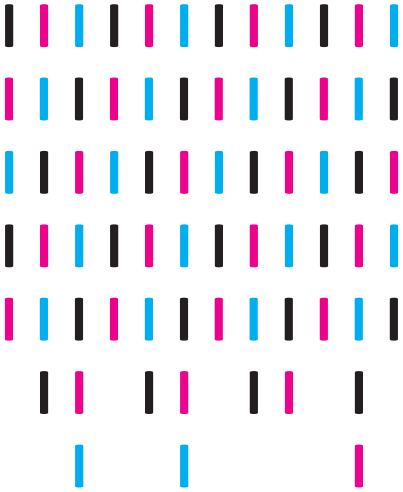

NEPS *SURVEY PAPERS*

Ariane Würbach, Uta Landrock,
Thorsten Schnapp, Johannes Ziesmer
and Michael Bergrab

NEPS TECHNICAL REPORT FOR WEIGHTING – THE ADDITIONAL CORONA SURVEY OF THE NATIONAL EDUCATIONAL PANEL STUDY

NEPS *Survey Paper* No. 89
Bamberg, Semptember 2021

NEPS
National Educational Panel Study

Survey Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS *Survey Paper* series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS *Survey Papers* are edited by a review board consisting of the scientific management of LIfBi and NEPS.

The NEPS *Survey Papers* are available at www.neps-data.de (see section “Publications“) and at www.lifbi.de/publications.

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg − Germany − contact@lifbi.de

NEPS Technical Report For Weighting – The Additional Corona Survey Of The National Educational Panel Study

Ariane Würbach[1], *Uta Landrock*[1], *Thorsten Schnapp*[1], *Johannes Ziesmer*[1], *Michael Bergrab*[1,2]

[1]*Leibniz Institute for Educational Trajectories, Germany*
[2]*University of Bamberg, Germany*

E-mail address:

statistik@lifbi.de

Bibliographic data:
Würbach, A., Landrock, U., Schnapp, T., Ziesmer, J., & Bergrab, M. (2021). *NEPS Technical Report for Weighting – the Additional Corona Survey of the National Educational Panel Study* (NEPS *Survey Paper* No. 89). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://doi.org/10.5157/NEPS:SP89:1.0.

NEPS Technical Report for Weighting – the Additional Corona Survey of the National Educational Panel Study

Abstract

The National Educational Panel Study (NEPS) launched an additional Corona survey to investigate the short, medium and long-term effects of the SARS-CoV-2 pandemic on educational pathways and skill development over the life course. The first part of the additional survey was administered as a web-based survey (CAWI) in May and June 2020. The second part is a new module, NEPS-C, integrated into the questionnaires of the main studies capturing and perpetuating the items of the former Corona-CAWI. Several questions regarding direct consequences on school and work but also on daily life are asked, along with questions capturing the perception on these changes as well as adaptation to the new circumstances. These additional data will enable researchers to trace individual responses to the pandemic and to explore differences in response patterns. All panel members of the Starting Cohorts 3 to 6 as well as the parents from the target persons of Starting Cohort 2 were invited within a panel support measure to complete the Corona-CAWI. Starting Cohort 1 was supplied with a corresponding NEPS-C module in the computer-assisted personal interview (CAPI-by-Phone) instrument of Wave 9. This report documents the figures of the panel samples at start, the current sample sizes and the number of target persons (or parents) being invited, complemented by those participating. In a first step, the structure of the panel samples and their changes in distribution are explored. Details on selectivity are presented, with regard to margin distributions from the Microcensus 2019. In a second step, the response propensity models for participation in the additional Corona survey are described. The weighting procedure including the post-stratification is illustrated in a third step. The report concludes with statistics for all weighting variables supplied and delivers some remarks on appropriate usage of weights.

Keywords

panel surveys, NEPS, SARS-CoV-2, additional Corona survey, non-response, cross-sectional weighting, longitudinal weighting, post-stratification

Contents

1. Synopsis

Table 1: Synopsis for weighting data of the additional Corona survey.

Study modes	
SC2 – SC6	web-based survey (CAWI)
SC1	computer-assisted personal interview (CAPI, CAPI-by-Phone)
Field periods	
SC2 – SC6	13.05.2020 – 22.06.2020
SC1	19.06.2020 – 30.08.2020
Gross samples	
SC2 – SC6	SC2 = 5,073 (54.3%)
	SC3 = 4,917 (59.1%)
	SC4 = 7,703 (46.9%)
	SC5 = 8,894 (49.7%)
	SC6 = 7,976 (46.5%)
	CAWI Total = 34,563 (47.6% of panel cohort at start)
SC1	at Wave 9 = 2,257 (65.8% of panel cohort at start)
Realized samples	
SC2 – SC6	SC2 = 1,587 (31.3%)
	SC3 = 1,031 (21.0%)
	SC4 = 1,700 (22.1%)
	SC5 = 2,859 (32.1%)
	SC6 = 2,678 (33.6%)
	CAWI Total = 9,855 (28.5% of gross sample)
SC1	at Wave 9 = 1,812 (80.3% of gross sample)
Weighted cases	
SC2 – SC6	SC2 = 1,452[a] (28.6%)
	SC3 = 1,021[b] (20.8%)
	SC4 = 1,700 (22.1%)
	SC5 = 2,857[c] (32.1%)
	SC6 = 2,678 (33.6%)
	CAWI Total = 9,708 (28.1% of gross sample)
SC1	at Wave 9 = 1,812 (80.3% of gross sample)

Note. [a] 126 cases were excluded due to child responses; [b] 10 cases from migrant supplement were excluded; [c] 2 foreign cases were excluded. Abbreviations are explained in the following text.

2. Prequel

The National Educational Panel Study (NEPS) is a study carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg. It is run by an interdisciplinary network led by Prof. Dr. Cordula Artelt. The target actictiviy of the NEPS is to collect longitudinal data on the development of competencies, educational processes, educational decisions, and returns to education in formal, nonformal, and informal contexts throughout the lifespan. The NEPS implemented a multi-cohort-sequence design and surveys six panel cohorts starting at different points in the life-course. Each of the panel cohorts is named according to the specific time in point they have been originally sampled. Overall, 60,000 persons were surveyed at start and followed in regular intervals. More detailed information is available in the documentation section on the homepage.[1]

In March 2020, the SARS-CoV-2 pandemic spread over Germany and along with this many constraints and changes came that had to be managed. The additional Corona-CAWI survey launched in May and June aims to investigate the short-term effects of the SARS-CoV-2 pandemic on educational pathways. For this purpose, several questions regarding direct consequences on school and work but also on active lifestyle have been asked, along with questions capturing the perception on these changes as well as adaptation to the new circumstances. Our data will enable researchers to trace individual responses to the pandemic and to explore differences in response patterns. The wider objective is to evaluate medium and long-term effects on educational trajectories, working careers and skill developments over the life course. On this occasion, a new module NEPS-C is integrated into the questionnaires of the main studies capturing and perpetuating the items of the Corona-CAWI.

The target persons of the additional Corona-CAWI study were all panel members of the Startings Cohorts 3 to 6, the parents from the target persons of Starting Cohort 2 as well as the participants from pilot and developmental studies that did not withdraw their panel consent or deceased. Starting Cohort 1 was supplied with a corresponding NEPS-C module in the CAPI instrument of Wave 9 integrating the set of questions of the Corona-CAWI. Weights are provided for the participants of the Corona-CAWI and the NEPS-C module only. This report documents the weighting of the additional Corona survey data, supplementing the data from the main studies of the NEPS Starting Cohorts 1 to 6. It refers to the Scientific Use Files (SUF):[2]

- https://doi.org/10.5157/NEPS:SC1:8.0.0,
- https://doi.org/10.5157/NEPS:SC2:9.0.0,
- https://doi.org/10.5157/NEPS:SC3:10.0.0,
- https://doi.org/10.5157/NEPS:SC4:11.0.0,
- https://doi.org/10.5157/NEPS:SC5:14.1.0,
- https://doi.org/10.5157/NEPS:SC6:11.1.0

.

[1] For general information on the NEPS, see Blossfeld et al., 2011.

[2] At the time of SUF releases further survey data was available though not being editied or published yet. Therefore, the release versions with the data of the additional Corona studies do not neccessarily match to the most current waves.

3. Description of the NEPS Starting Cohorts

The National Educational Panel Study (NEPS) is a study carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg. The NEPS surveys six panel cohorts starting at different points in the life-course. These six main samples implement the multi-cohort sequence design covering all important episodes in educational life. At start, 60,000 persons were surveyed and followed in regular intervals within their respective starting cohort (Blossfeld et al., 2011).[3] A detailed description of the panel cohorts and sampling strategies can be found in Aßmann et al., 2011. In general, SC1 and SC6 had been sampled in individual contexts, whereas SC2 to SC5 had been sampled in institutional contexts.

Starting Cohort 1

The study "Education from the very beginning" (SC1) run by the NEPS aims to generate detailed information on the development of infants, familial and institutional contexts. Starting Cohort 1 started as the last of the six panel cohorts in 2012 and comprises children born in Germany from February to July 2012. Access to this population had been gained via a register-based sample of addresses available at the level of municipalities. The selection of addresses was performed via a two-stage disproportional stratified sampling design. Municipalities were drawn as primary sampling units being explicitly stratified according to the BIK scale (i.e. classification of urbanization) in three strata proportional to the number of children being born in the first half of 2009. Addresses of newborns within 90 sampling points of 84 selected municipalities were sampled as secondary sampling units. Starting from a gross sample size of 8,483 individuals, 3,431 respondents gave their panel consent. For more detailed information on sampling of SC1 see also Bauer et al., 2013 and Würbach et al., 2016. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC1:8.0.0.

The families participating in SC1 were not invited for the Corona-CAWI because the regular main study (Wave 9) had been administered for March 2020. The main study was suspended for three months and continued in June 2020 as CAPI-by-Phone with an implemented questionnaire as additional Corona survey (NEPS-C module).

Starting Cohort 2

"From Kindergarten to Elementary School" (SC2) is the second study of the NEPS and started with a sample of Kindergarten children in the school year of 2010/2011 who were expected to begin schooling in the school year 2012/2013. A two-staged indirect sampling design was the preferred method of choice because at the time of sampling no frame information was available neither for Kindergarten children nor for Kindergarten institutions. The consequence was an alternative strategy sampling Kindergartens from a list of supplier Kindergartens provided by a sample of 212 elementary schools. The elementary school themselves were drawn by probability proportional to size sampling without replacement and an implicit stratification according to federal state, regional classification as well as organizing institutions. Within the sampled Kindergartens all children (and their parents) were asked for participation. From these, 2,996 respondents gave their panel consent.

[3] More detailed information is available in the documentation section on the homepage.

