NEPS
National Educational Panel Study

NEPS Working Papers

Christian Aßmann, Hans Walter Steinhauer, Sabine Zinn, & Solange Goßmann

Sampling and Weighting the Kindergarten Cohort Sample of the National Educational Panel Study

NEPS Working Paper No. 29

Bamberg, October 2013

SPONSORED BY THE
Federal Ministry of Education and Research

Working Papers of the German National Educational Panel Study (NEPS)
at the University of Bamberg

The NEPS Working Papers publish articles, expertises, and findings related to the German National Educational Panel Study (NEPS).
The NEPS Working Papers are edited by a board of researchers representing the wide range of disciplines covered by NEPS. The series started in 2011.

Papers appear in this series as work in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

Any opinions expressed in this series are those of the author(s) and not those of the NEPS Consortium.

The NEPS Working Papers are available at
http://www.uni-bamberg.de/neps/publikationen/neps-working-papers/

Contact: German National Educational Panel Study (NEPS) − University of Bamberg − 96045 Bamberg − Germany − contact.neps@uni-bamberg.de

Sampling and Weighting the Kindergarten Cohort Sample of the National Educational Panel Study

Christian Aßmann, Hans Walter Steinhauer, Sabine Zinn, and Solange Goßmann

E-mail address of lead author:
christian.assmann@uni-bamberg.de

Bibliographic data:

Aßmann, C., Steinhauer, H.W., Zinn, S. and S. Goßmann (2013). *Sampling and weighting the Kindergarten cohort sample of the National Educational Panel Study* (NEPS Working Paper No. 29). Bamberg: University of Bamberg, National Educational Panel Study.

Sampling and Weighting the Kindergarten Cohort Sample of the National Educational Panel Study

Abstract

The National Educational Panel Study (NEPS) surveys a cohort sample of Kindergarten children (Starting Cohort 2, SC2) and releases corresponding Scientific Use Files, see doi:10.5157/NEPS:SC2:2.0.0 for the actual version. This paper gives details on the applied indirect sampling procedure, the derivation of design weights, their successive adjustments, and the derivation of panel weights. Sampling of Kindergartens is based on an indirect approach because a sampling frame for direct sampling is available neither for Kindergarten children nor for Kindergarten institutions. Starting out with a sample of elementary schools, we draw on Kindergartens that were supplying these elementary schools with first-grade students in school year 2009/2010. After correcting for institutional and individual unit nonresponse each panel cohort member is assigned an adjusted design weight. Relying on these, for all panel waves cross-sectional and longitudinal weights are computed. Furthermore, weights are given for subgroups of the panel cohorts that are of special interest in our analysis. This concerns particularly the group of children continually taking part in the successive waves of the survey and the group of children and parents participating jointly. Analysis via probit regresions highlights the factors influencing the participation probability.

Keywords

sampling, sampling weights, response propensity analysis, Kindergarten cohort sample, National Educational Panel Study

1 Introduction

This paper documents the sampling procedures and the derivation of sampling weights for the sample of Kindergarten children cohort (Starting Cohort 2, SC2) of the National Educational Panel Study (NEPS), see doi:10.5157/NEPS:SC2:2.0.0 for the actual scientific use file. Because a sampling frame was available neither for Kindergarten children nor for Kindergarten institutions, the NEPS Kindergarten sample was established on the basis of an indirect sampling approach. The corresponding sampling design was based on elementary schools and their links to the Kindergartens (i.e., children transferring from Kindergarten to elementary school). Each sampled Kindergarten child was assigned a sampling weight, which was then adjusted for unit nonresponse. Nonresponse adjustments were based on analysis of participation descisions via binary limited dependent variable models, i.e. binary probit regressions. Along the distinct panel waves, for all participating children cross-sectional and longitudinal weights are provided. Cross-sectional weights are assigned to children relying on their participation in the different panel waves. Longitudinal weights are given to the group of children who have continually participated in the successive panel waves. As the information on children is enriched by interviewing one of their parents, additional weights are provided for the group of participating Kindergarten children for whom an interview with one parent was conducted. The remainder of this paper details the sampling procedure in Section 2. Section 3 details the derivation of the pure design weights. Section 4 describes the sample-specific response propensity analysis applied to SC2. At this point, we also explore in detail the derivation of weights accounting for the joint participation of children and parents. Section 5 gives insights into the trimming procedure that was applied to the weights to increase the statistical efficiency of weighted analysis. Finally, in Section 6 a summary of the provided weights is given and Section 7 concludes.

