

Starting Cohort 6: Adults (SC6)

SUF-Version 3.0.0

Data Manual [Supplement]:

Weighting

*Reiner Gilberg and Martin Kleudgen*

SPONSORED BY THE



Federal Ministry  
of Education  
and Research



Copyrighted Material

University of Bamberg, National Educational Panel Study (NEPS), 96045 Bamberg

<https://www.neps-data.de>

Principal Investigator: Prof. Dr. Hans-Günther Roßbach

Vice Managing Director: Prof. Dr. Sabine Weinert

Executive Director of Research: Dr. Jutta von Maurice

Executive Director of Administration: N.N.

Bamberg, 2013

# **Weighting the Second Wave of the Adult Cohort Sample of the National Educational Panel Study**

Data Release SC6 3.0.0

## **Technical Report**

Reiner Gilberg and Martin Kleudgen

infas Institut für angewandte  
Sozialwissenschaft GmbH

version: July 3, 2013

(This technical report has been prepared under the auspices of the NEPS methods group. It is an excerpt from the general reporting which infas provided to NEPS for documenting sampling and data survey. Its translation from German into English has been conducted by NEPS.)

The weighting of the second wave of the NEPS adult survey (conducted in 2010/2011) is mainly based on the weighting of its first wave (conducted in 2009/2010). The latter has been documented in the corresponding weighting report for the first wave of the adult survey (see infas, 2011; Aßmann & Zinn, 2011). All the basic steps of weighting design, nonresponse adjustment, and calibration are described there. They remain valid for the second wave of NEPS as well.

Because the second NEPS wave is a pure panel sample, the “old” weight from the previous wave could be carried over as a *starting weight* for a repeated run of the weighting routine in each case of the gross sample. Thus, this was also the final calibrated weight from the first NEPS wave ( $cdw\_NEPS\_W1$ )<sup>1</sup> for all participants of the first wave of the NEPS study.

After defining the entrance weights, a nonresponse adjustment of the weights was made once again with the predicted probability values from modeling the participation probability. Unlike in the first NEPS wave, there was no need to model panel participation separately, because only panel-willing cases were used.

## 1. Panel Weighting

The participation probability ( $p_{\text{Participation\_NEPS\_W2}}$ ) in the replication survey (W2) was estimated by way of a logistic regression model, which has a structure similar to the first wave. The basis was formed by all cases that had already participated in the first wave ( $n = 11.362$ ), in the following called *repeaters*.

The weights of the repeaters in the first wave (W1) were then, via participation probability, adjusted as follows, and they formed the temporary part weight ( $dw_{\text{adjusted\_repeater}}$ ):

$$dw_{\text{adjusted\_repeater}} = cdw_{\text{NEPS\_W1}} * (1 / p_{\text{Participation\_NEPS\_W2}})$$

At the same time, this part weight has to be used as the longitudinal weight for the longitudinal study from the first wave to the second wave. The users of the Scientific Use File can obtain it by multiplying the cross-sectional weight from the first wave by the inverse participation probability, which is included in the delivered weighting data set (variable: *prob\_wiederh\_w2*).

All parameters and results of the logistic regression analysis for repeaters are shown in the

---

<sup>1</sup> This is the final weight of the first wave, which has been calibrated according to ISCED97/MZ 2009.

following summary.

Table 1

*Results of the Logit Regression Model Measuring Panel Participation: Repeaters*

<b>Variable</b>	<b>Reference category</b>	<b>Odds's ratio</b>	<b>p-Value</b>
<b>Year of birth</b>	1980–1986		
1970–1979		1.20	0.06
1956–1969		1.38	0.00
1944–1955		1.04	0.74
<b>Gender</b>	Female		
Male		1.04	0.41
<b>Country of birth</b>	Born in Germany		
Born abroad		0.87	0.30
<b>Native language</b>	Non-German		
German		1.39	0.02
<b>Marital status</b>	Single		
Married		1.12	0.16
Separated		1.21	0.07
Widowed		1.20	0.30
<b>Household size</b>	Three and more persons		
One person		0.88	0.15
Two persons		0.89	0.06
<b>School education</b>	Lower secondary education (Realschule)		
No school-leaving qualifications or lower secondary education (Hauptschule)		0.80	0.00
Upper secondary education (Abitur/Fachhochschulreife)		1.36	0.00
Other		1.17	0.13
<b>School education of parents</b>	Lower secondary education (Realschule)		
Lower secondary education (Hauptschule)		1.19	0.01
Upper secondary education (Abitur/Fachhochschulreife)		1.10	0.18
Other		1.11	0.68
<b>Income</b>	€1,501–€3,500		
Up to €1,500		0.92	0.28

