



# National survey Weighting at the age of 5½ years

Thierry Siméon

## Contents

1. Weighting of the national survey at age 5½.....	2
2. Cross-sectional weights .....	4
3. Longitudinal weights .....	7

## 1. Weighting of the national survey at age 5½

The aim of the weighting presented here is to allow data from survey respondents to be used to infer characteristics of the target population (in the case of the ELFE cohort, the population of inference is that of infants born in 2011 in metropolitan France; at most a twin birth; excluding highly premature babies; with an adult mother). The total size of the target population is set at 753,500 families (and 764,000 children). In the ELFE surveys, to correct for biases introduced by nonparticipation (avoiding the assumption that participants and nonparticipants have identical characteristics), reweighting is used. This reweighting is based on the weighting method implemented for the survey wave at age 3½ and the subsequent waves: simultaneous calibration.<sup>1</sup>

In the case of the survey at age 5½, although the data collection method was the same for all respondents (computer-assisted telephone interviewing, or CATI), different questionnaires were used depending on the situation of the two parents and the child at the time of the survey (parents cohabiting or not, parent with custody of the child, whether or not the child had been placed in the care of an institution or foster family, whether the child was hospitalized or not), along with their responses to specific questions in the previous survey waves. In calculating the weights, two samples can thus be distinguished: children with at least one parent who responded to the survey (reference parent questionnaire completed) and children whose two cohabiting parents both responded to the survey (reference parent and cohabiting non-reference parent questionnaires completed). One set of weights, named PONDREF, is applicable only to the first of these samples (which represents all respondents and the entire target population).

Note that there is another category of respondents: non-cohabiting parents (in the case of shared custody for example). However, there are too few cases to produce a reliable weighting of this sub-population in the general context that is the focus here.

Finally, note that, contrary to the weights produced before the survey wave at age 3½, and given that the weighting of the survey at age 5½ is based on the new weighting method (with the proportion of twins – 2.75% of children, and thus 21,000 children, belonging to 10,500 families – set by the population margins), there was no need for to produce two distinct sets of weights, for families and for children, for this survey wave. The basic weighting can be used for analyses at either of these levels. To work at the family level, for example, this involves simply selecting a single child per family (e.g. using the constructed variable M00M1\_RANGALEA).

We apply the simultaneous calibration method to 13 variables.

### **They include:**

6 “contextual” variables (variables used for calibration since the initial survey wave in maternity units):

- Regional group of maternity, out of 5 (Ile-de-France Centre et Picardie / Northeast / Northwest / Southeast / Southwest)
- Primiparous mother (yes/no)
- Marital status (parents married/unmarried at the time of birth)
- Mother's age (18-24/25-29/30-34/35+)
- Mother's level of education (no schooling, primary, lower secondary vocational [CAP], secondary vocational [BEP]/three last grades of secondary school/higher education)
- Mother's immigrant status (yes/no)

To which we add the following variables, which are used to take attrition between waves into account, and which were selected beginning with the survey wave at age 3½:

---

<sup>1</sup> For a precise description of the methodology, please see “ELFE Survey: Weighting national survey data” by Thierry Siméon, available at <https://www.elfe-france.fr/fichier/rte/178/Cot%C3%A9%20recherche/Weighting-Elfe-surveys-general-document.pdf>

- Birth preparation sessions (yes/no)
- Father's activity status at the time of birth (in employment/other)
- Mother living with partner at birth (yes/no)
- Father's age (18-24/25-29/30-34/35+)
- Alcohol consumption during pregnancy (yes/no)
- Twin birth (yes/no)
- Mother's activity status at the time of childbirth (in employment/other)

Note that for these new variables, missing data (5% for father's age, below 2% for the rest) were imputed before the calibration. As for the methods used for previous survey waves, all of the totals for the ELFE population are drawn from vital records or the 2010 National Perinatal Survey.

Thus, two sets of weights are produced: one cross-sectional and one longitudinal.

Depending on the origin of the variables to be used in a planned analysis, the user chooses:

- **If the variables are exclusively from a single survey wave, to use the cross-sectional weights associated with that wave;**
- **If the variables are from multiple survey waves, to use the longitudinal weights associated with the latest wave used in the analysis.**

*Note: the delivered database may contain fewer individuals than are indicated in this document. Respondents have the right to request the destruction of their data. This phenomenon, which is rare, is assumed to have no impact on the weighting.*

## 2. Cross-sectional weights

The first available set of weights are the cross-sectional weights (A05E\_PONDREF). They apply to children whose reference parent responded to the questionnaire at age 5½, and whose questionnaire is considered complete (a total of 11,220 children).

