

ENQUÊTE NATIONALE 2 MOIS
Réalisée en 2011
PONDERATIONS AU TEMPS 1 (2 mois de l'enfant)



Ined
Inserm
EFS
InVS
Insee
DGS
DGPR
Drees
Cnaf

Hélène Juillard – Première version : mars 2014 – Dernière actualisation : **juillet 2015**

Plan

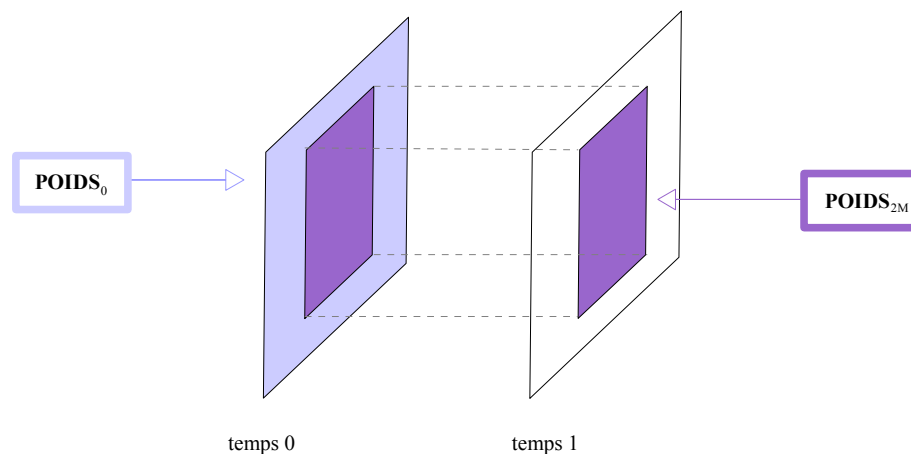
I – CONTEXTE (Cf. protocole).....	3
II – LA REPONDERATION.....	6
1) Descriptif de la non-réponse.....	6
2) Repondération en fonction de la non-réponse.....	8
3) Calage.....	9
III – UTILISATION DES POIDS ?.....	10

Quatre pondérations sont proposées : deux pondérations 'enfant' ([M02E_PONDREF](#), [M02E_PONDNONREF](#)) et deux pondérations 'famille' (information relative aux jumeaux non doublée, [M02F2_PONDREF](#), [M02F2_PONDNONREF](#)). Les résultats ci-après sont ceux des deux pondérations 'enfant', la méthode employée étant identique pour la famille. Attention, les pondérations 'enfant' et 'famille' ne sont pas faites pour être comparées.

En maternité (temps 0), l'échantillon final constitué d'environ 18 300 nourrissons, était issu d'un plan de sondage à probabilités d'inclusion inégales. Une pondération transversale POIDS_0 prenant en compte le plan de sondage et la non-réponse à différentes phases permettait d'inférer à l'ensemble de la population visée au travers de l'échantillon des 18 300 nourrissons (Cf. Pondérations en maternité).

Pour rappel, la **population d'inférence** ELFE est celle des nourrissons vivants, nés en 2011 en maternité métropolitaine, issus d'un accouchement au plus gémellaire, hors grands prématurés, ayant une mère majeure, en mesure de donner un consentement éclairé notamment dans l'une des langues proposées (français, anglais, arabe ou turc) et dont les parents ne résidaient pas temporairement en métropole.

Au temps 1 de l'étude (enquête à deux mois), la taille de l'échantillon des participants a diminué : environ 16 600 nourrissons dont la famille (mère et/ou père) a répondu à l'enquête. Il faut alors construire une nouvelle pondération POIDS_{2M} applicable à ce nouvel échantillon, c'est-à-dire aux nourrissons dont les familles participent **aux deux temps** d'enquête (maternité et 2 mois).



La nouvelle pondération POIDS_{2M} est à la fois "**transversale**" et "**longitudinale**". C'est-à-dire qu'elle permet de faire un portrait des nourrissons de la population Elfe au temps 1 ("transversale") et que l'on peut l'utiliser pour chercher des relations de cause à effet entre les temps 0 et 1 (longitudinale). Attention, la pondération transversale ne permet pas d'inférer sur les nourrissons ayant 2 mois en France. Elle est destinée aux nourrissons ayant 2 mois dans la population cible définie au temps 0 (par exemple, les enfants immigrés nés en 2011 n'en font pas partie).

* Pour les prochains temps d'enquête, les pondérations longitudinales et transversales ne se confondront plus.

I – CONTEXTE (Cf. protocole)

L'enquête à 6-8 semaines (2 mois) concerne les caractéristiques sociodémographiques de la famille, la place qu'y occupe l'enfant, les conditions de vie, le déroulement de la grossesse, la santé de l'enfant, son habitat, son environnement... Mères **et** pères (ou conjointe) sont interrogés à cette étape.

Si le mode de collecte est le même pour tous (téléphonique, par CATI), les questionnaires diffèrent selon la situation du couple et de l'enfant au moment de l'enquête (parents cohabitants ou non, parent détenant la garde de l'enfant, enfant placé ou non, enfant hospitalisé ou non).

Pour les pondérations, les questionnaires seront regroupés en deux catégories :

- la catégorie des questionnaires parent **réfèrent**
- la catégorie des questionnaires parent **non réfèrent**

Le questionnaire réfèrent est par définition celui comportant le plus d'informations sur l'enfant, il est très fréquemment rempli par la mère (protocole d'enquête du 2 mois) sauf si le père a la garde de l'enfant ou si la mère a demandé à ce que le père deviennent parent réfèrent**.

Le questionnaire non réfèrent complète le questionnaire réfèrent et est destiné à l'autre parent. Il s'agit à 99,99% de pères (ou conjointes), une seule mère ayant été repérée. Dans cette catégorie, deux sous-groupes :

- le parent est non réfèrent **cohabitant**
- le parent est non réfèrent **non cohabitant**

Afin de limiter la durée de l'entretien mère, une partie du questionnaire 'mère réfèrent' est posée au père (ou conjointe) **cohabitant(e)*****. Le père cohabitant répondait donc au questionnaire 'parent non réfèrent cohabitant' ainsi qu'à une partie du questionnaire 'mère réfèrent' (des variables concernant le ménage).

Cependant tous les pères cohabitants n'ont pas souhaité participer : une **reprise** a été proposée à la mère. Si le père cohabitant ne souhaitait pas participer, on recontactait la mère afin qu'elle complète la partie du questionnaire 'mère réfèrent' à laquelle le père cohabitant était normalement chargé de répondre. Cette reprise n'a pas toujours été effective (40%) entraînant une taille d'échantillon inférieure pour les **variables** de cette partie du questionnaire que nous appellerons **intriquées** (modules : *Etudes, Situation vis-à-vis du travail, Logement, Revenus du ménage, Conditions de vie*). Pour les couples non-cohabitants, ces mêmes **variables** sont renseignées par le réfèrent, et par le non réfèrent s'il a participé.