More than €3,500	1.05	0.40
------------------	------	------

<b>Federal State</b>	North Rhine-Westphalia	
Schleswig-Holstein	1.25	0.17
Hamburg	1.19	0.37
Lower Saxony	1.01	0.91
Bremen	1.29	0.41
Hesse	1.03	0.74
Rhineland-Palatinate	1.08	0.54
Baden-Württemberg	1.12	0.22
Bavaria	1.20	0.03
Saarland	1.12	0.60
Berlin	0.90	0.44
Brandenburg	1.16	0.35
Mecklenburg-Vorpommern	0.81	0.29
Saxony	1.29	0.05
Saxony-Anhalt	1.61	0.01
Thuringia	1.26	0.16

**BIK categories of municipal size** 500,000 and more inhab.  
(styp 1)

Less than 2,000 inhab.	1.38	0.14
2,000–5,000 inhab.	1.10	0.59
5,000–20,000 inhab.	1.10	0.39
20,000–50,000 inhab.	1.06	0.55
50,000–100,000 inhab. (styp 2/3/4)	1.15	0.21
50,000–100,000 inhab. (styp 1)	1.19	0.40
100,000–500,000 inhab. (styp 2/3/4)	1.00	0.99
100,000–500,000 inhab. (styp 1)	0.99	0.94
500,000 and more inhab. (styp 2/3/4)	0.86	0.13

**Attempts to contact target** 1 to 3 attempts

4 to 6 attempts	0.79	0.00
7 to 10 attempts	0.39	0.00
More than 10 attempts	0.15	0.00

**Pseudo R<sup>2</sup> (McFadden)** 0.1021

**Number of cases** 11,362

If we regard the ALWA survey as a starting basis for the panel of Starting Cohort 6, the ALWA survey could then also be called a *pre-wave*. Thus, there may be panel participants in the second wave who have already participated in a total of three survey sweeps and, consequently, for the first time there can also be cases that have skipped one wave.

These so-called temporary dropouts—that is, cases that did not take part in the second survey sweep or in the first wave—received their “old” weight from ALWA ( $cdw_{ALWA}$ ) as a starting weight. Modifying the weight could not be carried out by the logit model for repeaters but had to be determined via two more models for the nonparticipation probability of the skipped wave as well as for the participation probability of the current wave.

The participation probability for the first wave ( $p_{\text{Participation\_NEPS\_W1}}$ ), given the participation in ALWA, was already estimated in the first wave (see infas, 2011; Aßmann & Zinn, 2011). It has been carried over from the first wave, and the nonparticipation probability for the first wave ( $p_{\text{Nonparticipation\_NEPS\_W1}}$ ) has been derived from it:

$$p_{\text{Nonparticipation\_NEPS\_W1}} = 1 - p_{\text{Participation\_NEPS\_W1}}$$

To determine the participation probability for the second wave, given nonparticipation in the first wave, the participation probability ( $p_{\text{Participation\_tA\_NEPS\_W2}}$ ) for temporary dropouts ( $n = 833$ ) is estimated by another logistic regression model. Here, the number of variables in the model was reduced considerably compared to the repeaters model. This step was necessary because the number of cases of temporary dropouts had been too small and thus would have led to empty cells in the model.

The weights for temporary dropouts from ALWA were adjusted with the resulting probabilities as follows, and they formed the preliminary part weight ( $dw_{\text{adjusted\_tA}}$ ):

$$dw_{\text{adjusted\_tA}} = cdw_{ALWA} * (1 / (p_{\text{Nonparticipation\_NEPS\_W1}} * p_{\text{Participation\_tA\_NEPS\_W2}}))$$

All parameters and results of the logistic regression analysis for temporary dropouts are shown in the following summary.