2 Sampling

2.1 Population

The target population of the sample for SC2 focuses on children attending Kindergartens in Germany in the school year of 2010/2011 who are expected to begin schooling in the school year of 2012/2013. Hence, these children are approximately at the age of 4, as children in Germany are obliged to start attending elementary school between the age of 5 to 7, according to their date of birth, see Berendes et al. (2011) for details.

2.2 Structure of the Sample

Direct multistage sampling approaches were not applicable in this case as no frame was available for neither Kindergarten children nor Kindergarten institutions. However, Kindergartens and elementary schools are linked by children transferring from Kindergartens to elementary schools.[1] This link can be used to get access to the population of Kindergarten children by using an indirect sampling approach as introduced by Lavallée (2007). Hence, we established a sample of elementary schools to access the Kindergarten population. The elementary schools were drawn using a systemic probability proportional to size sampling. Then Kindergartens were sampled by probability proportional to size sampling without replacement. Within one Kindergarten all children were asked to participate. The procedure applied to sample Kindergartens resembles a two-stage indirect sampling approach. The principles of the implemented indirect sampling

[1] Note that 92% of all children aged between 3 and 5 years visit a day-care facility before attending elementary school, see Statistisches Bundesamt (2010) and Statistisches Bundesamt (2011).

approach developed by Lavallée (2007) can be stated as follows, see also Kiesl (2010) and Aßmann et al. (2011). Assume a population U_P that is linked in a defined manner to another population U_K. Further, there exists a sampling frame for the population U_P so that a sample s_P can be drawn. Via the definition of a link θ_{pk} between elements $p \in U_P$ and $k \in U_K$ relating to a certain measurement and a reference year access is gained to a sample s_K in the linked population U_K. Extensions of indirect sampling techniques to multistage sampling as well as their properties are discussed in Lavallée (2007) for two-stage sampling and in Kiesl (2010) for two-stage and three-stage sampling.

3 Derivation of Design Weights

For the indirect sampling of Kindergartens, elementary schools needed to be sampled in the first place. This sample of elementary schools then, in turn, served to establish a sample of first grade children in the school year 2012/2013. Therefore, let U_P be the population of $M_P(= 16,824)$ elementary schools with at least one class in Grade 1. A sample of $m_P(= 400)$ elementary schools was drawn via a systematic proportional to size (pps) scheme with the number S^1 of students in Grade 1 as measure of size, whereas the total number of students in Grade 1 in school year 2008/09 (i.e., the total measure of size) was

$$S^1 = \sum_{p=1}^{M_P} S_p^1 = 710,539, \tag{1}$$

where S_p^1 denotes the number of children in Grade 1 in school p, $p = 1, \ldots, M_P$. The resulting inclusion probabilities of elementary schools are given as

$$\pi_p = m_P \cdot \frac{S_p^1}{S^1}, \quad p = 1, \ldots, m_P. \tag{2}$$

Whereas the m_P schools served as the basis for the sample of first-grade children, only a simple random subsample of $\tilde{m}_P(= 212)$[2] schools forms the basis for the Kindergarten sample with corresponding inclusion probabilities $\tilde{\pi}_p = \frac{\tilde{m}_p}{m_p}\pi_p$. Based on the sample of $\tilde{m}_P$ elementary schools, each of these schools was asked to list Kindergartens that were supplying them with children in first grade in the school year of 2009/2010. These lists form the sets Ω_p, $p = 1, \ldots, \tilde{m}_P$, containing $|\Omega_p| = K_p$ Kindergartens for each sampled school, and thus relate to a subset of the population of Kindergartens U_K. Note that the sets Ω_p need not to be pairwise disjoint. This subset can be used to provide a sample of Kindergartens and Kindergarten children, respectively. The link function θ_{pk} between Kindergartens and schools was defined as the number of children transferring from Kindergarten k to school p, that is, $\theta_{pk} > 0$ if children moved from Kindergarten k to elementary school p and zero else. If all listed Kindergartens were surveyed, this setup would render to the following indirect weights of a Kindergarten k given as[3]

