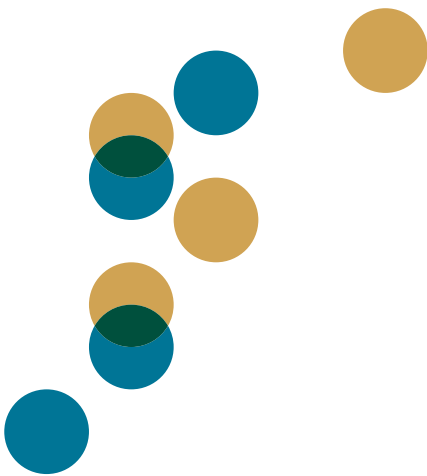


Working Paper

Examining the Diversity of Youth in Europe

A Classification of Generations and Ethnic Origins Using CILS4EU Data
(Technical Report)

Jörg Dollmann, Konstanze Jacob,
Frank Kalter



mannheimer zentrum
für europäische sozialforschung

UNIVERSITÄT
MANNHEIM

Jörg Dollmann
Konstanze Jacob
Frank Kalter

Examining the Diversity of Youth in Europe

A Classification of Generations and
Ethnic Origins Using CILS4EU Data
(Technical Report)

Arbeitspapiere – Working Papers
Nr. 156, 2014

Mannheimer Zentrum für Europäische Sozialforschung

Dollmann, Jörg:

Examining the Diversity of Youth in Europe. A Classification of Generations and Ethnic Origins Using CILS4EU Data (Technical Report) / Jörg Dollmann; Konstanze Jacob; Frank Kalter. – Mannheim: 2014
(Arbeitspapiere - Mannheimer Zentrum für Europäische Sozialforschung; 156)
ISSN 1437-8574

Not available in book shops.

Token fee: € 3,00

Purchase: Mannheimer Zentrum für Europäische Sozialforschung (MZES)

D – 68131 Mannheim

www.mzes.uni-mannheim.de

Editorial Note:

Jörg Dollmann is a research associate in the international project “Children of Immigrants Longitudinal Survey in Four European Countries” (CILS4EU) at the Mannheim Centre for European Social Research (MZES). His research interests are mainly social and ethnic educational inequalities as well as the role of institutional settings in creating these disparities.

Konstanze Jacob is a research associate in the international project “Children of Immigrants Longitudinal Survey in Four European Countries” (CILS4EU) at the Mannheim Centre for European Social Research (MZES). She is also a research and teaching associate at the Chair of General Sociology, University of Mannheim. Her research mainly focuses on migration and integration, intergenerational transmission, religion and religiosity and cultural capital theory.

Frank Kalter is a Professor of Sociology at the School of Social Sciences, University of Mannheim. He is currently the Director of the Mannheim Centre for European Social Research (MZES). His major research interests include migration, integration of ethnic minorities, and modelling of social processes.

Abstract

It is well known in migration research that the integration of immigrants and their descendants takes place over time and especially over generations. Also, it is common knowledge that immigrants stemming from different regions of the world show different paces and patterns of integration into their respective host societies. However, despite the importance of generational status and ethnic origin for studying integration processes, defining these constructs in migration surveys oftentimes is a challenging experience, mainly for three reasons. First, many surveys simply lack the collection of relevant information like the grandparents' countries of birth that enables researchers to categorise children of immigrants beyond the second generation. Second, regardless of the depth of the information collected, missing values on central variables necessary to define the generational status and ethnic origin may lead to the problem that the existing information cannot be used to reconstruct the migration history of respondents. Third, deviations from standard migration biographies assumed by existing typologies must be taken into account as well. This working paper provides an approach to identifying immigrants, their generational status and their ethnic origin, exemplified by using a recent survey conducted in Europe: the Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU). Using data from CILS4EU, we propose several general heuristics in order to construct a comprehensive generational status variable as well as a variable indicating the country of origin. We discuss and provide solutions for two of the problems mentioned above, the treatment of missing values as well as information about non-standard migration biographies. Subsequently, we present first descriptive results on the composition of the CILS4EU sample with respect to generational status and ethnic origin of the respondents. Besides demonstrating the procedure in CILS4EU when constructing the two central constructs generational status and ethnic origin, with this paper we would like to stimulate other migration surveys to consider similar heuristics in order to increase the comparability between the analyses and results originating from different data sources.

Contents

Introduction	7
<hr/>	
The Construction of the Generational Status Variable	8
Standard Classification Approach	10
Treatment of Missing Data	13
Non-Trivial Cases	21
Descriptive Results	23
<hr/>	
The Construction of the Country of Origin Variable	24
Standard Classification Approach	25
Treatment of Missing Data	28
Non-Trivial Cases	32
The Country of Origin Revisited – Using Additional Information	33
Descriptive Results	35
<hr/>	
Summary	36
<hr/>	
References	44
<hr/>	
Appendix	46

Introduction¹

Over the last years and decades there has been a continuous boom in research on the integration of children and adolescents with a migration background in many European countries. This is due to the fact that the descendants of immigrants represent a considerable and rapidly growing proportion of young people in contemporary Europe and that studying youths' situation in their formative years and along the life course is thus crucial to understanding the current state of our societies and predicting their future development. However, what is often lacking are adequate data to analyse these integration processes. And even if such data are available, two of the basic constructs for migration and integration research, generational status and country of origin of immigrants, are generated differently in different data sources, making it difficult to compare results from different surveys. Proposed in this working paper are several general heuristics for the construction of a comprehensive generational status variable as well as a variable indicating the country of origin. We would like to stimulate other migration surveys considering similar heuristics in order to increase the comparability between the analyses and results originating from different data sources.

The general considerations and heuristics are applied to the data of the Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) (Kalter et al. 2014a, 2014b) which aims to study integration processes of immigrant children in England, Germany, the Netherlands, and Sweden. In order to achieve this goal, comprehensive and comparative information on several integration dimensions was collected, like cognitive-cultural (e.g., language skills and use, cultural practices), structural (e.g., educational attainment, labour market inclusion), social (e.g., friendship patterns, romantic relationships), and emotional-cultural integration (e.g., sense of belonging, attitudes and norms).

The target population of the survey in each country encompasses young people with and without an immigrant background at around the age of 14 who were sampled using a school-based sampling approach.² Overall, 18,716 adolescents were interviewed in the first wave, with two subsequent annual waves. In order to identify children of immigrants from families with very distant and diverse migration histories, fine-grained measures capturing the countries of birth of the children, their parents and their grandparents were implemented in the survey instrument of the first wave. To validate and complement this information, we not only interviewed the respective child itself but also conducted, if possible, a cross-sectional interview with one of its parents parallel to the first wave of data collection.³

The working paper provides a suggestion on how this information on the countries of birth of the different actors can be used in order to clarify whether target persons have an immigrant background at all and to determine the *generational status* and the concrete *countries of origin* of the target persons in CILS4EU. The variables generated for the generational status and the ethnic origin of the children using the procedures as described in the following sections are provided in the CILS4EU-data available at the GESIS Data Archive for the Social Sciences (Kalter et al. 2014a, 2014b).

1 Financial support from NORFACE research programme on Migration in Europe - Social, Economic, Cultural and Policy Dynamics is acknowledged.

2 For details on the sampling process see Technical Report available at the GESIS Data Archive for the Social Sciences or on the project-website www.cils4.eu.

3 In one Federal State in Germany, some of the relevant information was only collected in the third wave. In Sweden, the parent interviews were partly conducted parallel to the second wave of data collection among the children.

The Construction of the Generational Status Variable

The concept of “generation” is widespread in integration research and is usually used to measure the success of immigrants’ integration processes into the receiving society over time (Crul and Vermeulen 2003, Farley and Alba 2002, Heath and Brinbaum 2007, Heath et al. 2008, Perlmann and Waldinger 1997, Rumbaut 1997, Warner and Srole 1945, Waters and Jimenez 2005). According to the considerations of Alba (1988: 213), for whom the concept of generation “refers to ancestral distance from the point of arrival in a society” of immigrants, we will construct the generational status variable by defining the target person of the survey as the reference point in order to determine the “ancestral distance” between the target person and his or her ancestors who arrived earlier in the receiving country for the first time. According to this approach, we will first consider the country of birth of the child, then the countries of birth of his or her parents and finally the countries of birth of his or her grandparents. The smallest possible “ancestral distance” applies to cases where the target person him- or herself migrated into the receiving society. In order to gain insights into the intra-generational integration processes over time of target persons with a personal migration history we will also consider his or her age at arrival in the receiving society (Barnett et al. 2012, Oropesa and Landale 1997, Parameshwaran 2013, Rumbaut 1997, 2004, Segeritz et al. 2010, Waters and Jimenez 2005).

In the following sections we will describe the construction of the generational status variable (variable name in the data: `generationG`) of the child following these general considerations.⁴ For the generational status variable, we will thereby use information provided by the child about his or her country of birth and age upon arrival as well as information provided in the parent interview about the biological parents’ countries of birth as well as those of the grandparents. Figure 1 provides an example of the question about the country of birth in the children’s questionnaire, which was implemented similarly in the parent questionnaire.⁵

Figure 1: Example of the questions about the respondent’s country of birth (here: from the child’s questionnaire)

Attention: Remember to check for a “Go to” instruction after you answer the question below.

3 In which country were you born?

<Survey country> → Go to **5**
 <Country 1>
 <Country 2>
 <Country 3>
 Other country → Please specify:

4 How old were you when you moved to <survey country>?

Age in years:

4 The Stata do-file generating all variables described in this Working Paper can be found at: http://www.mzes.uni-mannheim.de/publications/wp/wp-156_supp_stata_do.txt

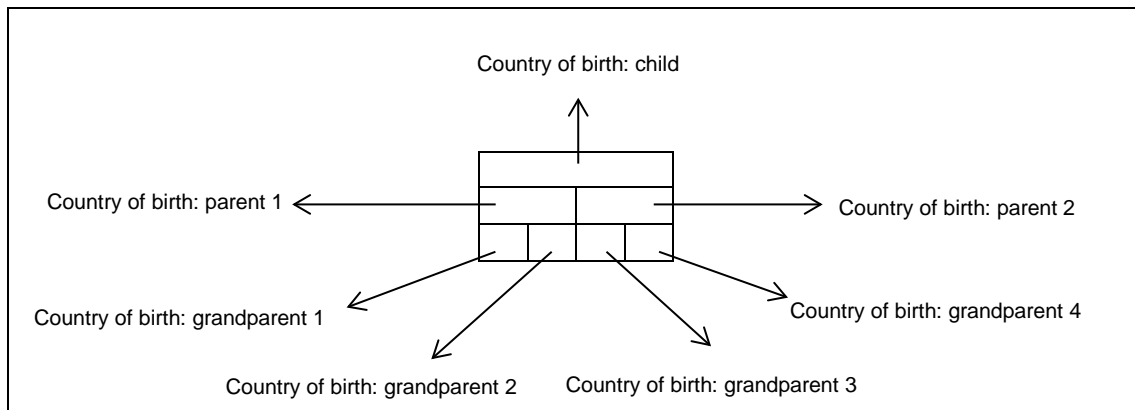
5 All survey instruments for children and parents are available at the GESIS Data Archive for the Social Sciences or on the project website www.cils4.eu.

For those cases where the parent information is not available (due to item non-response or refusal of the parents to participate in the survey) or where the information provided in the parent interview was not from the biological parents of the child, we will use information provided by the child about his or her biological parents' as well as grandparents' countries of birth. However, regarding the grandparents' countries of birth, we only asked the child whether they were born in the survey country or not – mainly due to time constraints of the survey and due to expectations that this information is less likely to be memorised by the children (Dagevos 1997).

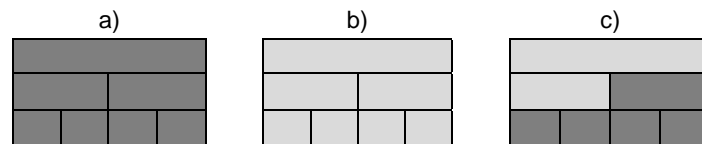
In one German Federal State we were not allowed to ask about third parties in the school-context, i.e., during the first and the second wave. Therefore, we asked respondents in the third wave about information on the countries of birth of the parents by means of telephone interviews. Furthermore, we asked whether the grandparents were born in the survey country or not. The questions are identical in wording to those used in the first wave of data collection. This information will be used in addition for the construction of the generational status as outlined below.