Table 2

*Results of the Logit Regression Model Measuring Panel Participation: Temporary Dropouts*

<b>Variable</b>	<b>Reference category</b>	<b>Odds's ratio</b>	<b>p-Value</b>
<b>Year of birth</b>	1980–1986		
1970–1979		1.18	0.47
1944–1969		1.13	0.53
<b>Gender</b>	Female		
Male		1.04	0.79
<b>Country of birth</b>	Born in Germany		
Born abroad		0.80	0.43
<b>Federal State</b>	North Rhine-Westphalia		
Schleswig-Holstein		0.57	0.22
Hamburg		0.31	0.15
Lower Saxony		1.40	0.28
Bremen		5.00	0.24
Hesse		0.88	0.71
Rhineland-Palatinate		0.61	0.24
Baden-Württemberg		0.70	0.21
Bavaria		0.80	0.38
Saarland		1.33	0.67
Berlin		0.74	0.53
Brandenburg		0.45	0.13
Mecklenburg-Vorpommern		1.27	0.78
Saxony		0.94	0.89
Saxony-Anhalt		0.38	0.05
Thuringia		0.77	0.65

Table 3

*Results of the Logit Regression Model Measuring Panel Participation: Temporary Dropouts (continued)*

<b>Variable</b>	<b>Reference category</b>	<b>Odds's ratio</b>	<b>p-Value</b>
<b>BIK categories of municipal size</b>	500,000 and more inhab. (styp 1)		
Less than 2,000 inhab.		2.16	0.23
2,000–5,000 inhab.		1.37	0.51
5,000–20,000 inhab.		1.03	0.93
20,000–50,000 inhab.		1.75	0.07
50,000–100,000 inhab. (styp 2/3/4)		3.04	0.00



50,000–100,000 inhab. (styp 1)	1.88	0.36
100,000–500,000 inhab. (styp 2/3/4)	1.55	0.12
100,000–500,000 inhab. (styp 1)	1.22	0.46
500,000 and more inhab. (styp 2/3/4)	1.38	0.33
<b>Attempts to contact target</b>	1 to 3 attempts	
4 to 6 attempts	0.86	0.45
7 to 10 attempts	0.54	0.01
More than 10 attempts	0.15	0.00
<b>Pseudo R<sup>2</sup> (McFadden)</b>	<b>0.1143</b>	
<b>Number of cases</b>	<b>833</b>	

To calculate a common entrance weight for the calibration ( $dw_{inweight}$ ), a convex combination of the modified part weights for repeaters and temporary dropouts was then carried out. This was accomplished by multiplying the part weight by the respective share of repeaters or temporary dropouts in the total sample size ( $n_{repeater}$  = number of cases of repeaters;  $n_{tA}$  = number of cases of temporary dropouts).

For repeaters:

$$dw_{inweight} = dw_{adjusted\_repeater} * (n_{repeater} / (n_{repeater} + n_{tA})).$$

And accordingly for temporary dropouts:

$$dw_{inweight} = dw_{adjusted\_tA} * (n_{tA} / (n_{tA} + n_{repeater})).$$

The resulting common entrance weight was trimmed before calibration at the 5th and 95th percentile in order to limit extreme outliers and, therefore, also the variance of the weights.

Table 4

*Trimmed Weights*

<i>n</i>	Mean	Minimum	Maximum	5th Percentile	95th Percentile
9.321	1	0.0699995	25.25758	0.0833901	13.04432

## 2. Calculation of Calibrated Weights

The trimmed entrance weights that are normalized according to the number of cases form the basis for the following calibration. As in the prior wave, the combination of gender and educational attainment, the combination of year of birth and educational attainment, Federal State, BIK category of municipal size, and country of birth were considered. The distributions mentioned here were adjusted according to the target values of the Microcensus 2010 by GREG-estimators (generalized regression estimation) (see Särndal & Lundström, 2005; Särndal, 2007).

A comparison of the Microcensus distribution 2010 (target) and the unweighted realized sample (actual) does not indicate any major differences. However, there are differences between the realized cases and the basic population, particularly pertaining to attributes of country of birth and education. These differences were equalized through calibration on the Microcensus distribution, which has been carried out as agreed. All weighted (target) and unweighted (actual) distributions that were adjusted through calibration are shown in the following tables.