In the school year 2012/2013, the cohort of Kindergarten children transitioned to elementary school. Children who transitioned to previously sampled schools were followed up within their institutional context together with their classmates augmenting the cohort sample. Besides that, there are previously sampled schools no children transitioned to. Students within these schools also augment the cohort sample. In total, 6,917 Grade 1 students (and their parents) gave their panel consent. Children who transitioned to other schools were tracked individually. By design, three subsamples can be differentiated: Group 1 is the group of students tested in Grade 1 in elementary schools, who were not tested in Kindergarten institutions in Wave 1 and Wave 2, these (target) persons form the augmentation sample of Wave 3. Group 2 is the group of Kindergarten children who were tested only in Kindergartens in Wave 1 and Wave 2, in Wave 3 they are assigned to the individual retracking field and are temporary dropouts by design until Wave 6. Group 3 is the group of Kindergarten children, who were tested in Kindergartens in Wave 1 and Wave 2 and transition to elementary schools surveyed by NEPS in Wave 3. These (target) persons belong to the longitudinal sample of Waves 1, 2, and 3 up to the current Wave 9. For more detailed information on sampling of SC2 Kindergarten children please refer to Hellrung et al., 2011 and Hellrung, Bockelmann, et al., 2013 and Steinhauer et al., 2015 as well as Steinhauer et al., 2016. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC2:9.0.0.

For SC2 Corona-CAWI the parents or legal guardians have been invited not the target persons themselves. However, in the following main study (Wave 10) the target persons will get the NEPS-C module to answer independently.

Starting Cohort 3

"Paths through lower secondary school" (SC3) covers Grade 5 students in school year 2010/2011 and their educational pathways. Access to the target population was gained via the schools. All officially recognized and state-approved educational institutions in Germany providing schooling for fifth grade students form the set of secondary schools. Based on a two-stage sampling design, schools were selected in a first stage as primary sampling units. Selection was done systematically using probability proportional to size sampling referring to the school frame of the school year 2008/2009, provided by the Federal Statistical Office of Germany (Federal Statistical Office, 2009a). The schools thus define clusters and units contained in a cluster are assumed to be more homogeneous than units in different clusters. Explicit stratification considers the overlap with Starting Cohort 4 schools concerning fifth graders. For an additional implicit stratification the characteristics as school type, federal state, regional classification and funding institution were employed. The final sample thus consisted of 365 schools, with 240 regular schools, 65 schools for students with special needs and 60 schools for the migration supplement.

Within the sampled regular schools, two classes were randomly selected on the second stage, when more than two classes were present and all classes otherwise. Within schools for students with special needs all classes were selected. On this second stage all students within the selected classes were invited to participate (main sample) and additional students were selected according to their migration background (migrant supplement).

By design, the initial SC3 sample consists of two samples (K5 in sum 6,112 individuals): the main sample of students in regular schools and special-needs schools and the aforementioned migrant supplement covering students with a migration background from Turkey and former

Soviet Union. For more detailed information on the initial sampling of the SC3 please refer to Bundt et al., 2011a, 2011b, 2011c as well as Steinhauer et al., 2015. In Wave 3, a refreshment sample of 2,205 individuals was added to the SC3 (K7). Based on the same sampling design this refreshment sample is similar to the main sample. Explicit stratification now had to consider the different timings in transitions to lower secondary education within two Federal states of Germany, because in Berlin and Brandenburg primary schooling lasts 6 years. Again, school type, federal state, regional classification and funding institution were used for implicit stratification. For more detailed information on sampling of the SC3 refreshment sample please refer to Hellrung, Hugk, et al., 2013; Hellrung, Meyer-Everdt, et al., 2013 and Steinhauer et al., 2015. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC3:10.0.0.

Starting Cohort 4

The study "School and vocational training" (SC4) follows educational pathways of Grade 9 students through higher secondary schools and vocational training. In a two-stage stratified sampling design access to the target population in different types of regular schools and special-need schools in the school year 2010/2011 was gained. All officially recognized and state-approved educational institutions in Germany providing schooling for fifth grade students form the set of secondary schools. The school type was used for explicit stratification. Additional implicit stratification was performed considering the following criteria: Federal state, regional classification, and funding. At first, a sample of 739 schools with students in Grade 9 has been drawn systematically with probability proportional to size. At the second stage, two classes within the sampled schools were selected randomly, if more than two classes were eligible. The panel sample of the SC4 at start comprises 16,425 individuals. For more detailed information on sampling of the SC4 please refer to Bundt et al., 2011d, 2011e as well as Steinhauer et al., 2015. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC4:11.0.0.

Starting Cohort 5

"From higher education to labor market" is the fifth starting cohort (SC5) of the NEPS. In SC5 first-year undergraduate students (German and non-German) are followed through their years of study and their entrance into working life and professional success, including college dropouts. As target population all students enrolled for the first time in the academic year 2010/2011 in public or state-approved institutions of higher education in Germany with focus on Bachelor's degree, a state examination (Staatsexamen) in medicine, law, pharmacy, and teaching, a diploma or Master's degree in Roman Catholic or Protestant theology or specific art and design degrees are defined. Excluded from target population were students attending universities, technical universities or universities of applied sciences run by Federal Ministries or Federal States for members of their public services. A stratified sampling strategy was applied in which study subjects were used for definition of clusters (Zinn et al., 2017). All students within a cluster were to be surveyed. An oversampling of teacher education students and students attending private universities and private universities of applied sciences was incorporated. In order to address the oversampling, a first stratification level according to higher educational institutions has been set up with four strata: Clusters linked to teacher education at public universities, all other fields of study at public universities, all fields of study at public universities of applied sciences, and all degree programs offered by private universities or private

universities of applied sciences. A second level of stratification was set up to address the heterogeneity of the students, considering the subject of study. Related subjects define different strata. As basis for implementing the sampling procedure information from the academic winter semester 2008/2009 provided by the Federal Statistic Office of Germany was used (Federal Statistical Office, 2009b). In Wave 1, 17,910 respondents gave their panel consent. For more detailed information on sampling of the SC5 see Steinwede and Aust, 2012 and Zinn et al., 2017. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC5:14.1.0.

Starting Cohort 6

"Adult education and lifelong learning" frames the sixth NEPS Starting Cohort with focus on adults and their educational and professional careers as well as their acquisition of competencies across adult life course. Surveying of the SC6 already started in 2007 with the adult survey "Working and Learning in a Changing World" (ALWA) in 2007/2008 conducted by the Institute for Employment Research (IAB), see Aust et al., 2011 and Aust et al., 2013. In this early study the birth years 1956-1986 were covered. The work continued on a subsample with panel willingness for NEPS in annual follow-up surveys (for 6,855 target persons). In the first survey conducted by the NEPS an augmentation sample – to cover also people born between 1944-1955 – as well as an refreshment sample – covering again the birth year range 1956-1986 – were surveyed (in sum 5,077 with panel consent). In the fourth survey wave in 2011/2012 (NEPS Wave 3) the sample was refreshed again, now covering the complete range of birth years 1944-1986 (5,208 with panel consent). For both refreshment samples as well as the augmentation sample the same target population of people living in private households in Germany as for the ALWA sample had been used, within the same communities and sampling points. Access to the target population had been gained via a register-based sample of addresses available at the level of municipalities. The selection of addresses was performed via a two-stage disproportional stratified sampling design. As for the SC1, the sampling points of the SC6 had been sampled at the level of municipalities. Within 240 municipalities 271 sampling points had been allocated according to the size of resident population of a municipality. Stratification had been incorporated according to federal states, administrative districts and classification of urbanization (BIK scale). Addresses of adults were sampled by means of systematic random sampling, as secondary sampling units. For more detailed information on sampling of SC6 see also Aust et al., 2011; Aust et al., 2013 and Hammon et al., 2016. The current Scientific Use File is available here: https://doi.org/10.5157/NEPS:SC6:11.1.0.

4. Design of the Corona-CAWI and NEPS-C Additional Study

In May 2020, an additional survey was launched to investigate the short, medium and long-term effects of the SARS-CoV-2 pandemic on educational pathways and skill developments over the life course. The first part of the additional survey was administered as a web-based survey (CAWI) from May 13th to June, 22th. The second part is a new module NEPS-C integrated into the questionnaires of the main studies capturing and perpetuating the items of the former Corona-CAWI. Several questions regarding direct consequences on school and work but also on active lifestyle are asked, along with questions capturing the perception as well as adaptation to the constraints and changes that came along with this. The data from the additional survey Corona-

CAWI and NEPS-C will enable researchers to trace individual responses to the pandemic and to explore differences in response patterns.

In the Corona-CAWI study all panel participants from our Starting Cohorts 2 to 6 (in sum 34,563) were invited in a panel support measure ("Sommerkarte") to complete a situation-specific questionnaire focussing explicitly on the SARS-CoV-2 pandemic. In these 6 weeks 9,855 individuals participated in the Corona-CAWI. The overall response rate is 28.5%, which is high with respect to the mode and the fact that CAWI is not the usual mode in most of the starting cohorts. Panel members of Starting Cohort 1 have been left out for the Corona-CAWI. Their main study was stopped and could not be administered as planned. Consequently, a modified study design with an adaption to CAPI-by-Phone design was relaunched for SC1 in June. Information from the NEPS-C module asked within the current study in 2020 were made accessible before SUF 9.0.0.

The additional Corona survey data will be made available to the research community in Scientific Use Files, as all main studies. The following Table 2 provides an overview of the sample sizes at start, the current sample sizes and wave, the number of target persons being invited to the Corona-CAWI (in sum 34,563) and those participating (in sum 9,855), in absolute and relative numbers. According to the AAPOR standard definition (The American Association for Public Opinion Research, 2016, p. 61 f.) all questionnaires returned are counted as realizations and the percentages given reflect the response rate. For SC1, all respondents in NEPS-C module are reported. Finally, the number of cases weighted is reported in the last column.

Table 2: Panel samples, participants and cases weighted in the additional Corona survey.

Starting cohort	Group	Panel sample	Current wave	Current sample	%	Invited	Participants	%	Weighted cases
1	All	3,431	9	2,257	65.8	2,257	1,812	80.3	1,812
2	All	9,337	10	5,095	54.6	5,073	1,587	31.3	1,452[a]
	Group 1	*6,341*	*10*	*3,855*	*60.8*	—	*1,231*	—	*1,127*
	Group 2	*2,420*	*10*	*851*	*35.2*	—	*223*	—	*203*
	Group 3	*576*	*10*	*389*	*67.5*	—	*133*	—	*122*
3	All	8,317	11	5,125	61.6	4,917	1,031	21.0	1,021[b]
	K5	*6,112*	*11*	*3,683*	*60.3*	—	*753*	—	*743*
	K7	*2,205*	*11*	*1,442*	*65.4*	—	*278*	—	*278*
4	All	16,425	12	7,911	48.2	7,703	1,700	22.1	1,700
5	All	17,910	16	8,933	49.9	8,894	2,859	32.1	2,857[c]
6	All	17,140	13	7,960	46.4	7,976	2,678	33.6	2,678
	ALWA	*6,855*	*13*	*3,647*	*53.2*	—	*1,267*	—	*1,267*
	NEPS 1	*5,077*	*13*	*2,044*	*40.3*	—	*683*	—	*683*
	NEPS 3	*5,208*	*13*	*2,266*	*43.5*	—	*728*	—	*728*

Note. "—" not applicable; [a] 126 cases were excluded due to child responses; [b] 10 cases from migrant supplement were excluded; [c] 2 foreign cases were excluded.

