

# Samples, Weights and Nonresponse: the Early Childhood Cohort of the National Educational Panel Study (Wave 4)

*Ariane Würbach*

# Samples, Weights, and Nonresponse: the Early Childhood Cohort of the National Educational Panel Study (Wave 4)

*Würbach, A.*

*Leibniz Institute for Educational Trajectories*

Technical Report referring to [DOI:10.5157/NEPS:SC1:4.0.0](https://doi.org/10.5157/NEPS:SC1:4.0.0)

**E-mail address of lead author:**

methoden@lifbi.de

---

## Samples, Weights, and Nonresponse: the Early Childhood Cohort of the National Educational Panel Study (Wave 4)

### 1 Prequel

The National Educational Panel Study (NEPS) surveys an Early Childhood cohort sample (Starting Cohort 1, SC1) and follows them over their transition to Kindergarten and beyond. The data are released via corresponding Scientific Use Files (SUF). The current SUF version is available under [DOI:10.5157/NEPS:SC1:4.0.0](https://doi.org/10.5157/NEPS:SC1:4.0.0).<sup>1</sup>

This report documents the weighting of Wave 4 and supplements the previous NEPS Survey Paper by Würbach, Zinn, and Aßmann (2016), which gives detailed information on the applied sampling procedure, the derivation of design weights, their successive adjustments, and the derivation of panel weights for the previous waves 1 to 3.

Table 1 summarizes the study numbers, the survey modes, the periods of the studies as well as the numbers of participants in each panel wave available in the current SUF.<sup>2</sup> In all waves, all parents of the panel cohort were asked to be interviewed (by CATI or CAPI<sup>3</sup>).

*Table 1: Survey overview for Starting Cohort 1.*

Wave	Study number	Survey mode	Time	Number of Participants
1	B04	CAPI	2012/13	3,481
2	B05	CATI/CAPI	2013	2,862
3	B91	CAPI	2014	2,609
4	B100	CAPI	2015	2,478

CATI: Computer-assisted telephone interview, CAPI: Computer-assisted personal interview.

All panel participants were invited for direct measurements (i.e., competence tests) in Wave 1, Wave 3 and Wave 4. In Wave 2, only a subsample of children was asked participating in the direct measurements (Würbach et al., 2016, Section 2.2). The accordant numbers are given in Table 2. This table details the used gross sample size, the number of participants in the interviews and in the direct measurements as well as the number of those who were actually weighted and available for analyses. The percentages given refer to the number of participants among the used gross sample.

<sup>1</sup>For general information on the NEPS, see Blossfeld, Roßbach, and von Maurice (2011). More detailed information is available in the documentation section on the [homepage](#).

<sup>2</sup>More details on the studies are given in the reports of the survey institute 'infas' *Institut für angewandte Sozialwissenschaft GmbH* which conducted the corresponding interviews and tests; see Bauer, Bech, Gilberg, and Kleudgen (2013), Aust and Bauer (2014a, 2014b), and Bauer et al. (2015).

<sup>3</sup>CATI: Computer-assisted telephone interview, CAPI: Computer-assisted personal interview.

*Table 2: Participation in direct competence measurements.*

Wave	Study number	Used gross sample	Participants	Analyzable and weighted cases	%
1	B04	3,481	3,481	3,121	89.7
2	B05	1,893	1,510	1,417	93.8
3	B91	3,281	2,609	n.a.	-
4	B100	3,143	2,478	n.a.	-

Note: "-" not applicable; "n.a." not (yet) available.

Across the distinct panel waves, for all participating units cross-sectional as well as longitudinal weights are provided. Furthermore, weights are given for individuals with additional information from competence measurements (currently available for Wave 1 and Wave 2).

The remainder of this supplement is structured as follows: Section 2 details the panel progress of the Starting Cohort 1 and the composition of the gross and net samples of the different waves is described. In Section 3 the nonresponse in Wave 4 is analyzed. Nonresponse models are estimated using logistic regressions. Finally, Section 4 concludes with a summary of the provided sampling weights and design information given in the corresponding weighting data set.

## 2 Panel progress

The following Table 3 completes the study summary of Starting Cohort 1 (Table 1) by detailing the composition of the distinct samples together with the numbers of nonrespondents and final dropouts. Final dropouts are separated into final dropouts due to refusal during the survey period and final dropouts between two consecutive waves.

