces rentrées différées peuvent faire progresser le taux de scolarisation de quelques points. L'interrogation de l'EFL s'étant déroulée au second trimestre de l'année scolaire 2010-2011 (janvier - février 2011), les enfants entrés à l'école maternelle en cours d'année sont pris en compte dans le taux de scolarisation obtenu à partir de l'EFL mais pas dans celui fourni par la Depp, qui lui est calculé à la rentrée de septembre.

Au niveau régional, l'EFL surestime dans quasiment chaque région le taux de scolarisation des enfants de 2 ans. Plus le taux de scolarisation est élevé et plus l'écart entre les sources est important, ce qui corrobore la théorie des rentrées différées. Le Nord-Pas-de-Calais affiche un taux de scolarisation supérieur à 41 % et l'écart entre les deux sources pour cette région est de l'ordre de 7 points.

3.2 Variable Mode de garde :

Comme nous l'avons repéré dans le tableau 2 sur la non-réponse, le problème majeur pour la variable mode de garde est la réponse multiple : 10,6 % de réponse multiple contre seulement 2,8 % de non-réponse à la question.

Afin d'apprécier la qualité de la réponse et orienter le choix des redressements, on utilise dans un premier temps deux variables issues de l'enquête mode de garde 2007 :

- la variable 'MdG 2007 - Cal' est le mode de garde de l'enfant pendant la journée de 8h à 19h du lundi au vendredi. Elle est construite à partir du calendrier des épisodes de garde. Cette variable est le mode de garde principal s'il est différent de l'école, sinon elle reprend le mode de garde secondaire. On obtient ainsi une variable donnant le mode de garde habituel en dehors de l'école (si l'enfant y va).
- La variable 'MdG 2007 - Qt Syn' est une variable de synthèse tirée de la question posée directement au parent sur le mode de garde principal de l'enfant (« Pour chacun de vos enfants non scolarisés ou scolarisés à temps partiel, pouvez-vous indiquer quel est actuellement son mode de garde principal ? »). Elle présente une plus grande ressemblance avec la question du mode de garde de l'EFL dans sa perception par le répondant. On peut alors penser que sa distribution sera plus proche de l'EFL que celle de la variable 'MdG 2007 - Qt Syn'. Mais cette question n'est posée que pour les enfants non scolarisés ou scolarisés à temps partiel. Ce qui ne correspond pas complètement au champ du tableau 28.

Correction de la réponse multiple :

La question concernant les modes de garde propose un choix entre les 5 modalités suivantes dans cet ordre :

- 1 : les parents
- 2 : une assistante maternelle
- 3 : une crèche (collective, familiale...)
- 4 : les grands-parents, famille
- 5 : autres.

Trois options¹ sont possibles pour corriger les réponses multiples :

- 1 – On ne les prend pas en compte et on considère ces réponses comme vides (EFL brut).

¹ Une quatrième option serait de conserver les réponses multiples. Mais cela introduirait le risque d'un biais car une seule réponse était explicitement demandée. De plus, cela pose le problème de la pondération des différentes modalités dans le cas d'une réponse multiple.

2 – On prend la première case remplie, car les modes de gardes apparaissent dans l'ordre décroissant de fréquence (EFL redr_2).

3 – Lorsqu'il y a une réponse multiple comportant la modalité « parent » et qu'au moins un des parents du ménage est inactif, le mode de garde habituel de l'enfant serait les parents.

Sinon, on prend en priorité le mode de garde payant. Si les deux modes de garde payants sont cochés (assistante maternelle et crèche), on choisit l'assistante maternelle. D'après l'enquête mode de garde 2007, dans la majorité des cas les enfants qui fréquentent ces deux modes de garde payants, passent plus de temps chez l'assistante maternelle qu'à la crèche.

Enfin, s'il n'y a aucun mode de garde payant, on prend la première case cochée car les modes de gardes apparaissent dans l'ordre décroissant de fréquence.

Ce redressement correspond à la colonne EFL redr_3.