$$w_k = \sum_{p=1}^{\tilde{m}_P} \frac{\theta_{pk}}{\theta_{+k}\tilde{\pi}_p}, \quad \text{with} \quad \theta_{+k} = \sum_{p=1}^{M_P} \theta_{pk}. \tag{3}$$

[2] Due to the intended sample size of approximately 3,000 Kindergarten children, 212 out of the 400 sampled elementary schools have been found sufficient to provide a list of Kindergartens, see Aßmann et al. (2012) for details.

[3] If a Kindergarten sends children only to a single elementary school, w_k corresponds to the Horwitz-Thompson weight given as $\frac{1}{\tilde{\pi}_p}$.

For illustration of this issue consider the population total of any variable Y in population U_K to be written as

$$t_Y = \sum_{k \in U_K} y_k = \sum_{k \in U_K} \left(\sum_{p \in U_P} \frac{\theta_{kp}}{\theta_{+k}} \right) y_k, \quad \text{where} \quad \sum_{p \in U_P} \frac{\theta_{kp}}{\theta_{+k}} = 1. \tag{4}$$

That is, the total of some variable in population U_K can be written as a total of the fractionized variable $\sum_{k \in U_K} \frac{\theta_{kp}}{\theta_{+k}} y_k$. This approach relates to the weight share method of Ernst (1989).

As the number of linked Kindergartens per school as well as the resulting number of Kindergarten children in the sample was prohibitively large to allow for a complete survey, a subsample of Kindergartens was added, that is, a subsample $s_{K|s_P}$ of all Kindergartens linked to sampled schools was drawn. Following Kiesl (2010), sampling of Kindergartens was done by school, using a probability proportional to size sampling without replacement. The measure of size was defined via the number of children transferring from Kindergarten k to school p. The subsamples of size k_p were drawn from each set Ω_p according to the following rule,

$$k_p = \begin{cases} 1 \text{ if } 0 < K_p \leq 6 \\ 2 \text{ if } 6 < K_p \leq 11 \\ 3 \text{ if } 11 < K_p \leq 19 \\ 4 \text{ else.} \end{cases} \tag{5}$$

As noted by Kiesl (2010), the resulting (conditional) inclusion probabilities of Kindergartens depend on the sample of elementary schools s_P, that is, for a given Kindergarten this probability can differ due to different subsamples $s_{K|s_P}$ and even different samples s_P. Calculation of inclusion probabilities is thus not feasible but allows for the construction of weights providing an unbiased estimator of population totals. According to Kiesl (2010), the resulting weights are given as

$$w_k = \sum_{p \in s_P} \frac{\theta_{pk}}{\theta_{+k}} \cdot \frac{I(k \in \Omega_p)}{\tilde{\pi}_p \cdot \pi_{k \in \Omega_p}}, \tag{6}$$

where $\pi_{k \in \Omega_p}$ refers to the sampling probability of Kindergarten k listed by elementary school p and $I(k \in \Omega_p)$ is 1 if $k \in \Omega_p$ and 0 else. As all Kindergarten children of the defined age group[4] were asked to participate, these weights correspond directly to children's weights. If a Kindergarten entered the sample only via a single elementary school, that is, it sent children only to one particular sampled elementary school, no summation over all elementary schools is necessary and the weights can be rendered directly to a two-stage sampling approach.

