

# **Pondération de l'enquête ELFE en maternité**



# SOMMAIRE

NOTE.....	1
NOTE SUPPLEMENTAIRE.....	2
INTRODUCTION.....	3
PLAN DE SONDAGE.....	5
PHASE 1 : POIDS 1, les maternités.....	6
1) Pondération des maternités incluses	
2) Description de la non-réponse	
3) Repondération des maternités en fonction de la non-réponse	
PHASE 2 : POIDS 2, le temps.....	10
Pondération pour représenter l'année 2011 et repondération en fonction de la non-réponse a certaines vagues	
PHASE 3 : POIDS 3, les nourrissons.....	13
1) Tirage exhaustif des nourrissons dans les maternités incluses	
2) Analyse de la non-réponse totale	
a) Définition de la non-réponse	
b) Traitement de la non-réponse partielle en vue de la construction de la pondération	
c) Description de la non-réponse	
3) Repondération des nourrissons participants dans les maternités participantes aux jours elfe	
POIDS DES NOURRISSONS : P1P2P3.....	21
1) Analyse du poids et troncature	
2) Estimation de l'effectif de la population à partir de l'état civil et de l'ENP	
CALAGE .....	24
1) Sources	
2) Choix des variables de calage	
3) Calage et troncature	
4) Tester la pondération sur quelques variables d'intérêt	
ANNEXES.....	31
BIBLIOGRAPHIE.....	34

## NOTE

L'enquête Elfe (Etude longitudinale française depuis l'enfance), a pour objectif de recueillir des informations sur la totalité des nouveau-nés d'un échantillon aléatoire de 349 maternités en France métropolitaine, sur 25 jours répartis tout au long de l'année 2011 en 4 temps d'enquête.

L'objectif de ce document est de fournir une pondération transversale de l'échantillon des participants constitué d'environ 18 300 nouveau-nés, permettant d'inférer à la population.

L'échantillon a été construit selon un **plan de sondage complexe** : il existe un degré relatif aux maternités et un autre degré relatif aux jours ; au croisement de ces deux degrés, tous les nourrissons nés dans les maternités et les jours sélectionnés sont recrutés. Les maternités tirées au sort sont issues d'un plan stratifié avec allocations proportionnelles à leurs tailles et les jours ont été choisis de façon à couvrir les quatre saisons de l'année 2011. Attention, ce plan ne peut être considéré comme un plan à deux degrés classique.

Après avoir pris en compte les poids initiaux dus au plan de sondage, les poids seront ajustés de la **non-réponse** qui s'observe à différents niveaux : lors des enquêtes en **maternité**, une partie des maternités n'a pas participé, de même qu'une partie des mères ayant accouché les jours d'enquête. On discernera deux types de non-réponse chez les maternités : celle des maternités n'ayant pas du tout participé et celles ayant participé partiellement (non-réponse sur les **jours** de participation). On a en notre possession les données de variables communes aux maternités participantes et à celles n'ayant pas participé. On dispose aussi, pour les **familles** non participantes, de plusieurs informations, communes à celles fournies directement par les mères volontaires en maternité. La méthode de repondération retenue sera la méthode des scores, méthode faisant appel à des groupes homogènes de réponse.

En second lieu, un **calage** sera effectué sur les marges de l'état civil et de l'ENP (enquête nationale périnatale), permettant à notre échantillon pondéré d'être cohérent géographiquement, socio-démographiquement et sur la situation familiale. La pondération finale, afin de l'apprécier, sera testée sur quelques variables d'intérêt.

Ce document présente uniquement la méthodologie utilisée pour la construction de la **pondération** au **premier temps** de l'enquête Elfe. L'estimation de la **variance** issue du plan de sondage utilisé pour l'enquête Elfe fait l'objet d'une autre note proposée en 2015 aux utilisateurs.

## NOTE SUPPLEMENTAIRE

- Deux pondérations sont proposées pour ce premier temps d'enquête : celle au niveau **enfant** (une ligne pour chaque jumeau) et celle au niveau **famille**. Seule la pondération au niveau enfant est illustrée dans ce document (celle au niveau famille suivant la même méthodologie).
- La pondération enfant M00E\_PONDVAL est valable sur la base livrée (une ligne = un enfant).  
La pondération famille M00F\_PONDVAL est valable sur une base au niveau famille (pour cela, utiliser la base enfant en supprimant une ligne de jumeau sur deux : si M00M1\_RANGALEA=1 alors supprimer).
- Certains individus n'ont pas de pondération (« . »), ceci est expliqué est dans la suite de ce document.
- La somme des poids proposés à l'utilisateur permet de retomber sur la taille de la population (764 000 pour la pondération au niveau enfant, 753 500 pour la pondération au niveau famille). En pratique, l'utilisateur peut normer ses poids en multipliant chaque poids par ce ratio : taille de l'échantillon avec pondération / taille de la population. Ainsi en sommant les poids, l'utilisateur retombera sur la taille de son échantillon.  
Par exemple, pour le niveau enfant :  $M00E\_PONDVALNORME = M00E\_PONDVAL \times (18\,207 / 764\,000)$ .

# INTRODUCTION

## ELFE, première étude française de l'enfance sur grande cohorte

L'unité mixte INED-INSERM-EFS ELFE a été créée en avril 2010. L'objectif de l'enquête ELFE est de comprendre comment les conditions périnatales et l'environnement affectent le développement, la santé et la socialisation des enfants. L'environnement familial, socioéconomique, physico-chimique et les interactions complexes entre ces différents facteurs seront pris en compte.



Ined  
Inserm  
InVS  
Insee  
DGS  
DGPR  
Drees  
Cnaf

Plus de **18 000 enfants** nés en 2011 en France métropolitaine constituent la cohorte ELFE (soit 1/43ème des naissances françaises). Ils seront suivis de la naissance à l'adolescence au travers de plusieurs entretiens, questionnaires et prélèvements. C'est un projet pluridisciplinaire impliquant une soixantaine d'équipes de recherche dans les domaines des sciences sociales, de la santé ou de l'environnement.

Quelle est la **population cible** ? Il s'agit de tous les nourrissons nés en 2011 en métropole à l'exception des grands prématurés (enquête Epipage 2) ; ont été inclus (**éligibles**) dans l'enquête les nourrissons nés durant certains jours sélectionnés (25 jours), dans certaines maternités tirées aléatoirement (349), sous certains critères : nés à partir de 33 semaines d'aménorrhées, nés au plus d'une grossesse gémellaire, mères non mineures, en mesure de comprendre les implications principales de la participation à l'étude notamment dans les langues proposées et résidant non temporairement en métropole.

## La pondération au temps 0

L'enquête ELFE est une enquête longitudinale dans laquelle l'unité statistique est l'enfant.

Une pondération, si elle est utilisée dans les analyses, doit permettre d'obtenir des résultats généralisables à l'ensemble de toute la population (et pas seulement à l'échantillon). Elle consiste à assigner à chacun des 18 300 nourrissons enquêtés, un poids statistique qui correspond au nombre d'enfants qu'il représente dans la population visée.

Le plan de sondage comporte plusieurs phases, celle des **maternités**, celles **des jours** ou vagues de participation et celle des **nourrissons**.

Pour représenter chaque saison, quatre périodes de l'année ont été sélectionnées : du 1er avril au 4 avril, du 27 juin au 4 juillet, du 27 septembre au 4 octobre et enfin du 28 novembre au 5 décembre : **25 jours\*** au total. Ces périodes ont été appelées « vague 1 », « vague 2 », « vague 3 » et « vague 4 » et chaque maternité parmi les maternités sélectionnées (sauf 4) sera invitée à participer à chacune de ces vagues. Certaines ne répondront à **aucune** : il s'agit de la première non-réponse totale traitée dans cette étude. Ensuite, la population ciblée étant celle de toute l'année 2011, il faut trouver

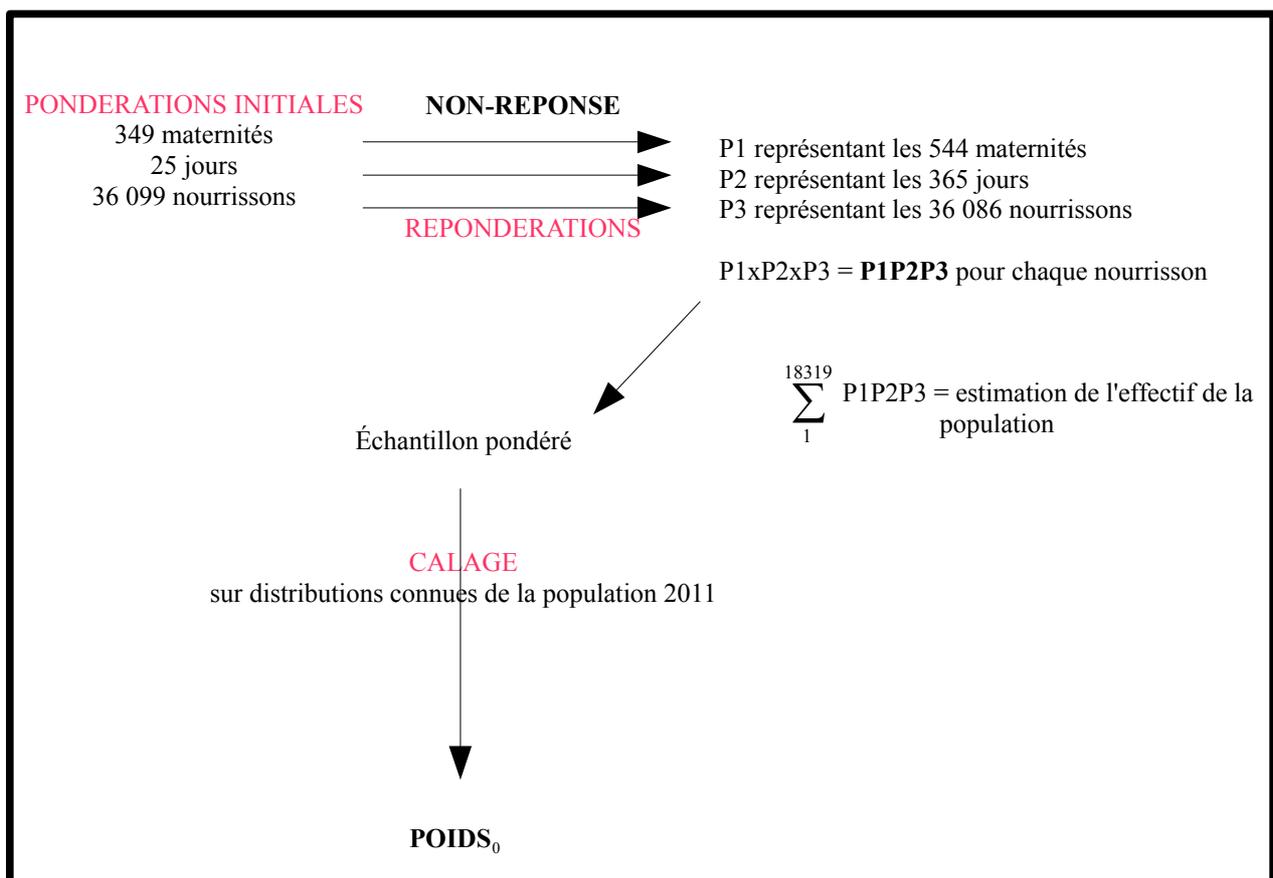
\* 1, 2, 3, 4 avril, 27, 28 juin, 1, 2, 3, 4 juillet, 27, 28, 29 septembre, 1, 2, 3, 4 octobre, 28, 29, 30 novembre, 1, 2, 3, 4, 5 décembre

un moyen de **représenter les 365 jours** de l'année (par exemple, par un coefficient de 365/25). Or il existe une non-réponse dans la participation par vague des maternités : certaines maternités n'ont pas participé à toutes les vagues mais **seulement à une, deux ou trois vagues**. La dernière non-réponse totale est relative aux mères abordées en maternité (peu de temps après l'accouchement) : 49 % d'entre elles n'ont **pas participé** à l'enquête. Pour ces dernières, quelques informations ont été récoltées dans un « dossier refus » et serviront à la repondération des participantes. Les participantes, quant à elles, ont répondu à un questionnaire en face-à-face portant sur le déroulement de leur grossesse, de leur accouchement, sur leur situation familiale.

