ELFE INCLUSION SURVEY
Carried out in 2011
WEIGHTINGS at TIME 0 (child birth)

Hélène Juillard , INED - April 2015

Outline

I– INTRODUCTION

The objective of the ELFE survey (French longitudinal study from childhood) is to collect information on all new-born infants from a random sample of 349 maternity units in metropolitan France, over 25 distributed throughout the year 2011 in 4 survey periods.

The final sample of around **18,300 infants**, or around 1 in 42 births in metropolitan France, was drawn from a sampling plan with unequal probabilities of inclusion.

To be used in analyses, a weighting must make it possible to generalize the results to the entire population (and not only to the sample). It consists in assigning a statistical weight to each of the 18,300 infants in the survey, corresponding to the number of children they represent in the target population (764,000 infants, and an estimated total of 753,500 families). The inference population consists of **infants born in 2011 in a maternity unit in metropolitan France, born along with at most a twin, excluding highly premature infants, mother was aged 18 or over and able to give informed consent, notably in one of the languages offered (French, English, Arabic, or Turkish), and whose parents did not reside temporarily in metropolitan France.**

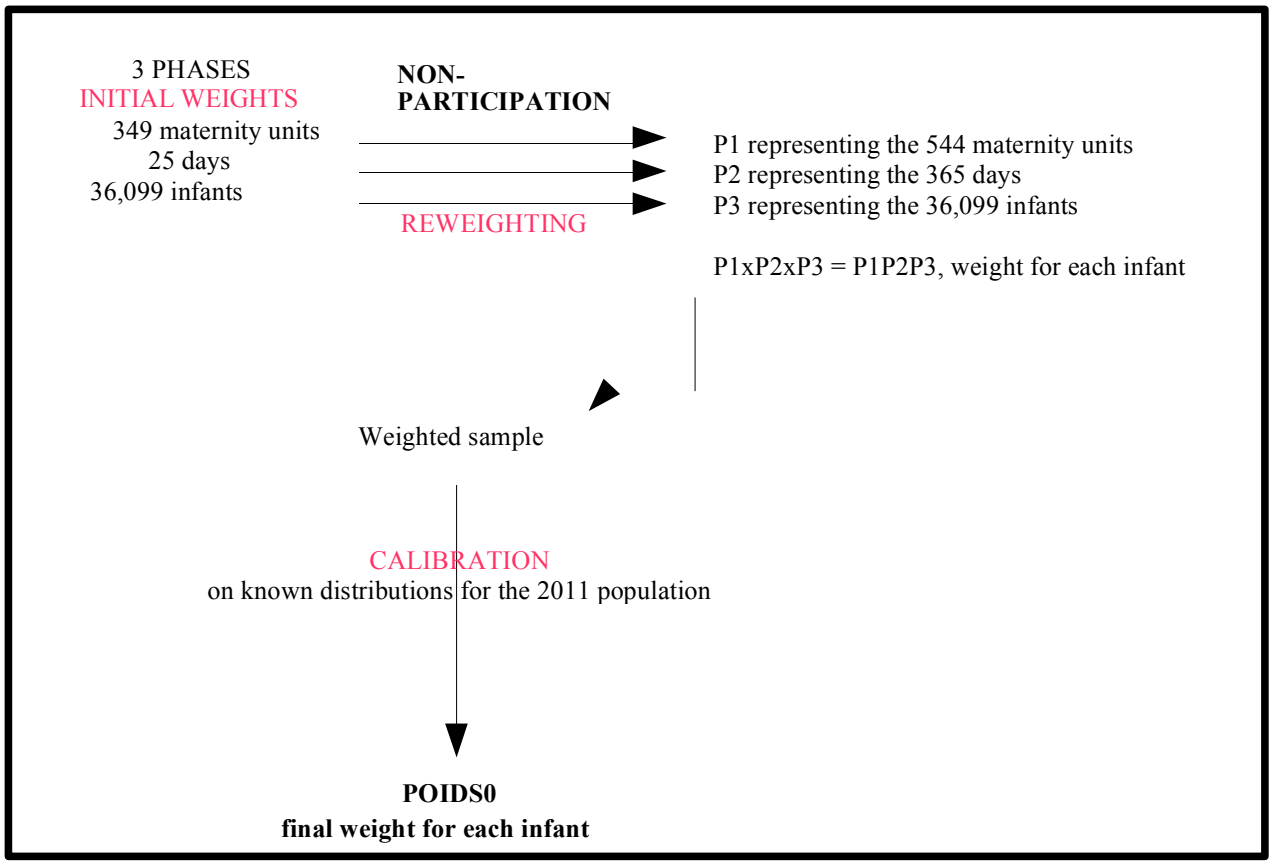
Two weightings are proposed: a child-level weighting (each twin considered separately) and a family-level weighting. The results below are from the child weighting. The method used for the family weighting is identical

The sample was constructed on the basis of a sampling plan with **several sampling phases**:[1] a phase for maternity units, another for days, and the last, an exhaustive one for infants. The randomly selected maternity units were drawn from a **stratified sampling plan** with allocations proportional to their sizes. To represent each season, four periods in the year were selected: from April 1st to 4th, from June 27th to July 4th, from September 27th to October 4th, and finally from November 28th to December 5th: 25 days in total.