4 Nonresponse Adjustment of Weights

Nonresponse accurring at the different stages of the sampling process was handled as follows. The refusal of elementary schools to participate in NEPS realizes before any fieldwork in Kindergarten institutions takes place and can be considered as independent of the surveying of Kindergarten children. Therefore, a replacement strategy was adapted to cope with nonresponse on the level of elementary schools. As the sampling of schools was based on implicit stratification according to Federal States, regional classification, and organizing institutions, a nonparticipating school was replaced by a school identical to the originally drawn school with regard to their implicit

[4] That is, children at about 4 years of age in school year 2010/2011.

stratification variables. The sampled Kindergartens may also refuse participation in the survey. To address this problem, for each sampled Kindergarten a set of replacement Kindergartens was defined upon the same Kindergarten list.[5] Definition was based on the smallest deviation between sampled Kindergarten and replacement Kindergartens with respect to the number of children transferring between Kindergarten and school. When a Kindergarten refused participation or failed to give explicit consent within a defined time range, the defined replacement Kindergartens were asked to participate. For each Kindergarten originally sampled there were two replacement Kindergartens, if $K_p \geq 3$.

Table 1: Response propensity model to adjust the initial sample to the panel cohort sample of Kindergarten children.

Variable	Parameter Estimate
Intercept	0.381***
Age group	
younger than median age (ref)	
older than/equal to median age	−0.046
Gender	
male (ref)	
female	−0.004
Language spoken at home	
no German (ref)	
German	0.439***
Place of living	
with both parents (ref)	
with one parent or others	0.110
Occupational status of parents	
both employed(ref)	
one employed	−0.096
none employed	−0.088
Missing indicator	
children's environment	−0.472***
personal characteristics	−1.852***
Random intercept	
on Kindergarten level	0.431
Sample size	4515

Notes: (i) The median age of children refers to the whole sample. (ii) The flags ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively.

4.1 Adjusting the Initial Sample to the Panel Cohort Sample

To establish a sample of children, all corresponding parents are asked to provide consent for them and their childdren to participate in the survey. Thus panel consent for children and panel consent of parents are coupled. All children and corresponding parents that provided panel consent form the panel cohort sample. To address potential selectivity within the panel cohort sample at the level of children, models regressing the panel-consent status on information

[5] Such processing has been feasible because the Kindergartens listed by a single school are similar with respect to regional aspects.

available for the gross sample of the Kindergarten children were estimated. For this purpose, the *glmer* function provided by *lme4* package (Bates et al., 2013) in R (R Development Core Team, 2013) was used with a binary probit link specification. The set of variables available includes year of birth, gender, language spoken in household, residence, and occupational status of the parents. In addition, a Kindergarten-specific random effect was considered to allow for a potential correlation among children attending the same Kindergarten. The empirical results for the corresponding random intercept model on the determinants considered are shown in Table 1. The results suggest that children speaking German at home have a higher propensity to participate in the survey. Effects of the opposite direction were found for children with information missing concerning personal characteristics (i.e., gender and year of birth) and information missing concerning the child's environment (i.e., language spoken at home, residence status, and occupational status of parents).

Table 2: Number of children participating in Wave 1 and 2, as well as number of children temporarily dropping-out from the sample.

Sample	Participants	Temporary Drop-outs	Final Drop-outs
Wave 1	2971	25	0
Wave 2	2763	233	0
Wave 1 & 2	2739	257	0

Table 3: Parameter estimates of the response propensity models for the participation of children in Wave 2 as well as for the joint participation of children and one parent in Wave 2.

Variable	Parameter Estimate		
	Only Children	Children & Parent	
		Children	Parent
Intercept	1.725***	1.458***	0.023
Age group			
younger than median age (ref.)			
older than/equal to median age	−0.191*	−0.249	−0.048***
Gender			
male (ref.)			
female	−0.012	−0.005	0.028
Language spoken at home			
no German (ref.)			
German	0.113	0.172***	0.818***
Place of living			
with both parents (ref.)			
with one parent or others	−0.180	−0.189***	−0.737*
Variance of random intercept *on Kindergarten level*	0.378		
Correlation between *children & parents*		0.219**	
Sample Size		2996	

Notes: (i) The median age of children refers to the whole sample. (ii) The flags ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively.

Table 4: Parameter estimates of the response propensity models for the participation of children in Wave 1 and 2 as well as for the joint participation of children and one parent in Wave 1 and 2.