5. Weighting of the Additional Corona Survey

In the NEPS Starting Cohorts, as in every panel study, we face non-response and attrition processes of the sampled individuals due to voluntariness. These processes might be at random or the refusals might be systematic. Especially in the latter case, these processes have to be accounted for. In the SUF releases we commonly provide design weights with adjustments for survey non-response. In order to do so, we trace for each target person the probabilities to participate in each survey wave as well as the probabilities to participate in consecutive waves. The processing in the non-response analysis is comparable between the Starting Cohorts. In a first step, individual participation propensities are estimated and the corresponding adjustment factors (inverse of the participation probability) are calculated. The initial design weights and the adjustment factors are then used to calculate the cross-sectional and the longitudinal weights. The models used for each Starting Cohort can be found in their respective weighting reports that are part of the supplemental material provided within each SUF release.

For weighting of the additional Corona survey we estimated two more models. First, we estimated a panel model to explore the selectivity up to now. Second, participation models are estimated conditional on the panel sample known at this point. The derived adjustment factors are used for calculation of the cross-sectional and longitudinal weights. The survey weights resulting from this multi-stage weighting strategy are post-stratified referring to data from the Official Statistics (Federal Statistical Office, 2020, Microcensus 2019).

In comparison with the regular weighting procedure, we used individual, household and regional characteristics at start of the survey instead of design and sampling information, solely. Previous effects of the design, i.e. sampling information, are assumed to diminish over time. Parsimony was the guidance for variable selection. Instead of a very large amount of variables, possibly leading to inflated non-response estimators and variances, we chose the models to be as concise as possible. This approach also permits a better comparability within and between the starting cohorts. On grounds of efficacy, highly correlated variables are excluded. Household structure is measured either by size of the household and the number of children under the age of 14, or as (single) parenthood and the number of children under the age of 14, to avoid multi-collinearity due to the number of adults being included already in both variants. Within the starting cohorts no further distinction is made between subsamples joining the main sample at later points. This means an important cut, however, as will be shown later the resulting weights do not differ that much between both concepts.

Explanatory variables with missing values less than or equal to 5% are imputed using multivariate imputation via chained equations (van Buuren & Groothuis-Oudshoorn, 2011) with imputation algorithm CART – classification and regression trees (Breiman et al., 1984; Doove et al., 2014). For explanatory variables with a proportion of missing values higher than 5% an extra category is added to account for the higher proportion of missing data potentially not at random.

Cross-sectional and longitudinal weights are provided in NEPS. Cross-sectional weights refer to respondents participating in the additional Corona survey (`w_tC`). Longitudinal weights refer to all target persons continuously participating in all previous studies, starting at Wave `x` (`w_txtoC`).[4] Furthermore, we provide post-stratified cross-sectional weights (`w_tC_cal`).[5]

[4] For SC4, Waves 4 and 6 were excluded. These were addressed only to panel members of the vocational track.

[5] In the SC6, the cross-sectional weights are post-stratified anyhow (`w_tC_rake`).

6. Panel Attrition and Selectivity

In a first step, the panel attrition of the NEPS Starting Cohorts will be explored. We compare the current panel samples in 2020 with the panel samples at start. In this respect we selected individual, household and regional characteristics as well as the accessibility in the first wave of participation. These variables do not change over time, or are fixed to the value from the first wave of surveying, respectively. Thus, we do not separate into subgroups due to different points in time the target person could enter the study, as known for several of the starting cohorts (SC2, SC3, SC6). All starting cohorts are now surveyed in individual contexts only, clustering within schools is skipped therefore. The models for panel attrition or continuance are build to be as close as possible between the starting cohorts.

We estimated backward stepwise binary probit models[6] considering *gender*, age (*year of birth* or *grouped*), migration background measured as German citizenship (*nationality*) or born in Germany (*home country*) versus other, educational attainment measured as highest ISCED (*H-ISCED*), household structure measured either by *household size* or *parenthood* and number of *children in household*, region is measured as *urbanization level* or *BIK scale* and *living in East or West Germany*, and if available *employment* and *occupational status*. Finally, the number of *contact attempts* needed to realize an interview in the first wave of surveying. An overview with respect to the variables used for each starting cohort is given in Table 3. In the following sections the panel models are given to illustrate the level of selectivity within our panel samples.

Table 3: Characteristics used for modeling panel continuance.

Characteristic	**SC1**	**SC2**	**SC3**	**SC4**	**SC5**	**SC6**
Gender	✓	✓	✓	✓	✓	✓
Year of birth[a]	✓	✓	✓	✓	✓	✓
Nationality[b]	✓	✓	✓	✓	✓	✓
Home country[c]					✓	
H-ISCED[d]	✓	✓	✓	✓	✓	✓
Household size[e]					✓	✓
Parenthood[f]	✓	✓	✓	✓		✓
Children in household[g]	✓	✓	✓	✓	✓	
Employment[h]	✓	✓	✓	✓	✓	✓
Occupational status[i]	✓	✓	✓	✓		
Currently studying[k]					✓	
Urbanization level[l]		✓	✓	✓		
BIK scale[m]	✓					✓
East/ West	✓	✓	✓	✓	✓	✓

[6] For SC5 a backward stepwise binary logit model is estimated.

Table 3: Characteristics used for modeling panel continuance (continued).

Characteristic	SC1	SC2	SC3	SC4	SC5	SC6
Contact attempts[n]	✓	✓	✓	✓	✓	✓

Note. [a] in years or categorized; [b] Germany/ other; [c] German/ other; [d] 0-2/ 3-4/ 5-6; [e] 2/ 3/ 4/ 5 and more; [f] single parent/ both parents; [g] 1/ 2/ 3 and more; [h] employed/ not employed; [i] low/ middle/ middle to high/ high/ unknown; [k] yes/ no; [l] urban/ semi-urban/ rural/ unknown; [m] less than 100,000 inhabitants/ 100,000 up to 500,000 inhabitants/ 500,000 and more inhabitants; [n] 0/ 1-3/ 4-10/ 11 and more.

6.1 Models for Panel Continuance

Starting Cohort 1

The probability of panel continuance in Starting Cohort 1 depends on the four individual characteristics gender, educational attainment, migration background and employment (Table 4). Biological fathers have a lower probability to remain in the panel sample compared to biological mothers. The higher the ISCED the higher is the panel willingness. Having another citizenship than German as well as being currently unemployed significantly decreases the probability for still being part of the panel sample compared to their counterparts.

Table 4: Panel continuance in Starting Cohort 1.

	Estimate
	(SE)
Constant	1.226***
	(0.095)
Gender: male (ref = ”female”)	-0.652***
	(0.161)
H-ISCED: 3-4 (ref = ”0-2”)	0.269***
	(0.082)
H-ISCED: 5-6 (ref = ”0-2”)	0.526***
	(0.086)
Nationality: other (ref = ”German”)	-0.236***
	(0.073)
Employment: not employed (ref = ”employed”)	-1.738***
	(0.062)
Observations	3,431

Note. *p<0.1; **p<0.05; ***p<0.01.

Starting Cohort 2

In Starting Cohort 2 the probability of remaining in the panel sample depends on various characteristics: Number of contact attempts, urbanization level, educational attainment, and the occupational status (Table 5). The more contact attempts were needed to realize an interview in Wave 1 the lower is the probability to remain in the panel sample. For highest ISCED and occupational status it can be noted that higher values correspond to higher probabilities for still being part of the panel sample. If the urbanization level or the ISCED is unknown, the probability to remain in the panel sample is significantly decreased.

Starting Cohort 3

Similar to before, Table 6 presents the variables the probability of remaining in the panel sample for Starting Cohort 3 is related with. These are the number of times the interviewee's parents where contacted, the region in terms of eastern and western parts of Germany, household structure with respect to the parents, gender, year of birth, the educational attainment and the occupational status of the participant's parents. Aside from students' parents that were not contacted (actual non-contact or due to filtering), a significant decrease in probability of remaining in the panel sample is visible for eleven and more attempted contacts in contrast to one, two or three approaches. This propensity is also decreased when the person's gender is male and when the person initially visited the school and thus lives or has lived in Eastern Germany. However, if both parents live in the same household it is more likely that a respondent remains in the panel sample compared to single parent households or persons with unknown structure. Additionally, increases in probabilities for staying in the panel sample are also visible for persons born in 1999, 2000 and later, compared to those being born 1998 or earlier. With the exception of 'unknown' categories, ceteris paribus increases in participation log odds and thus probabilities are again present in parental educational attainment, classified via the latest ISCED-97, and the parental occupational status, as determined by ISEI-08. For both, an increase exists if the respective status is increased as well.

Table 5: Panel continuance in Starting Cohort 2.

	Estimate
	(SE)
Constant	-0.206***
	(0.057)
Contact attempts: 4-10 (ref = ”1-3”)	-0.073**
	(0.031)
Contact attempts: 11 and more (ref = ”1-3”)	-0.284***
	(0.037)
Region: semi-urban (ref = ”urban”)	-0.016
	(0.034)
Region: rural (ref = ”urban”)	0.097*
	(0.054)
Region: unknown (ref = ”urban”)	-0.576***
	(0.036)
H-ISCED: 3-4 (ref = ”0-2”)	0.359***
	(0.054)
H-ISCED: 5-6 (ref = ”0-2”)	0.673***
	(0.059)
H-ISCED: unknown (ref = ”0-2”)	-0.346***
	(0.080)
Occupational status: middle (ref = ”low”)	0.214***
	(0.042)
Occupational status: middle to high (ref = ”low”)	0.289***
	(0.046)
Occupational status: high (ref = ”low”)	0.469***
	(0.051)
Occupational status: unknown (ref = ”low”)	0.116*
	(0.065)
Observations	9,337

Note. *p<0.1; **p<0.05; ***p<0.01.

Table 6: Panel continuance in Starting Cohort 3.

	Estimate
	(SE)
Constant	-0.446***
	(0.082)
Contact attempts: 0 (ref = ”1-3”)	0.231***
	(0.060)
Contact attempts: 4-10 (ref = ”1-3”)	0.017
	(0.039)
Contact attempts: 11 and more (ref = ”1-3”)	-0.224***
	(0.050)
Region: East (ref = ”West”)	-0.226***
	(0.036)
Parenthood: single parent (ref = ”both”)	-0.259***
	(0.045)
Parenthood: unknown (ref = ”both”)	-0.247
	(0.330)
Gender: male (ref = ”female”)	-0.099***
	(0.029)
Birthyear: 1999 (ref = ”1998 and before”)	0.400***
	(0.059)
Birthyear: 2000 and later (ref = ”1998 and before”)	0.568***
	(0.059)
H-ISCED: 3-4 (ref = ”0-2”)	0.340***
	(0.063)
H-ISCED: 5-6 (ref = ”0-2”)	0.564***
	(0.073)
H-ISCED: unknown (ref = ”0-2”)	-0.022
	(0.326)
Occupational status: middle (ref = ”low”)	0.220***
	(0.053)
Occupational status: middle to high (ref = ”low”)	0.315***
	(0.056)

Table 6: Panel continuance in Starting Cohort 3 (continued).