After taking into account the initial weights derived from the sampling frame, the weights are adjusted to take into account non-participation at different levels: a portion of the maternity units did not participate in the initial survey, and a portion of the mothers who gave birth on the survey days also did not participate. **Two types of non-participation** are distinguished for maternity units: units that did not participate at all, and units that participated partially (non-participation on some of the survey days). We have information on variables common to the participating and non-participating units. To deal with **mothers' non-participation**, we also have information on variables common to participating and non-participating mothers. A calibration was then performed against the civil register and the national perinatal survey (ENP), allowing our weighted sample to be matched to the target population in terms of geography and the mother's sociodemographic situation.

To construct this weighting, the sampling plan was divided into three phases: maternity units, days, and infants.

[1] Sampling plan constructed by Nicolas Razafindratsima (INED) and Hélène Sarter (InVS)



The 36,099 infants are those whose mothers were approached in the participating maternity units during the days of their participation.

II – THE WEIGHTING

1) Phase 1, the maternity units

The sample for this phase was constructed based on a stratification with allocation proportional to size. Note, then, the unequal initial weights $\frac{1}{\pi_i}$ depending on the **size of the maternity unit** (number of deliveries/year).

STRATUM h	Number of deliveries in the unit in 2008	Size of the stratum N_h	Size of the sample n_h	Probability of inclusion π_h	**poids-initial** $\frac{1}{\pi_h}$
1	[145, 699]	108	28	0.26	3.86
2	[700, 1009]	108	47	0.44	2.3
3	[1010, 1418]	109	66	0.60	1.65
4	[1422, 2187]	108	97	0.90	1.1
5	[2197, 5215]	111	111	1	1
Total		**544**	**349**		**544**

Size of the population: 544 maternity units — Size of the sample: 349 maternity units

Number of maternity units that agreed to participate: 320 — Participation rate: 91.69%

We have information on 4 variables that are common to participating and non-participating units: **region**, **legal status**, **stratum**, and **level of medical authorization**.

Comparison of participating and non-participating maternity units on 4 common variables	Number of maternity units	Number of maternity units that did not participate in any period (total NON-participation)	**NON-participation rate**	Chi-squared independence test or Fisher's exact test (p-value)
Total	**349**	**29**	**8.31%**	
Region				Sample per class too small
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%	
Size (number of births in 2008)				0.8989
[145, 699]	28	3	10.7%	
[700, 1009]	47	3	6.4%	
[1010, 1418]	66	4	6.1%	
[1422, 2187]	97	9	9.3%	
[2197, 5215]	111	10	9.0%	
Regional group				2.703e-13
Ile-de-France, Centre, Picardie	101	17	16.8%	
Southwest	69	7	10.1%	
Other	179	5	2.8%	
Authorization				0.1889
level 1	125	11	8.8%	
level 2	161	16	9.9%	
level 3	63	2	3.2%	
Legal status				0.0969
private non-profit	30	5	16.7%	
private for-profit	95	9	9.5%	
public	224	15	6.7%	

The weighted[2] scores method[3] (using the 4 variables mentioned above) was used to compensate the non-participation of the 29 non-participating maternity units (weight of zero) by increasing the weighting of the 320 participating maternity units by an adjustment factor $\frac{1}{\hat{p}_i}$.

$$P1_i = \frac{1}{\pi_i} \frac{1}{\hat{p}_i}$$

2) Phase 2, the survey days

In total, out of the 320 * 25 = 8,000 expected survey days in participating maternity units, 7,741 took place (96.76%).

2 Weighted by the initial weight of the stratum.

3 Method using logistic regression to create homogeneous participant groups, which serve as the basis for calculating mean response probabilities $\hat{p}_i$

Each infant was weighted according to its period *v* of birth (season). The first period included 4 days, the second 6 days, the third 7 days and the last 8 days:

$$P2_v = \frac{91}{\#\text{ of days in period}_v} \frac{\#\text{ of maternity units that participated in at least one period}}{\#\text{ of participating maternity units in period}_v}$$

This was done within each stratum of maternity units.