Les **poids corrigés** de la non-réponse à chacune des phases s'appelleront respectivement **P1, P2 et P3**. Chacune de ces corrections a été testée sous différentes méthodes qui supposent un mécanisme de non-réponse ignorable.

La multiplication des poids corrigés, notée **P1P2P3**, sera testée sur des variables de contrôle (d'intérêt) (définies par l'équipe) avant et après calage.

Les sources externes de ce dernier sont l'ENP 2010 (enquête nationale périnatale) et l'état-civil. Les nouveaux poids qu'il engendrera seront notés **POIDS<sub>0</sub>**.



## PLAN DE SONDAGE

Le plan de sondage des maternités a été construit par Nicolas Razafindratsima (Ined) et Hélène Sarter (InVS).

Il s'agit d'un plan aléatoire stratifié.

La base de sondage des maternités est constituée de la liste des maternités (publiques et privées) de France métropolitaine en **2008** : 542 maternités ont été répertoriées. Le plan de sondage est soumis à des contraintes budgétaires et logistiques : il existe une forte dispersion des maternités selon leurs tailles (nombre de naissances). Afin d'obtenir l'effectif visé de naissances sous ces contraintes, il faut sur-représenter les grandes maternités.

La variable de stratification est la **taille de la maternité**. Le tirage est systématique avec pour variables de stratification implicite le statut juridique, le niveau\* et la région en 5 postes (base de sondage triée préalablement au hasard puis selon les variables précédentes au sein de chaque strate explicite). La stratification implicite assure d'une certaine manière un échantillon proportionnel de maternités entre toutes les strates implicites.

**Finalement, 5 strates ont été construites à effectifs égaux avec allocation proportionnelle au nombre de naissances et selon un tirage systématique avec stratification implicite sur statut juridique, niveau et groupes de région.**

**Les 25 jours d'enquête (4, 6, 7 et 8 jours) ne sont pas issus d'un tirage aléatoire pour raisons logistiques mais sont choisis de manière à couvrir les 4 saisons et à ce que 12 des jours coïncident avec ceux de l'EDP (échantillon démographique permanent). Les mêmes 25 jours ont été associés à chaque maternité.**

**Les nourrissons de chaque maternité tirée sont tous enquêtés sous condition d'éligibilité.**

Pour une description plus précise du plan de sondage résultant de l'assemblage de ces trois phases (maternité, jour, nourrisson), voir la documentation concernant l'estimation de la variance issue du plan de sondage Elfe.

---

\* Il existe 3 niveaux de maternité : le niveau 1 pour les grossesses sans complications, le niveau 2 qui comporte une unité de néonatalogie et le niveau 3 disposant d'un service de réanimation néonatale. Les maternités de niveau 3 appartiennent toutes aux strates 4 et 5 (grandes maternités).

## PHASE 1 : POIDS 1, les maternités

### 1) Pondération des maternités incluses

Par allocations par strate proportionnelles au nombre de naissances (probabilités inégales), **342** maternités parmi **542** sont incluses dans l'enquête ELFE.

STRATES h	Nombre d'accouchements par maternité en 2008	Taille des strates $N_h$	Taille de l'échantillon $n_h$	Probabilité d'inclusion $\pi_h$	poids-initial $\frac{1}{\pi_h}$
1	[145, 699]	108	26	0.24	4,15
2	[700, 1009]	108	46	0.42	2,35
3	[1010, 1418]	109	65	0.59	1,68
4	[1422, 2187]	108	96	0.88	1,12
5	[2197, 5215]	109	109	1	1
<b>Total</b>		<b>542</b>	<b>342</b>		<b>542</b>

Tableau 1 : Poids initiaux issus de la stratification

$$n_h = 342 \frac{\text{Nombre de naissances en 2008 dans la strate } h}{\text{Nombre total de naissances en 2008}}, \quad \pi_h = \frac{n_h}{N_h}$$

Parmi les 342 maternités tirées, 4 n'ont pas été invitées à participer à l'étude ELFE. De plus, aux 342 maternités tirées ont été ajoutées 7 maternités. Parmi ces 7 maternités, deux ont fusionné (juridiquement) avec deux autres tirées dans la base de sondage, une qui devait fermer et qui faisait partie de la strate 5, une de strate 5 qui n'existait pas en 2008, deux maternités de niveau 3 (les deux seules de niveau 3 n'ayant pas été tirées) et enfin une maternité mystère. Les 4 qui n'ont pas été invitées à participer (et dont on ne bouge pas les poids) passeront automatiquement en non-réponse dans la partie suivante même s'il ne s'agit pas d'un refus à proprement dit.

Parmi les 7 maternités incluses, 2 maternités de strate 5 ne faisaient pas partie de la base de sondage, et auront un poids égal à 1 quelles que soient leurs caractéristiques, leurs probabilités d'inclusion étant totales (elles devraient être considérées comme faisant partie d'une strate supplémentaire mais dans les tableaux suivants elles seront insérées dans la strate 5). Les 5 autres maternités rajoutées faisaient partie de la base de sondage initiale. Pour les ajouter, on a considéré le plan de sondage **comme** en 2 temps. C'est-à-dire que dans un premier temps, on a tiré les 342 maternités et dans un second temps, les 5 maternités. Les nouveaux poids ont été calculé ainsi :

$$\pi_h' = P(\text{tirage}_1) + P(\text{tirage}_2 \text{ et } \neg \text{tirage}_1) \\ = \pi_h + P(\text{tirage}_2 | \neg \text{tirage}_1) P(\neg \text{tirage}_1) \quad \text{où } h=[1;5] \text{ l'indice de la strate d'appartenance}$$

Cette façon de faire suppose un tirage aléatoire au 2<sup>nd</sup> temps (ce qui est faux).

STRATES h	Taille des strates $N_h$	Taille de l'échantillon $n_h$	Probabilité d'inclusion $\pi_h$	pois-initial $\frac{1}{\pi_h}$
1	108	28	0.26	3,86
2	108	47	0.44	2,3
3	109	66	0.60	1,65
4	108	97	0.90	1,1
5	111	111	1	1
<b>Total</b>	<b>544</b>	<b>349</b>		<b>544</b>

Tableau 2 : Poids initiaux finaux

On a donc un effectif de maternités égal à **544** et un échantillon de **349** maternités.

Parmi ces 349 maternités, toutes n'ont pas participé : les participantes vont être repondérées en fonction des non-participantes.

Poids initial issu du plan de sondage $\frac{1}{\pi_h}$	Méthode	Min	Max	Max/min	Somme des poids	Moyenne	écart-type	CV
$\frac{1}{\pi_h}$	Allocation proportionnelle à la taille	1	3,86	3,86	544	1,56	0,81	51,90%

Tableau 3 : Statistiques du poids initial final

## 2) Description de la non-réponse

Parmi les 349 maternités de l'échantillon, 25 n'ont participé à aucune vague. De plus, 4 maternités tirées au sort n'ont finalement pas été invitées. Ce qui fait donc un total de 29 maternités non-participantes à prendre en compte dans cette phase de non-réponse.

Nous disposons de 4 variables pour caractériser les maternités participantes et les non-participantes : la taille de la maternité (sa strate d'appartenance), sa région, son niveau de médicalisation (autorisation) et son statut juridique.

Dans le tableau ci-dessous, on remarque de forts taux de non-réponses dans la région Ile-de-France (ou son groupe de régions), au niveau de médicalisation 2 et dans les maternités privées non-lucratives.

L'hypothèse d'indépendance est rejetée au seuil de 10 % pour la variable renseignant le *Statut juridique* de la maternité, et à 5 % pour la variable *Groupe de régions*.

<b>Variables caractérisant les maternités</b>	Nombre de maternités	Nombre de maternités qui n'ont participé à aucune vague (NON-réponse totale)	Taux de NON-réponse (en %)	Test d'indépendance du Khi-deux ou test exact de Fisher (p-value)
<b>Total</b>	<b>349</b>	<b>29</b>	<b>8,31%</b>	
<b>Taille</b> (nombre de naissances en 2008)				0.8989
[145, 699]	28	3	10,7%	
[700, 1009]	47	3	6,4%	
[1010, 1418]	66	4	5,9%	
[1422, 2187]	97	9	9,3%	
[2197, 5215]	111	10	9,0%	
<b>Région</b>				Effectifs par classe trop petits
Île-de-France	77	15	19,5%	
Champagne-Ardenne	7	0	0%	
Picardie	11	0	0%	
Haute Normandie	8	0	0%	
Centre	13	2	15,4%	
Basse Normandie	8	0	0%	
Bourgogne	10	0	0%	
Nord pas de Calais	24	1	4,2%	
Lorraine	11	0	0%	
Alsace	12	0	0%	
Franche-Comté	6	0	0%	
Pays de la Loire	15	0	0%	
Bretagne	21	2	9,5%	
Poitou Charentes	11	1	9,1%	
Aquitaine	15	0	0%	
Midi-Pyrénées	13	1	7,7%	
Limousin	2	0	0%	
Rhône-Alpes	37	6	16,2%	
Auvergne	3	0	0%	
Languedoc-Roussillon	16	0	0%	
PACA	28	1	3,6%	
Corse	1	0	0%	
<b>Groupe de régions</b>				2.703e-13
Île-de-France, Centre, Picardie	101	17	16,8%	
Sud-Est	69	7	10,1%	
Autres	179	5	2,8%	
<b>Autorisation</b>				0.1889
niveau 1	125	11	8,8%	
niveau 2	161	16	9,9%	
niveau 3	63	2	3,2%	
<b>Statut juridique</b>				0.0969
privé non lucratif	30	5	16,7%	
privé lucratif	95	9	9,5%	
public	224	15	6,7%	

Tableau 4 : Caractéristiques des maternités non-participantes

### 3) Repondération des maternités en fonction de la non-réponse

Il faut maintenant compenser la non-réponse de ces 29 maternités en rehaussant le poids des 320 maternités répondantes. Pour cela il faut déterminer la méthode et les variables à utiliser.