** Le questionnaire réfèrent diffère suivant qu'il s'agit d'un père ou d'une mère. Par exemple, le module *Histoire de grossesse* n'est renseigné que si l'enquêté est une mère.

*** Cet allègement ne concerne aucunement les parents non-cohabitants (dans ce cas le parent réfèrent répondait intégralement au questionnaire réfèrent).

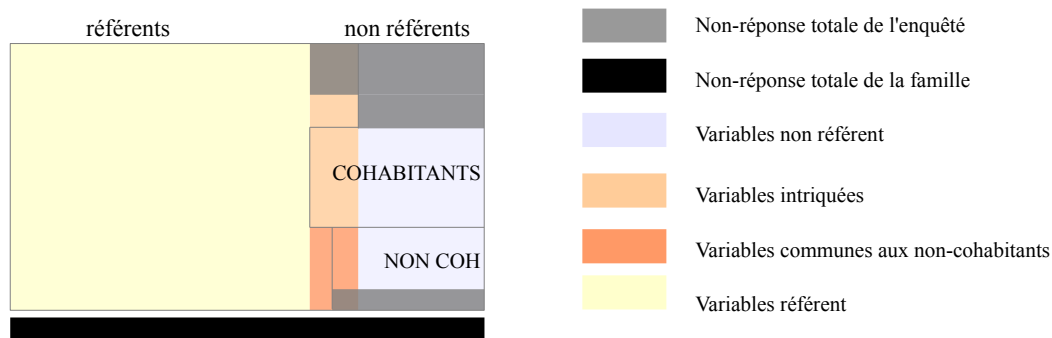


Schéma **simplifié** de la répartition des questionnaires et de la structure de la non-réponse totale (échelle non respectée)

Trois sortes de non-réponses sont à considérer :

- la **non-réponse totale** de la famille : si aucune personne du couple n'a participé
- l'**incomplétude de la famille** ou non-réponse partielle de la famille ou la non-réponse totale d'un enquêté : si l'un des parents ne participe pas
- l'**incomplétude du questionnaire** ou la non-réponse partielle de l'enquêté : si le parent décide de participer mais ne va pas jusqu'au bout du questionnaire (l'abandon en cours d'entretien, problèmes techniques, les rendez-vous pour reprise non honorés, ...). Ces questionnaires incomplets renseignent au minimum la situation familiale pour chaque individu. Les variables intriquées ne rentrent pas en compte dans cette définition de la complétude.

Questionnaires (unité=nourrisson)	Nombre de questionnaires complets	Nombre de questionnaires incomplets	Total
Variables référent (hors variables intriquées)	15872	660	16532
dont parent référent mère	15759	632	16391
dont parent référent père	113	28	141
Variables non-référent (hors variables intriquées)	12561	226	12787
dont parent non référent cohabitant	12423	211	12634
dont parent non référent non cohabitant	138	15	153
Variables intriquées + communes aux non-cohabitants*		≈ 14 500	

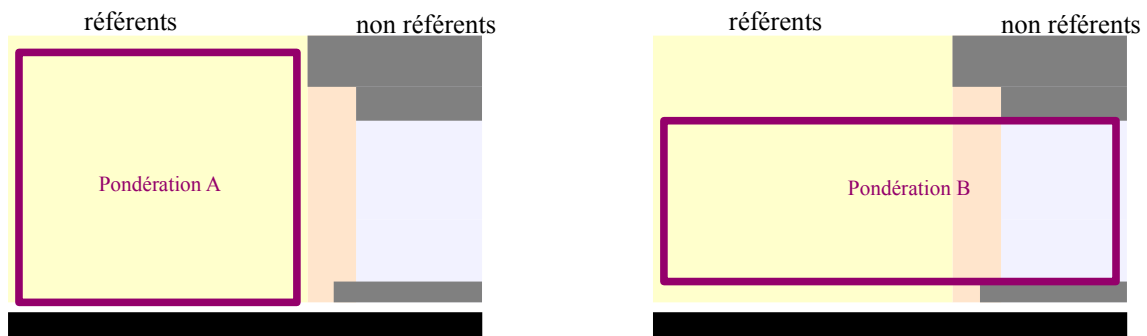
Grands groupes de variables et complétudes (mars 2014)

La première pondération poids_{2MA} présentée dans ce document est construite pour l'échantillon des nourrissons participants du 2 mois en excluant les familles dont le référent a rempli un questionnaire à statut incomplet. C'est-à-dire qu'un répondant sera défini comme un nourrisson pour lequel il existe **au moins un questionnaire complet référent**.

La seconde pondération poids_{2MB} est créée pour pouvoir aussi utiliser les variables propres au parent non référent pour lesquelles la première pondération n'est plus adéquate. Pour cette pondération, un répondant sera donc défini

♦ Au moins un des deux parents.

comme un nourrisson pour lequel il existe **un questionnaire complet non référent**.



Pourquoi avoir décidé de considérer les questionnaires incomplets comme des questionnaires non-répondants (c'est-à-dire des individus à poids nuls) ?

- ils ont des $POIDS_0$ supérieurs au poids moyen des complets (ce serait donc mettre un grand $POIDS_{2M}$ à une information partiellement vide),
- ils ne ressemblent donc pas aux complets (distribution sur quelques variables ci-dessous), mais ressemblent plutôt aux non-répondants totaux (vide) : le caractère incomplet n'est pas aléatoire. Donc les considérer comme répondants impliquerait que l'utilisateur impute les valeurs manquantes de manière non aléatoire.
- à noter : la non-réponse totale ou partielle à un temps t est liée à la non-réponse partielle au temps t-1 (« . » dans le tableau).

Complétude du questionnaire mère référent au temps 1 (mars 2014)	Complet	Incomplet	Vide
n	15759	632	1837
Répartition	86,40%	3,50%	10,10%
Répartition pondérée par $POIDS_0$	78,10%	5,50%	16,40%
$POIDS_0$ moyen	37,7	66,8	67,9
Distributions (quelques variables du temps 0)			
Mère active au moment de la grossesse			
- oui	86%	68,0%	62,2%
- non	12%	27,7%	30,9%
- .	2%	4,3%	06,9%
Nationalité de la mère			
- française	92,7%	81,0%	74,5%
- étrangère	6,3%	16,3%	21,7%
- .	1,0%	2,7%	03,8%
Présence du père à l'accouchement			
- oui	82,0%	69,6%	69,1%
- non	16,4%	26,7%	25,9%
- .	1,6%	3,6%	04,0%

Note : lorsque le questionnaire référent est incomplet, le questionnaire non-référent est vide.

Remarquons que la définition choisie de la non-réponse ne distingue aucunement les refus explicites de participer à l'enquête (temporairement ou définitivement) des autres cas (sans coordonnées téléphoniques, injoignables, rendez-vous non honorés, refus de la mère que le père participe...). Chacun des cas répond vraisemblablement à des mécanismes

différents, mais nous ferons l'hypothèse d'un seul pour les pondérations présentées ci-après.

