

README

Children of Immigrants Longitudinal Survey in 4 European Countries

Campus Use File (Germany) Waves 1-3

Release version: 1.0.0

Release date: 21 March 2022

Citation: Soiné, Hannah, Lena Arnold, Jörg Dollmann, Victoria Kerzner, and Markus Weißmann. 2022. Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) – Campus Use File (Germany). Readme. Waves 1-3, v1.0.0. Mannheim: Mannheim University.

Content

- 1 Introduction..... 1
- 2 Access information and publications..... 2
- 3 Data set and file overview 3
- 4 Sample selection..... 5
- 5 CUF-specific variables 6
 - 5.1 ID variables 6
 - 5.2 Grade terciles 6
 - 5.3 Religious affiliation..... 7
 - 5.4 Body mass index (BMI) 7
 - 5.6 Generational status..... 8
- 6 References..... 8

1 Introduction

This readme contains important information about the campus use file (CUF) of the Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) data. In the introduction, we briefly describe the aim of the CILS4EU project, give an overview of the data collection process and highlight the main differences between campus use file and scientific use files. In section two, you can find more information about how to access the data while section three gives an overview of the data sets and documentation included in the campus use file package. For data protection reasons, the CUF only contains a subsample of the CILS4EU data. Section four describes this subsample in more detail. Lastly, some variables from the scientific use file (SUF) were altered (e.g., aggregated) for release in the CUF. Section five provides more information about these CUF-specific variables.

The aim of the CILS4EU project collected data from youth in Europe with a specific focus on the **structural, cultural, social and emotional integration of the children of immigrants** in four countries: Germany, the Netherlands, Sweden, and England. In 2010, more than 18,000 students aged 14 were interviewed in their schools and approached again in the two following years. A three-stage sample design was used: In the first stage, schools enrolling grades with mainly 14-year-old students were sampled proportionally based on their size. Schools with a higher proportion of immigrants were oversampled. At the second stage, two classes with mainly 14-year-old students were randomly sampled per school. Lastly, all students of the sampled classes were part of the sample.

Data in the first wave was collected exclusively in the school context, the subsequent two data collections were a mix of in-school and in-home interviews. The majority of interviews were self-completion paper questionnaires, but in waves two and three, respondents were also approached via e-mail and phone. Please be sure to refer to the documentation materials (technical reports, questionnaires, and supplementary materials) of the SUF for detailed information about sampling, fieldwork, and response rates. They are included in the CUF data package as well as freely available on <https://www.cils4.eu/>.

The full SUF as well as a reduced version scientific use file is available to senior researchers and graduate students after filing a request. For more information about how to access these data, please refer to studies ZA5353 (full version) and ZA 5656 (reduced version) in the GESIS data catalogue (<https://www.gesis.org/en/home>). However, we do not want to limit access to the CILS4EU data to senior researchers. We believe that the data are a valuable resource for teaching and learning for undergraduate students, too. For that reason, we created the CUF, which is **available to all undergraduate and graduate students**, provided they are supervised or taught by a senior researcher with an institutional affiliation.

It is important to note that there are some **major differences between the SUF and CUF** of the CILS4EU data. First, the CUF only contains the German data from the CILS4EU project. Data from England, the Netherlands, and Sweden are not available in the CUF. Second, the CUF consists of a subsample of the German CILS4EU sample. For more information about sample selection, please refer to section four. Third, certain variables from the SUF are not included in the CUF. The documentation materials of the SUF contain a full list of variables available to researchers in the full version of the CILS4EU data. Lastly, certain variables in the CUF are modified versions of variables included in the SUF. For more information about their generation, please refer to section five.

Despite these modifications, which were necessary for data protection reasons, the CUF is a great tool for students to learn about and practice data analysis with real-world data. Since it contains three waves, it is especially well suited for learning about **longitudinal data analysis**. The CUF also contains social network data on classroom networks in waves one and two. These can be used for **social network analysis**. The CUF further includes **extensive documentation materials**, so students and teachers have all the information that is important for data analyses at their disposal. Lastly, due to the thematic focus on the integration of children of immigrants, the CUF data can be used by students to develop and work on research questions in the area of migration and integration in the German context.