In the following, we describe how the generational status variable is constructed using the information about these seven actors (child, two parents and four grandparents). For the ease of illustration later on, the ancestry pattern of a child can be displayed by the following scheme:

Figure 2: The country-of-birth ancestry-scheme



The rectangle representing the country of birth of each actor is displayed in light grey if the country of birth is the survey country, which is the country where the survey was conducted (i.e. United Kingdom,⁶ Germany, the Netherlands, and Sweden). In contrast, if the actor was born outside the respective survey country, the country is displayed in dark grey. The following examples illustrate three specific cases:



Example a) represents a case where all actors, the child, the parents and the grandparents, are foreign-born, whereas example b) illustrates the opposite – a case where all actors were born in the survey coun-

⁶ Although the survey was only conducted in England, births in Scotland, Wales, and Northern Ireland are also classified as births in the "survey country".

try. Example c) stands for a child who was born in the survey country, with one parent also born in the survey country and one parent born abroad, but with all four grandparents being foreign-born.

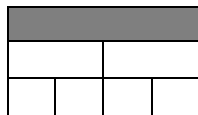
Standard Classification Approach

The construction of the generational status variable is geared to the common procedure in migration research (Rumbaut 2004), although we expand the commonly constructed categories by several extra categories. We thereby propose a very finely grained approach when defining the generation of the survey's target persons, distinguishing between the 1st, 1.25th, 1.5th, 1.75th, 2nd, 2.5th, 2.75th, Interethnic 2nd, 3rd, 3.25th, 3.5th, Interethnic 3rd, and 3.75th generations and between natives. This extensive categorisation is motivated by previous research in the field of migration and integration that shows different patterns of adaptation to the host society of certain generational status groups (Alba et al. 2002, Becker 2011, Farley and Alba 2002, Gresch and Kristen 2011, Ramakrishnan 2004, Segeritz et al. 2010).

The following diagrams, using the country-of-birth ancestry-scheme described above, illustrate how these different groups are constructed following the general considerations of a top-down approach, where we will first consider the country of birth of the child, then the countries of birth of the parents and finally the countries of birth of the grandparents. White rectangles mean that the country of birth of this actor is irrelevant for the assessment of the "ancestral distance from the point of arrival" (Alba 1988: 213) and therefore for the definition of the generation for this specific child.

1st Generation

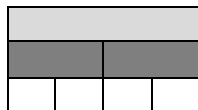
If a child is foreign-born and therefore itself migrated to the survey country, this child is defined as a 1st-generation immigrant, irrespective of the countries of birth of its ancestors.



By using information about the age of the child upon arrival, we further differentiate within the 1st generation between children who arrived after the age of 10 (1.25th generation), between the ages of 6 and 10 (1.5th generation), or before the age of 6 (1.75th generation) (Oropesa and Landale 1997, Parameshwaran 2013, Rumbaut 1997, 2004, Segeritz et al. 2010, Waters and Jimenez 2005).

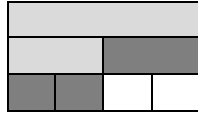
2nd Generation

Children born in the survey country, with both their parents being foreign-born, are members of the 2nd generation. The countries of birth of the grandparents are irrelevant.

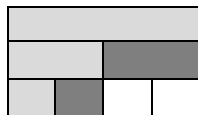


2.5th Generation

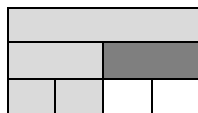
Children born in the survey country, with one parent also born in the survey country, but with the other parent being foreign-born, are classified as the 2.5th generation. For this generation, the parent born in the survey country needs to be a descendant of ancestors both of whom were foreign-born (i.e., this parent is a 2nd-generation immigrant), while the countries of birth of the parents of the foreign-born parent are irrelevant.

2.75th Generation

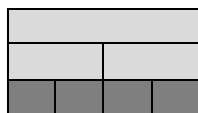
Children born in the survey country, with one parent born in the survey country and one parent born abroad, are classified as the 2.75th generational cohort, if one parent of the parent born in the survey country was also born in the survey country and the other born abroad. The countries of birth of the parents of the foreign-born parent are irrelevant.

Interethnic 2nd Generation

The interethnic 2nd generation comprises all children born in the survey country, again with one parent born in the survey country and one parent born abroad. In this case however, both parents of the parent born in the survey country were also born in the survey country. The countries of birth of the parents of the foreign-born parent are irrelevant since that parent is a 1st-generation immigrant.

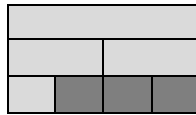
3rd Generation

The 3rd generation combines all children born in the survey country, with both parents also born in the survey country. However, in this group, all grandparents are foreign-born.



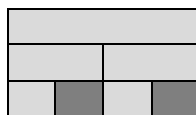
3.25th Generation

The 3.25th generation comprises all children born in the survey country, with both parents also born in the survey country. In addition, three grandparents are foreign-born, while one grandparent was born in the survey country. In this case, the child is a descendant of a parent from the 2nd generation and a parent with one parent born abroad and one parent born in the survey country.



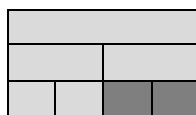
3.5th Generation

The 3.5th generation consists of children born in the survey country, with both parents also born in the survey country. In addition, two grandparents are foreign-born. However, each parent of the child has one foreign-born ancestor. In this case, the child is a descendant of parents both of whom have one parent born abroad and one parent born in the survey country.



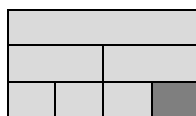
Interethnic 3rd Generation

The interethnic 3rd generation is represented by all children born in the survey country, with both parents also born in the survey country. In addition, both parents of one parent were also born in the survey country, while the other two grandparents of the child are foreign-born. This category comprises therefore children descending from a relationship between a 2nd-generation parent and a parent whose parents also were both born in the survey country.



3.75th Generation

The 3.75th generation comprises all children born in the survey country, with both parents also born in the survey country. In addition, one grandparent is foreign-born. In this sense, the 3.75th generation is to some degree comparable to the interethnic 3rd generation, as the child has one parent whose parents were both born in the survey country and one parent who is from the 2.5th, 2.75th, or Interethnic 2nd generation.



Natives

Finally, natives comprise the group where all actors were born in the survey country.

Treatment of Missing Data

Missing data is a well-known problem in social research (Allison 2001) and also appears when asking the respondents about their own and especially about their ancestor's countries of birth, since in this regard respondents often lack this knowledge (Dagevos 1997, Smith 1984). Figure 3 provides an overview on the number of missing values on the different countries of birth in CILS4EU for each of the 18,716 target persons, considering information from the children and their parents about the children's, their parents' and their grandparents' countries of birth.

Figure 3: The problem of missing data in the country-of-birth ancestry-scheme on different levels (n/%)

Child: 70/0.4			
Mother: 259/1.4		Father: 473/2.5	
Maternal Grandmother: 744/4.0	Maternal Grandfather: 952/5.1	Paternal Grandmother: 1,351/7.2	Paternal Grandfather: 1,577/8.4

The figure demonstrates that the problem of missing data increases with ancestral level, whereby grandparents' information is less likely available than parents' information, which is in turn less likely available than information on the child's country of birth. Furthermore, it is also worth noting that on the mother's side, the information is more complete than on the father's side. This may be due to the fact that the child is better informed about the mother's country of birth (e.g., in the case of separated parents where the child lives permanently with the mother), or that the biological mother is more frequently the respondent in the parent interview than the biological father.⁷

Referring to the common procedure of classifying generational status described above (Rumbaut 2004), it is evident that missing information on the countries of birth on different levels makes it difficult, or even impossible, to construct the generational status variable, which is problematic, especially for a survey focusing on children of immigrants. In order to use as much and to lose as little information on this central variable as possible, we develop some reasonable rules in order to replace missing values on the child's, its parents' and/or grandparents' countries of birth in a meaningful way. These rules and the rationale behind them will be described in the following. In addition, we will also report the number of cases in which data have been replaced according to these rules. Furthermore, we will show how we deal with those cases in which a meaningful replacement of the missing information is not possible. Finally, we will introduce a flag variable that provides an overview of the missing-data problem for a specific case.

⁷ 78.2 per cent of the parent interviews in CILS4EU were conducted with the biological mother and 20.2 per cent with the biological father.

Missing Replacement Rules

Given the problem of non-response on the child's, parents', and grandparents' countries of birth and the associated problem of construction the generational status variable, it is necessary to develop some basic rules in order to achieve reasonable assumptions about the missing information on the countries of birth on the different levels. In principle, the following rules are motivated by two basic assumptions. Firstly, we assume a *stability of residential patterns* before as well as after migration. If a person was born in the survey country, and information on their descendants is missing, it is likely that the descendant of the person born in the survey country was also born in the survey country. In contrast to this argument of residential stability of descendants of persons born in the survey country, one can also assume residential stability of ancestors of persons born outside the survey country. If a person was born outside the survey country, and information on their ancestors is missing, it is likely that the ancestor of the person born outside the survey country was also born outside the survey country. As can be seen from these two arguments, replacements of missing values in the direction of ancestors are only possible when the available information on the descendant indicates that this person was born abroad, while replacements in the direction of descendants are only reasonable if the available information on ancestors indicates that this person was born in the survey country.

Secondly, the rules outlined below are also justified by a general *homogamy* argument, which – in contrast to the first argument – is helpful to replace missing information on one actor with that on another actor at the same ancestral level, for example by using information on one parent to replace the missing information on the other parent. Given the literature on intermarriage and interethnic partner choices, with comparably low probabilities of interethnic partnerships all over Europe as opposed to ethnically homogeneous marriages, it can be assumed that persons seek spouses or partners who are similar to themselves, e.g. with respect to their migration history (Berthoud 2005, Çelikaksoy et al. 2010, Dale and Ahmed 2011; González-Ferrer 2006, Kalmijn and van Tubergen 2006, Kalter and Schroedter 2010, Kogan 2010, Muttaarak 2010, Van Tubergen and Maas 2007). Once information on one parent is missing, it is assumed that this missing information is probably similar to the information provided for the other parent.

In the following, six rules are displayed together with selected examples in order to illustrate the replacement procedure based on these two basic assumptions. However, it has to be noted that the “replacements” are not actual replacements in the data available to the scientific community (Kalter et al. 2014a, 2014b), but rather reasonable assumptions about the true values for missing information on the countries of birth of the different actors, with these assumptions being subsequently used to construct the generational status of the target person.

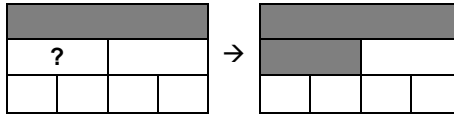
Besides the theoretical considerations based on an argument of residential stability and homogamy, we also checked the empirical validity of our assumptions. We thereby used information from actors without missing information and compared whether the logic behind the following rules can also be observed empirically. It turned out that, depending on the specific rule, between 78 and 98% of cases with complete information showed the same results as predicted with our theoretical assumptions (see table A.1 in the Appendix).⁸

8 Another possibility would have been to use Multiple Imputation in order to replace missing information on any of the necessary country of birth variables. We refrain from this approach for several reasons: First, imputation models are normally already oriented towards the final model of analysis, which we cannot foresee when providing a solution that fits all possible research interests. Second, we think that providing the scientific research community with several imputed data sets is simply not feasible and complicates the data dissemination procedure. However, it is reassuring that the high matching quotas in the Appendix show that we have replaced missing values not only with the most (theoretically) plausible, but also with the most probable values.

Rule 1: Top-Down Missing-Replacement if Descendant Is Foreign-Born

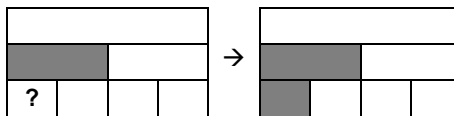
If a child is foreign-born, the missing parent(s) are also labelled as foreign-born. This also holds true for the parent-grandparent dyads with missing values on the grandparents (*residential stability argument*).

1. Foreign-born children with missing information (illustrated with a white rectangle and a question mark) on parent(s) → parent(s) is/are also foreign-born, regardless of the grandparents.⁹



In total, missing information on one or both parents has been replaced in n=109 cases.

2. Foreign-born parents, information on respective grandparents is missing → grandparents are foreign-born; regardless of whether the child was born in the survey country or not.

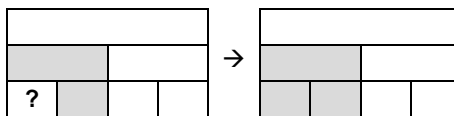


In total, missing information on one to four grandparents has been replaced in n=573 cases.

Please note, however, that there are 54 cases included in the figures above where the information of the child has been used to replace missing parent information and this parent information has then subsequently been used to replace the missing grandparent information.

Rule 2: Complement Grandparent-Dyads

If one grandparent is missing, and the other grandparent-information (on the partner of the missing grandparent) is available, we complemented the missing information with the available information (*homogamy argument*).



In total, missing information has been replaced in n=574 cases.