In order to avoid overly extreme values, the weights were truncated at 250. They were then adjusted to yield the required total of 764,000 children.

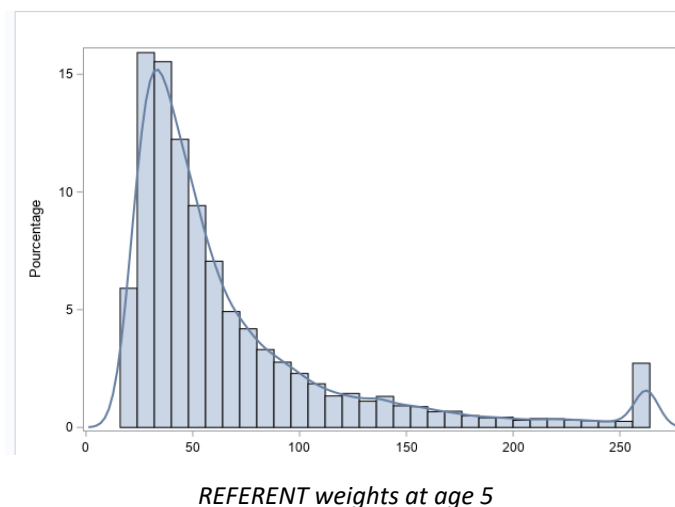
The results of this weighting are as follows:

Basic statistical characteristics			
Location		Variabilité	
Mean	68,10	Variability	54,28
Median	48,35	Standard deviation	2 946,00
Mode	262,55	Variance	246,06
		Interval	46,24
		Interquartile range	

Quantiles (Définition 5)	
Level	Quantile
100Max 100%	262,55
99%	262,55
95%	198,09
90%	139,69
75% Q3	79,55
50% Median	48,35
25% Q1	33,31
10%	26,40
5%	23,31
1%	19,63
0% Min	16,50

Note that the truncation of values above 250 applied to 2.6% of individuals.



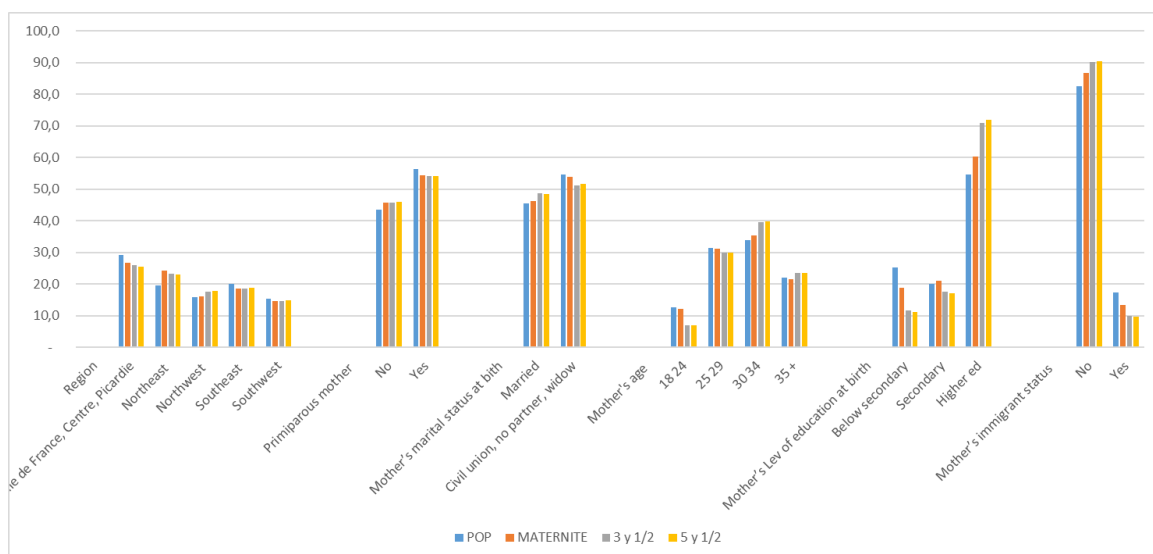
### Who are the respondents to this survey wave?

Two elements can be analysed: who the respondents to this survey wave are, and how the composition of the population of nonparticipants has changed over time. To do this, we can compare the distribution of all of the calibration variables across: the target population; respondents in maternity units; respondents at age 3½ (to estimate attrition before this age); and respondents at age 5½ (to estimate attrition since the last survey wave).