Tableau 7 : Comparaison des redressements pour la réponse multiple

Champs des enfants de moins de 4 ans non scolarisés ou scolarisés à temps partiel	MdG 2007 - Cal*	MdG 2007 - Qt Syn	EFL brut (reste 19 % de non-réponse, y compris les réponses multiples)	EFL redr_2 (reste 9 % de non-réponse)	EFL redr_3 (reste 9 % de non-réponse)
Parents	66	51	51	55	49
Assistante maternelle	18	27	28	26	29
Crèche	9	13	14	12	15
Famille	5	7	5	5	5
Autre	2	2	2	2	2

* Lorsque le mode de garde principal est l'école, on utilise le mode de garde secondaire pour construire la variable mode de garde

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de moins de 4 ans (nés en 2007 et après) non-scolarisés ou scolarisés à temps partiel.

Les répondants ont tendance à déclarer plus facilement un mode de garde qui est payant comme mode de garde habituel, même si l'enfant passe le plus de temps avec ses parents. Aussi, l'option 2 permet de se rapprocher du mode de garde obtenu à partir du calendrier dans l'enquête MdG2007, alors que l'option 3 se rapproche de la question de synthèse.

Choix retenu : option 3

Correction de la non-réponse :

Il y a 10 % de non-réponse sur la variable mode de garde.

On réalise un modèle polytomique calculant pour chaque enfant la probabilité d'avoir chaque mode de garde. Ce modèle est estimé grâce aux enfants ayant une réponse unique à la question du mode de garde habituel, les enfants avec une réponse multiple n'ont pas été retenus pour l'estimation du modèle. Celui-ci est calculé à partir des variables disponibles (la région découpée en 3 modalités [Paris, Île-de-France, Province], taux départemental de scolarisation des 2 ans, nombre de places en accueil collectif ou familial pour 100 enfants de moins de ans dans le département, nombre de places chez les assistantes maternelles agréées pour 100 enfants de moins de 3 ans dans le département, type de famille, mois et année de naissance de l'enfant, nombre d'enfants de moins de 10 ans du ménage, un des parents est homme ou femme au foyer, taille de l'unité urbaine, catégorie socioprofessionnelle de la personne de référence du ménage, la scolarisation de l'enfant).

Afin de décider quelle modalité imputer, on tire dans une loi uniforme entre 0 et 1 un seuil d'acceptation. Si ce seuil est dans l'intervalle de probabilité de la modalité 1 alors on accepte celle-ci, s'il est dans l'intervalle de probabilité de la modalité 2 alors on accepte la 2, etc. On affecte alors la valeur de la modalité choisie à la variable mode de garde.

Cette méthode permet de bien classer les réponses dans 50% des cas le champ des répondants, à comparer avec 20% (une chance sur 5) si l'imputation d'une des 5 modalités était totalement aléatoire.

On réalise des comparaisons sur des champs communs aux variables disponibles.

Tableau 8 : Comparaison des modes de gardes pour les enfants non scolarisés ou scolarisés à temps partiel.

Champs des enfants de moins de 4 ans non scolarisés ou scolarisés à temps partiel	MdG 2007 - Cal*	MdG 2007 - Qt Syn	EFL brut réponses simples sans la non-réponse	EFL brut réponses simples avec la non-réponse	EFL redressé
Parents	66	51	51	41	50,6
Assistante maternelle	18	27	28	23	28,0
Crèche	9	13	14	11	14,4
Famille	5	7	5	4	4,9
Autre	2	2	2	2	2,1
Non réponse	0	0	0	19	0

* Lorsque le mode de garde principal est l'école, on utilise le mode de garde secondaire pour construire la variable mode de garde

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de moins de 4 ans (nés en 2007 et après) non-scolarisés ou scolarisés à temps partiel.

Comme vu précédemment, le mode d'interrogation dans l'EFL ressemble à celui de la question de synthèse de l'enquête mode de garde de 2007. Sur un champ comparable, les résultats de l'EFL sont assez proches.