	Estimate
	(SE)
Occupational status: high (ref = ”low”)	0.532***
	(0.066)
Occupational status: unknown (ref = ”low”)	0.242***
	(0.075)
Observations	8,317

Note. *p<0.1; **p<0.05; ***p<0.01.

Starting Cohort 4

In Starting Cohort 4 the probability maintaining in the survey depends positively with descending magnitude on the following variables: Higher educational statuses (highest ISCED 3 or above), younger age (born after 1995 or later), or having a middle or high(er) occupational status. The probability is negatively influenced by the region, the parenthood, attempting to get in touch more than 4 or 11 times, respectively, apart from not being contacted which has a significant positive influence on panel willingness. Having an unknown parenthood status or an unknown ISCED decreases the probability to stay in the panel sample insignificantly. Compared with the other starting cohorts, gender is not a significant predictor for panel continuance in the SC4 (Table 7).

Table 7: Panel continuance in Starting Cohort 4.

	Estimate
	(SE)
Constant	-0.547***
	(0.053)
Contact attempts: 0 (ref = ”1-3”)	0.101**
	(0.039)
Contact attempts: 4-10 (ref = ”1-3”)	-0.043
	(0.028)
Contact attempts: 11 and more (ref = ”1-3”)	-0.290***
	(0.036)
Region: East (ref = ”West”)	-0.300***
	(0.030)
Parenthood: single parent (ref = ”both”)	-0.180***
	(0.033)
Parenthood: unknown (ref = ”both”)	-0.166
	(0.277)
Birthyear: 1995 (ref = ”1994 and before”)	0.297***
	(0.032)
Birthyear: 1996 and later (ref = ”1994 and before”)	0.465***
	(0.033)
H-ISCED: 3-4 (ref = ”0-2”)	0.328***
	(0.047)
H-ISCED: 5-6 (ref = ”0-2”)	0.557***
	(0.053)
H-ISCED: unknown (ref = ”0-2”)	-0.027
	(0.276)
Occupational status: middle (ref = ”low”)	0.139***
	(0.040)
Occupational status: middle to high (ref = ”low”)	0.282***
	(0.041)
Occupational status: high (ref = ”low”)	0.350***
	(0.045)

Table 7: Panel continuance in Starting Cohort 4 (continued).

	Estimate
	(SE)
Occupational status: unknown (ref = ”low”)	0.073
	(0.055)
Observations	16,425

Note. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Starting Cohort 5

In Starting Cohort 5, the probability of remaining in the panel sample depends on the number of contact attempts, the household size in Wave 15, gender, age, educational attainment, German citizenship, and on having children in the household. The more contact attempts were made the lower is the probability that the person is still member of the panel sample. Respondents in households with more than 2 persons are less likely to remain in the panel sample. Being male, having German citizenship, and having children all increase the probability to be part of the panel sample. The same is true for respondents with highest ISCED being 5-6 in comparison to those having highest ISCED of 0-2 (Table 8).

Starting Cohort 6

In Starting Cohort 6, the probability of remaining in the panel sample depends on the number of contact attempts, the familial situation, age, educational attainment, nationality, and employment status. The more contact attempts were made the lower is the probability that the person is still member of the panel sample. If the respondent is married, it is more likely that he or she remains in the panel sample compared to non married persons or persons with unknown marital status. The oldest age group has the lowest probability of still being part of the panel sample compared to the younger age groups. The higher the educational attainment the more likely it is to be a member of the panel sample. German citizens are more likely to stay than other nationalities. Employed respondents have a higher probability of being part of the panel sample than people not being employed (Table 9).

Table 8: Panel continuance in Starting Cohort 5.

	Estimate
	(SE)
Constant	-2.058^{***}
	(0.463)
Contact attempts: 4-10 (ref = ”1-3”)	-0.122^{***}
	(0.043)
Contact attempts: 11 and more (ref = ”1-3”)	-0.394^{***}
	(0.063)
Household size: 2 (ref = ”1 person”)	0.211^{***}
	(0.047)
Household size: 3 (ref = ”1 person”)	-0.770^{***}
	(0.063)
Household size: 4 (ref = ”1 person”)	-0.956^{***}
	(0.075)
Household size: 5 and more (ref = ”1 person”)	-1.059^{***}
	(0.104)
Gender: male (ref = ”female”)	0.395^{***}
	(0.041)
Birthyear: 1989 (ref = ”1988 and before”)	0.634^{***}
	(0.063)
Birthyear: 1990 (ref = ”1988 and before”)	0.743^{***}
	(0.054)
Birthyear: 1991 and later (ref = ”1988 and before”)	0.849^{***}
	(0.058)
H-ISCED: 3-4 (ref = ”0-2”)	-0.179
	(0.461)
H-ISCED: 5-6 (ref = ”0-2”)	2.591^{***}
	(0.461)
Nationality: other (ref = ”German”)	-1.175^{***}
	(0.135)
Children in household: yes (ref = ”no”)	2.013^{***}
	(0.081)
Observations	17,907

Note. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 9: Panel continuance in Starting Cohort 6.

	Estimate
	(SE)
Constant	-0.357^{***}
	(0.045)
Contact attempts: 4-10 (ref = ”1-3”)	-0.110^{***}
	(0.022)
Contact attempts: 11 and more (ref = ”1-3”)	-0.251^{***}
	(0.025)
Marital status: not married (ref = ”married”)	-0.048^{**}
	(0.022)
Marital status: unknown (ref = ”married”)	-0.635^{***}
	(0.047)
Birthyear: 1956-1969 (ref = ”1944-1955”)	0.224^{***}
	(0.025)
Birthyear: 1970-1979 (ref = ”1944-1955”)	0.100^{***}
	(0.031)
Birthyear: 1980-1986 (ref = ”1944-1955”)	0.044
	(0.036)
H-ISCED: 3-4 (ref = ”0-2”)	0.244^{***}
	(0.040)
H-ISCED: 5-6 (ref = ”0-2”)	0.422^{***}
	(0.041)
Nationality: other (ref = ”German”)	-0.326^{***}
	(0.040)
Employment: not employed (ref = ”employed”)	-0.078^{***}
	(0.024)
Observations	17,690

Note. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

6.2 Selectivity in NEPS Starting Cohorts

In a second step, we explore the shifts in the distribution of selected characteristics in more detail. Concretely, we compare the panel composition at start with the current composition of the panel sample as well as the participants with respect to educational attainment and the migration background (both of the parents, or the target itself, if available). These two characteristics were considered in the previous models for panel continuance and used for weighting and post-stratification for the most part.

Starting Cohort 1

As shown in Table 4, educational attainment, measured as highest ISCED of both parents, and migration background, either measured as nationality or as homecountry, are two of the main characteristics explaining panel continuance.

Table 10 displays the distribution of the highest ISCED among the SC1 respondents. At start of the study, the proportion of lower educated respondents is about 11% and decreases to less then 5% in the current panel sample. In the current NEPS-C study, among the participants only 4% have a lower educational attainment. We see also differences for respondents with intermediate education. There is a clear drop in the panel sample (from about 40% to 32%), though, their is no further decrease in participants for respondents with intermediate education. The figures from the corresponding Microcensus data from 2019, which have been used for calibration to external benchmark totals, are given in the last two columns for comparison.

With respect to migration background, we see a reduction of SC1 respondents having other than German nationality (14% at start, and 10% currently, see Table 11), as well as those being born abroad (24% at start and 18% currently, see Table 12). However, we see no differences between the panel sample and those actually participating. The distribution in the Microcensus data from 2019 used for calibration is displayed in the last two columns.

Starting Cohort 2

In Table 13, the distribution of the highest ISCED among the SC2 respondents is displayed. As in SC1, we see a reduction of the proportion of lower educated respondents from about 8% at start of the study to less then 5% in the current panel sample. In the current Corona-CAWI, only 1.5% of the participants have a lower educational attainment. For respondents with intermediate education we only see a decrease for participants of the Corona-CAWI, but to a lesser extent (from 44% to 38%). The corresponding figures from the Microcensus data from 2019, used for calibration, are given in the last two columns.

The proportion of respondents having another than German nationality decreases among the panel sample (from about 8% to 6%), and the participants of the Corona-CAWI (a further reduction to 4%), see Table 14. For comparison, the distribution in the Microcensus data from 2019 used for calibration is displayed in the last two columns.

Table 10: Sample composition of SC1 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	371	10.8	97	4.7	106	4.7	61	3.4	1,571,700	14.1
3-4	1,365	39.8	641	31.0	728	32.3	549	30.3	5,856,000	52.5
5-6	1,695	49.4	1,329	64.3	1,423	63.0	1,202	66.3	3,736,700	33.5
Total	3,431	100.0	2,067	100.0	2,257	100.0	1,812	100.0	11,164,400	100.0

Table 11: Sample composition of SC1 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%
German	2,947	85.9	1,870	90.5	2,033	90.1	1,657	91.4
Non-German	484	14.1	197	9.5	224	9.9	155	8.6
Total	3,431	100.0	2,067	100.0	2,257	100.0	1,812	100.0

Table 12: Sample composition of SC1 for homecountry.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
Germany	2,623	76.5	1,696	82.1	1,844	81.7	1,520	83.9	7,914,100	70.9
Other country	808	23.5	371	17.9	413	18.3	292	16.1	3,250,300	29.1
Total	3,431	100.0	2,067	100.0	2,257	100.0	1,812	100.0	11,164,400	100.0

Table 13: Sample composition of SC2 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	772	8.3	167	4.1	240	4.7	19	1.3	1,665,800	14.1
3-4	4,064	43.5	1,741	42.6	2,216	43.7	543	37.4	6,207,600	52.4
5-6	3,259	34.9	2,179	53.3	2,610	51.5	890	61.3	3,980,600	33.6
Unknown	1,242	13.3	0	0.0	0	0.0	0	0.0	NA	NA
Total	9,337	100.0	4,087	100.0	5,066	99.9	1,452	100.0	11,854,000	100.0

Table 14: Sample composition of SC2 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
German	7,401	79.3	3,675	89.9	4,488	88.6	1,378	94.9	9,415,700	79.4
Non-German	718	7.7	242	5.9	312	6.2	59	4.1	2,438,300	20.6
Unknown	1,218	13.0	170	4.2	266	5.3	15	1.0	NA	NA
Total	9,337	100.0	4,087	100.0	5,066	100.1	1,452	100.0	11,854,000	100.0

Starting Cohort 3

Selectivity in terms of reductions of specific characteristics is depicted in the overview of Table 15. Changes with respect to the highest ISCED are similar to the previously mentioned cohorts. Thus, a decrease in the amount of persons with a lower parental ISCED category is visible: Starting with 6.4% in the panel cohort, decreasing to 4.6% in the current sample and to 2.4% in the Corona-CAWI. The proportion of respondents whose parents have another than German nationality is almost stable (from about 4.5% to 4.4%), although a drop of roughly 2 percentage points to 2.1% within the Corona-CAWI is visible (Table 16). A slightly different shift in percentages is visible with regard to the home country (Table 17). First, there is an increase of about 9 and 12 percentage points between the panel sample (with 55.6%) to participants of the previous and the current study respectively with parents' home country being Germany. And an additional increase to 74.8% for the Corona-CAWI is also visible. This is accompanied by a decrease of proportions of respondents whose parents' home country is not Germany: Starting with 8.4% and finally being 4.7%.