Remarquons aussi que les demandes de sortie d'étude avec destruction des données diminuent les tailles d'échantillon de travail. Les chiffres dans ce document caractérisent un échantillon fixe.

II – LA REPONDERATION

Pour corriger le biais introduit par la non-réponse (sauf à supposer que les répondants et les non-répondants ont des caractéristiques identiques), on utilise ici la repondération.

1) Descriptif de la non-réponse

L'échantillon A est celui des parents dits référents : environ 15850.

L'échantillon B est celui des parents dits non référents: environ 12550. L'échantillon B est inclus dans l'échantillon A.

Quelques variables du temps 0 sur l'échantillon du temps 1	Part parmi les répondants au temps 0 (%)	Taux de NON- réponse pour chaque modalité de la variable considérée dans A (%)	Taux de NON- réponse pour chaque modalité de la variable considérée dans B (%)	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON- réponse pondéré par poids dans A (%)	Taux de NON- réponse pondéré par poids ₀ dans B (%)
Unité statistique : le nourrisson (mars 2014)						
Total	100% (18228)	12,90%	31,10%	100% (764000*)	21,00%	39,20%
Age mère						
- [18 ; 24]	11.9	22.2	50,3	13.9	33.0	59,5
- [25 ; 29]	30.8	13.2	30,6	31.0	20.6	37,9
- [30 ; 34]	35.0	9.8	25,0	32.9	16.9	32,4
- [35 ; 39]	17.3	10.8	28,4	16.6	17.6	35,0
- plus de 40 ans	4.2	14.5	38,0	4.6	23.0	46,4
- NR*	0.7	44.4	61,7	1.0	44.7	62,6
5 groupes de régions d'habitation à la naissance						
- Ile-de-France, Centre, Picardie	27.2	15.6	36,9	30.2	24.2	44,6
- Nord-Est	23.6	12.1	29,9	19.0	19.9	37,3
- Nord-Ouest	16.2	9.3	25,5	15.4	14.1	30,4
- Sud-Est (et Corse)	18.5	13.8	31,4	19.9	22.6	40,9
- Sud-Ouest	14.5	12.1	28,3	15.5	20.7	37,9

* 764 000 correspond à la somme des poids₀ (poids 'enfant' en maternité), c'est-à-dire la taille de la population.

* NR = non-réponse (partielle, c'est-à-dire à la question)

Quelques variables du temps 0 sur l'échantillon du temps 1	Part parmi les répondants au temps 0 (%)	Taux de NON- réponse pour chaque modalité de la variable considérée dans A (%)	Taux de NON- réponse pour chaque modalité de la variable considérée dans B (%)	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON- réponse pondéré par poids dans A (%)	Taux de NON- réponse pondéré par poids ₀ dans B (%)
Unité statistique : le nourrisson						
Groupe Social de la mère (variable brute)						
- 1 Agriculteurs, exploitants	0.3	3.6	19,6	0.3	4.2	19,0
- 2 Artisans, commerçants et chefs d'entreprise	3.1	14.0	36,0	2.9	15.2	36,4
- 3 Cadres et professions intellectuelles supérieures	15.9	6.0	18,1	11.0	6.1	18,4
- 4 Professions Intermédiaires	20.1	6.3	18,6	16.3	6.3	18,1
- 5 Employés	41.0	13.1	32,6	36.9	13.4	33,1
- 6 Ouvriers	1.7	18.5	42,9	1.8	19.1	44,2
- 7 Sans profession	4.7	26.9	53,2	8.6	46.4	68,2
- ne peut classer cette profession	11.9	24.2	50,9	20.4	42.7	64,7
- NR	1.2	36.7	55,1	1.7	47.8	62,6
Activité de la mère au moment de la grossesse						
- oui	83.1	10.1	26,5	72.3	12.9	29,5
- non	14.4	26.0	53,6	23.4	41.6	64,5
- NR	2.5	30.9	57,5	4.3	44.0	66,4
La mère vit en couple						
- oui	93.3	11.7	27,8	91.1	18.8	34,9
- non	5.4	28.9	84,1	7.5	44.2	88,8
- NR	1.3	33.9	52,6	1.5	39.9	56,5
Le père a assisté à l'accouchement						
- oui	80.3	11.1	26,3	76.3	17.4	32,6
- non	17.7	19.2	50,5	21.3	31.6	60,8
- NR	2.0	30.5	54,7	2.4	41.0	61,0
Activité du père à la naissance						
- oui	93.0	11.9	28,5	90.5	19.0	35,8
- non	3.5	21.9	52,7	4.8	33.2	61,5
- NR	3.5	32.0	78,9	4.7	46.3	83,2
Vacances d'au moins 4 jours pendant la grossesse						
- non	45.5	16.9	24,1	49.7	27.0	48,1
- oui	52.8	9.0	38,6	48.2	14.0	29,3
- NR	1.7	29.6	51,3	2.1	37.6	57,1
Séances de préparation à la naissance						
- oui	54.2	7.6	21,3	47.7	10.6	24,9
- non	43.7	18.6	42,4	49.7	29.7	51,9
- NR	2.1	32.5	52,5	2.6	44.3	60,3
Nationalité et statut immigré du couple						
- les 2 parents sont français	77.6	10.1	24,3	71.7	15.4	29,7
- Mère de nationalité étrangère et/ou immigrée, père français	4.7	17.6	40,7	5.9	27.9	50,8
- Père de nationalité étrangère et/ou immigré, mère française	6.0	16.9	49,9	6.2	26.3	57,6
- Les 2 parents de nationalité étrangère et/ou immigrés	6.1	31.9	65,6	9.5	44.5	71,8
- NR (pour au moins un des 2)	5,5	25,3	60,2	6,6	38,4	68,8

Pour traiter la non-réponse, nous disposons d'une information commune aux répondants et aux non-répondants. Pour cette analyse au temps 1, nous avons toute l'information issue du questionnaire rempli au temps 0 (commun aux répondants et aux non-répondants du temps 1 puisqu'ils étaient tous répondants au temps 0). Le tableau ci-dessus présente les taux de non-réponse (non-pondérés, puis pondérés par poids₀) de quelques variables choisies pour l'ajustement dans les pondérations.

Le test du chi-deux rejette l'hypothèse d'indépendance entre la variable de réponse et chacune des variables ci-dessus sur chacun des échantillons.

La modalité "NR" (non-réponse partielle) est intéressante et importante : intéressante car les taux de non-réponse totale à 2 mois sont beaucoup plus grands lorsque la mère n'avait pas répondu (NR) à la question posée en maternité. Importante car, pour chacune des variables d'ajustement choisies, cette modalité sera prise en compte dans le calcul de pondération.

Les distributions d'autres variables se trouvent en annexe 1. Pour une analyse beaucoup plus complète de la non-réponse à différents niveaux (maternité, département, région) et aux temps 0, 1 et 2, on pourra explorer la synthèse de S. Quemeneur : *La dimension géographique de la non-réponse*, 2013.