$P1 = \text{poids\_initial} * \text{coefficient d'ajustement}$ $P1 = 0$	pour les maternités participantes (à au moins 1 vague) pour les maternités non participantes (à aucune vague)
---	--

Plusieurs corrections ont été envisagées. On peut observer dans le tableau précédent des **différences** dans les taux de non-réponse, suivant les strates, les niveaux, les régions... Pour prendre en compte plusieurs de ces caractéristiques, la méthode des croisements et celle des scores ont été testées. Les maternités répondantes à l'intérieur de chaque groupe de réponse homogène sont repondérées.

La méthode des scores avec taux de réponse pondérés  $\hat{p}$  (par les poids initiaux relevant du plan de sondage) a été retenue : 10 groupes à effectifs égaux et toutes les variables utilisées (*Strate, Autorisation, Statut juridique, Groupe de régions*).

Coefficient d'ajustement de la non-réponse	Méthode	Min	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
$\frac{1}{\hat{p}}$	Score pondéré	1	1,27	1,27	349,47	1,09	0,1	9,46

Tableau 5 : Statistiques du coefficient d'ajustement de la non-réponse

P1= poids initial * coefficient d'ajustement	Méthode	Min	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
$\frac{1}{\pi_h} \frac{1}{\hat{p}}$	Score pondéré	1,02	4,09	4	544	1,7	0,93	52,6

Tableau 6 : Statistiques du poids initial corrigé de la non-réponse

La somme des poids P1 par méthode pondérée fait bien 544 (nombre total de maternités).

Pour cette première phase, on a donc une variabilité introduite d'abord par le plan de sondage stratifié (avec probabilités d'inclusion inégales) et ensuite par la non-participation des 29 maternités.

## PHASE 2 : POIDS 2, le temps

Il s'agit de la phase la plus compliquée. On se limite aux **320** maternités ayant participé à **au moins** une vague et non aux 349 maternités puisque la non-réponse totale des 29 maternités vient d'être traitée précédemment (phase 1).

**P2 = poids\_temps**

Il a été décidé 25 jours d'enquête, relatifs à 4 vagues (saisons).

Les jours ne sont pas tirés au hasard (contrainte logistique, synchronisation avec 12 jours de l'EDP) mais on pourrait faire l'hypothèse qu'ils sont issus d'un tirage aléatoire et donc qu'il s'agit d'une autre phase de tirage. Remarquons qu'il s'agit d'un second degré dans lequel il n'y a pas d'indépendance entre les 320 tirages puisqu'ils décernent les mêmes 25 jours à chaque unité primaire, la maternité.

<b>Participation 320 maternités</b>	Nombre de maternités		Taux de <b>NON-participation</b> partielle (en %)	Taux de participation (en %)	
À la vague 1	305		4,68	95,3	
À la vague 2	312		2,5	97,5	
À la vague 3	311		2,8	97,2	
À la vague 4	309		3,4	96,6	
<b>Total</b>	349	320		100	100
À 0 vague (NON- réponse totale)	29			8,31	
À 1 seule vague	4	4		1,16	1,25
À 2 vagues	6	6		1,74	1,87
À 3 vagues	19	19		5,51	5,94
À 4 vagues	291	291		84,35	90,94

Tableau 7 : Participation des maternités aux vagues et en nombre de vagues

Au total, sur les  $320 \times 25 = 8000$  jours d'enquête attendus chez les maternités participantes, 7741 ont été effectués (96,76%).

L'effet **journalier** des naissances : on sait que les accouchements sont moins nombreux en week-end qu'en semaine. Une étude menée par Xavier Thierry a montré que l'impact des jours de la semaine à l'intérieur de chaque combinaison de vague s'auto-annulait ou du moins était négligeable.

L'effet **saisonnier** des naissances : il existe, même s'il s'est beaucoup amoindri durant ces dernières années plus particulièrement avec la diffusion de la contraception depuis les années 60. Dans cette étude, il a juste été vérifié que les jours d'enquête ne tombaient pas sur un jour « spécial ». En France, comme dans d'autres pays, le pic des naissances est le 23 septembre (deux fois plus de conceptions à la Saint-Sylvestre qu'en temps normal) et ce jour ne fait pas partie des jours ELFE.

## **Pondération pour représenter l'année 2011 et repondération en fonction de la non-réponse a certaines vagues**

Plusieurs possibilités ont été envisagées :

- Pondérer chaque maternité  $i$  en fonction du **nombre de jours** de participation : faire l'hypothèse que les  $x$  jours de participation de chaque maternité (entre 4 et 25 jours, suivant les vagues auxquelles elles ont participé) représentent en **nombre de naissances** les 365 jours de l'année 2011 :

$$P2_{A_i} = \frac{365}{25} \frac{25}{x_i}$$

Cette pondération repose sur une hypothèse forte, surtout pour les maternités ayant participé à peu de vagues.

Si on considère le tirage des jours comme aléatoire, il s'agirait ici d'un tirage aléatoire simple (même probabilité pour chacun des jours).

Note : les maternités sont ouvertes toute l'année, pas besoin de prendre en compte les jours d'ouverture dans le calcul.

Cette possibilité suggère donc que chaque maternité s'auto-repondère en fonction de sa non-réponse à certaines vagues (elle repondère les naissances des vagues auxquelles elle participe).

- Pondérer chaque maternité en fonction du **nombre de naissances** durant les jours de participation (permet de se défaire de l'hypothèse précédente) :

$$P2 = \frac{\text{NB total de naissances dans la maternité durant l'année 2011}}{\text{NB de naissances dans la maternité durant les } x_i \text{ jours 2011}}$$

Poids différents pour chaque maternité. Cette pondération n'est aujourd'hui pas possible, l'information du nombre de naissances en 2011 par maternité n'étant pas disponible à ce jour.

- Pondérer par vague en fonction du nombre de jours : permet de prendre en compte l'aspect saisonnier, et la non-réponse des maternités par vague.

$$\begin{aligned}
 P2_{B,1} &= \frac{\text{NB total de maternités participant à au moins une vague}}{\text{NB de maternités participant à la vague 1}} \frac{90}{4} \\
 P2_{B,2} &= \frac{\text{NB total de maternités participant à au moins une vague}}{\text{NB de maternités participant à la vague 2}} \frac{91}{6} \\
 P2_{B,3} &= \frac{\text{NB total de maternités participant à au moins une vague}}{\text{NB de maternités participant à la vague 3}} \frac{92}{7} \\
 P2_{B,4} &= \frac{\text{NB total de maternités participant à au moins une vague}}{\text{NB de maternités participant à la vague 4}} \frac{92}{8}
 \end{aligned}$$

On peut l'imaginer comme un tirage aléatoire au sein de 4 strates (trimestres) avec l'hypothèse que les tirages des jours sont indépendants d'une strate à l'autre. Ce qui est faux, les jours de chaque vague ont été choisis, de façon regroupée et de façon à ce qu'ils ne soient pas voisins entre vagues.

- Pondérer par vague et par taille de maternité (strate) : c'est-à-dire comme ci-dessus mais spécifiquement à chaque strate, donc 4x5 possibilités.

$$\begin{aligned}
 P2_{C,1,strate1} &= \frac{\text{NB total de maternités participant à au moins une vague dans la strate 1}}{\text{NB de maternités participant à la vague 1 dans la strate 1}} \frac{90}{4} \\
 P2_{C,1,strate2} &= \frac{\text{NB total de maternités participant à au moins une vague dans la strate 2}}{\text{NB de maternités participant à la vague 1 dans la strate 2}} \frac{90}{4} \\
 &\dots
 \end{aligned}$$

La non-réponse d'une grande maternité (beaucoup de naissances) sera alors compensée par une grande maternité répondante et vice versa, au sein de chaque vague.

C'est cette dernière méthode qui a été choisie, permettant de prendre en compte l'aspect saisonnier dans la pondération.

## PHASE 3 : POIDS 3, les nourrissons

La phase 3 concerne celle des nourrissons : ils sont inclus exhaustivement sous critères d'éligibilité.

On observera plus de 49% de non-réponse à ce degré : participer à cette enquête, c'est-à-dire intégrer une cohorte pour 20 ans, est perçu comme un engagement fort.

### 1) Tirage exhaustif des nourrissons dans les maternités incluses

Tous les nourrissons éligibles des maternités incluses nés durant les jours ELFE sont inclus avec la **même** probabilité (ils ont tous un poids de base égal à 1).

Les mères souhaitant participer à l'enquête ont répondu à un questionnaire en face-à-face. Plusieurs informations ont pu être récoltées pour les mères non-répondantes au travers d'un « dossier refus ». Ces variables sont communes aux mères répondantes et non-répondantes : c'est ce qui permettra d'effectuer une pondération en fonction de la non-réponse prenant en compte les caractéristiques des mères.

Attention, les mères de jumeaux n'ont rempli qu'un questionnaire. L'unité statistique longitudinale étant le nourrisson, il faut doubler l'information mère dans la base servant à l'étude pour la pondération. Ceci, que la mère soit répondante ou non-répondante.

Taille de l'échantillon :  $n = 36\,099$  nourrissons dont 1242 jumeaux  
Taille du sous-échantillon des répondants :  $n_r = 18\,329$  nourrissons dont 576 jumeaux  
Taille du sous-échantillon des non-répondants :  $n_m = 17\,770$  nourrissons dont 666 jumeaux

Pour commencer, on observe 71 nourrissons nés hors des 25 jours d'inclusion : ils sont considérés hors champ dès le temps 0.

Taille de l'échantillon :  $n = 36\,028$  nourrissons dont 1240 jumeaux  
Taille du sous-échantillon des répondants :  $n_r = 18\,258$  nourrissons dont 574 jumeaux  
Taille du sous-échantillon des non-répondants :  $n_m = 17\,770$  nourrissons dont 666 jumeaux

Le poids de ce degré est calculé ainsi :

$$\mathbf{P3 = 1 * Coefficient d'ajustement * coefficient\_couverture}$$
$$\mathbf{= 0}$$

pour les nourrissons participants

pour les nourrissons non participants

### Coefficient de couverture

Il existe un défaut de couverture : certaines mères éligibles n'ont pas été approchées. Dans les faits, il était parfois impossible pour les enquêteurs d'aborder toutes les mères lorsqu'il y avait plusieurs naissances en même temps ou lorsque la mère quittait trop tôt la maternité.