Table 5

*Gender and Education (ISCED 97)*

<b>Gender and education</b>	<b>Actual distribution net sample</b>		<b>Target distribution basic population (Microcensus 2010)</b>	
	%	abs.	%	abs.
<b>Male</b>				
ISCED 1	0.48	45	1.58	744,484
ISCED 2	1.92	179	4.44	2,095,599
ISCED 3ca	2.74	255	1.94	918,490
ISCED 3b	17.58	1,639	24.06	11,364,786
ISCED 4ab	4.22	393	3.39	1,601,706
ISCED 5b	6.95	648	5.48	2,590,162
ISCED 5a	13.91	1,297	8.36	3,948,233
ISCED 6	1.45	135	0.88	415,862
<b>Female</b>				
ISCED 1	0.63	59	1.89	892,575
ISCED 2	4.66	434	6.57	3,102,092

ISCED 3ca	2.40	224	1.61	762,387
ISCED 3b	22.42	2,090	24.10	11,382,921
ISCED 4ab	6.28	585	4.24	2,002,132
ISCED 5b	1.26	117	4.03	1,901,064
ISCED 5a	12.26	1,143	6.97	3,291,538
ISCED 6	0.84	78	0.45	211,969
<b>Total</b>	<b>100.00</b>	<b>9,321</b>	<b>100.00</b>	<b>47,226,000</b>

*Note.* Educational attainment according to ISCED 97 in Microcensus evaluated by the Federal Statistical Office 2010:

ISCED 1 = no school-leaving qualifications after general leaving certificate or vocational degree, or leaving certificate after no more than 7 years of schooling;

ISCED 2 = leaving certificate of Haupt- or Realschule (lower secondary education) without vocational degree, or leaving certificate of Haupt- or Realschule and semiskilled training, vocational internship, or vocational preparatory year, or no general leaving certificate but semiskilled training, vocational internship, or vocational preparatory year;

ISCED 2ca = entrance qualification for higher education (universities and universities of applied sciences);

ISCED 3b = vocational education and training or vocational degree from full-time vocational school, vocational college, after 1 year at school for health-care professionals;

ISCED 4ab = entrance qualification for higher education (universities and universities of applied sciences) and vocational education and training/vocational degree from vocational school, vocational college, after 1 year at school for health-care professionals;

ISCED 5b = master's (crafts) or technician's certificate or comparable leaving certificate of school of continuing vocational training, certificate after 2 or 3 years at school for health-care professionals, leaving certificate of Fachakademie [type of school in Bavaria offering advanced vocational education and the possibility to obtain the entrance qualification for universities of applied sciences] or Berufsakademie [university of cooperative education] or leaving certificate of school for continuing vocational training in the former GDR or leaving certificate of college of public administration.

ISCED 5a = university of applied sciences, higher education;

ISCED 6 = doctoral degree.

Table 6

*Year of Birth and Education (ISCED 97)*

<b>Year of birth and education</b>	<b>Actual distribution net sample</b>		<b>Target distribution basic population (Microcensus 2010)</b>	
	%	abs.	%	abs.
<b>1975-1986</b>				
ISCED 1	0.28	26	0.80	375,452
ISCED 2	1.15	107	2.65	1,250,839
ISCED 3ca	2.37	221	2.38	1,123,346
ISCED 3b	5.25	489	9.78	4,618,870
ISCED 4ab	3.00	280	3.05	1,441,577
ISCED 5b	0.76	71	1.71	807,122