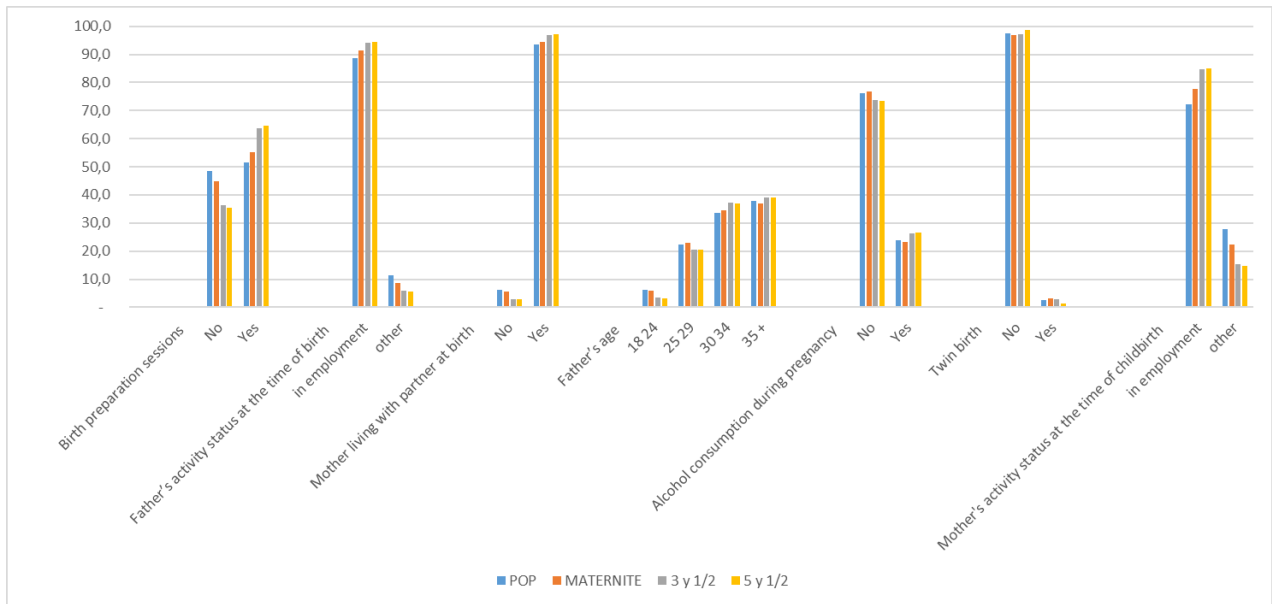
	POP	MATERNITE	3 y 1/2	5 y 1/2		POP	MATERNITE	3 y 1/2	5 y 1/2
<b>Region</b>					<b>Birth preparation sessions</b>				
Ile de France, Centre, Picardie	29,3	26,8	25,9	25,4	No	48,4	44,7	36,4	35,3
Northeast	19,5	24,2	23,2	23,0	Yes	51,6	55,3	63,6	64,7
Northwest	15,8	16,1	17,6	17,8					
Southeast	20,0	18,5	18,6	18,9	<b>Father's activity status at the time of birth</b>				
Southwest	15,5	14,5	14,7	14,9	in employment	88,7	91,4	94,1	94,4
					other	11,3	8,6	5,9	5,6
<b>Primiparous mother</b>					<b>Mother living with partner at birth</b>				
No	43,6	45,6	45,8	45,9	No	6,3	5,5	3,0	3,0
Yes	56,4	54,4	54,2	54,1	Yes	93,7	94,5	97,0	97,0
<b>Mother's marital status at bith</b>					<b>Father's age</b>				
Married	45,5	46,2	48,8	48,4	18 24	6,1	5,8	3,3	3,2
Civil union, no partner, widow	54,5	53,8	51,2	51,6	25 29	22,4	22,9	20,5	20,6
					30 34	33,5	34,5	37,1	37,0
<b>Mother's age</b>					35 +	37,9	36,8	39,1	39,2
18 24	12,6	12,1	7,0	6,9	<b>Alcohol consumption during pregnancy</b>				
25 29	31,4	31,1	30,0	29,9	No	76,2	76,8	73,7	73,5
30 34	33,9	35,3	39,6	39,7	Yes	23,8	23,3	26,3	26,5
35 +	22,0	21,6	23,5	23,6					
					<b>Twin birth</b>				
<b>Mother's Lev of education at birth</b>					No	97,4	96,9	97,1	98,7
Below secondary	25,2	18,8	11,5	11,1	Yes	2,6	3,1	2,9	1,4
Secondary	20,1	21,0	17,5	17,1					
Higher ed	54,6	60,2	71,0	71,8	<b>Mother's activity status at the time of childbirth</b>				
					in employment	72,1	77,7	84,6	85,2
<b>Mother's immigrant status</b>					other	27,9	22,3	15,4	14,8
No	82,6	86,7	90,1	90,4					
Yes	17,5	13,3	9,9	9,7					