On parle alors de **sous-couverture**, des individus de la population cible étant absents de la base de sondage. Or, le nombre de naissances éligibles par maternité est connu (travail effectué par Xavier Thierry), il a été récolté en salle d'accouchement.

Afin de corriger ce défaut, un coefficient a été calculé par région (nombre de nourrissons éligibles / nombres de nourrissons enquêtés). On affecte donc ce coefficient, légèrement supérieur à 1, à chaque nourrisson afin de rectifier l'erreur de sous-couverture :

Région	Mères éligibles	Mères abordées	Coefficient couverture (=éligibles/abordés)
Alsace 42	1276	1232	1.0357142857
Aquitaine 72	1597	1551	1.029658285
Auvergne 83	452	443	1.0203160271
Basse Normandie 25	794	785	1.0114649682
Bourgogne 26	1062	1061	1.0009425071
Bretagne 53	1897	1870	1.0144385027
Centre 24	1333	1292	1.0317337461
Champagne-Ardenne 21	778	770	1.0103896104
Corse 94	50	44	1.1363636364
Franche-Comté 43	665	668	1
Haute Normandie 23	1175	1162	1.0111876076
Ile-de-France 11	8642	7896	<b>1.0944782168</b>
Languedoc-Roussillon 91	1765	1719	1.026759744
Limousin 74	240	240	1
Lorraine 41	1205	1196	1.0075250836
Midi-Pyrénées 73	1020	972	1.049382716
Nord pas de Calais 31	2896	2822	1.0262225372
PACA 93	3014	2938	1.0258679374
Pays de la Loire 52	2231	2148	1.0386405959
Picardie 22	1112	1133	1
Poitou Charentes 54	842	825	1.0206060606
Rhône-Alpes 82	3484	3324	1.0481347774

Tableau 9 : Sous-couverture par région

Ce coefficient est en moyenne égal à 1,04%. Ce qui peut paraître petit mais représente environ 30 000 nourrissons sur la population de référence (sur toute l'année et toute la France métropolitaine) .

## 2) Analyse de la non-réponse totale

En moyenne, 50,8% des mères abordées ont accepté de participer à l'enquête. Ce qui correspond à 50,7% des nourrissons qui participent à l'enquête (puisque certaines mères ont des jumeaux).

Le taux de non-participation (49,3%) ne se répartit pas uniformément, nous allons regarder plus en détail quelles sont les caractéristiques individuelles des mères qui l'impacte.

Nous avons 7 variables communes aux répondantes et aux non-répondantes (en plus des 4 variables caractérisant la maternité d'accouchement).

- *Année de naissance* de la mère (par translation il s'agit de son âge) qui a été découpée en 6 modalités : [18 ; 22], [23 ; 24], [25 ; 29], [30 ; 34], [35 ; 39], plus de 40 ans
- *Age gestationnel* (en semaines) découpé en 3 modalités : [33 ; 37], [38 ; 40], plus de 40 semaines (les nourrissons de moins de 33 semaines sont appelés 'grands prématurés' et ne font pas partie du champ de l'enquête, ceux de moins de 37 semaines sont appelés 'prématurés')
- *Département d'habitation* de la mère regroupé par région, puis par groupe de régions : Ile-de-France, Centre, Picardie, Nord-Est, Nord-Ouest, Sud-Est, Sud-Ouest

- *PCS* (professions et catégories socioprofessionnelles) inspirée de la nomenclature en 8 postes :

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
9	Ne peut classer la profession

- *Activité au moment de la grossesse* : oui ou non
- *Indicatrice gémellaire* : a eu des jumeaux ou naissance unique (les triplets et autres ne faisant pas partie du champ de l'enquête)
- *Primiparité* (fait d'être pour la première fois parent) : oui ou non

### a) Définition de la non-réponse

Les mères ayant refusé de participer à l'enquête font partie de la non-réponse totale. Il y a une autre partie de la non-réponse à considérer : les mères ayant accepté de participer à l'enquête mais n'ayant **quasiment répondu à aucune question ou trop partiellement**.

Le questionnaire étant long, il faut donc sélectionner des variables 'importantes' pour lesquelles une trop grande non-réponse à ces questions entraînera un statut de non-répondante pour la mère. Or, le poids d'une non-répondante vaut zéro. L'étude étant longitudinale et les poids futurs dépendant du poids initial (au temps 0, c'est-à-dire en maternité), ceci équivaut à affecter un poids nul à ces mères pour tous les autres temps de l'enquête. Ce qui est difficilement envisageable.

Une vérification a été faite sur les variables *Indicatrice gémellaire, Activité, Primiparité, Age, Age gestationnel, Région d'habitation, CSP*. Il existe 128 mères pour lesquelles les valeurs sont manquantes pour 5 ou 6 de ces variables (sauf la région d'habitation qui est obligatoire).

Cette trop grande réponse partielle sur des variables importantes rend les unités inexploitable dans les calculs.

Cependant il paraît difficile d'attribuer un poids nul à ces unités longitudinales (au cas où ces mères participeraient au autres vagues).

A ce moment, il convient de préciser certaines évolutions dans la cohorte : les individus enquêtés ont la possibilité, à n'importe quel moment, de se retirer de l'étude et s'ils le désirent, de demander la destruction de leurs données. Une pondération étant adaptée à un échantillon fixe de répondants, il a été décidé que chaque pondération serait construite en ôtant de l'échantillon des répondants les individus ayant demandé la suppression de leurs données au moment du calcul de la pondération (par conséquent, les chiffres présentés dans les documents relatifs aux pondérations ne refléteront pas la participation au temps t de l'enquête, mais respecteront une définition précise de la non-réponse statistique).

Au 7 février 2014, 30 individus avaient demandé la suppression de leurs données depuis l'enquête en maternité.

Au 20 avril 2015, 51 individus avaient demandé la suppression de leurs données depuis l'enquête en maternité.

Taille de l'échantillon :  $n = 36\,028$  nourrissons dont 1240 jumeaux

Taille du sous-échantillon des répondants :  $n_r = 18\,207$  nourrissons dont 572 jumeaux

Taille du sous-échantillon des non-répondants :  $n_m = 17\,821$  nourrissons dont 668 jumeaux

Définition d'un répondant pour la pondération : individu ayant participé à l'enquête, respectant les critères de la population cible, respectant le plan de sondage (donc né dans l'une des maternités tirées au sort et né l'un des 25 jours) et n'ayant pas demandé la destruction de ses données au moment de la construction de la pondération.

## b) Traitement de la non-réponse partielle en vue de la construction de la pondération

La non-réponse partielle des variables communes aux répondantes et aux non-répondantes pose problème pour la repondération de la non-réponse totale. Dans notre cas, on peut soit la passer en modalité, soit l'imputer.

Pour l'ajustement en fonction de la non-réponse totale, la non-réponse partielle est passée en modalité (en considérant qu'un non-répondant à donnée manquante peut être représenté par un répondant à donnée manquante). Notons que la variable PCS a été corrigée à partir de variables des temps 0 et 1. Pour le calage futur, nous récupérons l'information identique issue de l'enquête à 2 mois disponible pour certaines variables.

L'enquête à 2 mois reprend certaines des questions posées à la mère en maternité. Parmi les variables auxquelles nous nous intéressons, par exemple, l'année de naissance de la mère est redemandée à 2 mois. Cela permet donc de résoudre **en partie** le problème de la non-réponse partielle au niveau des participantes.

Il s'agit d'une façon d'imputer les données manquantes en utilisant la **même** information récoltée au temps t+1. Si l'information n'a pas pu changer dans l'intervalle de temps, il semble sensé de l'utiliser (par exemple, l'indicatrice gémellaire). Cependant si l'information est relative à un état (ou une opinion) susceptible de changer avec le temps, la substitution par une donnée de l'enquête à 2 mois est problématique. La *CSP* constitue un caractère changeant. Néanmoins on peut raisonnablement penser que, aux erreurs de déclarations près, la situation professionnelle de la mère ne varie guère entre la date de l'accouchement et la fin du congé de maternité.

La variable recodée *PCS* est accessible aux chercheurs.

Il a été vérifié que les formulations des questions entre t et t+1 n'avaient pas changé.

Pour les variables *Indicatrice gémellaire*, *Age gestationnel* et *Primiparité*, l'information n'existe pas à 2 mois.

Pour la variable *Département habitation* de la mère, l'information des valeurs manquantes a été inférée à partir du département d'accouchement (maternité) pour les répondants et les non-répondants.

	Nombre de données manquantes	Après 'imputation'	Questionnaire maternité (temps 0) ou 2 mois (temps 1)
Age	127	57	Information disponible
Département habitation de la mère	282	0	
PCS	2400 (dont les données inclassables)	88	Information reconstituée en combinant des variables aux temps 0 et 1

Tableau 10 : Exemple de données manquantes

### c) Description de la non-réponse

L'unité à répondre étant le nourrisson, les informations des mères de jumeaux ont été doublées.

Distributions (sur les données non manquantes) • datant de février 2014	Nombre de nourrissons dans l'échantillon	Part des nourrissons dans l'échantillon	Nombre de nourrissons NON-répondants	Taux de NON-réponse des nourrissons	Test du chi-deux (p-value)
<b>Total</b>	36028	100,00%	17800	49,40%	
<b>Vague</b>					0.0003
-1	5300	14,7%	2510	47,3%	
-2	8913	24,8%	4321	48,3%	
-3	10344	28,7%	5197	50,1%	
-4	11471	31,8%	5772	50,3%	
<b>MERES</b>					
<b>Naissance</b>					0.0020
- unique	34659	96,5%	17068	49,2%	
- multiple	1240	3,5%	666	53,6%	
<b>Activité au moment de la grossesse</b>					<.0001
- oui	24946	72,1%	9802	39,3%	
- non	9638	27,9%	6978	72,4%	
<b>Est primipare</b>					<.0001
- oui	15633	44,2%	7416	47,3%	
- non	19731	55,8%	9940	50,3%	
<b>Age*</b>					<.0001
- [18 ; 22]	2597	7,3%	1564	60,2%	
- [23 ; 24]	2566	7,2%	1432	55,7%	
- [25 ; 29]	11361	31,9%	5738	50,4%	
- [30 ; 34]	11699	32,9%	5308	45,3%	
- [35 ; 39]	5798	16,3%	2645	45,5%	
- plus de 40 ans	1537	4,3%	776	50,3%	
<b>Age gestationnel</b>					0.0557
- [33 ; 37]	4370	12,2%	2249	51,4%	
- [38 ; 40]	24913	69,8%	12370	49,5%	
- plus de 40 semaines	6383	17,9%	3152	49,2%	
<b>5 groupes de régions (mères)</b>					<.0001
- Ile-de-France, Centre, Picardie	10202	28,3%	5287	51,7%	
- Nord-Est	7774	21,6%	3428	43,9%	
- Nord-Ouest	5986	16,6%	3039	50,7%	
- Sud-Est	6721	18,7%	3346	49,7%	
- Sud-Ouest	5316	14,8%	2671	50,2%	
<b>3 groupes de régions (mères)</b>					<.0001
- Ile-de-France, Centre, Picardie	10202	28,4%	5292	51,7%	
- Nord-Est	7774	21,6%	3428	43,9%	
- et le reste	18023	50,0%	9056	50,2%	
<b>PCS brute</b>					<.0001
- 1 Agriculteurs, exploitants	98	0,3%	42	42,9%	
- 2 Artisans, commerçants et chefs d'entreprise	971	2,9%	406	41,8%	
- 3 Cadres et professions intellectuelles supérieures	4105	12,2%	1198	28,2%	
- 4 Professions Intermédiaires	6132	18,4%	2462	40,1%	
- 5 Employés	13534	41,3%	6063	44,8%	
- 6 Ouvriers	823	3,1%	505	61,4%	
- 7 Sans profession	6711	21,8%	5860	87,3%	
- 9 Ne peut classer la profession	3407	9,5%	1235	36,2%	

\* Catégories calculées à partir de l'année de naissance de la mère (âge atteint en 2011).