Variable	Parameter Estimate		
	Only Children	Children & Parent	
		Children	Parent
Intercept	1.560***	1.374***	−0.123
Age group			
younger than median age (ref.)			
older than/equal to median age	−0.190*	−0.245***	−0.042
Gender			
male (ref.)			
female	−0.045	−0.035	0.055
Language spoken at home			
no German (ref.)			
German	0.196*	0.228**	0.794***
Place of living			
with both parents (ref.)			
with one parent or others	−0.202*	−0.197*	−0.741***
Variance of random intercept *on Kindergarten level*	0.295		
Correlation between *children & parents*		0.177**	
Sample Size		2996	

Notes: (i) The median age of children refers to the whole sample. (ii) The flags ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively.

Table 5: Number of parents participating in Wave 1 and Wave 2.

Sample	Participants	Temporary Drop-outs	Final Drop-outs
Wave 1	2340	452	204
Wave 2	2111	564	321
Wave 1 & 2	1963	773	260

4.2 Adjustments for Children's Participation in Wave 1 and in Wave 2

In the context of sample weight adjustment along the waves, two views on the sample and, therefore, two types of weights have to be differentiated: Cross-sectional weights relate to the sample of a specific wave and longitudinal weights refer to the sample of persons who continually participate in the successive panel waves. For Wave 1, by definition, cross-sectional weights and longitudinal weights are identical. Initially with Wave 2, they may start to differ.

The panel cohort sample differs from the sample of children actually participating in the successive waves of the panel. Dropouts occurred due to withdrawn declarations of participation consent or due to other (unswayable) reasons, for example, because of illness or bad weather conditions. In the first case, persons were specified as final dropouts. In the latter case, they were defined as temporary dropouts. Table 2 shows the number of children participating in Wave 1 and Wave 2, as well as the number of children temporarily dropping-out. In both waves, no children finally dropped out of the sample.

Table 6: Parameter estimates of the response propensity models for the joint participation of children and parents in Wave 1.

Variable	Parameter Estimate	
	Children	Parents
Intercept	2.223***	0.386***
Age group		
younger than median age (ref.)		
older than/equal to median age	−0.040	−0.048
Gender		
male (ref.)		
female	−0.162	0.061
Language spoken at home		
no German (ref.)		
German	0.404***	0.624***
Place of living		
with both parents (ref.)		
with one parent or others	−0.130*	−0.621***
Correlation between *children & parents*	0.027	
Sample Size	2996	

Notes: (i) The median age of children refers to the whole sample. (ii) The flags ***, **, and * denote significance at the 0.1%, 1%, and 5% level, respectively.

In Wave 1, the number of children temporarily dropping-out from the sample was so small (25 cases) that, for an accordant adjustment of sampling weights, an unconditional modeling was advisable. In accordance therewith, the design weights of the Kindergarten children were adopted only on the basis of the number $n_{np} = 25$ of temporary dropouts and the number $n_p = 2971$ of participants. We obtained the nonresponse adjusted weights of Wave 1 by multiplying the design weights w_k of Kindergarten children with the factor $(n_p + n_{np})/n_p$. In the cross-sectional sample of Wave 2, $n = 233$ children temporarily dropped-out from the sample and, in the longitudinal sample of Wave 1 and 2, $n = 257$ children. To assess whether the propensity of children to drop-out was driven by selection processes, we used propensity score models with probit specification. We studied the effects of age and gender of children, their place of living, and the language spoken at home. In addition, we accounted for the two-level structure of the data by specifying a random effect on the level of Kindergartens. Apart form few values missing in the variable 'language spoken at home' (eight cases) and in the variable 'place of living' (one case) the data at hand was complete. The few missing values were imputed by hot deck imputation. Besides the age of the child, none of the factors considered showed a significant effect on the propensity of a child to participate in Wave 2 (cp. Table 3). In contrast, the propensity of children to participate in both Waves in addition is significantly influenced by the language spoken at home and whether the child lives with both parents or not (cp. Table 4).