L'information commune aux répondants et non-répondants étant grande, il faut choisir quelles variables serviront à expliquer le comportement de non-participation au temps 1. En théorie ces variables doivent être corrélées avec la non-réponse mais aussi avec les variables d'intérêt de l'enquête.

2) Pondération en fonction de la non-réponse

L'arbre CHAID (arbre de décision non binaire pour des variables catégorielles ou continues) a été utilisé afin de visualiser rapidement quelles sont les variables qui caractérisent le plus la non-réponse. On observe en premières positions les variables *Groupe Social de la mère, Avoir suivi des séances de préparation à la naissance, Statut immigré des parents, Age mère, Vie en couple de la mère ...* variables qui se révèlent aussi significatives avec la régression. Ces dernières variables seront intégrées dans le modèle de régression logistique. L'annexe 2 présente 3 ajustements différents (ajout de variables 'santé'), le modèle 2 étant celui retenu pour cette pondération.

L'échantillon B est un sous-échantillon de A : après avoir contacté la mère, si possibilité et respect du protocole, on contactait le père. On peut donc considérer que l'échantillon B connaît la même phase de non-réponse que l'échantillon A à laquelle s'ajoute une seconde phase de non-réponse. En anticipant les accroissements du CV (Coefficient de Variation) des poids et de l'étendue des poids, l'ajustement sur l'échantillon B ne sera pas fait en 2 étapes (2 phases) mais en une seule : les variables de redressement seront quasiment les mêmes que celles utilisées pour la pondération A : sont ajoutées les variables *Père a assisté à l'accouchement* et *Activité du père au moment de l'accouchement*.

La méthode retenue pour traiter la non-réponse totale est celle des groupes de réponses homogènes créés à partir des probabilités estimées par la régression logistique. A partir des scores triés résultant de cette régression, 40 groupes de réponses homogènes ont été créés et pondérés par la pondération du temps 0.

$$poids_{2M} = poids_0 \times \text{coefficient d'ajustement de la non-réponse}$$

Ainsi, les répondants au temps 1 voient leurs poids augmenter afin de compenser la non-réponse au temps 1. Afin de réduire la dispersion des poids, les poids extrêmes qui se détachaient de la distribution ont été tronqués. Cela a pour effet de créer du biais (ce que l'on essaie de limiter en tronquant le moins possible) mais de diminuer la variance et l'étendue des poids.

poids ₀	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	11	18	20	75	114	200	18	764000	42	32	78

Statistiques du poids final au temps 0

P2M= poids ₀ x Coefficient d'ajustement de la non-réponse	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
P2M_A tronqué	11	19	21	74	103	250	18	700000	44	32	72
P2M_B tronqué	15	22	24	97	141	250	17	670000	54	42	79

Statistiques du poids ajusté de la non-réponse avant calage

3) Calage

Nous utilisons les mêmes variables de calage que celles utilisées pour la pondération au temps 0.

A noter : ces sources sont relatives à l'année 2011, tout comme les renseignements issus de l'échantillon (« à la naissance de l'enfant »). Ce calage à 2 mois est donc cohérent. Par ailleurs, on pourrait l'utiliser pour n'importe quel temps d'enquête t : cela voudrait dire que l'on cherche une cohérence entre notre échantillon au temps t (dont on connaît des caractéristiques de 2011) et des informations auxiliaires datant de 2011.

Variables de calage Distribution sur l'échantillon A des répondants du temps 1	Distribution sans aucune pondération	Distribution pondérée par poids ₀	Distribution pondérée par poids _{2MA} avant calage	Source
<i>âge de la mère à la naissance de l'enfant</i>				Etat civil (filtre sur métropole et mères majeures)
- [18, 24]	10,7	11,9	14,3	- 13,96%
- [25, 29]	30,9	31,4	31,9	- 31,22%
- [30, 34]	36,5	34,8	32,2	- 33,25%
- [35, 39]	17,8	17,4	18,8	- 16,90%
- + de 40 ans	04,1	04,5	04,8	- 04,67%
<i>groupe de régions de domicile à la naissance</i>				Etat civil (filtre sur métropole)
- Ile de France/Centre/Picardie	26,4	29,0	31,6	- 29,96%
- Nord-Est	23,8	19,2	18,4	- 19,15%
- Nord-Ouest	16,9	16,8	15,4	- 15,42%
- Sud-Est	18,3	19,5	19,9	- 19,03%
- Sud-Ouest	14,7	15,5	14,7	- 15,54%
<i>statut mère immigrée à la naissance de l'enfant</i>				Etat civil (filtre sur métropole)
- née en France	88,7	85,5	82,3	- 81,25%
- immigrée	11,3	14,5	17,7	- 18,75%
<i>état matrimonial de la mère à la naissance</i>				Etat civil (filtre sur métropole)
- né dans le mariage	46,8	45,7	45,5	- 45%
- né hors mariage	53,2	54,3	54,5	- 55%
<i>mère primipare à la naissance de l'enfant</i>				ENP (champ ELFE)
- oui	45,8	44,5	43,8	- 43,1%
- non	54,2	55,5	56,2	- 56,9%
<i>niveau d'étude de la mère à la naissance de l'enfant</i>				ENP (champ ELFE)
- non scolarisée/école primaire/collège/CAP ou BEP	16,5	22,1	24,8	- 27,79%
- 2nde/1ère/terminale générale, technique ou pro	19,8	18,6	20,2	- 19,88%
- études supérieures	63,7	59,2	55,0	- 52,33%

Les données manquantes (à très petits taux) pour ces variables ont été imputées par hotdeck aléatoire (avec tirage proportionnel au poids₀) afin que ces individus ne soient exclus du calage.

Le calage assure donc des distributions de l'échantillon pondéré ELFE identiques à celles de la colonne 'Source'. En pratique, la **troncature** des poids après calage (poids très grands ramenés à 250) modifie ces distributions (**biais**). La méthode du raking ratio est utilisée pour le calage, celle du logit tronqué ayant du mal à converger et n'empêchant pas la troncature qui suit.

poids _{2M}	n	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
A : après calage et troncature	15855	13	20	22	86	118	252	18	764000	48	37	77
B : après calage et troncature	12554	14	21	24	119	185	259	18	764000	61	52	85

Statistiques du poids final après calage et troncature

III – UTILISATION DES POIDS ?

Les variables du questionnaire face à face au temps 0 peuvent être pondérées par la pondération du temps 0, mais aussi par les pondérations A ou B.