Distributions (sur les données non manquantes)	Nombre de nourrissons dans l'échantillon	Part des nourrissons dans l'échantillon	Nombre de nourrissons NON-répondants	Taux de NON-réponse des nourrissons	Test du chi-deux (p-value)
<b>MATERNITES</b>					
<b>Strate</b>					0.0005
- 1	863	2,4%	402	46,5%	
- 2	2451	6,8%	1136	46,2%	
- 3	4750	13,2%	2422	50,8%	
- 4	9850	27,4%	4939	50,0%	
- 5	18085	50,2%	8872	49,0%	
<b>Statut juridique</b>					<.0001
- privé non lucratif	3166	8,8%	1403	44,2%	
- privé lucratif	8929	24,8%	4458	49,9%	
- public	23904	66,4%	11910	49,7%	
<b>Autorisation</b>					0.5734
- niveau 1	8191	22,8%	4015	48,8%	
- niveau 2	17159	47,7%	8520	49,5%	
- niveau 3	10649	29,5%	5236	49,1%	
<b>Régions (maternité)</b>					<.0001
- 11 Île-de-France	7875	21,9%	4023	50,9%	
- 21 Champagne-Ardenne	771	2,1%	407	52,7%	
- 22 Picardie	11220	3,1%	632	55,8%	
- 23 Haute Normandie	1160	3,2%	615	52,9%	
- 24 Centre	1288	3,6%	712	55,1%	
- 25 Basse Normandie	785	2,3%	460	58,5%	
- 26 Bourgogne	1049	2,9%	483	45,6%	
- 31 Nord pas de Calais	2820	7,8%	1206	42,7%	
- 41 Lorraine	1196	3,3%	516	43,1%	
- 42 Alsace	1229	3,4%	520	42,2%	
- 43 Franche-Comté	667	1,8%	235	35,2%	
- 52 Pays de la Loire	2144	5,9%	1041	48,4%	
- 53 Bretagne	1864	5,2%	910	48,7%	
- 54 Poitou Charentes	823	2,3%	406	49,2%	
- 72 Aquitaine	1547	4,30%	719	46,4%	
- 73 Midi-Pyrénées	969	2,7%	522	53,8%	
- 74 Limousin	240	0,7%	100	41,7%	
- 82 Rhône-Alpes	3322	9,2%	1640	49,3%	
- 83 Auvergne	440	1,2%	179	40,2%	
- 91 Languedoc-Roussillon	1717	4,8%	904	52,6%	
- 93 PACA	2932	8,1%	1511	51,4%	
- 94 Corse	41	0,1%	30	68,2%	

Tableau 11 : Caractéristiques des nourrissons non-participants

Exemple de lecture du tableau : durant la vague 1, 5300 nourrissons éligibles sont nés (14,7% de l'échantillon) dont 2510 ne participeront pas à l'enquête (47,3% des 5300).

On observe une plus grande participation des mères ayant accouché dans les maternités privées non lucratives et dans les maternités de petites tailles (strates 1 et 2). Remarquons qu'à l'inverse la non réponse au niveau maternité faisait apparaître une moindre participation des maternités du privé non lucratif; de même pour les maternités de la strate 1. L'enquête dans les maternités du Nord-Est de la France a particulièrement bien fonctionné. Concernant les caractéristiques des mères, les très jeunes ou âgées ont moins participé, de même que celles sans activité au moment de la grossesse ou sans profession.

### 3) Repondération des nourrissons participants dans les maternités participantes aux jours elfe

La méthode retenue pour traiter la non-réponse est celle des Groupes de Réponses Homogènes constitués à partir de la méthode des scores avec des taux de réponse non pondérés. A partir des scores triés résultant de la régression logistique, 35 groupes de réponses homogènes ont été créés.

Le modèle logistique prend en compte les caractéristiques de la mère et celles de sa maternité d'accouchement.  
 Variable expliquée : {participe, ne participe pas}  
 Variables explicatives : Age, PCS, Indicatrice gémellaire, Région d'habitation, Activité, Age gestationnel, Primiparité, , Strate, Statut juridique, Autorisation

Les groupes extrêmes (avec une grosse proportion de répondants ou énormément de non-répondants) ont des poids extrêmement petits ou grands. Ils ont été **regroupés** avec des groupes qui leur ressemblent (dans le sens de la participation ou non à l'enquête) et ainsi leurs poids ont été **répartis**. Si cela diminue la ressemblance des mères au sein du groupe, cela diminue aussi la dispersion des poids (équilibre entre biais et variance).

Coefficient d'ajustement de la non-réponse	Méthode	Min	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	Scores, non pondérés	1,3	7,4	5,7	36028	1,97	1,3	68

Tableau 12 : Statistiques du coefficient d'ajustement de la non-réponse

## POIDS DES NOURRISSONS : P1P2P3

Chaque nourrisson se voit affecté du poids corrigé de la maternité dans laquelle il est né, du poids temps de celle-ci et de son poids corrigé en fonction des caractéristiques de sa mère. Il est intéressant de regarder la construction de ce poids qui sera tronqué afin de diminuer sa variance. L'effectif de la population cible sera calculé, puis estimé à partir des poids P1P2P3 non tronqués.

### 1) Analyse du poids et troncature

P1P2P3	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	16	19	21	74	116	958	60	<b>783980</b>	43	39	91

Tableau 13 : Statistiques du P1P2P3

On ne veut pas d'une dispersion trop grande : il ne serait pas bon qu'un nourrisson puisse représenter 20 autres nourrissons tandis qu'un autre puisse en représenter 600.

Afin de réduire la dispersion des poids, quelques poids extrêmes qui se détachaient de la distribution ont été tronqués à 200. Cela a pour effet de créer du biais (ce que l'on essaie de limiter en tronquant le moins de poids possible) mais de diminuer l'étendue des poids (ainsi que sa variance). L'effectif estimé de la population diminue lui aussi (mais sera corrigé au niveau du calage).

P1P2P3 tronqué	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	16	19	21	74	115	200	12,5	768365	42	33	77

Tableau 14 : Statistiques du P1P2P3 tronqué

Profils des individus dont les poids très grands ont été tronqués : ce sont toutes des mères sans profession et en majorité habitant en Ile de France. Elles viennent de maternités petites (leurs P1 sont grands).

La somme des poids estime l'effectif total de la population. Nous allons nous intéresser à ce dernier.

## 2) Estimation de l'effectif de la population à partir de l'état civil et de l'ENP

Le nombre total de nourrissons répondant aux critères de l'enquête ELFE nés durant l'année 2011 et dans toute la France métropolitaine est **inconnu**.

Il sera donc estimé à partir des données de l'**état-civil 2011**, mises à disposition sur le site de l'INSEE en novembre 2011.

En partant du nombre total de nourrissons nés en France, il sera retiré certains effectifs répondant aux caractéristiques des mères non-éligibles. Cette façon de procéder est loin d'être précise car si certaines mères possèdent deux de ces caractéristiques, elles seront soustraites deux fois et ainsi de suite. Cette estimation demande donc de retrouver des effectifs ou pourcentages de chacun des critères de non-éligibilité. Ce sera possible pour l'âge de la mère (les moins de 18 ans étant exclues de ELFE), pour l'âge gestationnel (les grands prématurés étant exclus) l'**E.N.P.** sera consultée. Quant aux mères ne comprenant pas l'une des langues traduites du questionnaire, il ne nous est pas possible de les quantifier. La population des triplets est négligeable (0,01% selon l'échantillon de l'ENP).

Etapas :

- Total des naissances vivantes ayant eu lieu en France en 2011 : **823 394**

L'enquête cherchant à représenter la population de la métropole, il faut ôter les naissances des départements d'outre-mer.

- Total des naissances vivantes ayant eu lieu en France métropolitaine en 2011 : **792 996**

Il existe des naissances pour lesquelles le lieu d'accouchement n'a pas été précisé dans le bulletin statistique d'état civil : nous les garderons (**10 466** en métropole en 2011). Les naissances déclarées peuvent venir d'un établissement spécialisé ou à domicile (ou dans un autre lieu). Ce dernier point est plus embêtant sachant que certaines mères participant à l'enquête ont accouché à domicile et se sont ensuite rendues en maternité. La sous-population des mères ayant accouché à domicile étant hors-cible, ces mères participantes sont assez rares. Cette sous-population, dans l'état civil, sera donc elle aussi ôtée de l'effectif (**10 251**).

- Total des naissances vivantes ayant eu lieu en France métropolitaine dans un établissement spécialisé en 2011 : **782 745**

Sur les 792 996 naissances vivantes ayant eu lieu en France métropolitaine en 2011 , **0,5 %** ont des mères mineures (exclues de l'enquête). Le pourcentage a été calculé à partir de l'âge révolu.

- Total des naissances vivantes ayant eu lieu en France métropolitaine dans un établissement spécialisé en 2011 la mère étant majeure : **782 745 - 0,5% 782 745 = 778 831**

Il nous faut ôter la population des grands prématurés (moins de 33 semaines d'âge gestationnel) : l'**E.N.P.** (2010) l'estime à 1,9%. Ce taux est stable d'une année à l'autre.

- Total des naissances vivantes ayant eu lieu en France métropolitaine dans un établissement spécialisé en 2011 la mère étant majeure et le nourrisson n'étant pas un grand prématuré : **778 831 - 1,9% 778 831 = 764 033**

Sachant qu'il y a des mères qui ont accouché à domicile dans l'enquête et qu'il y a des mères exclues plusieurs fois dans

ce calcul (à la fois mineures et dont l'enfant est un grand prématuré), l'effectif de la population est **sous-estimé**.