*Table 3: Panel progress of Starting Cohort 1 by wave.*

Wave	Group	Panel Cohort			Status at the end of the wave			
		Total size	Not used	Used sample	Participants	Temporary dropout	Final dropout (in wave)	Final dropout (after wave)
1	All	–	–	8,483	3,481	0	5,002	50
2	<b>All</b>	<b>3,431</b>	<b>0</b>	<b>3,431</b>	<b>2,862</b>	<b>468</b>	<b>101</b>	<b>49</b>
	CATI	3,431	0	3,431	2,849	480	102	48
	CAPI	3,431	1,538	1,893	1,510	340	43	21
3	All	3,281	0	3,281	2,609	539	133	5
4	All	3,143	0	3,143	2,478	543	122	<sup>a</sup> 149

Note: <sup>a</sup> 143 units are declared as final dropouts because of not having participated for a period of two years.

### 3 Weighting Adjustments for Wave Participation

Systematic refusals may arise and for this, the (non-)response and attrition processes of the sampled individuals, has to be accounted for. Thus, for reasons of usability, commonly design weights are adjusted to account for nonresponse in the survey. For this purpose, the units' probabilities to participate in each survey wave are employed.<sup>4</sup> The processing in the non-response analysis with a comparison of the gross sample and the realized sample of Wave 1 is detailed in Würbach et al. (2016, Chapter 4).

Logistic regression models are used to estimate the individual participation propensities. On the basis of the estimated (non)response models participation probabilities are predicted and used as adjustment factors to derive cross-sectional and longitudinal survey weights.

#### 3.1 Modeling Wave 4 Participation

Directly on the onset of Wave 4, the panel cohort comprised 3,143 parents and children pairs. That is, 288 panel members withdrew their participation consent within Wave 2 or between the Waves 2 and 3, within Wave 3 or between the Waves 3 and 4. The Tables 4 and 5 give the corresponding variables and results for panel and wave participation.

Regarding panel willingness, as before in Wave 3 (Würbach et al., 2016, p. 19), the educational attainment, the employment status as well as the migration background show a clear effect on the participation probability. The higher the educational attainment, the higher is the participation probability. Being employed increases the probability to participate, whereas having a migration background decreases participation probability. Compared to the model for panel participation before Wave 3, in the current model the effect of gender is still high but not significant anymore. Since size of the household was excluded from the model<sup>5</sup>, the effect of the number of children in the household has become highly significant. That is, parents having two children or more have a higher propensity to participate.

The probability of attending the CATI is significantly influenced by age, the educational attainment, the migration background, the marital status, and the employment status. Older respondents (born before 1975), employed respondents, those with higher educational level and respondents without migration background are (still) more likely to participate. This also holds for married respondents opposed to single or divorced/widowed ones and respondents having two or more children. Again, the effect of the gender has become insignificant.

---

<sup>4</sup>In SC1 the target population are newborns but the respondents are their legal guardians. Hence, in this particular case it would be more appropriate to use the term realization probability instead of participation probability. Nevertheless, realization probability is not commonly used in the context of survey weighting, therefore it is waived.

<sup>5</sup>Household size was excluded due to multi-collinearity. High correlations with marital status of the interviewed person and the number of children in the household existed. Large values of *kappa* or the *condition index* suggest to remove the variable from the modeling, see Belsley, Kuh, and Welsch (1980, p. 104).

Table 4: Results of the logistic regression model for panel participation before Wave 4.

Value	Reference Category	Estimate	95%-CI
<i>Birth month</i>	February		
March		0.069	[-0.318,0.457]
April		0.031	[-0.332,0.394]
May		0.606	[0.203,1.008]
June/July		-0.185	[-0.785,0.414]
<i>Year of birth interviewed person</i>	1986 and later		
Before 1975		-0.170	[-0.606,0.266]
1976-1980		-0.171	[-0.507,0.165]
1981-1985		0.076	[-0.289,0.440]
<i>Gender of interviewed person</i>	Female		
Male		-0.769	[-1.573,0.035]
<i>Federal region</i>	East (including Berlin)		
West		0.196	[-0.193,0.585]
<i>BIK categories</i>	Less than 50,000 inhabitants		
50,000 up to 100,000 inhabitants		0.259	[-0.534,1.053]
100,000 up to 500,000 inhabitants		0.009	[-0.466,0.485]
500,000 or more inhabitants		-0.013	[-0.536,0.509]
<i>CASMIN of interviewed person</i>	1a, 1b, 2b		
1c, 2a		0.187	[-0.168,0.542]
2c		0.442	[0.041,0.842]
3ab		0.885	[0.472,1.297]
<i>Employment status of interviewed person</i>	Unemployed		
Employed		2.153	[1.803,2.503]
<i>Migration background of interviewed person</i>	No		
Yes		-0.498	[-0.752,-0.245]
<i>Marital status of interviewed person</i>	Single		
Married		-0.018	[-0.390,0.354]
Divorced/widowed		-0.184	[-0.855,0.488]
<i>Number of children in household</i>	1		
2		0.694	[0.357,1.031]
3		0.674	[0.259,1.088]
4+		0.768	[0.189,1.346]
Number of cases	3,431		

Table 5: Results of the logistic regression model for Wave 4 participation (CATI of parents).