From these results, nonresponse adjusted design weights for the samples of Wave 1 and 2 are achieved by multiplying the inverse of the predicted response propensities of the distinct models with the design weights w_k of the Kindergarten children.

4.3 Adjustment for the Joint Participation of Children and One Parent in Wave 1 and in Wave 2

To facilitate analyzing children data with information provided by one of their parents, we provide accordant 'group' weights. These weights refer to the subset of children for whom an interview with one parent could be realized. Generally, parents might refuse to participate in the survey, they might withdraw their panel consent, and they might temporarily drop-out from the survey. Subsequently, the first two situations are summarized under the category 'final drop-outs'. Table 5 gives the corresponding numbers of parents participating in Wave 1 and Wave 2. Analyzing the group of parents who temporality dropped-out from the sample and the group of parents who finally dropped-out did not reveal any significant differences between both groups. Therefore, subsequently all parents who dropped-out–for what reason ever–are considered as one group. To compute sampling weights for the children with parent information available, we use response propensity models with bivariate probit specification. This kind of approach allows modeling the conjoint–possibly, correlated–participation decision of children and parents. The accordant model setup requires to define two equations: one for the participation propensity of the child and one for the participation propensity of the parent. For model estimation, the *zelig* function with a bivariate binary probit link of the *R* package *ZeligChoice* (Owen et al., 2012) was employed. As factors describing the participation propensity of a child and of its parents we used in either case the age and the gender of the child, its place of living, as well as the language spoken at home. The few missing values in the variables 'language spoken at home' (one case) and 'child's place of living' (eight cases) were replaced by hot deck imputation. In Wave 1, the language spoken at home and whether or not a child lives with his/her parents showed a significant influence on the joint participation propensity of children and parents. In Wave 2 as well as in Wave 1 and 2, in this context additionally the age of the child played an important role. Table 6 as well as the second columns of the Tables 3 and 4 give the results of the corresponding analyses. The predicted response propensities p_k of the estimated models were used to compute the accordant sampling weights for the child-parent pairs by multiplying p_k^{-1} with the design weights w_k of the Kindergarten children.

5 Weight Trimming

To possibly increase the statistical efficiency of weighted analysis, the adjusted weights were trimmed. The general goal of weight trimming is to reduce sampling variance and, at the same time, to compensate for potential increase in bias. Trimming was performed using the so-called "Weight Distribution" approach (Potter, 1990). Here, design weights are assumed to follow an inverse beta distribution with a cumulative distribution function F_w. Parameters of the sampling weight distribution are estimated using the sampling weights, and a trimming level τ is computed whose occurrence probability is 1%, that is, $1 - F_w(\tau) = 0.01$. Sampling weights in excess of τ are trimmed to this level and the excess is distributed among the untrimmed weights. The parameters for the sampling weight distribution are then again estimated using the trimmed adjusted weights, and a revised trimming level $\tilde{\tau}$ is computed. The trimmed adjusted weights are compared to the revised level $\tilde{\tau}$. If any weights are in excess of $\tilde{\tau}$, they are trimmed to this level, and the excess is distributed among the untrimmed weights. This procedure is iteratively repeated until no weights are in excess of a newly revised trimming level. To ease statistical analysis, the trimmed sampling weights are standardized with mean 1.

6 Summary of Weights

The NEPS provides various kinds of sampling weights for the Kindergartens and the children as part of SC2. It provides sampling weights for the participants of the panel cohorts as well as for the participants of the all wave conducted for SC2. All Kindergarten and children weights are adjusted for unit nonresponse. As an addition, sampling weights are given for the sample for which an additional interview with one parent is available. All kinds of weights are provided 'purely' and in a trimmed and standardized form.[6] The following table lists the types of weights provided for SUF release version 2.0.0:

Type of weight	Label
Weights of Kindergartens	`w_i`
Weights of children as part of the panel cohort	`w_t`
Weights of children participating in Wave 1	`w_t1`
Weights of children participating in Wave 1 & realized parent interview in Wave 1	`w_tp1`
Weights of children participating in Wave 2	`w_t2`
Weights of children participating in Wave 2 & realized parent interview in Wave 2	`w_tp2`
Weights of children participating in Wave 1 and Wave 2	`w_t12`
Weights of children participating in Wave 1 and Wave 2 & realized parent interview in Wave 1 and 2	`w_tp12`