Variables d'intérêt	Échantillon du temps 0 sans pondération	Échantillon du temps 0 avec pondération poids ₀	Échantillon A au temps 1 sans pondération	Échantillon A au temps 1 avec pondération poids _{2MA}	Échantillon B au temps 1 sans pondération	Échantillon B au temps 1 avec pondération poids _{2MB}	Source de comparaison
							État civil 2011
Nationalité de la mère							
- française	- 91,7 %	- 87,1 %	- 93,2 %	- 89,0%	- 94,9 %	- 88,78%	- 86,6%
- étrangère	- 08,3 %	- 12,9 %	- 06,8 %	- 11,0%	- 05,1 %	- 11,2 %	- 13,4%
							ENP 2010 (échantillon)
Accouchement							
- travail spontané	- 70,8 %	- 70,0 %	- 71,1 %	- 70,6 %	- 71,6 %	- 70,1 %	- 66,6 %
- déclenchement	- 19,5 %	- 19,8 %	- 19,2 %	- 19,94%	- 18,8 %	- 19,8%	- 22,4 %
- césarienne	- 09,7 %	- 10,2 %	- 09,7 %	- 10,0 %	- 09,6 %	- 10,1 %	- 10,9 %

Utilisation des pondérations A et B sur des variables renseignées au temps 0

Au temps 1, pour des variables renseignées par le parent référent, on utilisera la pondération A :

%	Votre enfant a-t-il été vacciné pour le BCG ?			%	Bénéficiez-vous d'une couverture maladie complémentaire ?		
	oui	non	ne sait pas		oui	non	ne sait pas
non pondéré	25,1	73,6	1,2	94,2	5,6	0,2	
pondéré	29,1	69,3	1,5	90,4	9,2	0,4	

Utilisation transversale de la pondération A sur l'échantillon des **parents référents** (père ou mère)

Si l'échantillon sur lequel on souhaite travailler n'est pas très différent de l'échantillon A (B), la pondération A (B) pourra aussi être utilisée. Sinon, il faudra étudier les biais potentiels.

On peut donc utiliser la poids A pour les modules *Santé générale de la mère et antécédents familiaux* ou *Alimentation de l'enfant depuis la naissance* (questions posées uniquement aux mères, c'est-à-dire 99% de l'échantillon A) :

%		Tabagisme de la mère pendant la grossesse		Alimentation à la naissance		
		non	oui			
Tabagisme de la mère à 2 mois	oui	4,39	16,09	Lait maternel	Lait 1er âge	mixte
		4,31	18,63			
non	non	75,99	3,53	Lait 1er âge	29,9	4,8
		73,66	3,40			
Alimentation à 2 mois	mixte	5,9	0,5	mixte	0,5	1,2
		5,6	0,6			

Utilisation longitudinale de la pondération A sur l'échantillon des **mères** (référentes ou non)

La pondération A ne correspond pas exactement à une pondération 'mère' (il existe une centaine de pères référents) et l'échantillon B, à une pondération 'père' (1 mère non-cohabitante).

Dans les faits, si on s'intéresse aux déclarations du père, on pourra utiliser la pondération B sur des variables renseignées par le non-référent (à 99,99% des pères pour cette enquête) et on perdra l'information sur la centaine de pères référents (qui ne sont pas associés à une mère non référente et ne font donc pas partie de l'échantillon B).

Car utiliser la pondération A pour ces variables renseignées par le non-référent, ce serait utiliser un échantillon A avec certes tous les non-référents participants (puisque B inclus dans A), mais avec un pourcentage non négligeable de données manquantes pour lesquelles la pondération est positive (données manquantes due à la non participation de ces pères non-référents).

%	Tabagisme du père à 2 mois		Fréquence manquante
	oui	non	
Non pondéré sur l'échantillon A	35,23	64,77	3200
Poids A	37,95	62,05	195500
Non pondéré sur l'échantillon B	35,18	64,82	0
Poids B	39,47	60,53	0

Utilisation transversale de la pondération B sur l'échantillon des **pères non référents***

On pourra comparer des variables mère et père avec la pondération B :

%		Tabagisme du père à 2 mois	
		oui	non
Tabagisme de la mère à 2 mois	oui	13,7	5,0
		16,7	5,4
	non	21,5	59,8
		22,7	55,1

Utilisation transversale de la pondération B sur l'échantillon des **mères avec pères non référents**

Note : il se peut que la base de données qui vous est livrée contienne moins d'individus que ceux dénombrés dans ce document (les enquêtés ont la possibilité de demander la destruction de leurs données, le phénomène est supposé rare et l'impact sur la pondération, minime).

* En réalité, il y a 1 mère non référente dans l'échantillon B, donc 1 père référent.

Annexe 1

Quelques variables du temps 0 sur l'échantillon A du temps 1 Unité statistique : le nourrisson	Part parmi les répondants au temps 0 (%)	Taux de NON-réponse (%)	Test du Chi-deux (p-value)*	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON-réponse pondéré (%)
Total	100% (18228)	12.9%		100% (764000)	21,00%
MATERNITES					
Vague			0.7782		
-1	15.3	12.4		15.6	20.2
-2	25.3	13.2		25.4	21.6
-3	28.3	13.0		28.1	21.1
-4	31.2	12.9		30.9	20.7
Strate			0.1211		
- 1	2.5	14.0		6.6	16.8
- 2	7.0	11.3		10.7	15.0
- 3	13.0	12.4		16.8	19.8
- 4	27.0	12.5		24.7	22.6
- 5	50.5	13.5		41.2	22.7
Statut juridique			0.0001		
- privé non lucratif	9.7	9.7		8.5	16.2
- privé lucratif	24.5	12.9		25.9	19.7
- public	65.8	13.4		65.6	22.1
Autorisation			0.0057		
- niveau 1	22.9	11.8		30.4	16.2
- niveau 2	47.4	12.8		45.7	19.7
- niveau 3	29.6	14.0		23.9	22.1
ACCEPTATION TRANSMISSION DONNÉES					
Acceptation transmission données assurance maladie mère			<.0001		
- oui	94.8	12.0		94.3	19.7
- non	5.2	29.1		5.7	42.6
Acceptation transmission données sur l'enfant			<.0001		
- oui	93.4	11.8		92.8	19.3
- non	6.6	28.7		7.2	42.1
MERE					
Age mère			<.0001		
- [18 ; 24]	11.9	22.2		13.9	33.0
- [25 ; 29]	30.8	13.2		31.0	20.6
- [30 ; 34]	35.0	9.8		32.9	16.9
- [35 ; 39]	17.3	10.8		16.6	17.6
- plus de 40 ans	4.2	14.5		4.6	23.0
- NR**	0.7	44.4		1.0	44.7
5 groupes de régions d'habitation à la naissance			<.0001		
- Ile-de-France, Centre, Picardie	27.2	15.6		30.2	24.2
- Nord-Est	23.6	12.1		19.0	19.9
- Nord-Ouest	16.2	9.3		15.4	14.1
- Sud-Est (et Corse)	18.5	13.8		19.9	22.6
- Sud-Ouest	14.5	12.1		15.5	20.7
Groupe Social (variable brute)			<.0001		
- 1 Agriculteurs, exploitants	0.3	3.6		0.3	4.2
- 2 Artisans, commerçants et chefs d'entreprise	3.1	14.0		2.9	15.2
- 3 Cadres et professions intellectuelles supérieures	15.9	6.0		11.0	6.1
- 4 Professions Intermédiaires	20.1	6.3		16.3	6.3
- 5 Employés	41.0	13.1		36.9	13.4
- 6 Ouvriers	1.7	18.5		1.8	19.1
- 7 Sans profession	4.7	26.9		8.6	46.4
- ne peut classer cette profession	11.9	24.2		20.4	42.7
- NR	1.2	36.7		1.7	47.8

* Test du khi-deux sans la modalité de non-réponse NR (avec cette modalité le test est significatif pour toutes les variables).