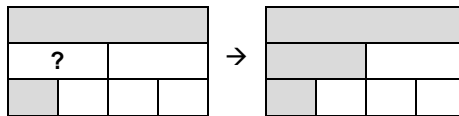
⁹ As before, white-coloured rectangles (without question mark) mean that the rule does not depend on the actual country of birth of this actor; in this example, the missing rectangle has been replaced with dark grey (foreign-born parent) irrespective of whether the grandparents (the parents of the missing actor) are foreign-born or were born in the survey country.

Rule 3: Bottom-up Missing Replacement if Ancestors Were Born in The Survey Country

Replace missing information on an actor with “ancestors born in the survey country”, as long as at least one ancestor was born in the survey country. This holds true for the grandparent(s)-parent-dyads/triads and the parent-child-dyads (*residential stability argument*).

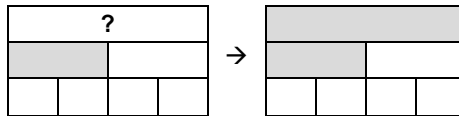
1. Grandparent(s) born in survey country, missing information on parent(s) → parent(s) were also born in survey country.

The replacement procedure for the parents holds true only if the child itself was born in the survey country (otherwise rule 1 would apply) or the child's information is missing.



In total, missing information on parents has been replaced in n=98 cases.

2. If the information on the country of birth of the child is missing, but at least one parent was born in the survey country, the child is assumed to also have been born in the survey country.

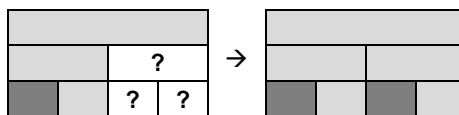


In total, missing information on children has been replaced in n=45 cases.

Please note, however, that there are 7 cases included in the figures above where the information on the grandparents has been used to replace missing parent information and this parent information has then subsequently been used to replace the missing child information.

Rule 4: Replace Missing Parent-Grandparent Triad by Complete Triad

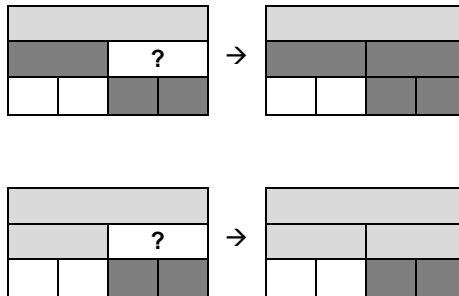
If a complete parent-grandparent triad is missing, and the other parent-grandparent triad is available, replace the missing information with the existing one (*homogamy argument*).



In total, missing information on complete triads has been replaced in n=99 cases.

Rule 5: Replace One Parent with Other Parent

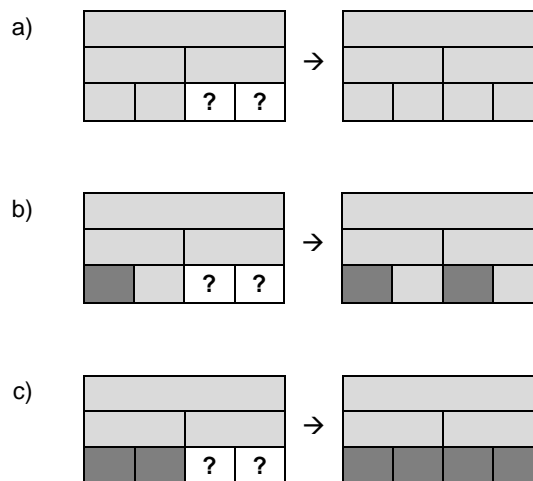
If information on a parent is missing, the missing information is replaced by the existing information on the other parent. This holds true only if the child was born in the survey country (otherwise rule 1 would apply) and the parents of the missing parent are foreign-born (otherwise rule 3 would apply) or any of this information is missing (*homogamy argument*).



In total, missing information on one parent has been replaced with information on the other parent in n=83 cases.

Rule 6: Replace Completely Missing Grandparent-Dyad with Known Grandparent-Dyad

If information on two grandparents (from one lineage) is missing, the information is replaced with existing information on both other grandparents. This holds true only as long as both parents of the child (i.e. the children of the grandparents) were born in the survey country or this information is missing. If they were born abroad, rule 1 would apply (*homogamy argument*).



In total, missing information on grandparents with information on other grandparents has been replaced in n=484 cases (scenario a: n=370; scenario b: n=42; scenario c: n=72).

Still Existing Problems

However, even when applying the rules listed above, 304 cases remain where a replacement of missing information on a child's and/or on its parents' and/or on its grandparents' countries of birth may not be feasible. The problem occurs when the parents were born abroad and the information on the child's country of birth is missing. The problem also exists when information on both parents and/or on all four grandparents is missing after application of the replacement rules listed above, but the child was born in the survey country. Finally, the problem occurs when no information about the country of birth of any actor is known. In order to use as much information as possible, we classify these different types of missing values separately in the final variable. These missing categories, together with the empirical distribution in the data, are displayed in the following

Parents foreign-born, no information on child available

?			

n=23

Child native-born, no information on parents and grandparents

?		?	
?	?	?	?

n=28

Child native-born, no information on parents – grandparents foreign-born

?		?	

n=19

Child and parents native-born, no information on grandparents

?	?	?	?

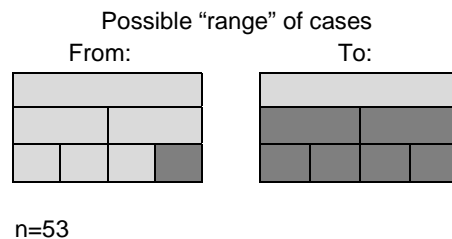
n=155

No information on any actor available

?			
?		?	
?	?	?	?

n=2

Finally, another source of uncertainties about the generational status persists. As already mentioned, in one German Federal State we were not allowed to ask for information on third parties during the school surveys. Although we included the questions about parents' and grandparents' countries of birth in the third wave of data collection, problems occur if a respondent did not participate in the third wave (unit non-response) or if the respondent did not provide information on the relevant questions (item non-response). In order to overcome this problem at least partially, we use information from the first wave, where we had asked the child – besides his or her country of birth – whether it had a so-called “immigrant background”. We explained what is meant by the phrase “immigrant background”, which is that at least one parent or grandparent was born abroad and moved to Germany later on. Given this information, this results in the following case if the child was born in Germany and stated having an “immigrant background”.

Child native-born, at least one ancestor foreign-born

If the child was born in the survey country and states not having an immigrant background (n=24), this child is classified as a native person. Finally and naturally, if the child was born outside Germany, it is classified as a 1st-generation immigrant.

Flag Variables for Missing Information

In order to provide some information about the extensiveness of the missing values for each case, we use a three-digit flag variable [genflag_missG].

Table 1: Composition of the flag variable indicating missing values

1 st digit	2 nd digit	3 rd digit
Information on child missing	Information on parent(s) missing	Information on grandparent(s) missing
0 – no information missing 1 – information missing	0 – no information missing 1 – information on one parent missing 2 – information on both parents missing	0 – no information missing 1 – information on one grandparent missing 2 – information on two grandparents missing 3 – information on three grandparents missing 4 – information on four grandparents missing

As can be seen from Table 1, the flag variable ranges from 0 to 124, where 0 means that information on all seven “country of birth”-variables is available, while 124 means that no information on these seven variables is available.

Table 2: Distribution of the missing flag variable over the countries

Flag value	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
<i>No information missing</i>										
0	3,495	81.0	4,448	88.7	4,085	93.6	4,429	88.1	16,457	87.9
<i>Information on at least one grandparent missing</i>										
1	186	4.3	107	2.1	84	1.9	142	2.8	519	2.8
2	253	5.9	206	4.1	95	2.2	150	3.0	704	3.8
3	51	1.2	10	0.2	7	0.2	25	0.5	93	0.5
4	143	3.3	49	1.0	33	0.8	121	2.4	346	1.9
<i>Information on at least one parent missing</i>										
10-14	109	2.5	85	1.7	36	0.8	112	2.2	342	1.8
<i>Information on at least both parents missing</i>										
20-24	45	1.0	86	1.7	9	0.2	45	0.9	185	1.0
<i>Information on at least the child is missing</i>										
100-124	33	0.8	22	0.4	14	0.3	1	0.0	70	0.4
Total	4,315		5,013		4,363		5,025		18,716	

Table 2 provides an overview of the distribution of the flag variable for missing values across countries. First of all, the flag variable indicates that in the vast majority of cases (87.9%) all information on all actors' countries of birth is available. Furthermore, the flag variable shows that the missing problem decreases with increasing flag values, meaning that missing values on the grandparents' countries of birth are more likely than on parents' countries of birth, whereas those are more likely than missing values on the child's level.

The flag variable can easily be used to exclude specific cases from the analyses, e.g. analyses with all cases having a flag variable smaller than 100 would mean that all cases are included where at least information on the child was available. Analyses with cases having a value on the flag variable smaller 20 would include all cases where information on at least one parent is available etc.

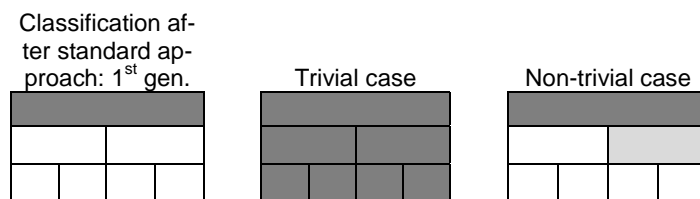
Non-Trivial Cases

Treatment of Non-Trivial Cases when Generating the Generational Status Variable

Besides the problem of missing values in central variables outlined in the previous section, a second problem arises from exceptional migration histories that do not perfectly fit into the standard classification approach. Given the generational status definition rules in this approach, the countries of birth of the ancestors of a foreign-born actor are regarded as being irrelevant for the definition of this person's generational status. However, using this definition and neglecting the information on ancestors of a foreign-born actor may lead to trivial and non-trivial cases. Given our "residential stability argument", migration is regarded as a unique event in the family; thus, all family members born after the migration event are assumed to have been born in the host country. Examples of trivial cases are those cases where one foreign-born child only has foreign-born ancestors. In contrast, a non-trivial case would imply that a foreign-born child has at least one grandparent who was born in the survey country. The following examples illustrate these trivial and non-trivial cases.

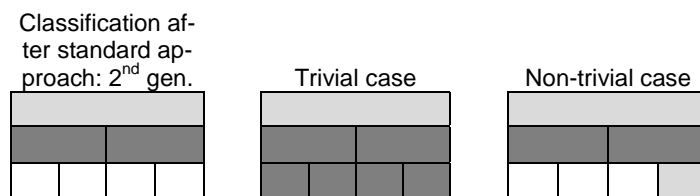
If the child is foreign-born, the child is classified as an immigrant of the first generation, regardless of whether one or both parents or one or more grandparent(s) were born in the survey country. There are 323 cases where the child is foreign-born, but at least one of the ancestors was born in the survey country.

Examples of trivial and non-trivial cases for the 1st generation with a parent born in the survey country:



If the child was born in the survey country, but both parents are foreign-born, the child is assumed to be an immigrant of the second generation, regardless of whether one or more grandparents were born in the survey country. In our data, there are 137 cases with this pattern of non-trivial information.

Examples of trivial and non-trivial cases for the 2nd generation:



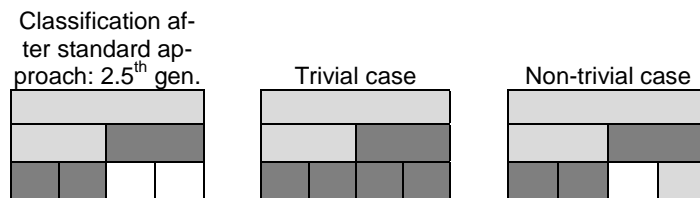
Such non-trivial cases can also occur in cases with missing information where a reasonable replacement of missing values is not possible (cf. the section above "Still existing problems"). In the following example, the information on the child is missing; the parents were all born in a foreign country, while some of the

grandparents were born in the survey country (missing pattern: “Parents foreign-born; no information on child”).



Other examples of non-trivial information are cases where the child was born in the survey country and one parent was also born in the survey country, while the other parent is foreign-born, and one or both parents of this foreign-born parent were born in the survey country. Then, the child is classified as 2.5th, 2.75th or interethnic 2nd-generation. The data set contains 306 cases with these kinds of non-trivial values.