Tableau 9 : Comparaison des modes de gardes pour les enfants scolarisés toute la journée

Champs des enfants de moins de 4 ans scolarisés toute la journée	MdG 2007 - Cal*	EFL brut réponses simples sans la non-réponse	EFL brut réponses simples avec la non-réponse	EFL redressé
Parents	82,4	61,5	40,6	60,9
Assistante maternelle	5	16,5	10,9	17,7
Crèche	1,2	4,3	2,8	5,3
Famille	5,2	9,4	6,2	8,8
Autre	6,2	8,3	5,5	7,3
Non réponse	0	0	34	0

* Lorsque le mode de garde principal est l'école, on utilise le mode de garde secondaire pour construire la variable mode de garde

Sources : Enquête Modes de garde et d'accueil des jeunes enfants en 2007, Drees, Enquête Famille et Logement 2011, Insee.

Champ : Enfants de moins de 4 ans (nés en 2007 et après) scolarisés toute la journée.

Peu de sources de comparaison sont disponibles sur le champ des enfants de moins de 4 ans scolarisés toute la journée. Seule la variable mode de garde principal obtenue à partir du calendrier des épisodes de garde de l'enquête mode de garde 2007 est disponible pour ce champ. Or cette variable est calculée à partir des épisodes de garde, et la question du mode de garde habituel n'étant pas posée sur ce champ dans l'enquête mode de garde 2007. Le mode de collecte explique ici les écarts importants.

Afin de corroborer la qualité des résultats obtenus avec l'EFL, nous les comparons avec ceux que l'on peut produire en utilisant l'échantillon national des allocataires (ENA) provenant de la Cnaf. Ce dernier permet notamment de dénombrer les familles allocataires du Complément de libre choix du Mode de Garde CMG assistante maternelle ou garde à domicile. Afin de retenir un champ plus proche de l'EFL, seuls les bénéficiaires qui ont perçu l'allocation au mois de décembre 2010 sont ici retenus. De plus, pour cette comparaison le champ est restreint aux familles ayant un enfant de moins de 3 ans.

Cet échantillon permet de réaliser des comparaisons au niveau régional du nombre de familles ayant recours à une assistante maternelle.

Tableau 10 : Comparaison des effectifs de familles ayant recours à une assistante maternelle pour un enfant de moins de 3 ans.

	Source EFL	Source ENA	Source ENA	Source ENA		
	Assistante maternelle	CMG assistante maternelle	CMG Garde à domicile	CMG Assistante maternelle ou garde à domicile	Nb d'enfants de moins de 3 ans (RP)	Nb d'enfant de moins de 3 ans dans EFL
Île-de-France	104 788	78 104	30 888	108 992	507 781	504 783
Champagne-Ardenne	13 333	12 558	52	12 610	48 360	50 643
Picardie	22 941	22 932	468	23 400	74 164	77 617
Haute-Normandie	26 274	21 970	572	22 542	70 233	68 461
Centre	37 448	34 528	598	35 126	90 302	94 873
Basse-Normandie	21 094	21 086	468	21 554	50 146	49 765
Bourgogne	18 998	17 940	364	18 304	53 354	53 633
Nord-Pas-de-Calais	40 270	38 766	1 118	39 884	164 556	174 759
Lorraine	21 441	21 606	520	22 126	80 182	81 406
Alsace	14 663	15 106	442	15 548	65 151	66 894
Franche-Comté	15 744	14 950	208	15 158	43 719	46 594
Pays de la Loire	66 797	64 454	1 352	65 806	136 851	142 792
Bretagne	48 851	49 452	962	50 414	113 197	117 811
Poitou-Charentes	21 619	22 022	468	22 490	56 833	57 346
Aquitaine	33 288	32 136	832	32 968	102 668	105 449
Midi-Pyrénées	28 706	26 520	832	27 352	94 987	95 620
Limousin	7 663	6 526	182	6 708	21 736	20 551
Rhône-Alpes	73 987	74 750	3 770	78 520	239 403	249 218
Auvergne	15 209	15 158	468	15 626	41 039	42 697
Languedoc-Roussillon	21 422	18 044	468	18 512	88 424	88 988
Provence-Alpes-Côte d'Azur	27 093	24 570	1 768	26 338	167 099	172 695
Corse	667	858	0	858	8 912	9 489
Total	682 295	634 036	46 800	680 836	2 319 097	2 372 081

Sources : Échantillon national des allocataires, Cnaf, Enquête Famille et Logement 2011, Insee.