ISCED 5a	5.17	482	4.07	1,921,433
ISCED 6	0.25	23	0.20	93,361
<b>1965-1974</b>				
ISCED 1	0.10	9	0.94	441,947
ISCED 2	1.08	101	2.46	1,161,747
ISCED 3ca	0.65	61	0.52	246,645
ISCED 3b	10.42	971	12.31	5,815,781
ISCED 4ab	2.86	267	2.25	1,064,096
ISCED 5b	2.22	207	2.60	1,227,183
ISCED 5a	6.77	631	4.22	1,994,299
ISCED 6	0.69	64	0.41	195,302
<b>1956-1964</b>				
ISCED 1	0.25	23	0.85	399,636
ISCED 2	1.53	143	2.53	1,194,871
ISCED 3ca	0.84	78	0.40	190,735
ISCED 3b	13.85	1,291	12.74	6,014,722
ISCED 4ab	3.54	330	1.56	735,693
ISCED 5b	2.65	247	2.71	1,277,624
ISCED 5a	8.86	826	3.53	1,669,186
ISCED 6	0.90	84	0.36	168,533
<b>1944-1955</b>				
ISCED 1	0.49	46	0.89	420,024
ISCED 2	2.81	262	3.37	1,590,234
ISCED 3ca	1.28	119	0.25	120,151
ISCED 3b	10.49	978	13.34	6,298,334
ISCED 4ab	1.08	101	0.77	362,472
ISCED 5b	2.57	240	2.50	1,179,297
ISCED 5a	5.37	501	3.50	1,654,853
ISCED 6	0.45	42	0.36	170,635
<b>Total</b>	<b>100.00</b>	<b>9,321</b>	<b>100.00</b>	<b>47,226,000</b>

*Note.* Please refer to Table 6 for detailed information on ISCED categories.

Table 7

*Federal State and BIK Categories of Municipal Size*

<b>Federal State</b>	<b>Actual distribution net sample</b>		<b>Target distribution basic population (Microcensus 2010)</b>	
	%	abs.	%	abs.
Schleswig-Holstein	2.99	279	3.37	1,593,000
Hamburg	2.04	190	2.30	1,085,000
Lower Saxony	10.28	958	9.50	4,487,000
Bremen	0.62	58	0.82	388,000
North Rhine-Westphalia	22.38	2,086	21.62	10,211,000
Hesse	7.82	729	7.46	3,522,000
Rhineland-Palatinate	4.86	453	4.84	2,284,000
Baden-Württemberg	12.29	1,146	12.95	6,118,000
Bavaria	15.48	1,443	15.40	7,272,000
Saarland	1.51	141	1.25	588,000
Berlin	3.51	327	4.46	2,108,000
Brandenburg	3.25	303	3.20	1,509,000
Mecklenburg-Vorpommern	1.51	141	2.07	979,000
Saxony	5.54	516	5.07	2,394,000
Saxony-Anhalt	3.05	284	2.88	1,358,000
Thuringia	2.86	267	2.82	1,330,000
<b>BIK categories of municipal size</b>				
less than 2,000 inhab.	2.17	202	1.92	909,000
2,000–5,000 inhab.	2.68	250	2.76	1,304,000
5,000–20,000 inhab.	8.05	750	7.81	3,686,000
20,000–50,000 inhab.	12.39	1,155	11.43	5,399,000
50,000–100,000 inhab. (styp 2/3/4)	9.10	848	7.82	3,692,000
50,000–100,000 inhab. (styp 1)	2.02	188	2.23	1,055,000
100,000–500,000 inhab. (styp 2/3/4)	15.72	1,465	14.84	7,007,000
100,000–500,000 inhab. (styp 1)	15.48	1,443	16.16	7,630,000
500,000 and more inhab. (styp 2/3/4)	8.41	784	9.08	4,288,000
500,000 and more inhab. (styp 1)	23.99	2,236	25.95	12,256,000
<b>Total</b>	<b>100.00</b>	<b>9,321</b>	<b>100.00</b>	<b>47,226,000</b>