Examples of trivial and non-trivial cases for the 2.5th generation:



Flag Variables for Non-Trivial Cases

These examples illustrate that non-trivial cases can be due to inconsistent information on one or both parents and/or on one to four grandparents, while the information on the child is always regarded as being consistent. Naturally, a non-trivial case due to inconsistent information on one actor (i.e. all actors are foreign-born, but one grandparent was born in the survey country) seems to be less problematic than a non-trivial case due to inconsistent information on six actors (child is foreign-born and therefore labelled as a 1st-generation immigrant, while its parents and grandparents were all born in the survey country). Therefore, we will introduce a two-digit flag variable to provide some information about the extensiveness of the non-trivial values for each case [genflag_ntG].

Table 3: Composition of the flag variable indicating trivial and non-trivial cases

1 st digit	2 nd digit
Information on parent(s) non-trivial	Information on grandparent(s) non-trivial
0 – all information trivial	0 – all information trivial
1 – information on one parent non-trivial	1 – information on one grandparent non-trivial
2 – information on both parents non-trivial	2 – information on two grandparents non-trivial
	3 – information on three grandparents non-trivial
	4 – information on four grandparents non-trivial

As can be seen from Table 3, the flag variable ranges from 0 to 24, where 0 means that all information for the child's, the parents' and the grandparents' countries of birth is trivial, while 24 means that no information is trivial. This would be the case if the child is foreign-born, while the parents as well as all grandparents were born in the survey country. As can be seen in Table 4, most cases are trivial cases (96% in

total). If there are non-trivial cases, these are most likely due to non-trivial information on at least one grandparent, but less likely due to non-trivial information on the parents.

Table 4: Distribution of the trivial/non-trivial flag variable over the countries

Flag value	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
<i>Only trivial cases</i>										
0	4,108	95.2	4,772	95.2	4,183	95.9	4,886	97.2	17,949	95.9
<i>Non-trivial cases with inconsistent values on grandparent level</i>										
1-4	138	3.2	181	3.6	125	2.9	78	1.6	522	2.8
<i>Non-trivial cases with inconsistent values on one parent</i>										
10-14	41	1.0	50	1.0	40	0.9	45	0.9	176	0.9
<i>Non-trivial cases with inconsistent values on both parents</i>										
20-24	28	0.7	10	0.2	15	0.3	16	0.3	69	0.4
Total	4,315		5,013		4,363		5,025		18,716	

Like the flag variable for the missing values, the flag variable for non-trivial cases can easily be used to exclude specific cases from the analyses, e.g. analyses with all cases having a value on the flag variable smaller than 10 would mean that all cases are included where at least information on both parents is consistent with the information on the child.

Descriptive Results

In the following Table 5, we will present the number of cases for each generation and for the natives, together with the number of cases that could not be classified due to missing information which could not be meaningfully replaced. The different generations are thereby clustered into different overarching groups depending on the countries of birth of the child and its parents.

Table 5: Descriptive results on generational status variable

	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
<i>Child foreign-born</i>										
Arrived at age 11+ (1.25 th gen)	199	4.6	103	2.1	36	0.8	222	4.4	560	3.0
Arrived at age 6-10 (1.5 th gen)	180	4.2	144	2.9	61	1.4	224	4.5	609	3.3
Arrived at age 0-5 (1.75 th gen)	183	4.2	266	5.3	170	3.9	179	3.6	798	4.3
No info on age upon arrival	47	1.1	22	0.4	28	0.6	28	0.6	125	0.7
<i>Parents foreign-born (2nd gen)</i>	543	12.6	1,232	24.6	671	15.4	1,017	20.2	3,463	18.5
<i>Parents foreign-born and native-born</i>										
One parent 2 nd gen (2.5 th gen)	257	6.0	179	3.6	62	1.4	112	2.2	610	3.3
One parent 2.5 th gen (2.75 th gen)	41	1.0	48	1.0	36	0.8	58	1.2	183	1.0
One parent native (Interethnic 2 nd gen)	225	5.2	336	6.7	293	6.7	371	7.4	1,225	6.6
<i>Parents native-born</i>										
All grandp. foreign-born (3 rd gen)	121	2.8	32	0.6	20	0.5	26	0.5	199	1.1
3 grandp. foreign-born (3.25 th gen)	32	0.8	14	0.3	10	0.2	18	0.4	74	0.4
2 grandp. foreign-born (3.5 th gen)	26	0.6	38	0.8	21	0.5	33	0.7	118	0.6
2 grandp. foreign-born (Interethnic 3 rd gen)	173	4.0	98	2.0	67	1.5	160	3.2	498	2.7
1 grandp. foreign-born (3.75 th gen)	203	4.7	310	6.2	329	7.5	373	7.4	1,215	6.5
No grandp. foreign-born (4 th + gen)	1,966	45.6	2,111	42.1	2,539	58.2	2,143	42.7	8,759	46.8
<i>Missing information, but immigrant background</i>										
Parents foreign-born; no info on child	8	0.2	10	0.2	4	0.1	1	0.0	23	0.1
Child native-born, no info on parents, grandp. foreign-born	10	0.2	2	0.0	2	0.1	5	0.1	19	0.1
Child native-born, at least one ancestor foreign-born	0	0.0	53	1.1	0	0.0	0	0.0	53	0.3
<i>Missing information, immigrant background unclear</i>										
Child native-born, no info on parents and grandp.	15	0.4	0	0.0	3	0.1	10	0.2	28	0.2
Child and parents native-born, no info on grandp.	84	2.0	15	0.3	11	0.3	45	0.9	155	0.8
No info on any actor	2	0.1	0	0.0	0	0.0	0	0.0	2	0.0
Total	4,315		5,013		4,363		5,025		18,716	

The Construction of the Country of Origin Variable

In contrast to the broad consensus in migration research on use of the straightforward approach when specifying the generational status of children of immigrants, the procedures when defining their ethnic origin are less consistent. Ethnic origin can in principle be defined by several indicators, where self-subscribed ethnic identity, nationality, language use and countries of birth of ancestors are the most relevant and most often used indicators (Dagevos 1997, Gresch and Kristen 2011, Jacobs et al. 2009, Smith 1984). Given the aim of the survey to study integration processes in children of immigrants, using information about the countries of birth of ancestors seems to be the most appropriate and stable way to define the ethnic origin of the children, especially as this is directly linked to the procedure for the generational status variable. Other indicators like self-subscribed ethnic identity and nationality may in contrast be sub-

ject to integration processes themselves, and are therefore a consequence rather than a predecessor of children's integration. For example, after several generations children with an immigrant background may solely identify with the receiving country, making it impossible to identify them as immigrants on the basis of the identity information.

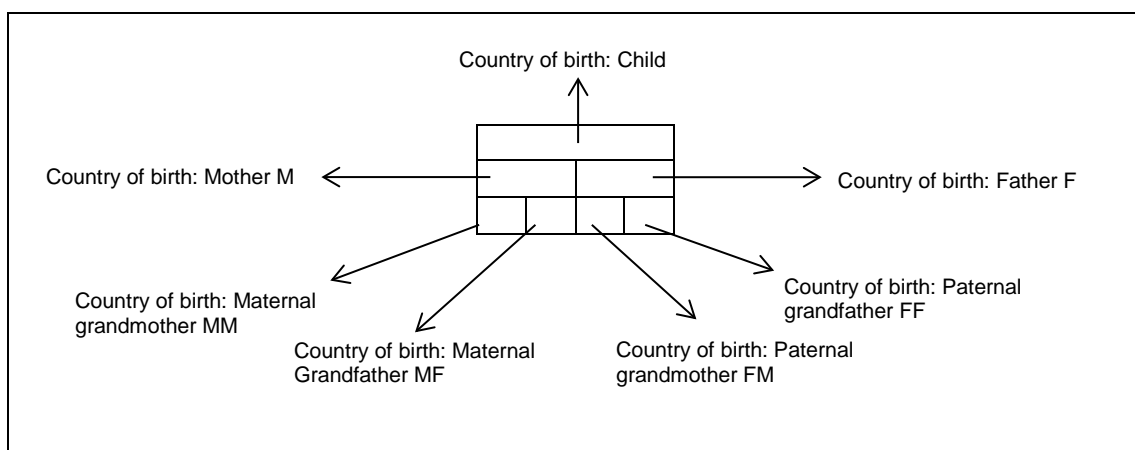
As outlined before, the current survey asks about the concrete countries of birth of the child, the parents and the grandparents. In order to define the ethnic origin through parentage, we will start at the grandparent level in order to define the heritage of the family before their migration to the respective survey country in CILS4EU. In contrast to the top-down approach for the generational status variable, we will thereby rely on a bottom-up approach, starting at the grandparent level in order to define the heritage of the respondent. This approach is applied to all possible immigrants, i.e. all persons that are defined as immigrants according to the classification rules in the previous section, plus the respondents with missing information for whom an immigrant background cannot be ruled out.

Standard Classification Approach

As for the generational status variable, we use information provided by the child about his or her country of birth, as well as information from the parents provided in the parent interview about the child's grandparents' countries of birth. As before, in those cases where the parent information is not available (due to item non-response, refusal of the parents to participate in the survey or when the interviewed parents are not the biological parents of the child), we use the information from the child's interview about his or her parents' countries of birth. Unfortunately, the information from the children about their grandparents' countries of birth cannot be used to construct the country of origin variable, as the children were only asked whether their grandparents were born in the survey country or not.

In order to construct the country of origin variable (`countorigG` and the national versions `countorig_enG`, `countorig_geG`, `countorig_nlg`, and `countorig_swG`), we use information about all seven actors (child, two parents and four grandparents). However, and in contrast to the generational status variable, the information for all actors is not dichotomous (born in survey country: yes/no). Instead, different countries of birth may be nominated for the various actors. This makes it necessary to define further rules about which information on which actor on which level should be considered most important when defining the country of origin variable. As some of these rules rely on a concrete identification of specific family members, the starting point is thereby to extend the country-of-birth ancestry-scheme from Figure 4 by specifying the different actors in more detail.

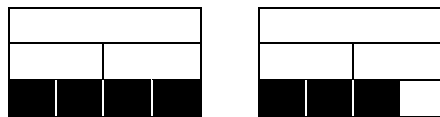
Figure 4: The extended country-of-birth ancestry-scheme



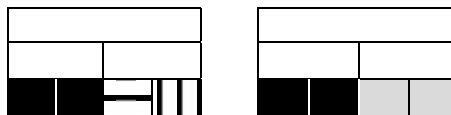
As can be seen from Figure 4, the squares representing the child's ancestors are now personalised, meaning that they represent a specific parent or a specific grandparent. How the information about the countries of birth of these actors is used to define the country of origin of the child will be illustrated in the following. According to the general consideration in the introduction, we will thereby start on the grandparent level.

Grandparent Level

At the grandparent level, we first define a simple *majority rule* in order to construct the country of origin variable, taking into account all foreign countries of birth. The majority rule is applied without considering which of the grandparents was born in which specific country. If four or three grandparents were born in one and the same foreign country (marked in black in the ancestry scheme below), this information is used to define the country of origin of the child. Like in the section about the generational status variable, information represented by white rectangles is irrelevant for the generation of the country of origin variable. So, if information about foreign countries of birth at the grandparent level is available, the child's as well as the parents' countries of birth are irrelevant for the country-of-origin variable. Furthermore, and due to the majority rule, the different country of birth of the single grandparent in the right example would be irrelevant, regardless of whether he or she was born in another foreign country or in the survey country.

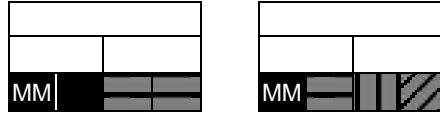


If two grandparents were born in the same foreign country, and the two other grandparents were born in two different countries (illustrated in the example on the left-hand side below by different patterned rectangles for the paternal grandmother and grandfather), the information from the first-mentioned grandparents is used in order to define the country of origin of the child. However, the majority rule only applies when defining a majority on different *foreign* countries of birth. If two grandparents were born in the same foreign country, and the two other grandparents were born in the survey country, then again the information from the first-mentioned grandparents is used in order to define the country of origin of the child, although there is no majority (cf. example on the right-hand side). In principle, we neglect information about grandparents born in the survey country in this and further examples, as this information does not contribute to the generation of the country of origin variable for respondents for whom we know that an immigrant background exists.

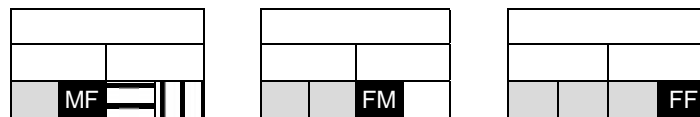


Whenever there is no majority on one foreign country of birth (e.g. two grandparents are foreign-born in one country (black), while two grandparents are foreign-born in another foreign country (grey with same pattern) as in the example below on the left-hand side; or all four grandparents were born in different foreign countries (grey with different patterns), as in the example below on the right-hand side, where the different countries are again represented by different patterned rectangles for the maternal grandfather and the paternal grandparents), we follow a *grandparent priority rule*. Following this rule, and in contrast to the majority rule outlined above, the concrete country of birth of a specific grandparent (maternal grandmother or grandfather, paternal grandmother or grandfather) is now relevant to the definition of the country of origin variable.