2 Access information and publications

The CILS4EU CUF can be used by students and teachers from anywhere as long as they adhere to the conditions – especially with regards to data security – that are specified in the **data use agreement**. This agreement is available from the GESIS website (<https://search.gesis.org>, search for “CILS4EU Campus Use File”) and must be filled out by the researcher who is teaching or supervising student(s) using the CUF. This can be in the form of methodological and/or substantial courses as well as group or individual capstone and thesis projects. Please note that only researchers with an institutional affiliation can apply for the data. There is no restriction for students, the CUF is available to both undergraduate and graduate students.

After handing in the data use agreement, the researcher who applied for the data will receive the CUF data and documentation materials as a ZIP-file via e-mail. They can then distribute it to their student(s). Only students who are listed on and have signed the data use agreement submitted by their teacher/supervisor are allowed to have access to the CUF!

Please cite the CUF data and documentation if you use them in your work. In the document *ZA5994_Citation_version1.0.0* you will find citation guidelines to help you do so. We strongly advise that any publications (e.g., theses, term papers) that are based on CILS4EU CUF data should **not be published in scientific journals in their original form**. As explained in greater detail in section four, the CUF is a non-random subsample of the German part of the CILS4EU data. As such, substantive findings from analyses of these data might be biased and are not generalizable.

If you wish to publish work that is based on CUF data, please apply for the full or reduced version SUF and repeat the analyses with these data. The SUF contains more than 18,000 observations across four countries and weights to adjust for uneven sampling probabilities. It also offers additional variables that are useful or necessary for certain analyses. The SUF is available to students who are in the final stages of their Master’s program and above. Information how to apply for the SUF data can be found on <https://www.cils4.eu/>.

3 Data set and file overview

The CUF ZIP-file contains many different data and documentation files. The following table provides an overview of the **files included in the CUF package** and their content. Column one contains the file type, column two the file name, and column three a description of the respective file. The codebook, data sets, citation guidelines, and this readme document are CUF-specific. The other documents are identical to the ones included in the SUF package, but contain important information relevant to the CUF data.

We provide separate data sets with thematically related variables. Each corresponds to a separately administered questionnaire or thematic module. Data sets can easily be combined using the person identifier `youthidCUF`. All data files are provided in `.dta`-format. This format is readable using statistical software like Stata or SPSS. Alternatively, the data can also be converted to other formats with different software (e.g. using the `haven` package in R or the `pandas` library in Python).

File type	File name	File content
Codebook	ZA5994_cod_cuf.pdf	This is the codebook of the CUF. All variables in all of the data sets included in the CUF are listed and described here. You will find the variable name, values, question number and wording (corresponds to number and wording in the master questionnaires), and – if applicable – additional information pertaining to that variable for all variables in the CUF.
Data set	w1_ya_ge_v1.0.0_cuf.dta	This is the CUF-version of the youth achievement data set. It contains respondents' scores on the language and cognitive ability tests implemented in wave one (8 variables).
Data set	w1_yc_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave one youth classmates data set. Each student nominated others in their class for a number of different criteria (e.g., best friends, live close). You can find these nominations and some supplementary information here (220 variables).
Data set	w1_yf_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave one youth friends data set. It contains data on several characteristics of up to five of the respondent's best friends and the respondent's interaction with them (81 variables).
Data set	w1_ym_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave one youth main data set, which corresponds to the biggest portion of the interview. It contains respondent's answers to a broad range of questions covering sociodemographic information, structural, cultural, and social integration, and well-being (231 variables).
Data set	w2_yc_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave two youth classmates data set. Similar to the wave one procedure, each student nominated others in their class for a number of different criteria. You can find these nominations and some supplementary information here (145 variables).
Data set	w2_ym_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave two youth main data set, which corresponds to the biggest portion of the interview. It contains respondent's answers to a broad range of questions covering sociodemographic information, structural, cultural, and social integration, and well-being (238 variables).