Value	Reference Category	Estimate	95%-CI
<i>Birth month</i>	February		
March		-0.050	[-0.351,0.252]
April		0.166	[-0.115,0.447]
May		-0.124	[-0.397,0.150]
June/July		-0.228	[-0.529,0.073]
<i>Year of birth interviewed person</i>	1986 and later		
Before 1975		0.353	[0.017,0.688]
1976-1980		0.160	[-0.107,0.427]
1981-1985		-0.038	[-0.348,0.273]
<i>Gender of interviewed person</i>	Female		
Male		-0.317	[-0.868,0.233]
<i>Federal region</i>	East (including Berlin)		
West		0.016	[-0.247,0.280]
<i>BIK categories</i>	Less than 50,000 inhabitants		
50,000 up to 100,000 inhabitants		-0.105	[-0.636,0.425]
100,000 up to 500,000 inhabitants		0.041	[-0.455,0.537]
500,000 or more inhabitants		0.037	[-0.443,0.518]
<i>CASMIN of interviewed person</i>	1a, 1b, 2b		
1c, 2a		0.397	[0.057,0.736]
2c		0.638	[0.314,0.963]
3ab		0.917	[0.565,1.268]
<i>Employment status of interviewed person</i>	Unemployed		
Employed		1.397	[1.174,1.620]
<i>Migration background of interviewed person</i>	No		
Yes		-0.519	[-0.730,-0.309]
<i>Marital status of interviewed person</i>	Single		
Married		0.373	[0.140,0.606]
Divorced/widowed		0.203	[-0.398,0.805]
<i>Number of children in household</i>	1		
2		0.626	[0.413,0.839]
3		0.745	[0.407,1.083]
4+		1.135	[0.612,1.657]
Number of cases	3,143		

### 3.2 Modeling Participation in Consecutive Waves

In addition to the cross-sectional weights, also weights for participation in consecutive waves, i.e. longitudinal weights, are provided. This is the longitudinal weight for participating in the parent interview in all four waves. For this purpose, a logistic regression model for attending all of the CATIs (in Wave 1 to 4) has been estimated. Table 6 gives the corresponding variables and results.

The coefficients of the longitudinal model for parent participation in CATI revives the picture that has emerged from modeling Wave 4 participation. Respondents born before 1981, those with higher educational level and respondents without migration background are even more likely to participate constantly in the parent interviews. Remarkable is, that the effect of the CASMIN (measure for educational attainment) has become much stronger. This again holds also for employed as well as married respondents, and respondents from large families. Though, the effect sizes are smaller for these groups. Alike in the cross-sectional model, the effect of the gender has become insignificant.

## 4 Summary and Use of Weights

The NEPS provides various kinds of weights for the Early Childhood cohort together with design information. Table 7 lists the design information and the summarizes all types of weights and their accordant label provided by SUF release version [DOI:10.5157/NEPS:SC1:4.0.0](https://doi.org/10.5157/NEPS:SC1:4.0.0). To ease statistical analysis, all weights apart from the pure design weight (Wave 1) are provided in a trimmed and standardized form (Würbach et al., 2016, Chapter 6). Standardized weights have mean one and sum up to the number of participants in the corresponding wave. Summary statistics for all kind of weights provided are given in Table 8. The summary statistics of weights for Wave 1, 2, and 3 differ slightly compared to the previous SUF because in the models size of the household was excluded.

Please refer to Würbach et al. (2016, Chapter 6) for advices regarding the usage of weights.



Table 6: Results of the logistic regression model for participation in Wave 1 to 4 (CATI of parents).

<b>Value</b>	<b>Reference Category</b>	<b>Estimate</b>	<b>95%-CI</b>
<i>Birth month</i>	February		
March		0.050	[-0.155,0.255]
April		0.113	[-0.100,0.325]
May		-0.007	[-0.185,0.172]
June/July		-0.129	[-0.410,0.151]
<i>Year of birth interviewed person</i>	1986 and later		
Before 1975		0.550	[0.278,0.821]
1976-1980		0.335	[0.087,0.582]
1981-1985		0.212	[-0.074,0.499]
<i>Gender of interviewed person</i>	Female		
Male		-0.413	[-0.887,0.061]
<i>Federal region</i>	East (including Berlin)		
West		0.203	[-0.086,0.491]
<i>BIK categories</i>	Less than 50,000 inhabitants		
50,000 up to 100,000 inhabitants		-0.102	[-0.517,0.314]
100,000 up to 500,000 inhabitants		-0.038	[-0.433,0.357]
500,000 or more inhabitants		-0.050	[-0.454,0.355]
<i>CASMIN of interviewed person</i>	1a, 1b, 2b		
1c, 2a		0.653	[0.324,0.982]
2c		1.020	[0.698,1.342]
3ab		1.315	[1.001,1.630]
<i>Employment status of interviewed person</i>	Unemployed		
Employed		1.222	[1.041,1.403]
<i>Migration background of interviewed person</i>	No		
Yes		-0.528	[-0.716,-0.340]
<i>Marital status of interviewed person</i>	Single		
Married		0.184	[0.013,0.356]
Divorced/widowed		-0.254	[-0.787,0.279]
<i>Number of children in household</i>	1		
2		0.580	[0.400,0.760]
3		0.547	[0.265,0.829]
4+		0.793	[0.315,1.270]
Number of cases	3,431		