According to the grandparent priority rule, we firstly attach most importance to the country of birth of the maternal grandmother whenever there is no majority on one foreign country of birth. The priority given to female ancestries is to some extent arbitrary (Smith 1984). However, there are only minor differences in the definition of the country of origin variable once we change this maternal priority rule into a paternal priority rule.¹⁰



However, if the maternal grandmother was born in the survey country, and the other actors were born in three different countries (no majority rule applicable), then the country of birth of the maternal grandfather is used to define the country of origin of the child. If the maternal grandfather was also born in the survey country, the information about the country of birth of the paternal grandmother is used to define the child's country of origin, regardless of the country of birth of the paternal grandfather (white square). Finally and naturally, if the three last-named actors were all born in the survey country, the information about the country of birth of the paternal grandfather of the child is used to define the country of origin of the child.

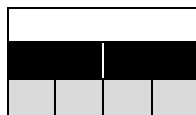


Whenever the information about the grandparents indicates that all of them were born in the survey country, we consider the information on the parent level to check whether these actors are foreign-born – and if so, in which country.

Parent Level

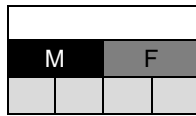
The considerations on the parent level are in line with the considerations described for the grandparent level. However, and as outlined above, these rules on the parent level only apply if the grandparents were all born in the survey country, as otherwise the rules on the grandparent level would have been applied.

If both parents were born in one and the same foreign country, this information is used to define the country of origin of the child.

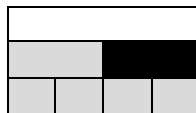


¹⁰ Using a father priority rule results in 683 differently assigned countries of origin (3.7% of the whole sample; 7.6% of all non-missing and non-survey countries of origin).

If both parents were born in different foreign countries, the country of birth of the mother is used to define the child's country of origin.

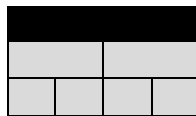


If one parent is foreign-born, while the other was born in the survey country, the country of origin of the child is defined by the country of birth of the foreign-born parent.



Child Level

Whenever information about the grandparents and the parents indicates that all of these actors were born in the survey country, information about the child's country of birth is used to define his or her country of origin variable.



Treatment of Missing Data

As already mentioned in the previous section, the questions asking about the countries of birth of the child, parents and grandparents in the parent and child questionnaires face the problem of missing data resulting from item or unit-nonresponse. Due to differently detailed information asked on the child and on the parent level, the problem of missing information is even more pronounced for the country of origin variable than for the generational status variable. Missing parent interviews lead to missing information on the countries of birth variables of all four grandparents, leaving only dichotomous variables with information about whether a grandparent was born in the survey country or not, which, though useful in generating the generational status variable, cannot be used to define the country of origin of the child.

Figure 5: The problem of missing data in the country-of-birth ancestry-scheme on different levels (n/%)

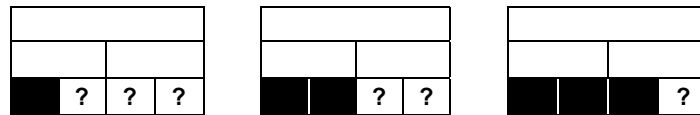
Child: 133/0.7			
Mother: 393/2.1		Father: 626/3.3	
Maternal grandmother: 4,582/24.5	Maternal grandfather: 4,811/25.7	Paternal grandmother: 5,971/31.9	Paternal grandfather: 6,206/33.2

Figure 5 shows the amount of missing and insufficient information for all seven actors. Compared to Figure 3 in the first part of this paper, the proportion of missing values on the grandparent level has intensified due to the fact that the concrete countries of birth of the grandparents' are unknown if the parent interview is missing. In total, between 25 and 33 per cent of the information on grandparents' countries of birth is missing, depending on the lineage and gender of the grandparents.

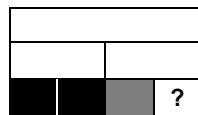
Given the bottom-up approach when defining the country of origin variable, a replacement of missing values according to specific rules as outlined in the first part of the paper, where a top-down approach was followed, is not appropriate. Furthermore, this procedure is also not necessary. In contrast to the generational status variable, where information about the country of birth (born in the survey country vs. foreign-born) of each of the seven actors is necessary in order to define the concrete generation of a respondent, the country of origin variable is designed such that in principle the information about one foreign country of birth of one of the actors is sufficient in order to define the country of origin of the respondent. Therefore, the rules from the standard classification approach described in the previous section are applied on the grandparent, parent, and child levels, taking into account also missing information on the countries of birth of different actors. In practice, missing information is treated as information indicating that the respective actor was born in the survey country, and will be used accordingly to define the country of origin of the child.

Missing Data on the Grandparent Level

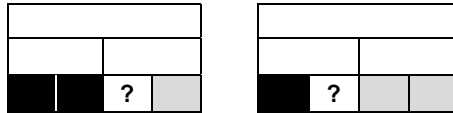
If all available information indicates that grandparents were born in the same country (black square), this information is used to define the country of origin of the child, regardless of the degree of missing values (white square with question marks).



If the existing information does not indicate a single foreign country of birth for the grandparents, we again follow the *majority rule*, taking into account possible missing data in order to construct the country of origin variable, considering all non-missing foreign countries of birth. One exemplary case is listed in the following.



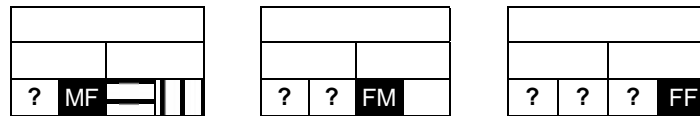
If information on some grandparents' countries of birth is missing, and existing information indicates that some grandparents were born in the survey country and some were foreign-born, then the information about the foreign countries of birth is used to define the country of origin of the child (cf. examples below) – even if there is a majority of information indicating that most of the grandparents were born in the survey country.



Whenever there is no majority on one foreign country of birth (e.g. two or three grandparents are foreign-born but in different foreign countries, while information on the countries of birth of the other grandparent(s) is missing, the *grandparent priority rule* is applied. According to this rule, we will, as before, firstly attach most importance to the country of birth of the maternal grandmother whenever there is no majority on one foreign country of birth.



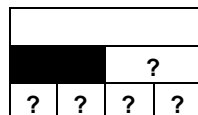
However, if information on the maternal grandmother is missing, and the other actors were born in three different countries (no majority rule applicable), then the country of birth of the maternal grandfather is used to define the country of origin of the child. If the maternal grandfather was also born in the survey country, the information about the country of birth of the paternal grandmother is used to define the child's country of origin. Finally, if only information about the paternal grandfather is available, this information is used to define the country of origin of the child.



Whenever the information on the grandparent level is completely missing, or the only existing information indicates that these grandparents were born in the survey country, we consider the information on the parent level to check whether these actors are foreign-born – and if so, in which country.

Missing Data on the Parent Level

If information on one parent is missing, the existing information on the other parent is used to define the country of origin variable of the child. If no information on the parent level indicates a foreign country of birth, the information on the child level is used.



Flag Variable for Missing Data

In order to gain some insight into the problem of the missing data when defining the country of origin variable, we again include a flag variable indicating missing values. In contrast to the flag variable introduced for the generational variable (Table 3), the flag variable for the country of origin has to account for missing as well as for insufficient information. Given the dichotomous variable about the countries of birth of the

grandparents provided by the child interviews, we know that the country of birth is not the survey country, but we do not know anything about the concrete countries of birth and therefore cannot use this information to construct the country of origin variable. Therefore, the flag variable [coflag_missG] indicates all cases with missing and/or insufficient information. As we use a bottom-up approach when defining the country of origin variable, where the highest priority is attached to the grandparent level and the least to the child level, the flag variable is in inversed order compared to the flag variable version for the generational status variable.

Table 6: Composition of the flag-variable indicating missing values

1 st digit	2 nd digit	3 rd digit
Information on grandparents missing	Information on parent(s) missing	Information on child missing
0 – no information missing	0 – no information missing	0 – no information missing
1 – information on one grandparent missing	1 – information on one parent missing	1 – information missing
2 – information on two grandparents missing	2 – information on both parents missing	
3 – information on three grandparents missing		
4 – information on four grandparents missing		

As can be seen from Table 6, the flag variable ranges from 421 to 0, where 421 means that all information is missing or insufficient to construct the country of origin variable, whereas 0 means that information about the countries of birth of all seven actors is available. Table 7 provides an overview of the distribution of the flag variable in the different countries. For 60% of the sample, information on all seven countries of birth is available. The vast majority of the remaining cases with missing information are due to missing information on at least one grandparent, while missing values for the parents and the child are negligible. Missing information about the grandparents' countries of birth is thereby in most cases due to missing parent interviews, where information about the concrete (foreign) country of birth of the grandparents was collected, while the child interviews were only helpful if the child indicated that the grandparent(s) was/were born in the survey country (and not in a non-specified foreign country).

Table 7: Distribution of the missing flag variable for the country of origin variable over the countries

Flag value	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
<i>No missing information</i>										
0	2,102	48.7	3,354	66.9	3,065	70.3	2,569	51.1	11,090	59.3
<i>Missing information on the child level</i>										
1	26	0.6	14	0.3	9	0.2	0	0.0	49	0.3
<i>Missing information on the parent level</i>										
10-21	16	0.4	14	0.3	6	0.1	30	0.6	66	0.4
<i>Missing information on the grandparent level</i>										
100-421	2,171	50.3	1,631	32.5	1,283	29.4	2,426	48.3	7,511	40.1
Total	4,315		5,013		4,363		5,025		18,716	

Non-Trivial Cases

Like for the generational status variable, we face some non-trivial cases when defining the country of origin variable. This holds true for those cases where the foreign-born actors were (at least partly) not born in one and the same country, but were born in different foreign countries. For instance, a child born in the survey country whose parents and grandparents were born in six different foreign countries is defined as having a Turkish country of origin, if the maternal grandmother is Turkish-born (application of the maternal priority rule on the grandparent level). However, there is obviously a difference between this hypothetical case defined as having a Turkish origin and a case in which all ancestors were born in Turkey, which would also be defined as having a Turkish origin due to the application of the majority rule on the grandparent level. To document the homogeneity or heterogeneity with respect to the foreign countries of birth of the different actors, we also constructed a flag variable.

Given the bottom-up strategy with the related classification rules, the child's country of origin is in the extreme case defined by the country of birth of one single grandparent (according to the maternal priority rule: the maternal grandmother), whereas all other six actors were born in six different foreign countries. Therefore, non-trivial cases with deviating foreign countries of birth in the family history can refer to the countries of birth of three other grandparents at maximum, the countries of birth of both parents and the child's country of birth. The flag variable [coflag_ntG] then simply indicates how many other foreign countries (compared to the country of the final country of origin variable) are nominated for which actors. Nominations of the survey country as an actor's country of birth are not seen as deviating cases and are not considered when constructing the flag variable. This scenario is already captured in the flag for the non-trivial cases referring to the generational status variable (see the previous section). The composition of the three-digit flag variable is summarized in table 8.

Table 8: Composition of the flag-variable indicating trivial and non-trivial cases

1 st digit	2 nd digit	3 rd digit
Information on grandparents inconsistent	Information on parent(s) inconsistent	Information on child inconsistent
0 – all information consistent 1 – information on one grandparent inconsistent 2 – information on two grandparents inconsistent 3 – information on three grandparents inconsistent	0 –all information consistent 1 –information on one parent inconsistent 2 –information on both parents inconsistent	0 –information consistent 1 –information inconsistent

As can be seen from Table 8, the flag ranges from 321 to 0, where 321 means that six actors were born in different foreign countries compared to the actor that defines the country of origin of the child (which is in this case the maternal grandmother, as no majority rule can be applied and therefore the priority rule is used). In contrast, 0 indicates that all seven actors were born in one and the same foreign country, or that some actors were born in a specific foreign country while others were born in the survey country or are missing. Table 9 provides an overview of the distribution of the flag variable in the different countries.