** NR = non-réponse (partielle, c'est-à-dire à la question)

Quelques variables du temps 0 sur l'échantillon A du temps 1 Unité statistique : le nourrisson	Part parmi les répondants au temps 0 (%)	Taux de NON-réponse (%)	Test du Chi-deux (p-value)	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON-réponse pondéré (%)
niveau d'étude de la mère - non scolarisée/école primaire/collège/CAP ou BEP - 2nde/1ère/terminale générale, technique ou professionnelle - études supérieures - NR	18.7 20.8 59.7 0.7	23.5 17.6 7.6 45.5	<.0001	27.3 19.8 51.9 1.0	36.2 26.3 10.5 46.1
Activité au moment de la grossesse - oui - non - NR	83.1 14.4 2.5	10.1 26.0 30.9	<.0001	72.3 23.4 4.3	12.9 41.6 44.0
Est primipare - non - oui - NR	53,7 45,1 1.2	13,1 12,4 34.5	0.1659	56,0 42,4 1.5	22,6 18,4 38.3
statut mère immigrée - née en France - immigrée - NR	85.5 13.0 1.5	10.8 24.9 31.9	<.0001	79.9 18.1 2.0	16.8 37.4 40.1
Nationalité de la mère - française (par naissance ou par acquisition) - étrangère ou apatride - NR	90.5 8.2 1.3	11.2 28.3 34.0	<.0001	85.6 12.7 1.7	17.6 41.1 41.7
état matrimonial - né dans le mariage - né hors mariage 2 - NR	44.9 52.4 2.7	11.4 13.7 26.4	<.0001	43.7 53.3 3.0	19.2 21.6 35.2
La mère vit en couple - oui - non - NR	93.3 5.4 1.3	11.7 28.9 33.9	<.0001	91.1 7.5 1.5	18.8 44.2 39.9
Bénéficiez-vous d'une couverture maladie complémentaire ? - oui - non - NR	88,8 7,7 3,5	11,1 26,8 31,1	<.0001	84,7 10,5 4,8	17,1 42,5 44,8
PERE					
Age père - moins de 24 ans - [25 ; 29] - [30 ; 34] - [35 ; 39] - plus de 40 ans - NR	5.5 21.6 32.8 21.6 13.3 5.0	22.1 13.1 10.0 10.4 12.9 31.6	<.0001	6.2 21.3 30.5 20.7 14.5 6.8	31.5 19.9 15.7 17.2 22.2 47.0
Groupe Social (variable brute) - 1 Agriculteurs, exploitants - 2 Artisans, commerçants et chefs d'entreprise - 3 Cadres et professions intellectuelles supérieures - 4 Professions Intermédiaires - 5 Employés - 6 Ouvriers - 7 Sans profession - ne peut classer cette profession - NR	1.5 8.5 20.2 11.9 35.3 12.6 1.7 5.5 2.7	7.9 15.2 6.2 7.2 13.3 16.8 23.6 25.0 29.5	<.0001	1.5 8.5 16.4 10.4 35.2 14.3 2.3 7.5 3.6	11.4 21.9 8.3 9.9 20.1 27.8 37.9 40.3 41.8
statut père immigré - né en France - immigré - NR	82.0 13.6 4.4	10.4 23.9 27.0	<.0001	77.2 17.5 5.3	16.1 36.6 40.1
Nationalité du père - français (par naissance ou par acquisition) - étranger ou ne sait pas - NR	87.7 8.6 3.8	10.9 26.9 28.0	<.0001	83.8 11.6 4.6	17.3 39.3 41.5

Quelques variables du temps 0 sur l'échantillon A du temps 1 Unité statistique : le nourrisson	Part parmi les répondants au temps 0 (%)	Taux de NON-réponse (%)	Test du Chi-deux (p-value)	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON-réponse pondéré (%)
Le père a assisté à l'accouchement			<.0001		
- oui	80.3	11.1		76.3	17.4
- non	17.7	19.2		21.3	31.6
- NR	2.0	30.5		2.4	41.0
Activité du père à la naissance			<.0001		
- oui	93.0	11.9		90.5	19.0
- non	3.5	21.9		4.8	33.2
- NR	3.5	32.0		4.7	46.3
GROSSESSE – ACCOUCHEMENT					
Naissance			0.0433		
- unique	96.5	12.9		96.5	21.0
- multiple	3.2	10.0		3.3	16.8
- NR	0.3	61.9		0.2	59.8
Enfant transféré			0.5819		
- oui	5.1	13.4		5.1	23.5
- non	94.5	12.7		94.6	20.7
- NR	0.4	61.4		0.3	61.7
Age gestationnel			0.5703		
- [33 ; 37]	11.6	13.3		11.9	22.7
- [38 ; 40]	68.8	12.7		69.6	21.0
- plus de 40 semaines	17.7	12.3		17.3	19.0
- NR	1.8	26.0		1.2	27.3
Réaction à la découverte de la grossesse			<.0001		
- heureuse que cette grossesse arrive maintenant	74.8	12.6		73.5	20.1
- aimé qu'elle arrive plus tôt	11.6	8.0		10.8	13.0
- aimé qu'elle arrive plus tard	9.5	15.5		10.6	25.9
- préféré ne pas être enceinte	2.4	20.2		3.1	35.4
- NR	1.6	37.1		2.0	46.7
Difficultés psychologiques persistantes pendant la grossesse			0.8963		
- oui	86.3	12.6		12.5	20.4
- non	12.4	12.7		85.85	20.7
- NR	1.4	33.9		1.7	40.7
Tabagisme avant la grossesse			0.0327		
- non	56.9	13.1		57.2	21.5
- oui	41.9	12.1		41.3	19.4
- NR	1.2	36.3		1.5	43.7
Tabagisme pendant la grossesse			<.0001		
- non	78.7	12.0		76.8	19.4
- oui	19.8	15.0		21.2	24.8
- NR	1.5	33.9		1.9	42.1
Statut pondéral de la mère (IMC avant grossesse)			<.0001		
- < 18,5	7.7	14.7		7.6	21.8
- [18.5, 25]	63.3	11.5		60.8	18.3
- [25, 30]	17.1	14.1		18.0	23.1
- > 30	9.8	14.4		10.9	24.3
- NR	2.0	35.9		2.7	50.9
Vacances d'au moins 4 jours pendant la grossesse			<.0001		
- non	45.5	16.9		49.7	27.0
- oui	52.8	9.0		48.2	14.0
- NR	1.7	29.6		2.1	37.6
Première alimentation de l'enfant à la naissance			<.0001		
- au lait maternel uniquement	59.5	10.1		57.1	16.9
- au lait premier âge uniquement	31.1	15.9		32.0	24.7
- allaitement mixte	8.0	19.2		9.1	30.1
- ne sait pas ou autre	0.6	11.6		0.6	21.9
- NR	0.8	43.7		1.1	44.6