Data set	w3_yf_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave three youth friends data set. It contains data on ethnic and educational background of up to three of the respondent's best friends (16 variables).
Data set	w3_ym_ge_v1.0.0_cuf.dta	This is the CUF-version of the wave three youth main data set, which corresponds to the biggest portion of the interview. It contains respondent's answers to a broad range of questions covering sociodemographic information, structural, cultural, and social integration, and well-being (142 variables).
Data set	w123_tr_ge_v1.0.0_cuf.dta	This is the CUF-version of the tracking data set. It contains general sample information and the participation status for all three waves (16 variables).
Other material	ZA5994_README_cuf.pdf	This is the present document that gives a short introduction of the CILS4EU CUF, how it can be accessed, what files it contains, and how it differs from the SUF. Every user should read this document before working with the CUF.
Technical report	ZA5353_TechnicalReport_Wave1.pdf	This is the technical report for wave one of the CILS4EU data. It contains information about sample selection, response rates, the development of the survey instruments (including the achievement test, for which the corresponding questionnaire is not available), and fieldwork procedures. It was not altered for inclusion in the CUF, it is the same report that is also included in the wave one SUF.
Technical report	ZA5353_TechnicalReport_Wave2.pdf	This is the technical report for wave two of the CILS4EU data. It contains information about response rates, the development of the survey instruments, and fieldwork procedures. It was not altered for inclusion in the CUF, it is the same report that is also included in the wave two SUF. Note that additional information, e.g. about sample selection, can be found in the wave one report.
Technical report	ZA5353_TechnicalReport_Wave3.pdf	This is the technical report for wave three of the CILS4EU data. It contains information about response rates, the development of the survey instruments, and fieldwork procedures. It was not altered for inclusion in the CUF, it is the same report that is also included in the wave three SUF. Note that additional information, e.g. about sample selection, can be found in the wave one report.
Questionnaire	w1_yc_master.pdf	These are the master questionnaires corresponding to the individual data sets listed above. These documents were not altered for inclusion in the CUF, they are identical to the ones included in the SUF. They consist of the English translation of all questions asked in respective questionnaire, including those that were only asked in specific countries (noted as subscript next to the question number). The question order in the master questionnaires corresponds to that used in the paper questionnaire, online and telephone questionnaires (in waves two and three) differed slightly. The question number included in the variable description in the CUF codebook corresponds to the question number in these master documents.
Questionnaire	w1_yf_master.pdf	
Questionnaire	w1_ym_master.pdf	
Questionnaire	w2_yc_master.pdf	
Questionnaire	w2_ym_master.pdf	
Questionnaire	w3_yf_master.pdf	
Questionnaire	w3_ym_master.pdf	

Other material	ZA5353_Sociometric FieldworkReport_Wave1.pdf	These reports contain further information about the classroom network data collected during waves one and two. They cover the fieldwork process, describe the network data in more detail, and explain important differences in the survey unit structure and nomination behavior on the individual and class level. These reports were not altered for inclusion in the CUF, they are the same reports that are also included in the SUF.
Other material	ZA5353_Sociometric FieldworkReport_Wave2.pdf	
Other material	Dollmann_et_al_2014.pdf	This is a working paper describing how generational status and country of origin was coded in the CILS4EU data and what problems arose with that. While the SUF version of the ethnic origin variable is not available in the CUF, the CUF-specific version of it still relies on the variables created by the procedure described here.
Other material	ZA5353_Examining_the_Diversity_revised.pdf	This is an updated version of the 2014 working paper described above that incorporates the information collected in the second data collection wave.
Other material	ZA5994_citation_cuf.pdf	This document offers guidelines how the CILS4EU CUF data and the accompanying documentation materials should be cited if you use them in your work.

4 Sample selection

As mentioned briefly in section one, the CUF only includes a subsample of the whole CILS4EU data set. The following four restrictions were made to the sample included in the full SUF data set to produce the CUF sample:

1. Only data from Germany (`country = 2`) is included in the CUF. No data sets are available for England, the Netherlands, and Sweden, where the CILS4EU survey was also conducted.
2. Only those respondents who participated in all three waves (`ps_w1_yRV != 19 & ps_w2_yRV != 19 & ps_w3_yRV != 19`) are included in the CUF. Be aware that not all respondents participated in all modules, which is why the number of observations might differ between the module-specific data sets included in the CUF.
3. Respondents from the federal state of Berlin are not included in the CUF because the questionnaires administered there differed slightly from those administered in the other federal states.
4. The smallest (less than ten students) and largest (more than 26 students) school classes were dropped in the creation of the CUF data set. This means that respondents from these classes are not included in the CUF data.

As a result of these changes, the CUF data contains a total of $N=2,498$ adolescents nested in 159 classes, which are again nested in 97 different schools in Germany. That means that 49.8% of the full wave one student sample realized in Germany ($N=5,013$) is included in the CUF.

5 CUF-specific variables

As mentioned in the introduction, some variables are included in the CUF, but not the SUF. These are aggregated or otherwise modified versions of variables present in the SUF which could not be included in the CUF in their original form for data protection reasons¹. These CUF-specific variables are described in this section of the readme. Note that some variables in the SUF could not be included in the CUF at all. Please refer to the codebooks of the SUFs available on the CILS4EU website <https://www.cils4.eu/> for a full list of variables available in these data.