Table 9: Distribution of the trivial/non-trivial flag variable over the countries

Flag value	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
<i>Only trivial cases</i>										
0	4,012	93.0	4,607	91.9	4,170	95.6	4,503	89.6	17,292	92.4
<i>Non-trivial cases with deviating values on the child level</i>										
1	76	1.8	31	0.6	13	0.3	81	1.6	201	1.1
<i>Non-trivial cases with deviating values on the parent level</i>										
10-21	139	3.2	169	3.4	106	2.4	305	6.1	719	3.8
<i>Non-trivial cases with deviating values on the grandparent level</i>										
100-321	88	2.0	206	4.1	74	1.7	136	2.7	504	2.7
Total	4,315		5,013		4,363		5,025		18,716	

As can be seen in Table 9, more than 92 per cent of the country of birth information is trivial, meaning that no actor was born in a different country compared to the final country of origin of the child. Additionally, information on the parent level is most likely to be non-trivial.

Like the flag-variable for the missing values, the flag variable for non-trivial cases can easily be used to exclude specific cases from the analyses, e.g. analyses with all cases having a value on the flag variable smaller than 100 would mean that all cases are included where at least no information on the grandparent level is inconsistent, meaning that all grandparents were born in the same foreign country or some in the survey country or are missing.

The Country of Origin Revisited – Using Additional Information

Even when applying the standard classification approach described above and the strategy to handle missing data, uncertainties about the country of origin persist in some cases. In some cases, we know that the child or the family has some kind of immigrant background, but the concrete country of origin cannot be defined due to missing information on all family members, while in other cases we cannot answer the question whether there is any immigration history at all in the family, let alone the question concerning the child's country of origin. In order to at least partly overcome the problem of missing information about the countries of birth of different actors, we use additional information to define the child's country of origin. Following the bottom-up approach, we will therefore start with additional information about the parents before we include additional information about the child. Although the natural starting point would be the grandparent level given the bottom-up approach, the only information collected on this level is actually their countries of birth, so no additional information such as their nationality or self-subscribed identity can be used.

In principle and as already mentioned in the introduction to this section, different information can be used to define the country of origin of the child if information on the child's, the parents', and/or grandparents' countries of birth is missing. The survey assesses information about the children's and parents' nationality and self-subscribed identity as possible indicators for the country of origin. Furthermore, for the German subsample where we were not allowed to ask about third parties, information about the "immigrant background" for the child is available. In addition, the language spoken at home was also assessed during the survey. However, as the information on the language cannot necessarily be attributed to a distinctive coun-

try of origin (for example, English and Russian are officially spoken in several countries), we desisted from using this information to define the child's country of origin. This leaves us with three indicators for clarifying the country of origin for children whose countries of birth are not sufficient to define their country of origin: nationality, identity and immigrant background (for the German subsample).

As a first step, we estimated how successful the different indicators nationality, identity and immigrant background are on the child and parent levels when defining the country of origin. More precisely, we checked the probability of successful identifications of the country of origin with these different indicators for those cases where the country of origin could be defined through information on the countries of birth. The results of these consistency checks demonstrate that the (first and second) nationality of parents as well as of the children is the best proxy for the country of origin of the child. If the first or second nationality of the respondent is different from the survey country, it equals the child's country of origin in 85% of the cases or even more, depending on which information from which actor is used. The identity of the parents as well as that of the child is a somewhat weaker predictor of the child's country of origin, resulting in a match in almost two thirds of the cases. The question about the immigrant background of the German subsample ranged with almost 80% concordance between these two indicators.

As a result, we use information about any foreign nationality of the parents and the children, as well as information on a possible foreign identity of the parents and the children to replace missing information on the child's country of origin. Furthermore, information about the immigrant background of the German subsample is also used. We start at the parent level with the first and second nationality, followed by the identity information (if the information did not indicate the survey country's nationality or if there was a foreign identity at all). Following the priority rules, we start with the mother's information, followed by the father's information. Due to the logic behind the patterns of missing values, where missing information on the target person's country of origin is in many cases due to missing parent interviews (and therefore missing information on the concrete countries of birth of the grandparents), only very few replacements were possible through this strategy. In total, 18 replacements were made using the additional information on the parent level.

In contrast, using additional information about the child's nationality and identity is more successful, as this information was provided through the child questionnaire. Here, we start again with the information about the (first and second) nationality which is not the survey country's nationality, followed by the child's identity whenever he or she indicates that he or she also feels he/she belongs to a group other than the survey country group. Lastly, for the German subsample where we asked directly for the migration background, this information is used in addition. Here, in 436 data cases have been replaced using this additional information.

As outlined in the introduction of this section, using information like self-subscribed ethnic identity or nationality may bias results, as these indicators may be subject to integration processes themselves, for example, citizenship might be a consequence of individual integration efforts. Therefore, and in order to get an insight into the information used when constructing the country of origin variable with this additional information, another flag variable is constructed. This flag variable [*coflag_aig*] represents the different predictive power of the different additional indicators, starting with the most predictive indicator with a value of 1 (parents' nationality) up to the least predictive indicator with a value of 5 (child's identity). A value of 0 indicates that the country of origin variable was constructed without using any additional information and therefore solely by using the information about the countries of birth. In contrast, a value of 6 specifies that no information was available in order to construct the country of origin variable. Table 10 provides an overview of the different match qualities associated with the different additional information used represented in the digits of the flag variable.

Table 10: Flag variable indicating the use of additional information

digit	match quality
0 – information on countries of birth of any ancestor used	-
1 – information on (any) parents' nationality used (first or second nationality)	86.5%
2 – information on child's nationality used (first or second nationality)	85.2%
3 – information on child's migration background used	77.0%
4 – information on parents' identity used (if an additional, non-survey-country identity exists)	69.2%
5 – information on child's identity used (if an additional, non-survey-country identity exists)	65.7%
6 – No information on any of the information above available	-

Given this flag, it is easy to include cases with more or less predictive power of the additional variables used, e.g. by including only those cases with a value of 0 to 2, where the country of origin was constructed using information about the countries of birth or the nationality of the child and his or her parents, which predicts the countries of birth of these actors in at least 85% of the cases in our sample. Table 11 shows the distribution of this flag variable.

Table 11: Distribution of the additional information flag variable over the countries

Flag value	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
0	3,711	86.0	4,772	95.2	4,206	96.4	4,652	92.6	17,341	92.7
1	0	0.0	1	0.0	1	0.0	2	0.0	4	0.0
2	72	1.7	50	1.0	6	0.1	3	0.1	131	0.7
3	0	0.0	13	0.3	0	0.0	0	0.0	13	0.1
4	3	0.1	5	0.1	1	0.0	5	0.1	14	0.1
5	103	2.4	47	0.9	41	0.9	96	1.9	287	1.5
6	426	9.9	125	2.5	108	2.5	267	5.3	926	5.0
Total	4,315		5,013		4,363		5,025		18,716	

As can be seen and is mentioned beforehand, the country of origin in the vast majority of cases is classified through the children's, parents' and grandparents' countries of birth. Given non-response on the parent level, main sources of additional information used to classify the country of origin were the children's interviews, where especially the self-subscribed identity was useful to define the country of origin variable. In contrast, additional parent information played only a minor role in the definition of a respondent's country of origin. However, in the vast majority of cases that could not be defined using the children's, parents', and grandparents' countries of birth, no additional information was available, making it impossible to define the concrete country of origin of the respondent.

Descriptive Results

In the following we will provide an overview of the country of origin variable of the CILS4EU sample. Table 12 lists all cases where a classification according to the standard classification approach or through the use of additional information is possible, together with the respective countries of origins. Furthermore, those cases are also listed for which even after the standard classification approach or after the use of additional information the country of origin is still unknown, but where it is clear that an immigrant background exists. Finally, the cases where it is unclear whether an immigrant background exists at all are also listed

in this table. As can be seen from this table, CILS4EU captures a wide variety of ethnic groups in each country and demonstrates the ethnic diversity in different European countries.

In addition to the extensive country of origin variable encompassing all possible countries of origin, we provide four reduced national classifications of the country of origin variable. These variables entail the largest countries of origin in the respective survey countries. Smaller immigrant groups are aggregated into categories according to broad geographical regions of origin due to privacy reasons. The frequencies of these classifications are shown in Table 13, Table 14, Table 15 and Table 16. As can be seen, not only within-country comparisons between different ethnic groups are possible, the data also provide the opportunity for between-country comparisons of identical ethnic groups, for example, of Turkish immigrants in Germany, the Netherlands and Sweden.

Summary

This working paper presents a comprehensive procedure for identifying children of immigrants according to their generational status and country of origin. Besides discussing and providing solutions for two central problems (missing values and non-standard migration histories), we demonstrate how these central variables are constructed in the Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU). In terms of children's generational status, missing values were surprisingly not a major problem, since missing information on any of the relevant actors never exceeds ten per cent, and complete information was gathered for 88 per cent of the whole sample. Additionally, we were able to show that our standard classification approach was perfectly appropriate for almost every child in CILS4EU, and non-trivial cases due to exceptional migration histories are a rare exception (less than five per cent). For the definition of the country of origin of the child, we face more serious missing-data problems, which is natural given the dichotomous question in the child questionnaire and consequently the need for parent interviews for information on the grandparent level. However, we successfully employed additional information from the questionnaires, like nationality and subjective identity, to at least partly clarify the ethnic origin of the child. An analysis of cases with complete information demonstrates that this approach is well suited, too. All in all, we hope to stimulate other researchers in the field of migration to implement similar measures in surveys on the integration of immigrant children and to apply similar heuristics in order to make results from different data sources more comparable in the future.

Table 12: Distribution of the country of origin

Country of origin	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
Africa	19	0.4	6	0.1	4	0.1	4	0.1	33	0.2
Afghanistan	16	0.4	28	0.6	30	0.7	31	0.6	105	0.6
South America	0	0.0	1	0.0	1	0.0	0	0.0	2	0.0
Albania	2	0.1	14	0.3	0	0.0	5	0.1	21	0.1
Algeria	6	0.1	9	0.2	4	0.1	4	0.1	23	0.1
Americas	19	0.4	13	0.3	0	0.0	0	0.0	32	0.2
Angola	8	0.2	6	0.1	4	0.1	0	0.0	18	0.1
Antigua and Barbuda	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
Caribbean	4	0.1	0	0.0	0	0.0	0	0.0	4	0.0
Azerbaijan	1	0.0	2	0.0	3	0.1	4	0.1	10	0.1
Argentina	2	0.1	5	0.1	1	0.0	0	0.0	8	0.0
Australia	5	0.1	3	0.1	6	0.1	4	0.1	18	0.1
Austria	0	0.0	37	0.7	7	0.2	10	0.2	54	0.3
Bahrain	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Bangladesh	52	1.2	2	0.0	0	0.0	18	0.4	72	0.4
Armenia	1	0.0	1	0.0	4	0.1	5	0.1	11	0.1