Starting Cohort 4

Table 18 displays the distribution of the highest ISCED among the SC4 respondents. We see the proportion of lower educated respondents (0-2) shrinking in the current sample (from 14.9% to 4.6%) and even more in the Corona study. In detail, in the current Corona-CAWI the lower ISCED (0-2) shrinks to about 3%. In return, persons in the middle categories of the ISCED (3-4) increase by almost 2%. The higher educated (5-6) are represented in even higher proportions, 26.5% in the current sample and 33.9% in the Corona study.

The proportion of respondents having another than German nationality shrinks to a very low level, 3.1% in the current sample and 2.9% in the Corona study, see Table 19 for more details. The last two columns represent the Microcensus data from 2019 used for calibration.

A similar finding can be seen in Table 20: Respondents with home country other than Germany are represented to a lesser extent in the Corona-CAWI (6.1%) compared to the current sample (7.0%). In contrast again, respondents with home country Germany are higher represented than in the current sample, 68.1% in the Corona study compared to 60.1% in the current sample.

Table 15: Sample composition of SC3 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	531	6.4	160	4.2	235	4.6	24	2.4	1,571,700	14.1
3-4	2,859	34.4	1,431	37.2	1,878	36.6	341	33.4	5,856,000	52.5
5-6	1,924	23.1	1,318	34.3	1,646	32.1	443	43.4	3,736,700	33.5
Unknown	3,003	36.1	934	24.3	1,366	26.7	213	20.9	NA	NA
Total	8,317	100.0	3,843	100.0	5,125	100.0	1,021	100.1	11,164,400	100.0

Table 16: Sample composition of SC3 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
German	4,954	59.6	2,776	72.2	3,567	69.6	792	77.6	9,415,700	79.4
Non-German	373	4.5	163	4.2	228	4.4	21	2.1	2,438,300	20.6
Unknown	2,990	36.0	904	23.5	1,330	26.0	208	20.4	NA	NA
Total	8,317	100.1	3,843	99.9	5,125	100.0	1,021	100.1	11,854,000	100.0

Table 17: Sample composition of SC3 for homecountry.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
Germany	4,625	55.6	2,604	67.8	3,325	64.9	764	74.8	7,914,100	70.9
Other country	702	8.4	334	8.7	469	9.2	48	4.7	3,250,300	29.1
Unknown	2,990	36.0	905	23.5	1,331	26.0	209	20.5	NA	NA
Total	8,317	100.0	3,843	100.0	5,125	100.1	1,021	100.0	11,164,400	100.0

Table 18: Sample composition of SC4 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	1,175	14.9	205	4.2	355	4.6	53	3.1	116,400	43.2
3-4	6,451	81.6	1,769	36.6	2,747	35.8	632	37.2	139,200	51.6
5-6	281	3.6	1,457	30.2	2,034	26.5	577	33.9	14,100	5.2
Unknown	0	0.0	1,397	28.9	2,532	33.0	438	25.8	NA	NA
Total	7,907	100.1	4,828	99.9	7,668	99.9	1,700	100.0	269,700	100.0

Table 19: Sample composition of SC4 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
German	6,954	87.9	3,293	68.2	4,903	63.9	1,214	71.4	2,172,400	63.6
Non-German	261	3.3	140	2.9	240	3.1	50	2.9	496,800	36.4
Unknown	696	8.8	1,395	28.9	2,525	32.9	436	25.6	NA	NA
Total	7,911	100.0	4,828	100.0	7,668	99.9	1,700	99.9	2,669,200	100.0

Table 20: Sample composition of SC4 for homecountry.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
Germany	5,222	66.6	3,105	64.3	4,606	60.1	1,161	68.3	171,600	63.6
Other country	2,620	33.4	328	6.8	537	7.0	103	6.1	98,100	36.4
Unknown	0	0.0	1,395	28.9	2,525	32.9	436	25.6	NA	NA
Total	7,842	100.0	4,828	100.0	7,668	100.0	1,700	100.0	269,700	100.0

Starting Cohort 5

In Table 21, the distribution of the highest ISCED among the SC5 respondents is displayed. As expected for this cohort of students, the proportion of the group with the highest ISCED category (5-6) increases significantly from panel start to the current study. There are only small differences in the distribution of the categories between the current sample and the Corona-CAWI participants, with a small tendency for the highest ISCED category to participate more. The corresponding figures from the Microcensus data from 2019, used for calibration, are given in the last two columns.

The proportion of respondents having another than German nationality decreases among the panel sample (from about 3.4% to 2.0%), and the participants of the Corona-CAWI (a further reduction to 1.3%), see Table 22. The share of people born in Germany increases slightly over time, from 77.2% to 79.9% for the sample and to 82.1% for the participants of the Corona-CAWI. The detailed numbers, together with the distribution in the Microcensus data from 2019 used for calibration, are shown in Table 23.

Starting Cohort 6

With respect to the composition of the sample regarding nationality and country of birth we can see that the share of German citizens as well as the share of respondents born in Germany increases in the course of the NEPS study from 93%, respectively 90% at the beginning up to 97%, respectively 96% in the Corona study (Tables 25 and 26). The same applies for educational attainment: Respondents with lower educational attainments show higher panel attrition than respondents with higher educational attainments. The share of respondents with ISCED 0-2 and ISCED 3-4 decreases from 7% to 2%, respectively from 49% to 35%, whereas the share of respondents with ISCED 5-6 increases from 44% to 63% (Table 24).

Table 21: Sample composition of SC5 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	36	0.2	5	0.1	6	0.1	1	0.0	550,800	12.9
3-4	7,706	43.0	811	12.4	1,160	13.0	318	11.1	2,241,600	52.6
5-6	10,165	56.8	5,714	87.5	7,775	87.0	2,538	88.8	1,465,500	34.4
Total	17,907	100.0	6,530	100.0	8,941	100.1	2,857	99.9	4,257,900	100.0

Table 22: Sample composition of SC5 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
German	17,300	96.6	6,418	98.3	8,763	98	2,819	98.7	3,389,600	79.6
Non-German	607	3.4	112	1.7	178	2	38	1.3	868,800	20.4
Total	17,907	100.0	6,530	100.0	8,941	100	2,857	100.0	4,258,400	100.0

Table 23: Sample composition of SC5 for homecountry.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
Germany	13,825	77.2	5,274	80.8	7,144	79.9	2,346	82.1	3,239,500	76.1
Other country	4,082	22.8	1,256	19.2	1,797	20.1	511	17.9	1,018,400	23.9
Total	17,907	100.0	6,530	100.0	8,941	100.0	2,857	100.0	4,257,900	100.0

Table 24: Sample composition of SC6 for highest ISCED.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
0-2	1,281	7.2	301	4.3	354	4.4	60	2.2	6,426,100	13.9
3-4	8,628	48.8	2,873	40.7	3,273	41.1	944	35.3	26,496,600	57.5
5-6	7,781	44.0	3,878	55.0	4,330	54.4	1,673	62.5	13,176,800	28.6
Total	17,690	100.0	7,052	100.0	7,957	99.9	2,677	100.0	46,099,500	100.0

Table 25: Sample composition of SC6 for nationality.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
German	16,493	93.2	6,730	95.4	7,580	95.3	2,597	97	40,367,200	87.6
Non-German	1,197	6.8	322	4.6	377	4.7	80	3	5,730,000	12.4
Total	17,690	100.0	7,052	100.0	7,957	100.0	2,677	100	46,097,200	100.0

Table 26: Sample composition of SC6 for homecountry.

Category	Panel sample	%	Participation previous study	%	Current sample	%	Participation Corona study	%	Microcensus 2019	%
Germany	15,900	89.9	6,583	93.3	7,406	93.1	2,563	95.7	36,730,100	79.7
Other country	1,790	10.1	469	6.7	551	6.9	114	4.3	9,369,400	20.3
Total	17,690	100.0	7,052	100.0	7,957	100.0	2,677	100.0	46,099,500	100.0

7. Participation in Additional Corona Survey

This chapter explores the participation willingness of the NEPS panel members in the additional Corona survey.[7] We model participation probability conditional on the current panel samples being invited. The models for participation use the same variables as for panel continuance, though being updated to the current status if time varying, as for example employment. Except for SC1 and SC2, where the parent information was the only source, for SC3 up to SC6 the target information is used, respectively.[8] We added information on previous participation willingness: A dummy for participation in the previous wave as well as a grouped indicator for participation in all waves before, thereby considering the point in time the target person entered the study. Again, we do not separate into subgroups due to different points in time the target person could enter the study for SC2, SC3, and SC6. Clustering within institutions is skipped as well.

To estimate the individual participation propensities backward stepwise probit regression models had been used with the exception of SC5, where a logit regression model was estimated.[9] The models for participation willingness in the additional Corona survey are given in the following sections.

Starting Cohort 1

The propensity to participate in the NEPS-C-module is increased for respondents who participated in the previous wave already, biological fathers, and those with higher educational attainment (Table 27). Though, the significance in the positive gender effect might be attributable to the fact that the overall amount of fathers (59) is much smaller compared to mothers (1753). A strong negative effect is found for respondents with an overall lower response rate.

[7] We decided to not include a further correction step for availability of an electronical device and/ or accessability of internet, because completion of the CAWI was not restricted to own devices, neither was invitation to completion restricted to those with valid e-mail address.

[8] For few variables considered we face an information gap from previous surveying to early summer this year. This may lead to some bias in time-varying variables because of unobserved changes. However, empirically we observe only few changes in biography from wave to wave. For this reason, we consider these effects to be negligible and see no indication for further treatment.

[9] There was no improvement using complementary log-log regression (cloglog) models instead of probit or logit regression models.

Table 27: Participation in NEPS-C-Module for Starting Cohort 1.

	Estimate
	(SE)
Constant	-0.345^{*}
	(0.187)
Participation: less waves (ref = ”all waves”)	-0.790^{***}
	(0.106)
Participation: almost all (ref = ”all waves”)	-0.342^{***}
	(0.085)
Gender: male (ref = ”female”)	0.630^{***}
	(0.242)
H-ISCED: 3-4 (ref = ”0-2”)	0.484^{***}
	(0.140)
H-ISCED: 5-6 (ref = ”0-2”)	0.708^{***}
	(0.136)
Participation Wave 8: yes (ref = ”no”)	0.881^{***}
	(0.118)
Observations	2,257

Note. *p<0.1; **p<0.05; ***p<0.01.

Starting Cohort 2

Again, participation propensity is increased for respondents that participated in the previous wave and for those with higher educational attainment (Table 28). Item non-response in parenthood and nationality as well as an overall lower response rate significantly decreases the participation propensity.

Table 28: Participation in Corona-CAWI for Starting Cohort 2.