Champ : Pour EFL : Familles ayant au moins un enfant de moins de 3 ans (nés en 2008 et après) gardés par une assistante maternelle.

Pour l'ENA : Familles bénéficiaires du CMG en décembre 2010 pour la garde d'un enfant de moins de 3 ans.

Les résultats obtenus dans l'EFL comprennent les assistantes maternelles non agréées qui ne donnent pas droit au CMG. D'après l'enquête mode de garde en 2007, elles représentent 2 % de l'ensemble des assistantes maternelles.

Au niveau national, l'EFL fournit un nombre de familles qui recourent à une assistante maternelle comme mode de garde habituel de leur enfant plus élevé de 5% que le nombre de familles bénéficiaires d'un CMG assistante maternelle fin décembre 2010.

Au niveau régional, les écarts demeurent assez faibles, à l'exception de la plus grosse région, l'Île de France, pour laquelle l'EFL dépasse le nombre de bénéficiaires d'un CMG assistante maternelle de 33%. Or, la région Île de France se singularise par un recours très important à la garde à domicile. Cette modalité n'était pas proposée dans l'EFL, il est probable que les familles utilisant une garde à domicile aient choisi la modalité assistante maternelle pour définir le mode de garde habituel de leur enfant de moins de 3 ans. Les

ordres de grandeur obtenus à partir de l'EFL du nombre de familles ayant recours à une assistante maternelle semblent cohérents avec le nombre de familles allocataires du CMG (assistante maternelle ou garde à domicile) de l'ENA. Toutefois ces comparaisons sont à prendre avec précaution, car le mode de collecte ainsi que les unités statistiques ne sont pas les mêmes, d'une part un fichier de familles allocataires et d'autre part une enquête auto-administrée.

Tableau 11 : Nombre de places en accueil collectif et familial pour 100 enfants de moins de 3 ans, et proportion d'enfants de moins de 3 ans gardés habituellement en crèche.

		Nombre de places en accueil collectif et familial pour 100 enfants de moins de 3 ans (PMI)	Proportion d'enfants de moins de 3 ans gardés habituellement en crèche (EFL)
11	Île-de-France	24,2	21,6
21	Champagne-Ardenne	12,0	12,7
22	Picardie	7,4	8,3
23	Haute-Normandie	9,7	8,4
24	Centre	11,9	8,4
25	Basse-Normandie	8,1	7,5
26	Bourgogne	12,1	11,1
31	Nord-Pas-de-Calais	8,7	9,1
41	Lorraine	10,1	11,9
42	Alsace	15,7	17,7
43	Franche-Comté	9,8	9,5
52	Pays de la Loire	9,9	10,0
53	Bretagne	9,0	9,5
54	Poitou-Charentes	10,4	11,4
72	Aquitaine	15,5	15,9
73	Midi-Pyrénées	18,0	19,2
74	Limousin	13,3	17,1
82	Rhône-Alpes	15,8	15,4
83	Auvergne	11,8	10,3
91	Languedoc-Roussillon	15,1	17,8
93	Provence-Alpes-Côte d'Azur	18,9	23,4
94	Corse	17,4	18,3
	ensemble	15,3	15,1

Sources : Enquête PMI 2010, Drees,
Enquête Famille et Logement 2011, Insee.

Champ : Enfants de moins de 3 ans (nés en 2008 et après).

L'enquête PMI réalisée par la Drees fournit le nombre de places offertes en accueil collectif et familial pour 100 enfants de moins de 3 ans par région. Cet indicateur est construit en excluant les jardins d'enfants qui accueillent les enfants de plus de 3 ans, en revanche, la totalité des places des haltes-garderies sont incluses, bien que ces structures puissent accueillir des enfants de plus de 3 ans.

Ces chiffres donnent des ordres de grandeur de la part d'enfants accueillis en structures collectives ou familiales, et les résultats obtenus par l'EFL sont cohérents avec ces ordres de grandeur.

Notons que les unités statistiques sont de nature différente entre ces deux enquêtes, d'une part une capacité d'accueil et d'autre part une proportion d'enfants effectivement accueillis.