Quelques variables du temps 0 sur l'échantillon A du temps 1 Unité statistique : le nourrisson	Part parmi les répondants au temps 0 (%)	Taux de NON-réponse (%)	Test du Chi-deux (p-value)	Part pondérée parmi les répondants au temps 0 (%)	Taux de NON-réponse pondéré (%)
Hypertension - non - HTA chronique - HTA pendant une grossesse antérieure uniquement - NR	95.3 1.0 1.3 2.4	12.7 13.8 11.1 21.5	0.7792	95.4 1.1 1.4 2.1	20.8 24.3 21.2 24.5
Hypertension artérielle apparue pendant la grossesse - non - oui avec protéinurie - oui sans protéinurie - NR	93.5 1.4 2.0 3.1	12.7 13.7 11.9 19.8	0.8150	93.5 1.6 2.1 2.7	20.8 25.9 21.4 23.4
Diabète - non - type 1 - type 2 - diabète gestationnel (grossesses antérieures) - NR	93.4 0.5 0.5 2.6 3.1	12.6 19.3 10.8 15.2 20.7	0.0543	93.3 0.5 0.5 2.8 2.8	20.6 35.2 23.0 24.3 25.1
Diabète gestationnel - non - oui - NR	87.6 7.1 5.3	12.6 14.1 17.8	0.0550	87.7 7.4 5.0	20.5 24.1 25.0
Amniocentèse - non - oui - inconnu - ne sait pas - NR	89.2 3.6 5.2 0.9 1.0	12.6 12.1 15.7 15.8 32.5	0.0242	89.2 4.0 5.5 0.9 0.7	20.7 18.5 25.0 24.8 33.4
Déclaration de la grossesse - grossesse non déclarée - au 1er trimestre - au 2ème trimestre - au 3ème trimestre - NR	0.7 93.2 4.0 0.8 1.3	16.3 12.1 22.1 27.1 36.7	<.0001	0.8 91.3 5.1 1.1 1.7	25.7 19.4 34.7 46.2 46.0
Séances de préparation à la naissance - oui - non - NR	54.2 43.7 2.1	7.6 18.6 32.5	<.0001	47.7 49.7 2.6	10.6 29.7 44.3
Accouchement : début du travail - travail spontané - déclenchement - césarienne avant début du travail - NR	69.3 19.1 9.5 2.1	12.4 14.1 12.3 24.2	0.2015	68.8 19.4 10.0 1.8	20.4 22.4 20.6 29.5
Type d'accouchement - Voie basse spontanée - Forceps, spatules, ventouses - Césarienne - Ne sait pas - NR	66.6 12.2 18.2 1.9 1.0	12.7 12.8 12.9 13.8 32.5	0.2407	66.9 11.5 18.9 1.9 0.7	20.8 19.9 21.6 21.7 33.4

Annexe 2

Pondération du 2 mois : phase d'ajustement de la non-réponse

Comparaison entre 3 ajustements différents (ajout de variables 'santé') sur des variables des temps 0 et 1.

Attention, il s'agit des distributions avant troncature et calage.

Modèle 1 : Age mère, Groupe de régions, PCS mère, Vie en couple mère, nationalité/immigré du couple

Modèle 2 : {modèle 1} + Séances préparation à la naissance + Avoir pris des vacances pendant la grossesse

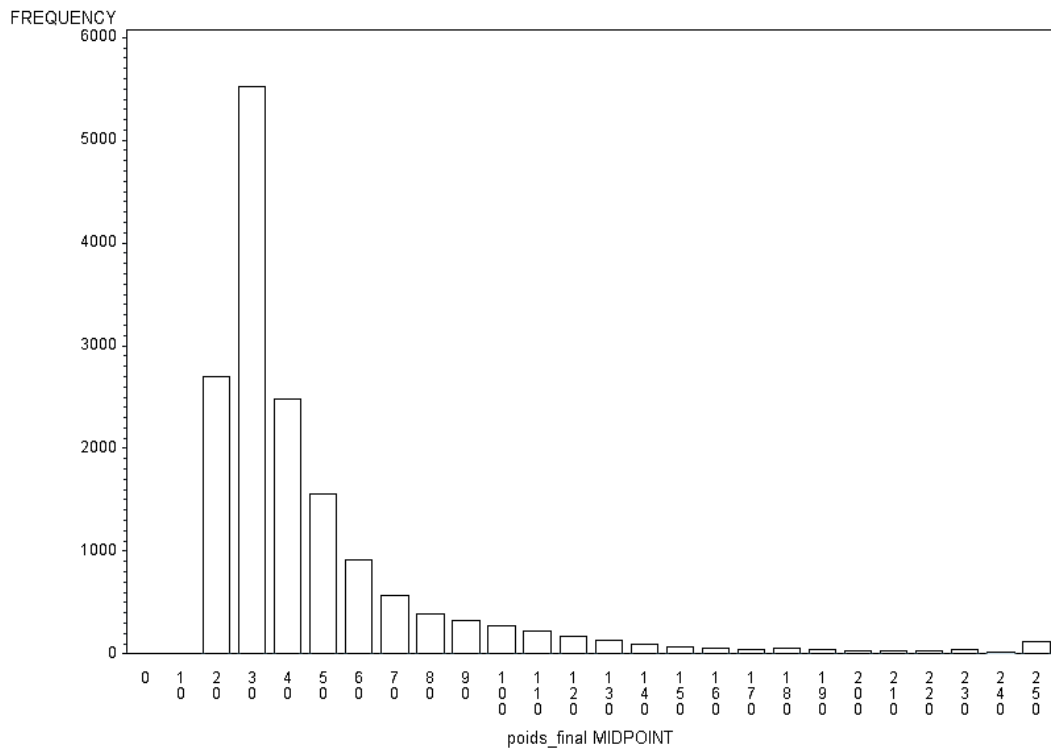
Modèle 3 : {modèle 2} + IMC + Tabagisme pendant la grossesse + Accouchement + Alimentation à la naissance