La population des nourrissons éligibles ELFE représente donc environ  $764\ 033 / 792\ 996 = 96,3 \%$  de la population des nourrissons en France métropolitaine.

# CALAGE

Le calage est une méthode de redressement : par repondération des individus, elle permet que l'échantillon pondéré soit cohérent avec des informations auxiliaires disponibles sur un certain nombre de variables. Cette information doit correspondre à **toute** la population ELFE.

En général, on souhaite caler notre échantillon sur des variables les plus corrélées possible avec notre variable d'intérêt. Ici, les variables d'intérêt sont multiples et proviennent de plusieurs domaines : sociologie, environnement, médical... Et la pondération devant être unique, on ne pourra effectuer différents calages (un calage socio-démographique et un calage santé-environnement) pour différentes pondérations.

Le calage ne prenant pas en compte les données manquantes, il faudra imputer les variables présentées ci-dessous.

## 1) Sources

Pour rappel, les critères d'éligibilité d'un nourrisson à l'enquête ELFE sont : naissance métropolitaine en maternité, âge gestationnel supérieur à 33 semaines, naissance au plus gémellaire, mère non mineure et en mesure de donner un consentement éclairé notamment dans l'une des langues proposées.

L'INSEE met à disposition sur son site des données de fichiers d'**état civil** des naissances chaque année. Il s'agit de **toutes** les naissances de France, comme nous l'avons vu précédemment en calculant l'effectif de la population ELFE. Il faudrait donc pour les variables de calage sélectionnées, respecter les critères d'éligibilité au niveau de l'état civil, ce qui n'est possible que pour le caractère métropolitain des naissances.

**L'Enquête Nationale Périnatale (ENP)** a lieu régulièrement (1995, 1998, 2003, 2010) en France. Elle vise à connaître l'état de santé et les soins périnataux des enfants, des mères, leurs caractéristiques, les facteurs à risques et par sa répétition, permet de suivre les évolutions entre enquêtes. L'ENP 2010 a eu lieu du **15 au 21 mars 2010** dans **toutes** les maternités (publiques et privées) de métropole ainsi que trois départements d'outre-mer (Guadeloupe, Guyane et La Réunion) : 15 933 nourrissons constituent l'échantillon. Il nous a été possible de travailler sur le sous-échantillon respectant les critères d'éligibilité ELFE : **14 492** nourrissons (filtre sur l'âge gestationnel, l'âge de la mère, l'indicatrice gémellaire et les naissances métropolitaines).

Avantages du calage sur l'état civil : exhaustivité des naissances en 2011 en France

Limites du calage sur l'état civil : impossible de retrouver les critères d'éligibilité ELFE

Avantages du calage sur l'ENP : possibilité de retrouver les critères d'éligibilité ELFE

Limites du calage sur l'ENP : échantillon

## Caler ELFE sur l'ENP ?

Dans l'ENP, il n'y a pas de non-réponse totale au niveau des maternités (une seule a refusé de participer mais certaines informations sur ses naissances ont pu être récupérées grâce au certificat du huitième jour). La non-réponse totale au niveau des mères (mères ne pouvant ou ne souhaitant pas participer à l'enquête) et la non-réponse partielle sont elles aussi corrigées par le certificat de santé du huitième jour pour les items correspondants. 4,1% des dossiers sont estimés comme incomplets ou vides (pourcentage fondé sur 10 questions).

Ainsi l'échantillon ENP n'a quasiment pas de non-réponse totale, contrairement à l'échantillon ELFE (49%).

En revanche, les maternités ELFE sont enquêtées 25 jours sur 4 périodes de l'année alors que les maternités ENP le sont 7 jours sur une seule période. L'échantillon est donc plus petit et les effets saisonniers ne sont pas pris en compte.

Il n'est pas **évident**, à ce niveau, qu'il faille caler l'enquête ELFE sur les résultats de l'échantillon ENP. Cependant, des comparaisons ont été faites entre l'ENP et l'état-civil 2009 montrant des distributions voisines sur plusieurs variables entre les deux sources. Il a été décidé par l'équipe de caler sur les deux sources.

## 2) Choix des variables de calage

On voudrait pouvoir caler sur un grand nombre de variables. Tout d'abord, le calage demande un nombre **limité** de variables (avec un nombre limité de modalités) pour être performant et ensuite, on ne connaît pas beaucoup de distributions de référence (avec mêmes définitions des variables et même champ que ELFE).

Le choix s'est porté sur les variables : **Age**, **Région**, **état matrimonial**, **statut immigré**, **niveau d'étude** et **Primiparité**.

L'âge et la région sont regroupées en 5 et 6 modalités, le niveau d'étude en 3 modalités et les autres variables sont binaires.

Le calage sur l'**Age** issu de l'état civil va permettre d'augmenter les poids des mères très jeunes ou âgées, qui malgré la repondération ajustée de la non-réponse sont encore sous-représentées. De la même façon, le calage sur les groupes de **régions** nous assurera une bonne représentation géographique.

La **Primiparité** et le **Statut matrimonial** sont des variables permettant de caractériser la situation familiale, dimension importante dans cette enquête.

On choisit de caler sur le **Niveau d'étude** qui est une caractéristique peu susceptible de changer dans le temps après accouchement.

La population des immigrés étant plus vaste que celle des étrangers (le fait d'acquérir la nationalité française concerne une sous-population), il a été décidé de caler sur le **Statut immigré**. De la même manière, définir une population par son pays de naissance est un état qui ne change pas, contrairement à un statut juridique. C'est un découpage plus répandu au niveau de la statistique démographique pour les comparaisons internationales (statut d'étranger n'est pas le même entre pays).

Pour la **PCS** : étant donné qu'elle a été améliorée dans ELFE (utilisation du 2 mois), sa distribution n'est plus comparable à celle de l'ENP qui a gardé cette variable brute et, comme la collecte de l'échantillon de l'ENP n'a pas de

raison d'être plus optimale que celle de ELFE, il n'y pas de raison de déformer nos poids pour coïncider avec la distribution de l'ENP.

Quant à l'*Activité*, on observe une différence de l'ordre de 10% entre les 2 distributions due à la fois à la différence de formulation de la question (activité au moment de la grossesse, activité à la fin de la grossesse), et aux erreurs d'échantillonnage d'ELFE et de l'ENP. De plus, l'échantillon a précédemment été repondéré en fonction de la variable de statut professionnel *Activité*.

Aucun des calages (testés et ayant pour source l'ENP) prenant en compte une des variables de santé (*Age gestationnel, Début du travail, Identité gémellaire, Alimentation à la naissance, Transfert...*) ou plusieurs (Cf. Annexe) ne permet une globale meilleure estimation de ces variables (et le nombre de variables de calage reste limité), le calage ne s'effectuera pas sur une des variables santé. Rappelons que les variables *Age gestationnel* et *Identité gémellaire* ont été prises en compte dans l'ajustement de la non-réponse.

On aura vérifié que les définitions des modalités des variables de calage sont les mêmes sur l'échantillon et sur la population.

Variables de calage	Distribution avant pondération	Distribution après pondération et avant calage	Source
<i>âge de la mère</i> *			Etat civil (filtre sur métropole et mères majeures)
- [18, 22]	- 05,71%	- 07,78%	- 06,86%
- [23, 24]	- 06,27%	- 07,39%	- 07,10%
- [25, 29]	- 31,07%	- 32,00%	- 31,22%
- [30, 34]	- 35,32%	- 32,63%	- 33,25%
- [35, 39]	- 17,42%	- 16,04%	- 16,90%
- + de 40 ans	- 04,21%	- 04,15%	- 04,67%
<i>Groupe de régions de domicile</i>			Etat civil (filtre sur métropole)
- Ile de France/Centre/Picardie	- 26,96%	- 30,4%	- 29,96%
- Nord-Est	- 23,84%	- 19,9%	- 19,15%
- Nord-Ouest	- 16,17%	- 14,4%	- 15,42%
- Sud-Est	- 18,52%	- 19,4%	- 19,03%
- Sud-Ouest	- 14,51%	- 15,8%	- 15,54%
<i>statut mère immigrée</i>			Etat civil (filtre sur métropole)
- née en France	- 86,66%	- 82,2%	- 81,25%
- immigrée	- 13,34%	- 17,8%	- 18,75%
<i>état matrimonial</i>			Etat civil (filtre sur métropole)
- né dans le mariage	- 46,17%	- 45,7%	- 45%
- né hors mariage	- 53,83%	- 54,3%	- 55%
<i>mère primipare</i>			ENP (champ ELFE)
- oui	- 45,65%	- 44,7%	- 43,1%
- non	- 54,35%	- 55,3%	- 56,9%
<i>niveau d'étude de la mère</i>			ENP (champ ELFE)
- non scolarisée/école primaire/collège/CAP ou BEP	- 18,91%	- 23,7%	- 27,79%
- 2nde/1ère/terminale générale, technique ou professionnelle	- 20,93%	- 23,9%	- 19,88%
- études supérieures	- 60,16%	- 52,3%	- 52,33%

Tableau 15 : Distributions des variables de calage sur l'échantillon des répondants

Pour une variable comme la *Primiparité*, le choix de l'ENP comme source vient par défaut : le bulletin statistique de l'état civil renseigne très mal le rang génésique (ainsi que les catégories professionnelles).

\* Catégories calculées à partir de l'année de naissance de la mère (âge atteint en 2011).

Nous avons donc des variables de calage caractérisant la situation familiale comme la *primiparité* ou la *situation matrimoniale*. Un calage géographique en 5 groupes de régions. Et plusieurs variables socio-démographiques : *Age, Statut immigrée* et *Niveau d'étude*.

Imputation de la non-réponse partielle pour le calage

Il faut qu'il n'y ait plus de non-réponse partielle dans les variables de calage de l'échantillon : si tel était le cas, les unités associées ne seraient pas prises en compte dans le calage. Les données manquantes (à très petits taux) ont été imputées par hot deck aléatoire afin de ne pas déformer la distribution de la variable.

### 3) Calage et troncature

Pour utiliser la macro CALMAR, il faut soit donner les effectifs (sur la population) des modalités de la variable de calage, soit en donner les pourcentages et dans ce cas indiquer l'effectif total de la population. La taille estimée de la population N=764 000 calculée précédemment a été utilisée pour chacune des pondérations.

La méthode du raking ratio a d'abord été testée. On veut qu'en moyenne les rapports de poids  $\frac{w_k}{d_k}$  soient proches de 1.

Avec les méthodes de calage tronqué (ici, logit), on essaie donc de limiter la dispersion des rapports de poids par les bornes LO et UP. Ces méthodes n'ont pas été retenues, les rapports de poids étant proches des anciens, les CV augmentant très légèrement et les variables d'intérêt ne bougeant pas.