Table 7: Variables included in the weighting data for SC1 version 4.0.0 of the SUF.

Variable	Applies to	Content
<i>Identifier</i>		
ID_t	all targets	Identifier for target person
<i>Design information</i>		
psu	all targets	Primary Sampling Unit (Point number)
stratum	all targets	Stratification variable according to sampling frame
px80101_R	all targets	Federal State according to sampling frame
<i>Design weights adjusted for initial nonresponse</i>		
w_t1ext*	3,481 cases	Design weight for parents participating in Wave 1 (unstandardized)
w_t1	3,481 cases	Cross-sectional weight for parents participating in Wave 1
w_t1comp	3,121 cases	Cross-sectional weight for children participating in Wave 1 with direct measurements
w_t2	2,862 cases	Cross-sectional weight for parents participating in Wave 2
w_t2comp	1,417 cases	Cross-sectional weight for children participating in Wave 2 with direct measurements
w_t12comp	1,362 cases	Longitudinal weight for children participating in Wave 1 and 2 with direct measurements
w_t3	2,609 cases	Cross-sectional weight for parents participating in Wave 3
w_t123	2,427 cases	Longitudinal weight for parents participating in Wave 1, 2, and 3
w_t4	2,478 cases	Cross-sectional weight for parents participating in Wave 4
w_t1234	2,171 cases	Longitudinal weight for parents participating in Wave 1 up to Wave 4

\*The superscript ext indicates that this weight can be used to extrapolate to the target population.

Table 8: Summary statistics for all weights provided.

Label of weight	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.
w_t1ext	26.346	40.683	49.752	93.062	67.924	656.658
w_t1	0.285	0.440	0.538	1.000	0.734	4.788
w_t1comp	0.272	0.433	0.549	1.000	0.762	4.809
w_t2	0.251	0.418	0.537	1.000	0.781	4.824
w_t2comp	0.310	0.483	0.576	1.000	0.810	4.842
w_t12comp	0.268	0.438	0.575	1.000	0.848	4.838
w_t3	0.232	0.390	0.531	1.000	0.854	4.825
w_t123	0.220	0.379	0.529	1.000	0.904	4.841
w_t4	0.236	0.404	0.530	1.000	0.796	4.837
w_t1234	0.210	0.381	0.514	1.000	0.860	4.855

For further information on weighting please contact [methoden@lifbi.de](mailto:methoden@lifbi.de).

**Acknowledgements** This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Kindergarten, DOI:10.5157/NEPS:SC1:4.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network.

## References

- Aust, F., & Bauer, A. (2014a). *Methodenbericht NEPS Startkohorte 1 - Haupterhebung 2013 B05 (CATI-Erhebung)* (Tech. Rep.). Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.
- Aust, F., & Bauer, A. (2014b). *Methodenbericht NEPS Startkohorte 1 - Haupterhebung 2013 B05 (CAPI-Erhebung)* (Tech. Rep.). Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.
- Bauer, A., Bech, K., Gilberg, R., & Kleudgen, M. (2013). *Methodenbericht NEPS Startkohorte 1 - Haupterhebung 2012/2013 B04* (Tech. Rep.). Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.
- Bauer, A., Hess, D., Malina, A., Aust, F., Blome, A., & Cramer, R. (2015). *Methodenbericht NEPS-Startkohorte 1 - Haupterhebung 2014 B91* (Tech. Rep.). Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.
- Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.
- Blossfeld, H.-P., Roßbach, H. G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (Vol. 14). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Würbach, A., Zinn, S., & Aßmann, C. (2016). *Samples, Weights, and Nonresponse: the Early Childhood Cohort of the National Educational Panel Study (Wave 1 to 3)* (NEPS Survey Paper No. 8). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from [https://www.neps-data.de/Portals/0/Survey%20Papers/SP\\_VIII.pdf](https://www.neps-data.de/Portals/0/Survey%20Papers/SP_VIII.pdf)