**Le chef du bureau
de la jeunesse et de la famille**



Thibaut de Saint Pol

ANNEXE

Tableau récapitulatif de l'appariement

	Enfants du tableau 13 nés en 2007 ou après (appariés à l'EAR ou tout le temps dans le logement)	Enfant de l'EAR	Enfants du tableau 28	Conditions particulières	Variables d'appariement	champ	non réponse
Enfants d'égo ou de son conjoint dans le logement de moins de 4 ans (année de naissance T13)							
étape 1.1	38638		38638	aucune	prénom + cabefl	Oui	
étape 1.1	1457		1457	aucune	prénom modulo 2 caractères + cabefl	Oui	
étape 1.2	76		38	deux prénoms sur une seule ligne en T28	prénom + cabefl	Oui	
étape 1.3	239		239	comparaison des chaînes de caractères au sein des prénoms	prénom modulo 2 caractères + cabefl	Oui	
étape 1.4	736		736	nbT13 = nbT28 = 1	cabefl	Oui	
étape 1.4	170		170	nbT13 = nbT28 = 2	cabefl + sco	Oui	
étape 1.5	17		17	élargissement au logement	prénom modulo 2 caractères + cabfl	Oui	
étape 1.6	3413		0	nbT28 = 0	cabefl	Oui	Oui
étape 1.6	71		0	nbT28 = 0 et pas d'appariement avec EAR		Oui	Oui
étape 1.7	37		37	cas repérés à la main		Oui	
étape 1.8	92		0	enfants restant dans le T13		Oui	Oui
Enfants dans le logement de moins de 4 ans (année de naissance EAR)							
étape 2.1		201	201	nbEAR = nbT28 = 1	cabfl	Oui	
étape 2.1		102	102	nbEAR = nbT28 = 2	cabfl + sco	Oui	
étape 2.2		579	0	nbT28 = 0		Non	
étape 2.2		0	2635	nbEAR = 0 et appariement avec des + 4 ans du t13 sur le prénom		Non	
étape 2.2		0	504	enfant dans le logement mais déjà apparié par le biais d'un autre questionnaire		Non	
étape 2.2		0	254	présence d'enfants dans le logement (EAR) de plus de 4 ans		Non	
étape 2.2		0	594	aucun enfant recensé dans le logement		Non	
étape 2.3		2	2	cas à la main		Oui	
Total dans le champ	44 946	305					

Dictionnaire des codes : Table enfant_mdg_sco

Variables	Commentaires	Format
CABBI_ENF	Code à barres du BI de l'enfant	\$10.
CABEFL	Code à barres du questionnaire de l'EFL	\$10.
CABFL	Code à barres du logement	\$10.
E28anai	Année de naissance	N8.
E28mdg	Mode de garde habituel redressé	\$5.
E28mdg_X	Mode de garde habituel brut	\$5.
E28sco	Scolarisation redressée	\$3.
E28sco_X	Scolarisation brute	\$3.
ident	Identifiant de l'enfant	\$23.
I_e28mdg	Indicateur de redressement du mode de garde	\$1.
I_e28sco	Indicateur de redressement de la scolarisation	\$1.
Poids_men_fam_enf	pondération	N8.

E28anai Année de naissance de l'enfant

Cette variable correspond à l'année de naissance de l'enfant dans l'EFL (E13_14anai), si cette dernière est inconnue alors l'année de naissance est celle provenant du recensement (anai_r).

Modalités	Signification
2007 à 2011	Année de naissance

Numérique

E28mdg Mode de garde habituel redressé

Modalités	Signification
1	Parents
2	Assistante maternelle
3	Crèche (collective, familiale...)
4	Grands-parents, famille
5	Autres

Caractère

E28sco Scolarisation

Modalités	Signification
1	Non scolarisé
2	Scolarisé uniquement le matin
3	Scolarisé toute la journée

Caractère

I_e28mdg Indicateur de redressement du mode de garde

Modalités	Signification
0	Modalité d'origine
1	Réponse biffée ou réponse multiple
2	Imputation de la non-réponse

Caractère

I_e28sco

Indicateur de redressement de la scolarisation

Modalités	Signification
0	Modalité d'origine
1	Réponse biffée ou réponse multiple
2	Imputation de la non-réponse

Caractère