5.1 ID variables

There are three ID variables in the CUF. One uniquely identifies respondents (`youthidCUF`), one school classes (`classidCUF`), and one the schools (`schoolidCUF`). Equivalent variables are available in the SUF (`youthid`, `classid`, and `schoolid`). The CUF-specific versions follow the same logic as the SUF IDs: The `schoolid` is a four-digit number that specific to the school that the respondent went to. The `classid` is a six-digit number specific to the class the respondent attended. It consists of the `schoolid` and two more digits specific to the class within a school. The `youthid` is the `classid` of the student's class and two more digits specific to the student within the class. To prevent merging the CUF and SUF data, a random four-digit number between 1000 and 6000 was assigned as the `schoolidCUF`. The `classidCUF` and `youthidCUF` variables were then changed accordingly², so that the first four digits are the new `schoolidCUF`.

5.2 Grade terciles

There are two different types of grades which CILS4EU respondents provided information about: First, the grades in their latest school report in math (SUF: `y1_gradem_ge`, `y2_gradem_ge`, `y3_gradem_ge`), German (SUF: `y1_gradesc_ge`, `y2_gradesc_ge`, `y3_gradesc_ge`), and English (SUF: `y1_gradee_ge`, `y2_gradee_ge`, `y3_gradee_ge`). Questions about these were asked in all three waves. Second, the grades in their school-leaving certificate in math (`y3_ggradem_ge`), German (`y3_ggradesc_ge`), English (`y3_ggradee_ge`), and overall (`y3_ggradeo_ge`). Questions about these were only asked in wave three. While numeric values for grades are provided in the SUF, we aggregated the information into terciles in the CUF. The variables `y*_grade*_geCUF` in the CUF indicate whether a student is in the top 33% ("1st tercile"), middle 33% ("2nd tercile"), or bottom 33% ("3rd tercile") of their class with regards to that grade type and subject. It is important to note that the grades two, three, and four account for around 90% of all grades reported by students. Also, the grade distribution varies by class: Some teachers use only two of the six possible grades, while others use all of them. If teachers use few different grades or if the number of respondents in a class is small, the creation of terciles becomes difficult. If there are only two different grades reported by students within a class, the group cannot be split into terciles. Instead, we e.g. have a class with ten students in the 1st tercile and three students in the 3rd tercile, but no one in the 2nd tercile. Furthermore, even if there are at least three different grades assigned to students within one class, due to the uneven distribution of grades the size of terciles within a class is uneven as well. For example, there could be ten students in the first tercile (received the grade two), four students in the second tercile (received the grade three), and one student (received the grade six) in the third tercile in a specific class.

¹ Data processing for the CUF was done using version 16.1 of the statistical software Stata.

² Plus, in three cases the last two digits of the `youthid` were changed to create the `youthidCUF` because they ended in a unique combination of digits that would have otherwise made it possible to identify them in the SUF.

5.3 Religious affiliation

The variables *y1_rell1CUF*, *y2_rell1CUF*, and *y3_rell1CUF* contain information about the respondent's religious affiliation. They are based on the variables *y1_rell1*, *y2_rell1*, and *y3_rell1* in the SUF. The CUF and SUF versions differ because “Buddhism”, “Hinduism”, “Judaism”, and “Sikhism” were removed as separate categories. They are included in the “Other religion” category in the CUF-specific variables.

5.4 Body mass index (BMI)

While height and weight of the respondents is included in the SUF, we combine and aggregate this information for inclusion in the CUF. To do so, we use the variables *y1_heightRV* and *y1_weightRV* to calculate every respondent's body mass index (BMI). For that purpose, height was converted from meter into centimeters and the following formula was applied: $BMI = \frac{weight}{height^2}$

In a second step, the BMI-values were split into five categories indicating in which percentile of the population the respondent ranges with their BMI compared to children of the same sex and age. For example, the value of five, which is labeled as “severely underweight”, means that only five percent of children in the same sex and age group have this or a lower BMI and 95 percent have a higher BMI. The other cutoff points are 10 (“underweight”), 50 (“normal weight”), 85 (“overweight”), and 95 (“severely overweight”).