Table 8

*Year of Birth*

<b>Year of birth</b>	<b>Actual distribution net sample</b>		<b>Target distribution basic population (Microcensus 2010)</b>	
	<b>%</b>	<b>abs.</b>	<b>%</b>	<b>abs.</b>
1944	1.72	160	1.95	919,000
1945	1.43	133	1.42	671,000
1946	1.64	153	1.69	797,000
1947	1.83	171	1.89	892,000
1948	1.75	163	2.03	957,000
1949	2.31	215	2.17	1,023,000
1950	2.10	196	2.25	1,062,000
1951	2.29	213	2.26	1,065,000
1952	2.44	227	2.28	1,075,000
1953	2.11	197	2.30	1,087,000
1954	2.64	246	2.38	1,125,000
1955	2.31	215	2.38	1,123,000
1956	3.30	308	2.48	1,170,000
1957	3.11	290	2.56	1,210,000
1958	3.27	305	2.57	1,215,000
1959	4.14	386	2.69	1,272,000
1960	3.80	354	2.80	1,323,000
1961	3.48	324	2.82	1,332,000
1962	3.80	354	2.80	1,323,000
1963	3.68	343	2.94	1,389,000
1964	3.84	358	3.00	1,417,000
1965	3.89	363	3.02	1,428,000
1966	3.45	322	3.11	1,470,000
1967	2.97	277	2.94	1,388,000
1968	2.84	265	2.83	1,336,000
1969	2.48	231	2.71	1,278,000
1970	2.40	224	2.59	1,221,000
1971	1.97	184	2.41	1,139,000
1972	1.91	178	2.18	1,031,000
1973	1.49	139	1.98	933,000
1974	1.37	128	1.95	923,000
1975	1.32	123	1.97	931,000
1976	1.23	115	1.99	940,000
1977	1.46	136	2.01	950,000
1978	1.35	126	2.04	962,000
1979	1.47	137	2.03	957,000
1980	1.38	129	2.18	1,031,000
1981	1.37	128	2.12	1,003,000
1982	1.46	136	2.15	1,013,000
1983	1.71	159	2.10	991,000

1984	1.46	136	2.02	953,000
1985	1.65	154	1.98	935,000
1986	2.36	220	2.05	966,000

Table 9

*Country of Birth*

<b>Country of birth (migration experience)</b>	<b>Actual distribution net sample</b>		<b>Target distribution basic population (Microcensus 2010)</b>	
	%	abs.	%	abs.
Born abroad (with migration experience)	8.30	774	17.48	8,257,000
Born in Germany (without migration experience)	91.70	8,547	82.52	38,969,000
<b>Total</b>	<b>100.00</b>	<b>9,321</b>	<b>100.00</b>	<b>47,226,000</b>

The statistical values, the measure of effectiveness, and the effective number of cases of the calibrated weight are displayed in the following table. The measure of effectiveness E is based on the weighting factor's variance—the larger its variance, the stronger the influence on the sampling error as well. The measure of effectiveness specifies—expressed as a percentage of the realized number of cases—how large the effective number of cases is for a passive attribute that does not correlate with the active attributes in the case of weighting (thus, it is a worst-case criterion, because the active attributes used for calibration should be defined in such a way that will result in a high correlation with the passive attributes—that is, attributes that are not used for calibration).

The effective number of cases is consistent with the number of respondents that would have produced the same sampling error under a simple random sampling design given the variance of the attribute in the sample. The measure of effectiveness illustrates the proportion from  $n$  to  $n'$  as a percentage. The effectiveness is calculated as follows:

$$E = (n' / n) * 100, \quad \text{where: } n' = (\sum_i g_i)^2 / \sum_i g_i^2$$

Table 11

*Trimmed Weights*

	<b>Mean</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Effectiveness</b>	<b>Effective number of cases</b>
Weight Wave 2 (ISCED_gew_w2)	1	0.0639018	11.81254	55.67%	5,189

In consideration of the multilevel weighting concept with design weighting, nonresponse adjustment, and calibration, and the resulting variance lag of the weights, the located measure of effectiveness has to be considered as good.

## References

- Aßmann, C., & Zinn, S., (2011). Starting Cohort 6: Adults (SC6), SUF-Version 1.0.0 [Data Manual, Supplement C], Weighting. Retrieved from: [https://www.neps-data.de/Portals/0/Neps/Datenzentrum/Forschungsdaten/SC6/1-0-0/SC6\\_1-0-0\\_DM\\_SUP\\_C\\_Weighting\\_EN.pdf](https://www.neps-data.de/Portals/0/Neps/Datenzentrum/Forschungsdaten/SC6/1-0-0/SC6_1-0-0_DM_SUP_C_Weighting_EN.pdf)
- infas (2011). Etappe 8 Erwachsene, Methodenbericht zur Haupterhebung, 1. Welle 2009/2010, Bonn.
- Särndal, C. E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. New York: Wiley.
- Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey Methodology (Statistics Canada)*, 33(2), 99–119.