	Estimate
	(SE)
Constant	-1.006^{***}
	(0.144)
Participation: less waves (ref = ”all waves”)	-0.575^{***}
	(0.054)
Participation: almost all (ref = ”all waves”)	-0.308^{***}
	(0.053)
Parenthood: single parent (ref = ”both”)	-0.118^{*}
	(0.062)
Parenthood: unknown (ref = ”both”)	-0.527^{***}
	(0.059)
H-ISCED: 3-4 (ref = ”0-2”)	0.504^{***}
	(0.129)
H-ISCED: 5-6 (ref = ”0-2”)	0.714^{***}
	(0.129)
Nationality: other (ref = ”German”)	-0.148
	(0.091)
Nationality: unknown (ref = ”German”)	-0.654^{***}
	(0.137)
Participation Wave 9: yes (ref = ”no”)	0.215^{***}
	(0.064)
Observations	5,095

Note. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Starting Cohort 3

Characteristics in relationship with participation in the NEPS-C-module for students of Starting Cohort 3 (Table 29) are somewhat similar although not the same to characteristics of the previously described cohorts. Like before, there is an increase in current participation for respondents who already participated in the previous wave and a decrease with an overall lower response rate. New relationships refer to the students' gender and school enrollment. If a student's gender is stated as male (compared to female) and the person is still in school, in contrast to not being enrolled anymore, the propensity of taking part in the additional Corona survey is decreased; if this school status is unknown there is an increase present.

Table 29: Participation in Corona-CAWI for Starting Cohort 3.

	Estimate
	(SE)
Constant	-1.080^{***}
	(0.079)
Participation: less waves (ref = "all waves")	-0.237^{***}
	(0.060)
Participation: almost all (ref = "all waves")	-0.046
	(0.048)
Gender: male (ref = "female")	-0.295^{***}
	(0.041)
In school: yes (ref = "no")	-0.249^{***}
	(0.058)
In school: unknown (ref = "no")	0.127^{***}
	(0.047)
Participation Wave 10: yes (ref = "no")	0.569^{***}
	(0.059)
Observations	5,125

Note. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Starting Cohort 4

Due to perfect multi-collinearity between the factor variables school and college we skipped the first one. The results present similar participation dependencies: Participating in the previous wave, having a high(er) ISCED status, and being not in school influences the participation in the Corona-CAWI significant positively. Male people are less likely to participate in the Corona-CAWI as well as people having another than German nationality or an unknown one. Hence, if the school information is unknown the participation propensity is also decreased (Table 30).

Table 30: Participation in Corona-CAWI for Starting Cohort 4.

	Estimate
	(SE)
Constant	-1.446^{***}
	(0.092)
Gender: male (ref = ”female”)	-0.385^{***}
	(0.033)
H-ISCED: 3-4 (ref = ”0-2”)	0.324^{***}
	(0.054)
H-ISCED: 5-6 (ref = ”0-2”)	0.342^{***}
	(0.100)
Nationality: other (ref = ”German”)	-0.596^{***}
	(0.126)
Nationality: unknown (ref = ”German”)	-0.063
	(0.067)
In school: no (ref = ”yes”)	0.204^{***}
	(0.074)
In school: unknown (ref = ”yes”)	-0.426^{***}
	(0.094)
Participation Wave 11: yes (ref = ”no”)	0.540^{***}
	(0.044)
Observations	7,911

Note. *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Starting Cohort 5

Similar to the other starting cohorts, participation probability is mainly driven by previous participation and gender, with participation in the previous Wave 15 having a positive effect, and negative effects for people participating seldomly or often. Males tend to participate less, compared to females. Additionally, being employed increases the participation probability (Table 31).

Table 31: Participation in Corona-CAWI for Starting Cohort 5.

	Estimate
	(SE)
Constant	-0.657^{***}
	(0.069)
Participation: seldom (ref = ”very often/ always”)	-2.099^{***}
	(0.727)
Participation: often (ref = ”very often/ always”)	-1.325^{***}
	(0.076)
Gender: male (ref = ”female”)	-0.208^{***}
	(0.049)
Employment: not employed (ref = ”employed”)	-0.233^{***}
	(0.069)
Employment: unknown (ref = ”employed”)	-0.764^{***}
	(0.146)
Participation Wave 15: yes (ref = ”no”)	0.419^{***}
	(0.069)
Observations	8,933

Note. *p<0.1; **p<0.05; ***p<0.01.

Starting Cohort 6

The probability of participating in the Corona study in Starting Cohort 6 depends on the previous participation behavior, the living area, education, and nationality (Table 32). Respondents who participate in most of the previous waves or in the wave directly before the Corona study are more likely to participate in the Corona study than persons with less participations or people not participating in the previous wave. The higher the education the higher the probability to participate in the Corona study. People without German nationality or living in the eastern part of Germany are less likely to participate than people with German nationality or living in the western part of Germany.

Table 32: Participation in Corona-CAWI for Starting Cohort 6.

	Estimate
	(SE)
Constant	-0.727^{***}
	(0.082)
Participation: seldom (ref = ”very often/ always”)	-0.572^{***}
	(0.061)
Participation: often (ref = ”very often/ always”)	-0.304^{***}
	(0.039)
Participation Wave 11: yes (ref = ”no”)	-0.342^{***}
	(0.066)
East/ West: East (ref = ”West”)	-0.226^{***}
	(0.037)
H-ISCED: 3-4 (ref = ”0-2”)	0.384^{***}
	(0.084)
H-ISCED: 5-6 (ref = ”0-2”)	0.660^{***}
	(0.083)
Nationality: other (ref = ”German”)	-0.378^{***}
	(0.077)
Observations	7,960

Note. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Starting Cohort 6: Focal group of employed target persons

For SC6 we provide additional weights for respondents who reported to be currently employed during the Corona study. The participation model of employed respondents (Table 33) is comparable to the participation model of the entire sample of SC6: The more often one participated in the past the more likely is the participation in the Corona study, too. The same is true for the participation in the previous Wave 11. German citizens and people living in the western part of Germany are more likely to participate than non-Germans and people living in the eastern part of Germany. Similar to the findings of the other starting cohorts higher educational attainments lead to higher participation probabilities. Further influencing factors are age and employment status reported in the previous Wave 11. The participation probabilities of the younger age groups are higher than those of the oldest age group. Respondents who were employed in the previous Wave 11 are more likely to participate than unemployed respondents or respondents with unknown employment status.

Table 33: Participation in Corona-CAWI for Starting Cohort 6, focal group: Employed target persons.

	Estimate
	(SE)
Constant	−1.609***
	(0.124)
Participation: seldom (ref = ”very often/ always”)	−0.491***
	(0.072)
Participation: often (ref = ”very often/ always”)	−0.272***
	(0.045)
Participation Wave 11: yes (ref = ”no”)	−0.259***
	(0.073)
East/ West: East (ref = ”West”)	−0.179***
	(0.043)
Birth year: 1956-1969 (ref = ”1944-1955”)	0.839***
	(0.065)
Birth year: 1970-1979 (ref = ”1944-1955”)	0.857***
	(0.072)
Birth year: 1980-1986 (ref = ”1944-1955”)	0.878***
	(0.078)
H-ISCED: 3-4 (ref = ”0-2”)	0.351***
	(0.112)
H-ISCED: 5-6 (ref = ”0-2”)	0.591***
	(0.111)
Nationality: other (ref = ”German”)	−0.364***
	(0.087)
Employment Wave 11: not employed (ref = ”employed”)	−1.026***
	(0.065)
Employment Wave 11: unknown (ref = ”employed”)	−0.247**
	(0.097)
Observations	7,960

Note. *p<0.1; **p<0.05; ***p<0.01.

8. Post-Stratification

Survey weights are adjusted to external benchmark totals from official statistics at the individual level of response. The post-stratification step is made with respect to individual levels characteristics as well as household and regional information, based on the German Microcensus data from 2019 (Federal Statistical Office, 2020).

Iterative proportional fitting or raking (Deville et al., 1993), as this type of calibration method is called as well, is conducted for all starting cohorts by utilizing the respective R function *rake* provided by Lumley, 2011, 2020 within the R package *survey*. Depending on the dimensionality of the sample and population margins available, theoretical adjustment of cells can be achieved by minimizing the squared distance between the sample and population information provided (Deming & Stephan, 1940). Facilitating the method of Lagrange multipliers as a strategy to solve the inherent optimization problem, allows to derive and practically employ an iteratively alternating ratio, i.e. an adjustment factor. This factor is not identical but approximates and converges to the least squares solution and in turn minimizes distances of specified sample and population characteristics, see Stephan, 1942 and Fienberg, 1940. After a sufficiently large amount of iterations, the adjustment factor's final value can be transformed into the desired calibration weight, thus compensating for sample imbalances.

Please note, that calibration at the current point in time is different from calibration at start of surveying. In calibration at start, the panel samples of the institutional Starting Cohorts 2, 3, and 4 are adjusted to external benchmark totals from official school statistics, see Steinhauer and Zinn, 2016a, 2016b, 2016c. And characteristics used for calibration are restricted to information known for the gross sample, i.e. sample information: SC2 – in Wave 1 gender, language spoken at home, federal state, in Wave 3 and 6 gender and federal state; SC3 – in Wave 1 school type, gender, federal state, in Wave 3 gender and federal state; SC4 – in Wave 1 school type and federal state.

Starting Cohort 6 is the only one being post-stratified on a regular basis, see Hammon et al., 2016. Calibration in Waves 1 to 3 to external benchmark totals taken from the Microcensus 2009 and 2010 was performed by infas (Institute for Applied Social Sciences). In Wave 2 gender, birth year, and educational attainment (ISCED-97) were used for calibration. Since Wave 3 and up to now, place of living (federal state), classification of urbanization (BIK scale), birth year and country of birth were added. As of Wave 4 sampling weights are calibrated to values of the Microcensus by the NEPS research unit Statistical Survey Methods.

In Table 34 the respondent characteristics used for calibration are listed. The dimensionality of information that is used for calibration is larger. However, characteristics with numerous categories are grouped to avoid zero entries as much as possible, e.g. for birth year or BIK scale. Despite these efforts, for some variables only uni- or bivariate margin totals could be considered.

We removed H-ISCED and employment for calibration of Starting Cohorts 3, 4, and 5, because these characteristics are not decisive in these age groups. For calibration of SC2 also employment is skipped and nationality was used instead of the country of birth, due to availability. Please keep in mind that the parents or legal guardians have been invited for participation in Corona-CAWI not the target persons themselves. This will be recuperated in the next main study in 2021. For SC1 and SC2 calibration gender was excluded due to the overall reduced

amount of fathers participating on the one hand, and on the other hand for SC2 there is no sufficiently precise knowledge on which parent responded. The official statistics were restricted to households with children younger than 14 years for SC1 and to households with children in household for SC2. BIK scale and employment were used for calibration in SC1 and SC6 only due to availability.

Calibration induces tangible effects on survey weights, e.g. larger maximum values (Table 36 in the following chapter). This is attributable to the contexts in which the panel cohorts originally were sampled. Calibration is now considered to resample the population total.