Calage avant troncature	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	11	17	20	74	113	281	26	764000	41,9	34	81

Tableau 16 : Statistiques du P1P2P3 après calage

Une centaine d'individus s'éloignent de la distribution (à droite). Ils ont été tronqués à 200 (créant du biais mais diminuant l'étendue et la variance des poids).

Regardons les profils des individus dont les poids trop grands ont été tronqués : ce sont des mères qui avaient déjà des poids assez grand avant calage. On retrouve des mères sans profession et/ou de statut immigré et/ou de niveau d'étude le plus bas : des classes sous-représentées avant calage.

Troncature après calage	Min	P5	P10	P90	P95	Max	Max/min	Somme des poids	Moyenne	écart-type	CV (%)
	11	18	20	74	113	200	18	759088	41,7	32	78

Tableau 17 : Statistiques du P1P2P3 après calage et troncature

La troncature diminue l'effectif total de la population estimé : il suffit d'appliquer un facteur correctif (764 000/somme des poids après troncature) à chaque poids et ainsi en sommant les poids finaux, on retrouvera bien les 764 000 nourrissons : **POIDS<sub>0</sub>** .

#### 4) Tester la pondération sur quelques variables d'intérêt

Pour apprécier une pondération, il est intéressant de choisir quelques variables et de comparer leurs distributions après pondération et après calage aux distributions connues sur **toute** la population. Cela suppose d'avoir la possibilité de connaître cette dernière information.

Ces variables sont choisies par l'équipe ELFE comme étant des variables « importantes » : nationalité de la mère, situation d'emploi du père, diabète gestationnel, statut pondéral de la mère, première alimentation du nourrisson...

Les sources sont l'état civil et l'ENP.

On observe un petit nombre de non-réponses partielles sur ces variables (dans ELFE et dans l'ENP) que l'on supposera ignorable.

Il se peut, dans cette enquête pluridisciplinaire, que suivant le calage choisi, certaines variables d'intérêt soient plus corrélées avec les variables de calage (socio-démographiques) et donc mieux estimées.

Variables d'intérêt (Nombre de valeurs manquantes dans ELFE)	Échantillon sans pondération	Échantillon pondéré avant calage	Échantillon pondéré après calage	Source p2 [IC* à 95%]
				État civil 2011
Nationalité de la mère (238)				
- française (par naissance ou par acquisition)	- 91,7%	- 87,58%	- <b>87,06%</b>	- 86,6%
- étrangère ou apatride	- 08,3%	- 12,41%	- 12,94%	- 13,4%
				ENP sur le champ Elfe
Age gestationnel (330)				
- [33 ; 37]	- 11,84%	- 11,89%	- 11,94%	- 12,32% [11,8-12,9]
- [38 ; 40]	- 70,09%	- 70,42%	- <b>70,47%</b>	- 69,15% [68,4-69,9]
- plus de 40 semaines	- 18,07%	- 17,69%	- 17,58%	- 18,53% [17,9-19,2]
situation d'emploi du père (650)				
- oui	- 91,53%	- 88,65%	- <b>88,35%</b>	- 87,36% [86,8-87,9]
- non	- 8,47%	- 11,35%	- 11,65%	- 12,64% [12,1-13,2]
statut pondéral de la mère (IMC avant grossesse) (364)				
- < 18,5	- 07,88%	- 07,97%	- 7,94%	- 08,20% [07,7-08,7]
- [18.5, 25]	- 64,60%	- 62,93%	- 62,65%	- 64,57% [63,8-65,4]
- [25, 30]	- 17,50%	- 18,28%	- <b>18,40%</b>	- 17,33% [16,7-18,0]
- > 30	- 10,01%	- 10,82%	- <b>11,00%</b>	- 09,89% [09,4-10,4]
professionnel de santé ayant suivi la grossesse (329)				
- gynécologue-obstétricien	- 66,45%	- 66,25%	- 65,78%	- 67,02% [66,2-67,8]
- généraliste	- 05,12%	- 04,98%	- <b>05,17%</b>	- 04,59% [04,2-04,9]
- sage-femme	- 12,04%	- 12,82%	- <b>12,87%</b>	- 11,58% [11,0-12,1]
- plusieurs de ces professionnels	- 16,39%	- 15,95%	- 16,17%	- 16,81% [16,2-17,4]

\* Intervalle de confiance approximé par  $\left[ \hat{p} \pm 1.96 \sqrt{\frac{1}{n} \hat{p}(1-\hat{p})} \right]$

Nombre de séances de préparation pendant la grossesse (8703)				
- de 4	- 16,61%	- 18,29%	- 18,33%	- 19,66% [16,0-17,3]
- [4, 8]	- 79,04%	- 77,27%	- <b>77,25%</b>	- 76,15% [75,4-76,9]
- + de 8	- 04,36%	- 04,43%	- 04,41%	- 04,19% [03,9-04,5]
première alimentation de l'enfant à la naissance (136)				
- au lait maternel uniquement	- 59,95%	- 58,01%	- 57,86%	- 59,84% [59,0-60,6]
- au lait premier âge uniquement	- 31,37%	- 32,18%	- <b>32,22%</b>	- 30,87% [30,1-31,6]
- allaitement mixte	- 08,07%	- 09,16%	- <b>09,22%</b>	- 08,32% [07,9-08,8]
- inconnu car enfant transféré	- 00,62%	- 00,65%	- 00,69%	- 00,97% [00,8-01,1]
Identité gémellaire (58)				
- 0	- 96,83%	- 96,70%	- 96,73%	- 97,25% [97,0-97,5]
- 1	- 03,17%	- 03,30%	- <b>03,27%</b>	- 02,75% [02,5-03,0]
hypertension artérielle (566)				
- non	- 96,52%	- 96,20%	- <b>96,14%</b>	- 95,35% [95,0-95,7]
- oui avec protéinurie	- 01,45%	- 01,65%	- 01,69%	- 01,91% [01,7-02,1]
- oui sans protéinurie	- 02,03%	- 02,15%	- 02,16%	- 02,74% [02,5-03,0]
Accouchement : début du travail (déclenchement, césarienne) (382)				
- travail spontané	- 70,82%	- 70,12%	- <b>70,10%</b>	- 66,63% [65,8-67,4]
- déclenchement	- 19,51%	- 19,79%	- 19,95%	- 22,44% [21,7-23,1]
- césarienne avant début du travail	- 09,67%	- 10,08%	- 09,95%	- 10,93% [10,4-11,4]

Tableau 18 : Distributions des variables d'intérêt

En **gras**, les proportions ELFE supérieures à celles issues des sources.

Pour la variable *Nombre de séances de préparation pendant la grossesse*, il y a trop de données manquantes pour pouvoir comparer les distributions.

Prenons la variable *Nationalité de la mère* : d'après l'état civil, 13,4% des mères en France sont étrangères. Dans notre échantillon (18228 nourrissons), 8,3% des mères sont étrangères. En repondérant en fonction des caractéristiques des non-participants, ce taux passe à 12,3%. Puis en calant (notamment sur le *statut immigré*), il augmente jusqu'à 12,9%. Ceci est un exemple montrant que la repondération et le calage vont 'favoriser' les variables de type socio-démographiques.

Concernant les variables santé : comme dit précédemment, aucune de ces variables prises pour calage n'améliorait globalement leurs estimations. Certaines de ces variables, sans pondération aucune, ont des distributions déjà proches de celles de l'ENP : *suivi par un professionnel de la santé* et *première alimentation*. La plupart ne voient pas leurs distributions changer au fil des pondérations.

Rappelons que la distribution de l'ENP est issue d'un échantillon, nous ne ferons donc pas un test d'ajustement qui permettrait de tester si la distribution ELFE est en adéquation avec celle de l'ENP.

## TABLEAUX

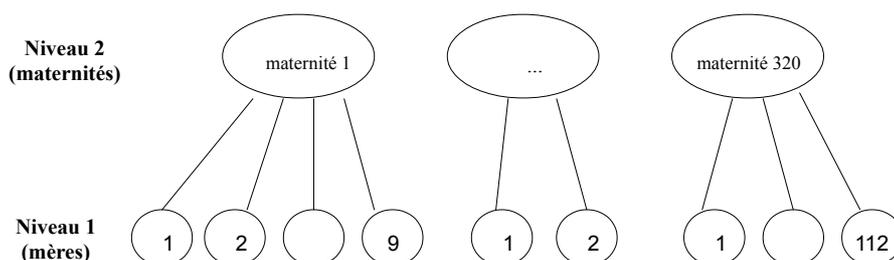
Tableau 1 : Poids initiaux issus de la stratification.....	6
Tableau 2 : Poids initiaux finaux.....	7
Tableau 3 : Statistiques du poids initial final.....	7
Tableau 4 : Caractéristiques des maternités non-participantes.....	8
Tableau 5 : Statistiques du coefficient d'ajustement de la non-réponse.....	9
Tableau 6 : Statistiques du poids initial corrigé de la non-réponse.....	9
Tableau 7 : Participation des maternités aux vagues et en nombre de vagues.....	10
Tableau 9 : Sous-couverture par région.....	14
Tableau 10 : Exemple de données manquantes.....	17
Tableau 11 : Caractéristiques des nourrissons non-participants.....	19
Tableau 12 : Statistiques du coefficient d'ajustement de la non-réponse.....	20
Tableau 13 : Statistiques du P1P2P3.....	21
Tableau 14 : Statistiques du P1P2P3 tronqué.....	21
Tableau 15 : Distributions des variables de calage sur l'échantillon des répondants.....	26
Tableau 16 : Statistiques du P1P2P3 après calage.....	27
Tableau 17 : Statistiques du P1P2P3 après calage et troncature.....	27
Tableau 18 : Distributions des variables d'intérêt.....	29

# ANNEXES

## a) Analyse multiniveau

L'analyse multi-niveaux permet de prendre en considération les variables individuelles et les variables contextuelles, **simultanément**. Elle a pour objectif de reconnaître les liens qui existent entre individus d'un même groupe, de comprendre les influences entre individus et leurs groupes d'appartenance. L'idée de base étant que les individus d'un même groupe ont tendance à avoir des comportements qui se ressemblent, en comparaison aux individus des autres groupes : effet de grappe, variance intra petite des groupes. Cette homogénéité interne au groupe ne permet pas de vérifier l'**hypothèse d'indépendance des observations**, hypothèse nécessaire aux outils standards de l'analyse statistique. On peut observer des covariances entre les observations partageant le même contexte.

**Variable à expliquer : le fait de participer ou pas à l'enquête ELFE.** Il s'agit en fait de circonscrire les caractéristiques propres aux mères et les caractéristiques propres à la maternité qui pourraient donner lieu à un comportement particulier en matière de participation. Certaines maternités sont-elles plus répondantes que d'autres ? Est-ce dû à leurs caractéristiques (statut juridique, niveau, nombre d'accouchements par an) ? Quelles sont les caractéristiques individuelles des mères qui répondent ?