Échantillon A distribution % (écart relatif %)	Non pondéré	Poids 1	Poids 2	Poids 3	Source (ENP 2010)
Nationalité mère (temps 0)					
française	93,3 (07,7)	86,8 (00,2)	87,4 (00,9)	87,7 (01,3)	86,6
étrangère	06,7 (50,0)	13,2 (01,5)	12,6 (06,0)	12,5 (06,7)	13,4
Lieu naissance mère (temps 0)					
-France	88,7 (09,2)	80,4 (01,0)	81,3 (00,1)	81,6 (00,5)	81,2
-étranger	11,3 (39,9)	19,6 (04,3)	18,7 (00,5)	18,4 (02,0)	18,8
Niveau d'étude de la mère					
- non scolarisée / école primaire /collège / CAP ou BEP	16,5 (40,6)	24,6 (11,5)	25,3 (09,0)	25,3 (09,0)	27,8
- 2nde/1ère/terminale générale, technique ou professionnelle	19,8 (00,5)	20,2 (01,5)	20,1 (01,0)	20,1 (01,0)	19,9
- études supérieures	63,7 (21,8)	55,2 (05,5)	54,5 (04,2)	54,6 (04,4)	52,3
Séances Préparation					
oui	58,5 (20,9)	52,3 (08,1)	51,6 (06,6)	48,6 (00,4)	48,4
non	41,5 (19,6)	47,7 (07,6)	48,4 (06,2)	51,4 (00,4)	51,6
Diabète gestationnel					
- non	92,6	92,2	92,3	92,4	92,6
- oui	07,4	07,8	07,7	07,6	07,4
Type d'accouchement					
- Voie basse spontanée	71,1	70,1	69,9	69,5	66,6
- Forceps, spatules, ventouses	19,2	19,6	19,8	20,6	22,4
- Césarienne	09,7	10,2	10,3	09,9	10,9
Statut pondéral de la mère (IMC avant grossesse)					
- < 18,5	07,7	07,7	07,7	08,3	08,2
- [18,5, 25]	65,4	62,9	62,6	62,7	64,6
- [25, 30]	17,2	18,4	18,5	18,5	17,3
- > 30	09,8	11,0	11,2	10,5	9,9
Première alimentation de l'enfant à la naissance					
- au lait maternel uniquement	62,0 (03,7)	60,7 (01,5)	59,7 (00,2)	56,5 (05,5)	59,8
- au lait premier âge uniquement	30,5 (01,3)	29,9 (03,2)	31,1 (00,6)	34,0 (10,0)	30,9
- allaitement mixte	07,5 (09,6)	09,3 (12,0)	09,2 (10,8)	09,5 (14,5)	08,3
Alimentation 2 mois					
- lait maternel	32,0	30,3	29,9	28,7	-
- 1er âge	60,4	61,6	62,3	63,7	-
- mixte	07,6	08,1	07,8	07,6	-
Vaccination BCG					
- oui	25,1	29,9	29,6	29,2	-
- non	73,6	68,5	68,8	69,1	-
- ne sait pas	01,2	01,6	01,6	01,6	-
Allergie au lait de vache					
- oui	01,4	01,5	01,5	01,5	-
- non	98,6	98,5	98,5	98,5	-

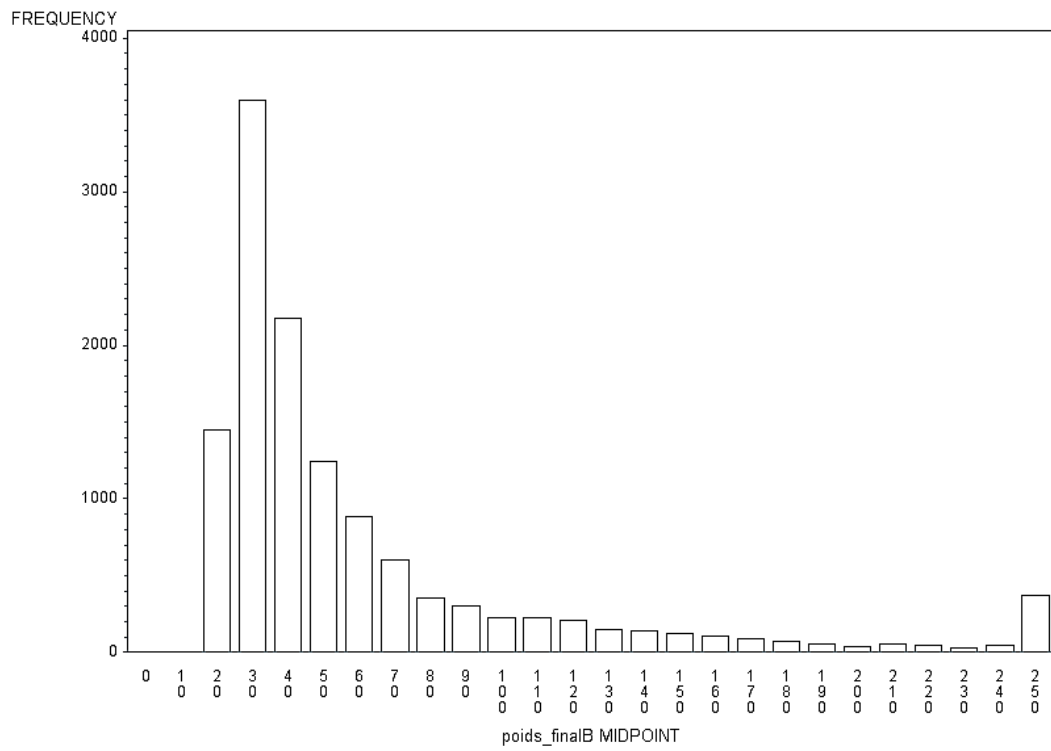
La mère a été allaitée					
- oui	48,7	51,1	50,6	49,8	-
- non	48,5	45,7	46,3	46,9	-
- ne sait pas	02,7	03,2	03,2	03,2	-
Personnes handicapées dans le ménage					
- oui	01,9	02,3	02,2	02,3	-
- non	98,1	97,7	97,8	97,7	-
Santé de l'enfant					
- en bonne santé	87,1	87,8	87,8	87,7	-
- plutôt en bonne santé	11,9	11,0	11,1	11,1	-
- plutôt en mauvaise santé	00,6	00,7	00,7	00,7	-
- en mauvaise santé	00,3	00,4	00,4	00,4	-
Tabagisme mère pendant la grossesse					
- non	80,4	79,5	78,8	78,8	-
- oui	19,6	20,5	21,2	21,2	-
Tabagisme mère à 2 mois					
- non	79,6	78,6	77,8	77,6	-
- oui	20,4	21,4	22,2	22,4	-

Annexe 3

Poids final A



Poids final B



On remarquera sur chaque graphe, une dernière barre se détachant de la distribution : c'est le résultat de la troncature. La quantité de 'grands poids' croît à chaque nouvelle enquête (à valeur de troncature fixée, ici 250).

Bibliographie :

Davezies, L., d'Haultfoeuille, X., 2009. Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête. Document de travail de l'INSEE, n° G 2009/06.

Kalton G, Flores-Cervantes I., 2003, Weighting Methods. *Journal of Official Statistics* ;19:81-97.

Kreuter, F., Olson, K., 2011, Multiple Auxiliary Variables in Nonresponse Adjustment. *Sociological Methods Research*, 40(2), 311-332.

Little, R. J., Vartivarian, S., 2003, On weighting the rates in non-response weights. *Statistics in Medicine* 22(9): 1589-1599.

Quemeneur, S., 2013, *La dimension géographique de la non-réponse*, Elfe.

Wun, L.-M., T. Ezzati-Rice, 2007, Assessment of the Impact of Health Variables on Nonresponse Adjustment in the Medical Expenditure Panel Survey (MEPS). *Proceedings of the Survey Research Methods Section*, ASA.