Table 34: Respondent characteristics used for calibration.

Starting cohort	Birth year	Gender	Home country[a]	Nationality[b]	Federal state	BIK scale[c]	ISCED[d]	Employment
1	✓[e]		✓		✓	✓	✓	✓
2	✓[e]			✓	✓		✓	
3	✓	✓	✓		✓			
4	✓	✓	✓		✓			
5	✓	✓	✓		✓			
6	✓[e]	✓	✓		✓	✓	✓	✓

Note. [a] Germany/ other; [b] German/ other; [c] Less than 100,000 inhabitants/ 100,000 up to 500,000 inhabitants/ 500,000 inhabitants and more; [d] 0-2/ 3-4/ 5-6; [e] categorized.

9. Provision of Weights and Remarks

All survey weights are trimmed to increase statistical efficacy, i.e. to reduce sampling error and possible inflation of standard errors. Individual weights above a predefined threshold are not assigned to this maximum but are iteratively relocated to the remaining distribution, thus employing the *Weight Distribution* approach supposed by Potter (1990). Furthermore, to ease usage in statistical analyses, all weights are standardized to sum up to the sample size with mean equal to 1. Cross-sectional as well as longitudinal weights are constrained to all target persons participating, i.e. in cross-sectional weights all participants in Corona-CAWI or NEPS-C-module, and in longitudinal weights all target persons continuously participating up to the Corona-CAWI or NEPS-C-module are considered. Table 35 gives an overview for all weighting variables provided for each starting cohort. The accordant summary of all non-zero survey weights is reported in Table 36.

For Starting Cohort 6 additional weights are provided for the focal group of employed participants in the Corona-CAWI. This applies to 1,799 respondents (weights: `w_t4toC_F`, `w_tC_F_cal`, `w_tC_F_rake`). The cross-sectional weights are adjusted to external benchmark totals from the employed population of Germany. Weights for this focal group are supplied only for participating panel members of Starting Cohort 6 being employed (full-time, part-time) at the time of the Corona-CAWI.

Table 35: Weighting variables corresponding to the additional Corona survey.

Variable	Applies to	Content
SC1		
w_tC	1,812	Cross-sectional weight for persons participating in NEPS-C-Module
w_t1toC	1,404	Longitudinal weight for persons participating in Wave 1 to 9 and in NEPS-C-Module
w_tC_cal	1,812	Calibrated cross-sectional weight for persons participating in NEPS-C-Module
SC2		
w_tC	1,452	Cross-sectional weight for persons participating in Corona-CAWI
w_t3toC	851	Longitudinal weight for persons participating in Wave 3 to 9 and in Corona-CAWI
w_t6toC	1,101	Longitudinal weight for persons participating in Wave 6 to 9 and in Corona-CAWI
w_tC_cal	1,452	Calibrated cross-sectional weight for persons participating in Corona-CAWI
SC3		
w_tC	1,021	Cross-sectional weight for targets participating in Corona-CAWI
w_t1toC	398	Longitudinal weight for targets participating in Wave 1 to 9 and in Corona-CAWI
w_t3toC	574	Longitudinal weight for targets participating in Wave 3 to 9 and in Corona-CAWI
w_tC_cal	1,021	Calibrated cross-sectional weight for targets participating in Corona-CAWI
SC4		
w_tC	1,700	Cross-sectional weight for targets participating in Corona-CAWI
w_t1toC	969	Longitudinal weight for targets participating in Wave 1 to 10 and in Corona-CAWI
w_tC_cal	1,700	Calibrated cross-sectional weight for targets participating in Corona-CAWI

Table 35: Weighting variables corresponding to the additional Corona survey (continued).

Variable	Applies to	Content
SC5		
w_tC	2,857	Cross-sectional weight for targets participating in Corona-CAWI
w_allCAWI	1,574	Longitudinal weight for targets participating in all CAWI and in Corona-CAWI
w_allWaves	835	Longitudinal weight for targets participating in Wave 1 to 15 and in Corona-CAWI
w_tC_cal	2,857	Calibrated cross-sectional weight for targets participating in Corona-CAWI
SC6		
w_tC_rake	2,678	Raked cross-sectional weight for targets participating in Corona-CAWI
w_t4toC	2,219	Longitudinal weight for targets participating in Wave 4 to 11 and in Corona-CAWI
w_tC_cal	2,678	Calibrated cross-sectional weight for targets participating in Corona-CAWI
SC6 focal group		
w_tC_F_rake	1,799	Raked cross-sectional weight for targets participating in Corona-CAWI (employed respondents)
w_t4toC_F	1,466	Longitudinal weight for targets participating in Wave 4 to 11 and in Corona-CAWI (employed respondents)
w_tC_F_cal	1,799	Calibrated cross-sectional weight for targets participating in Corona-CAWI (employed respondents)

For completion of this report, we want to give some remarks and hints on usage:

1. Interpretation of the uncalibrated weights has to be done in starting cohort perspective as usually done, i.e. considering the specific study designs.
2. Calibration heavily depends on the characteristics used for post-stratification. The accordance with population totals is ensured only for the characteristics considered (Table 34). That is, calibrated weights do not necessarily return corrected numbers for any other characteristic. Furthermore, this approach may lead to artifacts in distributions of characteristics not being corrected for.

3. Cross-sectional weights apply to scientific issues considering this specific substudy. Longitudinal weights, however, should be used when the dependent variable extends over several studies beforehand, not only the Corona-CAWI or NEPS-C-module.
4. For pure data description without the purpose of representativeness no correction, e.g. by weighting, is needed.
5. We strictly advice against sub-setting from participants. The sum of weights won't match anymore and the non-response adjustment as well as the calibration step is likely to be different for a selection of target persons, e.g. when excluding those with migration background.
6. Please consider correct treatment of item non-response, e.g. by multiple imputation using `mice` (van Buuren & Groothuis-Oudshoorn, 2011), if the number of analyzed cases does not match to the number of participants and those with weights provided.
7. Sub-setting can also be a result of skip patterns due to filter instructions. For example, from all participants only those being currently employed or those with children are considered. Such subgroups need own weighting. Exemplarily, we supply additional weights for the focal group of employed respondents in Starting Cohort 6.

If, for one or more of the aforementioned reasons, weights cannot be applied, the recommended strategy depends on the type of analysis considered: In the case of descriptive analyses, the only option is to adjust the interpretation of findings. In this case, only statements about the sample are possible but no generalized statements regarding the target population as such. The present report as well as the preceding weighting reports provide indications which characteristics have led to biases in the composition of the sample. These should be kept in mind when reporting unweighted descriptive findings.

In the case of statistical modeling (such as regression analyses) it has to be examined prior in advance whether the design is *informative*, i.e. if the sampling design and the non-response bears information that is predictive for the model of interest. Is this the case, we recommend applying a model-based weighting approach as suggested by Snijders and Bosker, 2012. This implies the inclusion of all sampling variables and additionally those used for non-response correction as covariates into the analysis model. These variables affect the chance of being part of the current sample (cp. participation models in Chapter 7). In a final step the models with and without these additional covariates should be compared to explore whether the inclusion of these variables is decisive. Relevant indicators are changes in the significance of effects, smaller AIC (BIC), or changes in the estimates versus differences in standard errors only.

According to our findings we can assume that respondents from specific groups were harder to reach, or to motivate respectively, e.g. respondents with migration background and respondents with lower educational attainment. We see these selection processes in the panel development anyhow. However, the Corona-CAWI was conducted as an additional survey in the frame of panel maintenance which might have induced a lower perceived importance compared to the main studies, possibly explaining the overall lower priority in responding.

Table 36: Summary statistics for all weights provided.

Label of weight	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.
SC1						
w_tC	0.252	0.432	0.549	1	0.784	4.820
w_t1toC	0.267	0.450	0.543	1	0.751	4.858
w_tC_cal	0.000	0.083	0.135	1	0.441	52.146
SC2						
w_tC	0.072	0.307	0.576	1	1.099	4.967
w_t3toC	0.134	0.443	0.709	1	1.235	4.150
w_t6toC	0.115	0.388	0.679	1	1.165	4.508
w_tC_cal	0.000	0.122	0.279	1	0.686	64.610
SC3						
w_tC	0.004	0.044	0.158	1	0.929	5.778
w_t1toC	0.145	0.375	0.655	1	1.184	4.578
w_t3toC	0.049	0.354	0.579	1	1.209	4.822
w_tC_cal	0.001	0.040	0.143	1	0.868	21.966
SC4						
w_tC	0.076	0.493	0.759	1	1.165	3.939
w_t1toC	0.100	0.565	0.812	1	1.198	3.417
w_tC_cal	0.000	0.006	0.266	1	0.948	44.726
SC5						
w_tC	0.008	0.316	0.880	1	1.262	3.921
w_allCAWI	0.090	0.301	0.633	1	1.173	4.900
w_allWaves	0.140	0.475	0.770	1	1.288	3.923
w_tC_cal	0.007	0.335	0.736	1	1.276	10.261
SC6						
w_tC_rake	0.003	0.085	0.296	1	1.034	17.431
w_t4toC	0.013	0.153	0.386	1	1.049	5.491
w_tC_cal	0.000	0.000	0.058	1	0.459	53.626

Table 36: Summary statistics for all weights provided (conntinued).

Label of weight	Min.	Lower Quart.	Median	Mean	Upper Quart.	Max.
SC6 focal group						
w_tC_F_rake	0.000	0.001	0.015	1	0.387	23.999
w_t4toC_F	0.012	0.144	0.387	1	1.046	5.491
w_tC_F_cal	0.001	0.038	0.137	1	0.642	46.956

Postambel

We used the freely available statistic software `R` (R Core Team, 2020, p. 3.6.3) and `RStudio` (RStudio Team, 2020, p. 1.2.5042) and the following packages for data preparation, imputation and analyses: `caret` (Kuhn, 2020, pp. 6.0–86), `corrplot` (Wei & Simko, 2020, p. 0.84), `dplyr` (Wickham et al., 2020, p. 1.0.1), `haven` (Wickham & Miller, 2020, p. 2.3.1), `Hmisc` (Frank E. Harrell Jr., 2020, pp. 4.4–0), `kableExtra` (Zhu, 2019, p. 1.1.0), `knitr` (Xie, 2021, p. 1.33), `lme4` (Bates et al., 2015, p. 1.1.23), `margins` (Leeper, 2020, p. 0.3.23), `memisc` (Elff, 2020, p. 0.99.27.3), `mice` (van Buuren & Groothuis-Oudshoorn, 2011, p. 3.11.0), `miceadds` (Robitzsch & Grund, 2020, pp. 3.10–28), `rcompanion` (Mangiafico, 2020, p. 2.3.25), `reshape2` (Wickham, 2007, p. 1.2.5042), `survey` (Lumley, 2020, p. 4.0). This document is a LaTeX Sweave file (`.Rnw`), the standard for production of weighting reports in NEPS.