The trimmed and standardized form of the weights are marked by the suffix `_std`. Subsequently, the distribution of the different kinds of trimmed and standardized weights is summarized:

Type of weight	Number of units	Min.	Lower quart.	Median	Mean	Upper quart.	Max.
`w_i_std`	279	0.11	0.56	0.79	1	1.17	3.58
`w_t_std`	2996	0.10	0.50	0.75	1	1.18	3.91
`w_t1_std`	2971	0.10	0.50	0.75	1	1.18	3.91
`w_tp1_std`	2322	0.10	0.50	0.73	1	1.19	3.97
`w_t2_std`	2763	0.10	0.54	0.80	1	1.27	4.22
`w_tp2_std`	1986	0.13	0.72	1.01	1	1.78	5.10
`w_t12_std`	2739	0.11	0.55	0.81	1	1.27	4.24
`w_tp12_std`	1832	0.15	0.80	1.16	1	1.89	5.23

7 Conclusion

This paper provides an overview concerning the sampling procedures applied for establishing a cohort sample of Kindergarten children. Further nonresponse adjustments based on selectivity analyses taking the cluster structure and correlation between parents and children into account are discussed. The results highlight the factors influencing the participation decision, where typical factors like language spoken at home have an impact on the participation probability.

[6] The sampling weights are standardized in such a way that their mean is 1.

References

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., ... Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: challenges and solutions. In H.-P. Blossfeld, H.-G. Roßbach, & J. v. Maurice (Eds.), *Education as a lifelong process* (Vol. 14, pp. 51–65). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0181-8

Aßmann, C., Steinhauer, H. W., & Rässler, S. (2012). Aspekte der Stichprobenziehung in der erziehungswissenschaftlichen Forschung. In S. Maschke & L. Stecher (Eds.), *Enzyklopädie der Erziehungswissenschaft Online (EEO), Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung* (pp. 1–15). Weinheim und Basel: Beltz Juventa. doi: 10.3262/EEO07120215

Bates, D., Maechler, M., & Bolker, B. (2013, April). *lme4: Linear mixed-effects models using S4 classes.* http://cran.r-project.org/web/packages/lme4/index.html. Retrieved from `http://CRAN.R-project.org/package=lme4` (Version 0.999999-2)

Berendes, K., Fey, D., Linberg, T., Wenz, S. E., Roßbach, H.-G., Schneider, T., & Weinert, S. (2011). Kindergarten and elementary school. In H.-P. Blossfeld, H.-G. Roßbach, & J. v. Maurice (Eds.), *Education as a lifelong process* (Vol. 14, pp. 203–216). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/s11618-011-0187-2

Ernst, L. (1989). Weighting issues for longitudinal household and family estimates. *Panel Surveys*, 135-159.

Kiesl, H. (2010). *Selecting kindergarten children by three-stage indirect sampling: JSM 2010, Session #420, Section on Survey Research Methods.*

Lavallée, P. (2007). *Indirect sampling.* New York and NY: Springer.

Owen, M., Imai, K., Lau, O., & King, G. (2012, October). *ZeligChoice: Add-on package for Zelig.* http://cran.r-project.org/web/packages/ZeligChoice/index.html. Retrieved from `http://CRAN.R-project.org/package=lme4` (Version 0.7-0)

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In American Statistical Association (Ed.), *Proceedings of the Survey Research Methods Section* (pp. 225–230). Retrieved 25.09.2012, from `http://www.amstat.org/sections/srms/Proceedings/papers/1990_034.pdf`

R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing.* Vienna and Austria. Retrieved from `http://www.R-project.org/`

Statistisches Bundesamt. (2010). *Bevölkerung und Erwerbstätigkeit, Bevölkerungsfortschreibung* [Fachserie 1, Reihe 1.3].

Statistisches Bundesamt. (2011). *Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03.2011* [Statistiken der Kinder- und Jugendhilfe].