3) Phase 3, the infants

The 36,099 eligible infants born within the participating maternity units on one of the survey days were invited to be part of the cohort. The parents of more than 18,300 infants consented to the inclusion of their child(ren) in the ELFE study.

Number of individuals approached: 36,099 infants

Participation rate: 50.8%

Among the 18,329 participants in the study, 71 were born outside the 25 survey days: they were thus were removed from the sample to which the weighting corresponds. Moreover, 51 individuals requested the destruction of their data (between the survey and the time when the weighting was constructed).

Size of the sample of participants and non-participants: 36,028 infants

Size of the sub-sample of participants: 18,207 infants

We have 11 variables that are common to participants and non-participants: **mother's age**, her **district of residence**, her **socio-economic class**, indicator of her **activity** status at the time of the pregnancy, a **primiparity** indicator, a **twin birth** indicator, the infant's **gestational age**, and the **4 variables** characterizing the maternity unit.

Note: responses from the survey at 2 months were used to reduce the proportion of missing values for the variables age, activity, and SOC from the survey in maternity units.

Comparison of the characteristics of participating and non-participating infants on the 11 shared variables (on non-missing data)[4]	Number of infants in the sample	Proportion of infants in the sample	Number of NON-participants	**Rate of infants' NON-participation**	Chi-square test (p-value)
Total	36,028	100.00%	17,800	49.40%	
Period					0.0003
-1	5,300	14.7%	2,510	47.3%	
-2	8,913	24.8%	4,321	48.3%	
-3	10,344	28.7%	5,197	50.1%	
-4	11,471	31.8%	5,772	50.3%	
MOTHERS					
Birth					0.0020
- single	34,659	96.5%	17,068	49.2%	
- multiple	1,240	3.5%	666	53.6%	

[4] Table constructed in February 2014

Activity at the time of pregnancy					<.0001
- yes	24,946	72.1%	9,802	39.3%	
- no	9,638	27.9%	6,978	72.4%	
Was primipara					<.0001
- yes	15,633	44.2%	7,416	47.3%	
- no	19,731	55.8%	9,940	50.3%	
Age					<.0001
- under 22 years	2,597	7.3%	1,564	60.2%	
- [23 ; 24]	2,566	7.2%	1,432	55.7%	
- [25 ; 29]	11,361	31.9%	5,738	50.4%	
- [30 ; 34]	11,699	32.9%	5,308	45.3%	
- [35 ; 39]	5,798	16.3%	2,645	45.5%	
- over 40 years	1,537	4.3%	776	50.3%	
Infant's gestational age					0.0557
- [33 ; 37]	4,370	12.2%	2,249	51.4%	
- [38 ; 40]	24,913	69.8%	12,370	49.5%	
- more than 40 weeks	6,383	17.9%	3,152	49.2%	
3 groups of regions of residence (mothers)					<.0001
- Île-de-France, Centre, Picardie	10,202	28.3%	5,287	51.7%	
- Northeast	7,774	21.6%	3,428	43.9%	
- Northwest	5,986	16.6%	3,039	50.7%	
- Southwest	6,721	18.7%	3,346	49.7%	
- Southwest	5,316	14.8%	2,671	50.2%	
Socio economic class					<.0001
- 1 Farmer	98	0.3%	42	42.9%	
- 2 Self-employed (non-farming)	971	2.9%	406	41.8%	
- 3 Manager or higher-level intellectual occupation	4,105	12.2%	1,198	28.2%	
- 4 Intermediate occupation	6,132	18.4%	2,462	40.1%	
- 5 Clerical or sales worker	13,534	41.3%	6,063	44.8%	
- 6 Manual worker	823	3.1%	505	61.4%	
- 7 No occupation	6,711	21.8%	5,860	87.3%	
- 9 Cannot classify occupation	3,407	9.5%	1,235	36.2%	
MATERNITY UNITS					
Stratum					0.0005
- 1	863	2.4%	402	46.5%	
- 2	2451	6.8%	1,136	46.2%	
- 3	4,750	13.2%	2,422	50.8%	
- 4	9,850	27.4%	4,939	50.0%	
- 5	18,085	50.2%	8,872	49.0%	
Legal status					<.0001
- private non-profit	3,166	8.8%	1,403	44.2%	
- private for-profit	8,929	24.8%	4,458	49.9%	
- public	23,904	66.4%	11,910	49.7%	
Authorization					0.5734
- level 1	8,191	22.8%	4,015	48.8%	
- level 2	17,159	47.7%	8,520	49.5%	
- level 3	10,649	29.5%	5,236	49.1%	

The weighted[5] scores[6] method (using the 11 variables above) was used to compensate the non-participation of the 17,800 non-participating infants (weight zero) by increasing the weighting of the 18,207 participating infants by an adjustment factor $\frac{1}{\hat{q}_j}$.