Acknowledgments

This paper uses data from the National Educational Panel Study (NEPS):

- Starting Cohort Newborns, https://doi.org/10.5157/NEPS:SC1:8.0.0,
- Starting Cohort Kindergarten, https://doi.org/10.5157/NEPS:SC2:9.0.0,
- Starting Cohort Grade 5, https://doi.org/10.5157/NEPS:SC3:10.0.0,
- Starting Cohort Grade 9, https://doi.org/10.5157/NEPS:SC4:11.0.0,
- Starting Cohort First-Year Students, https://doi.org/10.5157/NEPS:SC5:14.1.0,
- Starting Cohort Adults, https://doi.org/10.5157/NEPS:SC6:11.1.0.

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.

Additionally, survey data from previous waves not being edited and/ or published yet was used for weight preparation.

10. References

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., & Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions. In H.-P. Blossfeld, H. G. Roßbach, & J. von Maurice (Eds.), *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (pp. 51–65). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/s11618-011-0181-8

Aust, F., Gilberg, R., Hess, D., Kleudgen, M., & Steinwede, A. (2011). *Methodenbericht NEPS Etappe 8: Befragung von Erwachsenen - Haupterhebung 1. Welle 2009/2010* (Tech. Rep.). infas Institut für angewandte Sozialwissenschaft GmbH. Bonn. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC6/1-0-0/Methodenbericht_SC6_W2_B72.pdf

Aust, F., Hess, D., Kleudgen, M., Malina, A., & Steinwede, A. (2013). *Methodenbericht NEPS Startkohorte 6 - Haupterhebung 2011-2012 (B68)* (Tech. Rep.). infas Institut für angewandte Sozialwissenschaft GmbH. Bonn. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC6/5-0-0/Methodenbericht_SC6_W4_B68.pdf

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, A., Bech, K., Gilberg, R., & Kleudgen, M. (2013). *Methodenbericht NEPS Startkohorte 1 - Haupterhebung 2012/2013 B04* (tech. rep.). infas Institut für angewandte Sozialwissenschaft GmbH. Bonn. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC1/1-0-0/SC1_MB_1.pdf

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). VS Verlag für Sozialwissenschaften.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (Eds.). (1984). *Classification and regression trees*. Wadsworth Publishing.

Bundt, S., Gomolka, J., Martin, G., & Knoll, S. (2011a). *Methodenbericht: NEPS Startkohorte 3 - Haupterhebung – Herbst/Winter 2010 (A28)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/1-0-0/Methodenbericht_SC3_W1_PAPI.pdf

Bundt, S., Gomolka, J., Martin, G., & Knoll, S. (2011b). *Methodenbericht: NEPS Startkohorte 3 – Förderschulen - Haupterhebung – Herbst/Winter 2010 (A56)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/1-0-0/Methodenbericht_SC3_W1_PAPI_F%C3%B6rderschulen.pdf

Bundt, S., Gomolka, J., Martin, G., & Knoll, S. (2011c). *Methodenbericht: NEPS Startkohorte 3 – Migranten-Zusatzstudie - Haupterhebung – Herbst/Winter 2010 (A63)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/1-0-0/Methodenbericht_SC3_W1_PAPI_Supplemental.pdf

Bundt, S., Gomolka, J., Martin, G., & Knoll, S. (2011d). *Methodenbericht: NEPS Startkohorte 4 - Haupterhebung – Herbst/Winter 2010 (A46, A67, A83)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/Methodenbericht_A46_A67_A83.pdf

Bundt, S., Gomolka, J., Martin, G., & Knoll, S. (2011e). *Methodenbericht: NEPS Startkohorte 4 – Förderschulen - Haupterhebung – Herbst/Winter 2010 (A60, A86)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/Methodenbericht_A60_A86.pdf

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, *11*(4), 427–444. https://doi.org/10.1214/aoms/1177731829

Deville, J.-C., Särndal, C.-E., & Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, *88*(423), 1013–1020. https://doi.org/10.1080/01621459.1993.10476369

Doove, L., Buuren, S. V., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis*, *72*, 91–104. https://doi.org/10.1016/j.csda.2013.10.025

Elff, M. (2020). *Memisc: Management of survey data and presentation of analysis results* [R package version 0.99.27.3]. https://CRAN.R-project.org/package=memisc

Federal Statistical Office. (2009a). *Bildung und Kultur - Allgemeinbildende Schulen: Schuljahr 2008/2009* (Fachserie 11 Reihe 1). Federal Statistical Office. Wiesbaden. https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00006815/2110100097004.pdf

Federal Statistical Office. (2009b). *Bildung und Kultur - Studierende an Hochschulen: Wintersemester 2008/2009* (Fachserie 11 Reihe 4.1). Federal Statistical Office. Wiesbaden. https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00006844/2110410097004.pdf

Federal Statistical Office. (2020). Special evaluation.

Fienberg, S. E. (1940). An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics*, *41*(3), 907–917. https://doi.org/10.1214/aoms/1177696968

Frank E. Harrell Jr., e. a., Charles Dupont. (2020). *Hmisc: Harrell miscellaneous* [R package version 4.4-0]. https://CRAN.R-project.org/package=Hmisc

Hammon, A., Zinn, S., Aßmann, C., & Würbach, A. (2016). *Samples, weights, and nonresponse: The adult cohort of the national educational panel study (wave 2 to 6)* (NEPS Survey Paper No. 7). Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg, Germany.

Hellrung, M., Bockelmann, J., Schneider, C., Waschk, A., & Hillen, P. (2013). *Methodenbericht: NEPS Startkohorte 2 - Haupterhebung – Frühjahr 2013 (A14, A14_A)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/3-0-0/NEPS_FieldReport_SC2_W3_PAPI.pdf

Hellrung, M., Hugk, N., Waschk, A., & Hillen, P. (2013). *Methodenbericht: NEPS Startkohorte 3 – Förderschulen - Haupterhebung – Herbst/Winter 2012 (A58)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/3-0-0/Methodenbericht_SC3_W3_PAPI_F%C3%B6rderschulen.pdf

Hellrung, M., Meyer-Everdt, M., Waschk, A., & Hillen, P. (2013). *Methodenbericht: NEPS Startkohorte 3 - Haupterhebung – Herbst/Winter 2012 (A30, A30_A)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/3-0-0/Methodenbericht_SC3_W3_PAPI.pdf

Hellrung, M., Waschk, A., Oberlein, J., & Hillen, P. (2011). *Methodenbericht: NEPS Startkohorte 2 - Haupterhebung – Winter/Frühjahr/Sommer 2011 (A12)* (tech. rep.). IEA Data Processing and Research Center (IEA DPC). https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/1-0-0/NEPS_FieldReport_SC2_W1_PAPI.pdf

Kuhn, M. (2020). *Caret: Classification and regression training* [R package version 6.0-86]. https://CRAN.R-project.org/package=caret

Leeper, T. J. (2020). *Margins: Marginal effects for model objects* [R package version 0.3.23].

Lumley, T. (2011). *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons.

Lumley, T. (2020). Survey: Analysis of complex survey samples [R package version 4.0].

Mangiafico, S. (2020). *Rcompanion: Functions to support extension education program evaluation* [R package version 2.3.25]. https://CRAN.R-project.org/package=rcompanion

Potter, F. J. (1990). A study of procedures to identify and trim extreme sampling weights. In A. S. Association (Ed.), *Proceedings of the survey research methods section* (pp. 225–230). http://www.amstat.org/sections/srms/Proceedings/papers/1990%5Ctextunderscore%20034.pdf

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Robitzsch, A., & Grund, S. (2020). *Miceadds: Some additional multiple imputation functions, especially for 'mice'* [R package version 3.10-28]. https://CRAN.R-project.org/package=miceadds

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, Inc. Boston, MA. http://www.rstudio.com/

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Steinhauer, H. W., Aßmann, C., Zinn, S., Goßmann, S., & Rässler, S. (2015). Sampling and Weighting Cohort Samples in Institutional Contexts. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *9*(2), 131–157. https://doi.org/10.1007/s11943-015-0162-0

Steinhauer, H. W., & Zinn, S. (2016a). *NEPS Technical Report for Weighting: Weighting the sample of Kindergarten Children and Grade 1 Students of the National Educational Panel Study (Waves 1 to 4)* (Technical Report). Leibniz Institute for Educational Trajectories. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/4-0-0/SC2_4-0-0_W.pdf

Steinhauer, H. W., & Zinn, S. (2016b). *NEPS Technical Report for Weighting: Weighting the Sample of Starting Cohort 3 of the National Educational Panel Study (Waves 1 to 5)* (Technical Report). Leibniz Institute for Educational Trajectories. Bamberg. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/5-0-0/SC3_5-0-0_W.pdf

Steinhauer, H. W., & Zinn, S. (2016c). *NEPS Technical Report for Weighting: Weighting the sample of Starting Cohort 4 of the National Educational Panel Study (Wave 1 to 6)* (NEPS Survey Paper No. 2). Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXIII.pdf

Steinhauer, H. W., Zinn, S., Gaasch, C., & Goßmann, S. (2016). *NEPS Technical Report for Weighting: Weighting the Sample of Kindergarten Children and Grade 1 Students of the National Educational Panel Study (Wave 1 to 3)* (Working Paper No. 66). Leibniz Institute for Educational Trajectories, National Educational Panel Study. Bamberg. https://www.neps-data.de/de-de/projekt%C3%BCbersicht/publikationen/nepsworkingpapers.aspx

Steinwede, J., & Aust, F. (2012). *Methodenbericht NEPS Startkohorte 5 - CATI-Haupterhebung Herbst 2010 (B52)* (tech. rep.). infas Institut für angewandte Sozialwissenschaft GmbH. Bonn. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/3-0-0/NEPS_FieldReport_SC5_W1_CATI.pdf

Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *The Annals of Mathematical Statistics*, *13*(2), 166–178. https://doi.org/10.1214/aoms/1177731604

The American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (tech. rep.). AAPOR. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. https://www.jstatsoft.org/v45/i03/

Wei, T., & Simko, V. (2020). *R package 'corrplot': Visualization of a correlation matrix* [R package version 0.84]. https://github.com/taiyun/corrplot

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. http://www.jstatsoft.org/v21/i12/

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation* [R package version 1.0.1]. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Miller, E. (2020). *Haven: Import and export 'spss', 'stata' and 'sas' files* [R package version 2.3.1]. https://CRAN.R-project.org/package=haven

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). Leibniz Institute for Educational Trajectories. Bamberg. https://www.neps-data.de/Portals/0/Survey%20Papers/SP_VIII.pdf

Xie, Y. (2021). *Knitr: A general-purpose package for dynamic report generation in r* [R package version 1.33].

Zhu, H. (2019). *Kableextra: Construct complex table with 'kable' and pipe syntax* [R package version 1.1.0]. https://CRAN.R-project.org/package=kableExtra

Zinn, S., Steinhauer, H. W., & Aßmann, C. (2017). *Samples, Weights, and Nonresponse: the Student Sample of the National Educationalonal Panel Study (Wave 1 to 8)* (NEPS Survey Paper No. 18). Leibniz Institute for Educational Trajectories, National Educational Panel Study. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/8-0-0/SC5_W_8-0-0_en.pdf