A few different phenomena can be seen in the graphs below:

- There are particular populations with a significant proportion of nonparticipants from the outset (where the blue bar is taller than the orange bar, the category includes a substantial proportion of nonparticipants beginning with the survey in maternity units): young parents, parents with low levels of education, immigrant parents, and unemployed parents.
- A certain population left the survey between the first and second waves (where the grey bar is shorter than the orange bar) – broadly, the same as in the earlier phase, but with greater attrition.
- Finally, a certain population left the survey in the last wave, at age 5½ (where the yellow bar is shorter than the grey bar). Fortunately, this is a very small group.

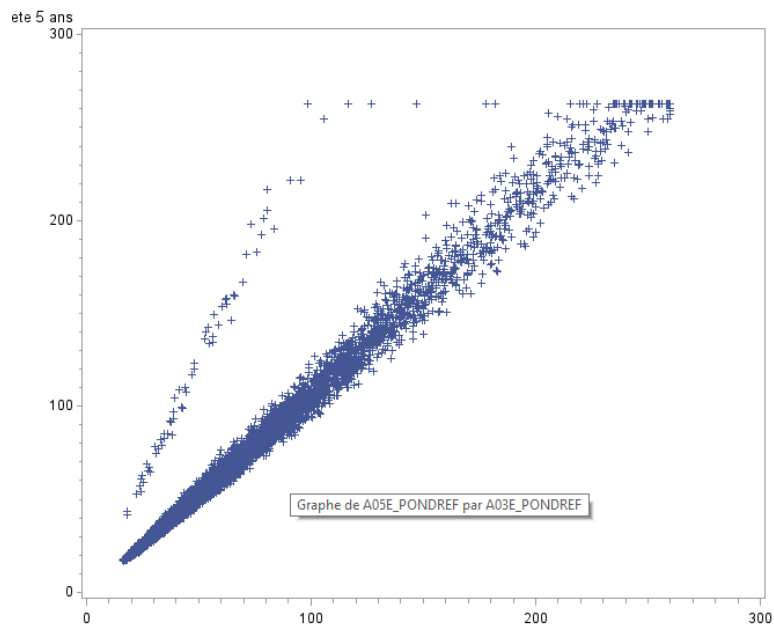


*Distribution (%) of calibration variables across: the target population; respondents in maternity units; respondents at age 3½ respondents at age 5½*



*Distribution (%) of calibration variables across: the target population; respondents in maternity units; respondents at age 3½ respondents at age 5½*

Note that the proportions of the population represented by (the weights assigned to) the 10,441 children whose reference parent participated in the survey waves at ages 5½ and 3½ are very strongly correlated. This ensures a certain continuity in analyses of the survey data: variability between measurements at the ages of 3 and 5 can be understood to result from variability in the phenomenon, and not from large variations in the weighting.



*Graph of A05E\_PONDREF by A03E\_PONDREF : Weighting of the survey wave at age 3 to work at the child level on the sub-sample of completed reference parent questionnaires*

However, the weights assigned to around 2% of children (fewer than 200 represent 26,000 in the target population) changed by more than 20% between these two survey waves. These 200 children, who represent those whose probability of participation in the survey is decreasing even faster than that of the rest of the population, are predominantly twins (3% of the population, but 70% of these children). Their share of participants

decreased by 50% between these two waves, in contrast to the earlier interval, and requiring them to represent 3% of the population led to a considerable increase in the weight assigned to each of them.

### 3. Longitudinal weights

The second set of weights is longitudinal (A05E\_LONGPONDREF). It applies to the children whose reference parent participated in all of the previous survey waves and completed the questionnaire at age 5½, a total of 9,798 children.

The results of this weighting are as follows (the maximum weight was predefined at 300):

Mesures statistiques de base				Quantiles (Définition 5)	
				Level	Quantile
Location		Variabilité		100Max 100%	322,93
Mean	77,98	Variability	68,29	99%	322,93
Median	52,55	Standard deviation	4 664,00	95%	246,34
Mode	322,93	Variance	306,56	90%	168,74
		Interval	53,91	75% Q3	89,68
		Interquartile range		50% Median	52,55
				25% Q1	35,76
				10%	27,57
				5%	24,09
				1%	19,88
				0% Min	16,37