Table 12: Distribution of the country of origin – continued

Country of origin	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
Barbados	5	0.1	0	0.0	0	0.0	1	0.0	6	0.0
Belgium	4	0.1	9	0.2	29	0.7	1	0.0	43	0.2
Bermuda	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Plurinational State of Bolivia	0	0.0	2	0.0	0	0.0	7	0.1	9	0.1
Bosnia and Herzegovina	0	0.0	37	0.7	5	0.1	127	2.5	169	0.9
Brazil	7	0.2	19	0.4	1	0.0	6	0.1	33	0.2
Bulgaria	3	0.1	3	0.1	2	0.1	13	0.3	21	0.1
Myanmar	3	0.1	0	0.0	0	0.0	3	0.1	6	0.0
Burundi	0	0.0	0	0.0	4	0.1	4	0.1	8	0.0
Belarus	0	0.0	2	0.0	0	0.0	1	0.0	3	0.0
Cambodia	0	0.0	1	0.0	0	0.0	2	0.0	3	0.0
Cameroon	2	0.1	3	0.1	0	0.0	0	0.0	5	0.0
Canada	11	0.3	2	0.0	8	0.2	2	0.0	23	0.1
Cape Verde	0	0.0	0	0.0	23	0.5	0	0.0	23	0.1
Asia	0	0.0	1	0.0	0	0.0	1	0.0	2	0.0
Sri Lanka	46	1.1	16	0.3	6	0.1	8	0.2	76	0.4
Chile	0	0.0	3	0.1	3	0.1	40	0.8	46	0.3
China	46	1.1	12	0.2	24	0.6	26	0.5	108	0.6
Colombia	5	0.1	3	0.1	6	0.1	13	0.3	27	0.1
Congo	6	0.1	8	0.2	3	0.1	13	0.3	30	0.2
Croatia	1	0.0	29	0.6	5	0.1	21	0.4	56	0.3
Cuba	1	0.0	2	0.0	2	0.1	6	0.1	11	0.1
Cyprus	13	0.3	1	0.0	2	0.1	2	0.0	18	0.1
Czechoslovakia	0	0.0	8	0.2	0	0.0	1	0.0	9	0.1
Czech Republic	1	0.0	36	0.7	0	0.0	3	0.1	40	0.2
Benin	0	0.0	1	0.0	0	0.0	1	0.0	2	0.0
Denmark	4	0.1	4	0.1	2	0.1	74	1.5	84	0.5
Dominica	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Dominican Republic	1	0.0	6	0.1	7	0.2	1	0.0	15	0.1
Ecuador	2	0.1	3	0.1	1	0.0	4	0.1	10	0.1
Ethiopia	0	0.0	2	0.0	8	0.2	21	0.4	31	0.2
Eritrea	2	0.1	10	0.2	5	0.1	22	0.4	39	0.2
Estonia	1	0.0	1	0.0	1	0.0	26	0.5	29	0.2
Fiji	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Finland	0	0.0	1	0.0	3	0.1	310	6.2	314	1.7
Aland Islands	0	0.0	0	0.0	0	0.0	1	0.0	1	0.0
France	19	0.4	19	0.4	13	0.3	9	0.2	60	0.3
Djibouti	1	0.0	1	0.0	0	0.0	3	0.1	5	0.0
Georgia	0	0.0	3	0.1	0	0.0	3	0.1	6	0.0
Gambia	4	0.1	2	0.0	0	0.0	12	0.2	18	0.1
Occupied Palestinian Territories	1	0.0	18	0.4	0	0.0	27	0.5	46	0.3
Germany	44	1.0	2,111	42.1	84	1.9	87	1.7	2,326	12.4
Ghana	48	1.1	10	0.2	9	0.2	6	0.1	73	0.4
Gibraltar	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Greece	5	0.1	52	1.0	6	0.1	17	0.3	80	0.4
Grenada	7	0.2	0	0.0	0	0.0	0	0.0	7	0.0
Guatemala	0	0.0	0	0.0	0	0.0	1	0.0	1	0.0
Guinea	1	0.0	1	0.0	0	0.0	0	0.0	2	0.0
Guyana	5	0.1	0	0.0	6	0.1	0	0.0	11	0.1
Honduras	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
China, Hong Kong Special Administrative Region	3	0.1	0	0.0	1	0.0	1	0.0	5	0.0
Hungary	0	0.0	22	0.4	3	0.1	24	0.5	49	0.3
Iceland	1	0.0	0	0.0	0	0.0	2	0.0	3	0.0
India	281	6.5	9	0.2	14	0.3	18	0.4	322	1.7
Indonesia	3	0.1	2	0.0	163	3.7	2	0.0	170	0.9
Islamic Republic of Iran	5	0.1	22	0.4	10	0.2	79	1.6	116	0.6
Iraq	7	0.2	44	0.9	34	0.8	228	4.5	313	1.7
Ireland	121	2.8	1	0.0	3	0.1	1	0.0	126	0.7

Table 12: Distribution of the country of origin – continued

Country of origin	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
Israel	0	0.0	0	0.0	2	0.1	2	0.0	4	0.0
Italy	45	1.0	164	3.3	19	0.4	28	0.6	256	1.4
Cote d'Ivoire	2	0.1	1	0.0	1	0.0	3	0.1	7	0.0
Jamaica	147	3.4	0	0.0	0	0.0	1	0.0	148	0.8
Japan	6	0.1	0	0.0	0	0.0	5	0.1	11	0.1
Kazakhstan	0	0.0	64	1.3	1	0.0	0	0.0	65	0.4
Jordan	0	0.0	2	0.0	2	0.1	2	0.0	6	0.0
Kenya	41	1.0	4	0.1	0	0.0	2	0.0	47	0.3
Republic of Korea	7	0.2	3	0.1	1	0.0	15	0.3	26	0.1
Kuwait	3	0.1	0	0.0	1	0.0	5	0.1	9	0.1
Kyrgyzstan	0	0.0	1	0.0	0	0.0	0	0.0	1	0.0
Lao People's Democratic Republic	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Lebanon	2	0.1	59	1.2	5	0.1	95	1.9	161	0.9
Latvia	5	0.1	3	0.1	0	0.0	10	0.2	18	0.1
Liberia	0	0.0	2	0.0	2	0.1	1	0.0	5	0.0
Libya	5	0.1	2	0.0	0	0.0	2	0.0	9	0.1
Lithuania	12	0.3	1	0.0	1	0.0	4	0.1	18	0.1
Luxembourg	0	0.0	2	0.0	0	0.0	1	0.0	3	0.0
Malawi	3	0.1	0	0.0	0	0.0	0	0.0	3	0.0
Malaysia	8	0.2	0	0.0	1	0.0	5	0.1	14	0.1
Mali	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
Malta	5	0.1	0	0.0	0	0.0	0	0.0	5	0.0
Mauritius	14	0.3	1	0.0	2	0.1	0	0.0	17	0.1
Mexico	1	0.0	4	0.1	0	0.0	3	0.1	8	0.0
Mongolia	0	0.0	0	0.0	0	0.0	2	0.0	2	0.0
Republic of Moldova	1	0.0	2	0.0	0	0.0	0	0.0	3	0.0
Montenegro	2	0.1	4	0.1	0	0.0	10	0.2	16	0.1
Morocco	5	0.1	37	0.7	274	6.3	22	0.4	338	1.8
Mozambique	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Nepal	10	0.2	0	0.0	0	0.0	0	0.0	10	0.1
Netherlands	6	0.1	22	0.4	2,539	58.2	12	0.2	2,579	13.8
Netherlands Antilles	1	0.0	0	0.0	38	0.9	0	0.0	39	0.2
Curacao	0	0.0	0	0.0	10	0.2	0	0.0	10	0.1
Aruba	0	0.0	0	0.0	71	1.6	0	0.0	71	0.4
Bonaire, Saint Eustatius and Saba	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
New Zealand	10	0.2	0	0.0	2	0.1	2	0.0	14	0.1
Nicaragua	0	0.0	1	0.0	0	0.0	0	0.0	1	0.0
Nigeria	71	1.7	5	0.1	4	0.1	5	0.1	85	0.5
Norway	1	0.0	1	0.0	1	0.0	63	1.3	66	0.4
Pakistan	327	7.6	14	0.3	9	0.2	5	0.1	355	1.9
Panama	0	0.0	0	0.0	0	0.0	1	0.0	1	0.0
Papua New Guinea	0	0.0	0	0.0	2	0.1	0	0.0	2	0.0
Paraguay	0	0.0	1	0.0	0	0.0	1	0.0	2	0.0
Peru	1	0.0	1	0.0	2	0.1	11	0.2	15	0.1
Philippines	15	0.4	7	0.1	12	0.3	17	0.3	51	0.3
Poland	49	1.1	262	5.2	13	0.3	75	1.5	399	2.1
Portugal	19	0.4	30	0.6	4	0.1	8	0.2	61	0.3
Romania	5	0.1	38	0.8	5	0.1	18	0.4	66	0.4
Russian Federation	10	0.2	201	4.0	4	0.1	15	0.3	230	1.2
Rwanda	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
Saint Kitts and Nevis	6	0.1	0	0.0	0	0.0	0	0.0	6	0.0
Saint Martin (French part)	0	0.0	0	0.0	1	0.0	0	0.0	1	0.0
Saint Vincent and the Grenadines	5	0.1	0	0.0	0	0.0	0	0.0	5	0.0
Saudi Arabia	1	0.0	1	0.0	0	0.0	2	0.0	4	0.0
Senegal	0	0.0	2	0.0	1	0.0	5	0.1	8	0.0
Serbia	6	0.1	117	2.3	7	0.2	54	1.1	184	1.0
Seychelles	0	0.0	0	0.0	0	0.0	2	0.0	2	0.0
Sierra Leone	6	0.1	0	0.0	1	0.0	0	0.0	7	0.0
Singapore	6	0.1	0	0.0	1	0.0	0	0.0	7	0.0

Table 12: Distribution of the country of origin – continued

Country of origin	England		Germany		Netherlands		Sweden		Total	
	N	%	N	%	N	%	N	%	N	%
Slovakia	3	0.1	4	0.1	2	0.1	2	0.0	11	0.1
Viet Nam	6	0.1	14	0.3	7	0.2	39	0.8	66	0.4
Slovenia	0	0.0	7	0.1	1	0.0	5	0.1	13	0.1
Somalia	39	0.9	5	0.1	12	0.3	114	2.3	170	0.9
South Africa	25	0.6	4	0.1	4	0.1	4	0.1	37	0.2
Zimbabwe	24	0.6	1	0.0	0	0.0	1	0.0	26	0.1
Spain	15	0.4	33	0.7	19	0.4	10	0.2	77	0.4
Sudan	2	0.1	0	0.0	1	0.0	3	0.1	6	0.0
Suriname	0	0.0	0	0.0	229	5.3	0	0.0	229	1.2
Sweden	2	0.1	1	0.0	0	0.0	2,143	42.7	2,146	11.5
Switzerland	3	0.1	10	0.2	8	0.2	2	0.0	23	0.1
Syrian Arab Republic	2	0.1	15	0.3	5	0.1	89	1.8	111	0.6
Thailand	4	0.1	15	0.3	9	0.2	42	0.8	70	0.4
Togo	1	0.0	2	0.0	0	0.0	3	0.1	6	0.0
Tonga	2	0.1	0	0.0	0	0.0	0	0.0	2	0.0
Trinidad and Tobago	7	0.2	0	0.0	1	0.0	0	0.0	8	0.0
United Arab Emirates	2	0.1	1	0.0	0	0.0	0	0.0	3	0.0
Tunisia	3	0.1	10	0.2	9	0.2	12	0.2	34	0.2
Turkey	12	0.3	890	17.8	277	6.4	136	2.7	1,315	7.0
Turkmenistan	0	0.0	0	0.0	0	0.0	1	0.0	1	0.0
Turks and Caicos Islands	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Uganda	9	0.2	2	0.0	0	0.0	4	0.1	15	0.1
Ukraine	3	0.1	31	0.6	1	0.0	1	0.0	36	0.2
The Former Yugoslav Republic of Macedonia	1	0.0	27	0.5	5	0.1	38	0.8	71	0.4
USSR	0	0.0	3	0.1	0	0.0	0	0.0	3	0.0
Egypt	5	0.1	3	0.1	15	0.3	6	0.1	29	0.2
United Kingdom of Great Britain and Northern Ireland	1,966	45.6	17	0.3	26	0.6	26	0.5	2,035	10.9
Channel Islands	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Guernsey	1	0.0	0	0.0	0	0.0	0	0.0	1	0.0
United Republic of Tanzania	5	0.1	0	0.0	1	0.0	5	0.1	11	0.1
United States of America	0	0.0	19	0.4	12	0.3	30	0.6	61	0.3
Uruguay	0	0.0	0	0.0	0	0.0	6	0.1	6	0.0
Uzbekistan	0	0.0	0	0.0	0	0.0	4	0.1	4	0.0
Bolivarian Republic of Venezuela	1	0.0	2	0.0	1	0.0	0	0.0	4	0.0
Yemen	2	0.1	0	0.0	0	0.0	0	0.0	2	0.0
Socialist Federal Republic of Yugoslavia	1	0.0	16	0.3	6	0.1	11	0.2	34	0.2
Serbia and Montenegro	0	0.0	2	0.0	0	0.0	0	0.0	2	0.0
Zambia	6	0.1	0	0.0	0	0.0	0	0.0	6	0.0
Arabian Country	1	0.0	5	0.1	0	0.0	2	0.0	8	0.0
Former German Eastern Territories	0	0.0	15	0.3	0	0.0	0	0.0	15	0.1
Kosovo-Albania	0	0.0	0	0.0	0	0.0	98	2.0	98	0.5
Kurdistan	1	0.0	6	0.1	0	0.0	39	0.8	46	0.3
Kashmir	2	0.1	0	0.0	0	0.0	0	0.0	2	0.0
Sinti and Roma	0	0.0	0	0.0	0	0.0	2	0.0	2	0.0
<i>Immigrant background exists, but unclear which country of origin</i>	340	7.9	113	2.3	97	2.2	222	4.4	772	4.1
<i>Unclear whether immigrant background exists</i>	86	2.0	12	0.2	11	0.3	45	0.9	154	0.8
Total	4,315		5,013		4,363		5,025		18,716	

Table 13: Country of origin – national classification (England)