Age	Category	BMI-Range Boys	BMI-Range Girls
13	Severely underweight	< 15.6	< 15.2
	Underweight	15.6 – 16.2	15.2 - 15.6
	Normal weight	> 16.2 – 21.6	> 15.6 - 22.0
	Overweight	> 21.6 - 24.5	> 22.0 - 24.4
	Severely overweight	> 24.5	> 24.4
14	Severely underweight	< 16.1	< 16.2
	Underweight	16.1 - 16.7	16.2 - 17.0
	Normal weight	> 16.7 - 22.5	> 17.0 - 23.1
	Overweight	> 22.5 - 25.7	> 23.1 - 26.0
	Severely overweight	> 25.7	> 26.0
15	Severely underweight	< 17.0	< 16.9
	Underweight	17.0 - 17.8	16.9 - 17.6
	Normal weight	> 17.8 - 23.0	> 17.6 - 23.1
	Overweight	> 23.0 - 25.9	> 23.1 - 27.7
	Severely overweight	> 25.9	> 27.7
16	Severely underweight	< 17.8	< 16.9
	Underweight	17.8 - 18.5	16.9 - 17.8
	Normal weight	> 18.5 - 23.6	> 17.8 - 22.7
	Overweight	> 23.6 - 26.0	> 22.7 - 24.2
	Severely overweight	> 26.0	> 24.2
17	Severely underweight	< 17.6	< 17.1
	Underweight	17.6 - 18.6	17.1 - 17.8
	Normal weight	> 18.6 - 23.6	> 17.8 - 23.3
	Overweight	> 23.6 - 25.8	> 23.3 - 25.7
	Severely overweight	> 25.8	> 25.7
18	Severely underweight	< 17.6	< 17.6
	Underweight	17.6 - 18.6	17.6 - 18.3
	Normal weight	> 18.6 - 23.9	> 18.3 - 23.4
	Overweight	> 23.9 - 26.8	> 23.4 - 25.0
	Severely overweight	> 26.8	> 25.0

The classification tables with information on the BMI distribution by age and gender were taken from the online “BMI-calculator for children” by German health insurance company BKK24 that is available online under <https://www.bkk24.de/lbl/ratgeber-gesundheit/artikel-lesen/bmi-rechner-fuer-kinder.html> (accessed 01.02.22). They are based on data collected by Coners et al. (1996). The table above is an overview of these data by age and gender for the ages present in the CUF data.

5.6 Generational status

The variables `y1_generationGCUF`, `y2_generationGCUF`, and `y3_generationGCUF` indicate whether the respondent has no migration background, immigrated to Germany themselves (“1st generation”), or is a child of immigrants (“2nd generation”). The starting point for the creation of these variables are the variables `y1_generationG`, `y2_generationG`, and `y3_generationG` in the full version of the SUF. Please consult the two working papers by Dollmann et al. (2014, 2015) that are included in the CUF package for a detailed description of how the generational status variables in the SUF were created. In the CUF-specific variables, we combined categories one to four of the SUF-variables into “1st generation”, categories five to seven into “2nd generation” and the remaining non-missing categories into “no migration background”. In other words, respondents who are foreign-born and migrated to Germany themselves are categorized as “1st generation” in the CUF variable (regardless of migration age). Respondents who were born in Germany, but have at least one parent who was born abroad are categorized as “2nd generation”. If the respondent and both parents were born in Germany, they are categorized as “no migration background”. This means that 3rd generation immigrants whose grandparents are foreign born are included in this category in the CUF variable. Missing values were taken from the SUF unchanged.

6 References

Coners, H., Himmelmann, G. W., Hebebrand, J., Hesecker, H., Remschmidt, H., Schäfer, H. (1996). Perzentilkurven für den Body-Mass-Index zur Gewichtsbeurteilung bei Kindern und Jugendlichen ab einem Alter von 10 Jahren. *Der Kinderarzt* 8: 1002-1007.

Dollmann, J., Jacob, K., Kalter, F. (2014), Examining the Diversity of Youth in Europe. A Classification of Generations and Ethnic Origins Using CILS4EU Data. *MZES Working Paper 156*. Available at <https://www.cils4.eu/>.

Dollmann, Jörg and Jacob, Konstanze (2015). Examining the Diversity of Youth in Europe. A Classification of Generations and Ethnic Origins Using CILS4EU Data. A Wave 2-Update, v2.3.0. Mannheim: University of Mannheim. Available at <https://www.cils4.eu/>.