There is some **undercoverage**: some eligible mothers were not approached (a mean of 4%). However, the number of eligible births per maternity unit is known. In order to correct for this issue, a coefficient was calculated by region (number of eligible infants / number of infants included in the survey) and assigned to each infant.

$$P3_j = \frac{1}{q_j}\ coeff_{undercoverage}$$

5 Weighted by the initial weight of the stratum

6 Method using logistic regression to create homogeneous participant groups, which serve as the basis for calculating mean response probabilities $\hat{q}_i$

4) Calibration

The weighting of each infant *j* is thus affected by the corrected weight of the maternity unit *i* where they were born, the associated time weighting, and its weight as corrected according to the characteristics of non-participating infants.

$$P1P2P3_j = P1_i \; P2_v \; P3_j$$

In order to ensure coherence with a few selected variables that are available for the entire population, we performed a calibration on distributions from vital records and the national perinatal survey (ENP). This calibration modifies the weights P1P2P3j.

Calibration on **age** (civil register filtered for metropolitan France and age ≥ 18) increases the weight of the youngest and oldest mothers, who, despite the reweighting adjusted for non-participation, remain under-represented. Likewise, calibration on groups of **regions** (civil register filtered on metropolitan France) ensures appropriate geographical representation. **Primiparity** (ENP) and **marital status** (civil register filtered for metropolitan France) characterize the infant's family situation, an important dimension of the ELFE study.
We also chose to calibrate on **level of education** (ENP), a characteristic that is unlikely to change after childbirth. As the population of immigrants is larger than that of non-French citizens (with a sub-population having acquired French nationality), the decision was made to calibrate on **immigrant status** (civil registered filtered for metropolitan France).

$$\text{Calibration}\square\, P1P2P3_j\square = POIDS_{0j}$$

Calibration variables	Distribution before weighting P1P2P3	Distribution after weighting P1P2P3 and before calibration	Source
mother's age			Civil register (filtered for and mother's age □18)
- [18, 22]	- 05.71%	- 07.86%	- 06.86%
- [23, 24]	- 06.27%	- 07.45%	- 07.10%
- [25, 29]	- 31.07%	- 32.15%	- 31.22%
- [30, 34]	- 35.32%	- 32.50%	- 33.25%
- [35, 39]	- 17.42%	- 15.88%	- 16.90%
- + than 40 years	- 04.21%	- 04.17%	- 04.67%
Group of regions of residence			Civil register (filtered for
- Île-de-France/Centre/Picardie	- 26.96%	- 30.4%	- 29.96%
- Northeast	- 23.84%	- 19.6%	- 19.15%
- Northwest	- 16.17%	- 14.3%	- 15.42%
- Southwest	- 18.52%	- 20.0%	- 19.03%
- Southwest	- 14.51%	- 15.7%	- 15.54%
mother's immigrant status			Civil register (filtered for
- born in France	- 86.66%	- 82.2%	- 81.25%
- immigrant	- 13.34%	- 17.8%	- 18.75%
marital status			Civil register (filtered for
- parents married at childbirth	- 46.17%	- 45.8%	- 45%
- parents unmarried	- 53.83%	- 54.2%	- 55%
mother primipara			ENP* (ELFE coverage)
- yes	- 45.65%	- 44.8%	- 43.1%
- no	- 54.35%	- 55.2%	- 56.9%
mother's level of education			ENP* (ELFE coverage)
- no schooling/primary/middle school/lower secondary	- 18.91%	- 23.7%	- 27.79%
- grades 10, 11. 12	- 20.93%	- 23.9%	- 19.88%
- higher education	- 60.16%	- 52.3%	- 52.33%

The calibration thus ensures distributions of the weighted ELFE sample identical to those in the 'Source' column. The raking ratio method was used.

5) Description of weights

In order to limit the dispersion of the weights (which impacts the variance of our estimates), some weights were truncated at 200.

Min	P5	P10	P90	P95	Max	Max/min	Sum of weights	Mean	Standard deviation	Coefficient of variation (%)
11	18	20	75	114	201	15	764,000	42	32	77.6

Note: it is possible that the database that is delivered or analysed will contain fewer individuals than the number indicated in this document, as respondents are entitled to request the destruction of their data. (This phenomenon is assumed to be rare and its impact on weighting minimal.)