La structure hiérarchique de l'échantillon des mères issu d'un sondage par **grappes** nécessite des méthodes spécifiques de régression : par exemple, l'analyse multi-niveaux. Est-ce que les mères appartenant à une même maternité auront **tendance à se ressembler** et peut-être à se ressembler dans leurs comportements face à l'enquête ?

Niveau	Variables explicatives	effectif
Niveau 2 : maternités	- région (5 groupes) - nombre d'accouchements (strate) - autorisation (niveau 1, 2 ou 3) - statut juridique	320
Niveau 1 : mères	- primiparité - âge de la mère - âge gestationnel - région d'habitation de la mère - indicatrice gémellaire - activité au moment de la grossesse - CSP	36028

Modèle à intercept aléatoire :

$$Y_{ij} = (\alpha + v_j) + \beta X_{ij} + e_{ij} = (\alpha + \beta X_{ij}) + (v_j + e_{ij})$$

variable expliquée {participe, ne participe pas}

variables explicatives : variables niveau 1 et variables niveau 2

Les variables *Age gestationnel* et *Primiparité* n'étant pas significatives ont été retirées du modèle, de même que le *Groupe de région d'habitation* de la mère (corrélé au *groupe de région* de la maternité).

Les paramètres non aléatoires estimés par le multiniveau sont assez proches du logistique simple. On s'intéresse à  $\sigma_v^2$  la variance inter maternités, qui, dans notre exemple, est égale à 0,2235 et pour laquelle il existe un test, dont le résultat, ici justifie l'utilisation d'un modèle multi-niveaux. Certaines variables significatives avec le modèle simple ne le sont plus si l'on prend en compte l'effet groupe : les variables *Strate* et *Statut juridique*. C'est en comparant les écarts-type des estimations des deux modèles que l'on perçoit ce changement : ils doublent ou triplent avec l'effet maternité.

On peut calculer l'ICC (intraclass correlation coefficient) qui mesure la part de la variance totale attribuable aux variations du niveau groupe (donc la part de la variance expliquée par la structure maternité).

$$ICC = \frac{\sigma_v^2}{\sigma_v^2 + \frac{\pi^2}{3}}$$

où v le terme aléatoire de niveau 2. Il est égal dans ce modèle à **6,36%**.

Donc, **selon notre modèle**, il existe un petit effet significatif des maternités sur la participation ou non à l'enquête.

**b) Comparaison entre deux calages (recherche d'un impact sur les variables 'santé')**

Calage a : *Age, Région, état matrimonial, statut immigré, niveau d'étude et Primiparité*

Calage b : variable du calage a + *Transfert à la naissance, Début du travail, première alimentation*

Les distributions sont calculées avant la troncature qui suit le calage. Pour calculer le biais relatif  $\frac{(\hat{p}-p)}{p}$ , on considère que la valeur p observée par l'ENP est la 'vraie' valeur (celle observée sur toute la population).

Distributions % (biais relatif %)	Non pondéré	Calage a	Calage b	Source (ENP 2010)
<b>Transfert</b>				
non	93.3 (01,4)	93.1 (01,2)	91.8 (00,2)	92,02
oui	06.7 (16,0)	06.9 (13,4)	08.2 (02,8)	07,98
<b>Début du travail</b>				
- travail spontané	70,82 (06,3)	70.0 (05,1)	66.7 (00,1)	66,63
- déclenchement	19,51 (13,1)	19.8 (11,8)	22.5 (00,3)	22,44
- césarienne avant début du travail	09,67 (11,5)	10.2 (06,7)	10.9 (00,3)	10,93
<b>Première alimentation</b> de l'enfant à la naissance				
- au lait maternel uniquement	59,95 (00,2)	57.7 (03,6)	59.8 (00,1)	59,84
- au lait premier âge uniquement	31,37 (01,6)	32.4 (05,0)	30.9 (00,1)	30,87
- allaitement mixte	08,07 (03,0)	09.2 (10,6)	08.3 (00,2)	08,32
- inconnu car enfant transféré	00,62 (36,1)	00.7 (27,8)	00.9 (07,2)	00,97
NB de séances de préparation pendant la grossesse				
- de 4	16,55 (15,8)	19.7 (00,2)	19.7 (00,2)	19,66
- [4, 8]	79,09 (03,9)	76.2 (00,1)	76.2 (00,1)	76,15
- + de 8	04,36 (04,1)	04.2 (00,2)	04.2 (00,2)	04,19
Statut pondéral de la mère (IMC avant grossesse)				
- < 18,5	07,88 (03,9)	07.8 (04,9)	07.8 (04,9)	08,20
- [18.5, 25]	64,60 (00,0)	62.4 (03,4)	62.3 (03,5)	64,57
- [25, 30]	17,50 (01,0)	18.5 (06,8)	18.6 (07,3)	17,33
- > 30	10,01 (01,2)	11.2 (13,2)	11.3 (14,3)	09,89
Age gestationnel				
- [33 ; 37]	11,84 (03,9)	12.3 (00,2)	12.6 (02,3)	12,32
- [38 ; 40]	70,09 (01,4)	69.2 (00,1)	69.4 (00,4)	69,15
- plus de 40 semaines	18,07 (02,5)	18.5 (00,2)	18.0 (02,9)	18,53
Professionnel de santé ayant suivi la grossesse				
- gynécologue-obstétricien	66,45 (00,9)	66.1 (01,4)	66.2 (01,2)	67,02
- généraliste	05,12 (11,5)	05.2 (13,3)	05.1 (11,1)	04,59
- sage-femme	12,04 (04,0)	12.8 (10,5)	12.8 (10,5)	11,58
- plusieurs de ces professionnels	16,39 (02,5)	15.9 (05,4)	16.0 (04,8)	16,81
Identité gémellaire				
- 0	96,83 (00,4)	96.7 (05,4)	96.5 (00,8)	97,25
- 1	03,17 (15,3)	03.3 (00,6)	03.5 (27,3)	02,75
Hypertension artérielle pendant la grossesse				
- non	96,52 (01,2)	96.1 (00,8)	95.9 (00,6)	95,35
- oui avec protéinurie	01,45 (24,1)	01.7 (11,0)	01.8 (05,8)	01,91
- oui sans protéinurie	02,03 (25,9)	02.2 (19,7)	02.3 (16,1)	02,74
Diabète gestationnel				
- non	92.8 (00,1)	92.6 (00,3)	92.4 (00,5)	92,86
- oui traité par insuline	01.9 (03,3)	02.2 (19,6)	02.3 (25,0)	01,84
- oui traité par régime	05.3 (00,0)	05.2 (01,9)	05.3 (00,0)	05,30
Nationalité de la mère				
- française (par naissance ou par acquisition)	91,7 (05,9)	87.0 (00,5)	87.0 (00,5)	86,6
- étrangère ou apatride	08,3 (38,1)	13.0 (03,0)	13.0 (03,0)	13,4
Activité père				
- non	08.5 (32,5)	11.6 (07,9)	11.6 (07,9)	12,6
- oui	91.5 (04,7)	88.4 (01,1)	88.4 (01,1)	87,4
Activité mère				
- non	14,9	25,0	25,1	-
- oui	85,1	75,0	74,9	-

# BIBLIOGRAPHIE

- Aragon, Y., Boistard, H., Goga-Cardot, C. et Ruiz-Gazen, A., 2012, Cours de sondage, M2 Statistique et économétrie
- Blondel, B. et Kermarrec, M., 2011, Enquête nationale périnatale 2010 : les naissances 2010 et leur évolution depuis 2003, INSERM
- Caron, N., La correction de la non-réponse par repondération et par imputation, INSEE
- Chaix B. et Chauvin P., 2002, L'apport des modèles multiniveau dans l'analyse contextuelle en épidémiologie sociale : une revue de littérature, *Revue d'épidémiologie et de santé publique*
- Chauvet, G., 2012, Données Manquantes dans les Enquêtes, ENSAI (diapo)
- Chauvet, G., 2011, Méthodes de redressement, ENSAI (diapo)
- Courgeau, D., 2004, Du groupe à l'individu, synthèse multiniveau, Editions de l'INED
- Dai J, Li Z, Roche D, 2006, Hierarchical Logistic Regression Modeling with SAS GLIMMIX. Proceedings of the Thirty-first Annual SAS Users Group International Conference. Cary, North Carolina: SAS Institute Inc.
- Entwisle, Mason et Hermalin, 1986, The multilevel Dependence of Contraceptive use on Socioeconomic Development and Family Planning Program Strength, *Demography*, Vol.23, n°2
- Goldberg, M., Guéguen, A., Sitta, R., Zins, M., 2013, Le point de vue de l'épidémiologie : être représentatif ou universel ?
- Grilli L. et Testa M.R., 2006, L'influence des différences de fécondité dans les régions européennes sur la taille idéale de la famille, *Population* vol.61
- Haziza, D., 2005, Inférence en présence d'imputation simple : un survol, *Journal de la société statistique de Paris*, tome 146, n°4
- Haziza, D., Estimation non paramétrique des probabilités de réponse dans les enquêtes (diapo)
- Grondin, Lavallée, Théorie et application des enquêtes longitudinales, *Statistique Canada*, Décembre 2008, cours
- Kish, L., 1992, Weighting for Unequal Pi, *Journal of Official Statistics*, 8(2), p. 183-200
- Kott, 2012, Pourquoi les poids de sondage devraient être intégrés dans la correction de la non-réponse totale fondée sur des groupes de réponse homogènes
- Little, R. et Vartivarian, S., 2005, Does Weighting for Nonresponse Increase the Variance of Survey Means?, *Survey Methodology*
- Massé, E., 2006, Insee-Méthodes n°116, 2006 - 4ème partie : Aspects statistiques
- Razafindratsima, Sarter et l'équipe Elfe, 2008, Évaluation et traitement de la non-réponse dans la cohorte Elfe : résultats des études pilotes, *Recueil du Symposium 2008 de Statistique Canada*
- Razafindratsima, N., Kishimba, N., 2004, La déperdition dans la cohorte Cocon entre 2000 et 2002, *Population*, Vol.59
- Razafindratsima, N., 2010, Les pondérations de la vague 2 de l'enquête ERFI (notes internes)
- Razafindratsima, N., 2003, Calcul des pondérations de l'enquête « Biographie et entourage » en petite et grande couronne (notes internes)
- Régnier-Loilier A., Rohrbasser J.M., Y a-t-il une saison pour faire des enfants ?, *Population et sociétés*, n°474
- Sautory, O., 1993, La macro CALMAR : redressement d'un échantillon par calage sur marges, INSEE
- Singer, 1998, Using SAS PROC MIXED to fit Multilevel Models, Hierarchical Models, and Individual Growth Models, *Journal of Educational and Behavioral Statistics*, Vol. 24, n°4

Apport des cohortes à la connaissance de la santé, ADSP, 2012

<http://www.statsoft.fr/concepts-statistiques/modeles-chaid/modeles-chaid.htm>