	N	England %	% Immig.
<i>Country of origin – largest groups</i>			
United Kingdom of Great Britain and Northern Ireland	1,966	45.6	
Pakistan	327	7.6	17.0
India	281	6.5	14.6
Jamaica	147	3.4	7.6
Ireland	121	2.8	6.3
Nigeria	71	1.7	3.7
Bangladesh	52	1.2	2.7
<i>Country of origin – aggregated</i>			
Eastern Africa ¹	145	3.4	7.5
Western Africa ²	62	1.4	3.2
Other Africa ³	87	2.0	4.5
Latin America and the Caribbean ⁴	63	1.5	3.3
Northern America and Oceania ⁵	49	1.1	2.6
Southern Asia ⁶	79	1.8	4.1
Eastern Asia ⁷	62	1.4	3.2
Other Asia ⁸	95	2.2	4.9
Eastern Europe ⁹	75	1.7	3.9
Southern Europe ¹⁰	103	2.4	5.4
Other Europe ¹¹	104	2.4	5.4
<i>Immigrant background exists, but unclear which country of origin</i>	340	7.9	
<i>Unclear whether immigrant background exists</i>	86	2.0	
Total	4,315		
1	Eritrea 2, Djibouti 1, Kenya 41, Malawi 3, Mauritius 14, Mozambique 1, Somalia 39, Zimbabwe 24, Uganda 9, United Republic of Tanzania 5, Zambia 6		
2	Gambia 4, Ghana 48, Guinea 1, Cote d'Ivoire 2, Sierra Leone 6, Togo 1		
3	Africa 19, Algeria 6, Angola 8, Cameroon 2, Congo 6, Libya 5, Morocco 5, South Africa 25, Sudan 2, Tunisia 3, Egypt 5, Arabian Country 1		
4	Caribbean 4, Argentina 2, Barbados 5, Brazil 7, Colombia 5, Cuba 1, Dominica 1, Dominican Republic 1, Ecuador 2, Grenada 7, Guyana 5, Mexico 1, Netherlands Antilles 1, Peru 1, Saint Kitts and Nevis 6, Saint Vincent and the Grenadines 5, Trinidad and Tobago 7, Turks and Caicos Islands 1, Bolivarian Republic of Venezuela 1		
5	Americas 19, Australia 5, Bermuda 1, Canada 11, Fiji 1, New Zealand 10, Tonga 2		
6	Afghanistan 16, Sri Lanka 46, Islamic Republic of Iran 5, Nepal 10, Kashmir 2		
7	China 46, China, Hong Kong Special Administrative Region 3, Japan 6, Republic of Korea 7		
8	Azerbaijan 1, Bahrain 1, Armenia 1, Myanmar 3, Cyprus 13, Occupied Palestinian Territory 1, Indonesia 3, Iraq 7, Kuwait 3, Lao People's Democratic Republic 1, Lebanon 2, Malaysia 8, Philippines 15, Saudi Arabia 1, Singapore 6, Viet Nam 6, Syrian Arab Republic 2, Thailand 4, United Arab Emirates 2, Turkey 12, Yemen 2, Kurdistan 1		
9	Bulgaria 3, Czech Republic 1, Republic of Moldova 1, Poland 49, Romania 5, Russian Federation 10, Slovakia 3, Ukraine 3		
10	Albania 2, Croatia 1, Gibraltar 1, Greece 5, Italy 45, Malta 5, Montenegro 2, Portugal 19, Serbia 6, Spain 15, The Former Yugoslav Republic of Macedonia 1, Socialist Federal Republic of Yugoslavia 1		
11	Belgium 4, Denmark 4, Estonia 1, France 19, Germany 44, Iceland 1, Latvia 5, Lithuania 12, Netherlands 6, Norway 1, Sweden 2, Switzerland 3, Channel Islands 1, Guernsey 1		

Table 14: Country of origin – national classification (Germany)

	N	Germany %	% Immig.
<i>Country of origin – largest groups</i>			
Germany	2,111	42.1	
Turkey (incl. 6 cases from Kurdistan)	896	17.9	32.3
Former Soviet Union ¹	310	6.2	11.2
Poland	262	5.2	9.4
Former Yugoslavia ²	239	4.8	8.6
Italy	164	3.3	5.9
Lebanon	59	1.2	2.1
Greece	52	1.0	1.9
<i>Country of origin – aggregated</i>			
Northern Africa ³	66	1.3	2.4
Other Africa ⁴	79	1.6	2.8
Latin America and the Caribbean ⁵	53	1.1	1.9
Northern America and Oceania ⁶	37	0.7	1.3
Southern Asia ⁷	91	1.8	3.3
Western Asia ⁸	82	1.6	3.0
Other Asia ⁹	55	1.1	2.0
Eastern Europe ¹⁰	126	2.5	4.5
Southern Europe ¹¹	77	1.5	2.8
Other Europe ¹²	129	2.6	4.7
<i>Immigrant background exists, but unclear which country of origin</i>	113	2.3	
<i>Unclear whether immigrant background exists</i>	12	0.2	
Total	5,013		
1	Azerbaijan 2, Armenia 1, Belarus 2, Georgia 3, Kazakhstan 64, Kyrgyzstan 1, Republic of Moldova 2, Russian Federation 201, Ukraine 31, USSR 3		
2	Bosnia and Herzegovina 37, Croatia 29, Montenegro 4, Serbia 117, Slovenia 7, The Former Yugoslav Republic of Macedonia 27, Socialist Federal Republic of Yugoslavia 16, Serbia and Montenegro 2		
3	Algeria 9, Libya 2, Morocco 37, Tunisia 10, Egypt 3, Arabian Country 5		
4	Africa 6, Angola 6, Cameroon 3, Congo 8, Benin 1, Ethiopia 2, Eritrea 10, Djibouti 1, Gambia 2, Ghana 10, Guinea 1, Cote d'Ivoire 1, Kenya 4, Liberia 2, Mauritius 1, Nigeria 5, Senegal 2, Somalia 5, South Africa 4, Zimbabwe 1, Togo 2, Uganda 2		
5	South America 1, Argentina 5, Plurinational State of Bolivia 2, Brazil 19, Chile 3, Colombia 3, Cuba 2, Dominican Republic 6, Ecuador 3, Mexico 4, Nicaragua 1, Paraguay 1, Peru 1, Bolivarian Republic of Venezuela 2		
6	Americas 13, Australia 3, Canada 2, United States of America 19		
7	Afghanistan 28, Bangladesh 2, Sri Lanka 16, India 9, Islamic Republic of Iran 22, Pakistan 14		
8	Cyprus 1, Occupied Palestinian Territory 18, Iraq 44, Jordan 2, Saudi Arabia 1, Syrian Arab Republic 15, United Arab Emirates 1		
9	Cambodia 1, Asia 1, China 12, Indonesia 2, Republic of Korea 3, Philippines 7, Viet Nam 14, Thailand 15		
10	Bulgaria 3, Czechoslovakia 8, Czech Republic 36, Hungary 22, Romania 38, Slovakia 4, Former German Eastern Territories 15		
11	Albania 14, Portugal 30, Spain 33		
12	Austria 37, Belgium 9, Denmark 4, Estonia 1, Finland 1, France 19, Ireland 1, Latvia 3, Lithuania 1, Luxembourg 2, Netherlands 22, Norway 1, Sweden 1, Switzerland 10, United Kingdom of Great Britain and Northern Ireland 17		

Table 15: Country of origin – national classification (Netherlands)

	Netherlands		
	N	%	% Immig.
<i>Country of origin – largest groups</i>			
Netherlands	2,539	58.2	
Turkey	277	6.4	16.1
Morocco	274	6.3	16.0
Suriname	229	5.3	13.3
Indonesia	163	3.7	9.5
Netherlands Antilles ¹	120	2.8	7.0
Germany	84	1.9	4.9
<i>Country of origin – aggregated</i>			
Africa ²	119	2.7	6.9
Latin America and the Caribbean ³	35	0.8	2.0
Northern America and Oceania ⁴	30	0.7	1.8
Southern Asia ⁵	69	1.6	4.0
Western Asia ⁶	58	1.3	3.4
Other Asia ⁷	57	1.3	3.3
Southern Europe ⁸	77	1.8	4.5
Western Europe ⁹	57	1.3	3.3
Other Europe ¹⁰	67	1.5	3.9
<i>Immigrant background exists, but unclear which country of origin</i>			
	97	2.2	
<i>Unclear whether immigrant background exists</i>			
	11	0.3	
Total	4,363		
1	Netherlands Antilles 38, Curacao 10, Aruba 71, Bonaire, Saint Eustatius and Saba 1		
2	Africa 4, Algeria 4, Angola 4, Burundi 4, Cape Verde 23, Congo 3, Ethiopia 8, Eritrea 5, Ghana 9, Cote d'Ivoire 1, Liberia 2, Mali 1, Mauritius 2, Nigeria 4, Rwanda 1, Senegal 1, Sierra Leone 1, Somalia 12, South Africa 4, Sudan 1, Tunisia 9, Egypt 15, United Republic of Tanzania 1		
3	South America 1, Antigua and Barbuda 1, Argentina 1, Brazil 1, Chile 3, Colombia 6, Cuba 2, Dominican Republic 7, Ecuador 1, Guyana 6, Honduras 1, Peru 2, Saint Martin (French part) 1, Trinidad and Tobago 1, Bolivarian Republic of Venezuela 1		
4	Australia 6, Canada 8, New Zealand 2, Papua New Guinea 2, United States of America 12		
5	Afghanistan 30, Sri Lanka 6, India 14, Islamic Republic of Iran 10, Pakistan 9		
6	Azerbaijan 3, Armenia 4, Cyprus 2, Iraq 34, Israel 2, Jordan 2, Kuwait 1, Lebanon 5, Syrian Arab Republic 5		
7	China 24, China, Hong Kong Special Administrative Region 1, Kazakhstan 1, Republic of Korea 1, Malaysia 1, Philippines 12, Singapore 1, Viet Nam 7, Thailand 9		
8	Bosnia and Herzegovina 5, Croatia 5, Greece 6, Italy 19, Portugal 4, Serbia 7, Slovenia 1, Spain 19, The Former Yugoslav Republic of Macedonia 5, Socialist Federal Republic of Yugoslavia 6		
9	Austria 7, Belgium 29, France 13, Switzerland 8		
10	Bulgaria 2, Denmark 2, Estonia 1, Finland 3, Hungary 3, Ireland 3, Lithuania 1, Norway 1, Poland 13, Romania 5, Russian Federation 4, Slovakia 2, Ukraine 1, United Kingdom of Great Britain and Northern Ireland 26		

Table 16: Country of origin – national classification (Sweden)

	Sweden		
	N	%	% Immig.
<i>Country of origin – largest groups</i>			
Sweden	2,143	42.7	
Former Yugoslavia ¹	364	7.2	13.9
Finland	310	6.2	11.9
Iraq	228	4.5	8.7
Turkey (incl. 39 cases from Kurdistan)	175	3.5	6.7
Somalia	114	2.3	4.4
Lebanon	95	1.9	3.6
Syrian Arab Republic	89	1.8	3.4
Germany	87	1.7	3.3
Islamic Republic of Iran	79	1.6	3.0
Poland	75	1.5	2.9
Denmark	74	1.5	2.8
Norway	63	1.3	2.4
<i>Country of origin – aggregated</i>			
Eastern Africa ²	64	1.3	2.5
Northern Africa ³	51	1.0	2.0
Other Africa ⁴	57	1.1	2.3
Latin America and the Caribbean ⁵	102	2.0	3.9
Northern America and Oceania ⁶	38	0.8	1.5
Southern Asia ⁷	80	1.6	2.9
Western Asia ⁸	52	1.0	2.0
South Eastern Asia ⁹	110	2.2	4.2
Other Asia ¹⁰	55	1.1	2.1
Eastern Europe ¹¹	78	1.6	3.0
Southern Europe ¹²	70	1.4	2.7
Other Europe ¹³	105	2.1	4.0
<i>Immigrant background exists, but unclear which country of origin</i>			
	222	4.4	
<i>Unclear whether immigrant background exists</i>			
	45	0.9	
Total	5,025		

- 1 Bosnia and Herzegovina 127, Croatia 21, Montenegro 10, Serbia 54, Slovenia 5, The Former Yugoslav Republic of Macedonia 38, Socialist Federal Republic of Yugoslavia 11, Kosovo-Albania 98
- 2 Burundi 4, Ethiopia 21, Eritrea 22, Djibouti 3, Kenya 2, Seychelles 2, Zimbabwe 1, Uganda 4, United Republic of Tanzania 5
- 3 Algeria 4, Libya 2, Morocco 22, Sudan 3, Tunisia 12, Egypt 6, Arabian Country 2
- 4 Africa 4, Congo 13, Benin 1, Gambia 12, Ghana 6, Cote d'Ivoire 3, Liberia 1, Nigeria 5, Senegal 5, South Africa 4, Togo 3
- 5 Barbados 1, Plurinational State of Bolivia 7, Brazil 6, Chile 40, Colombia 13, Cuba 6, Dominican Republic 1, Ecuador 4, Guatemala 1, Jamaica 1, Mexico 3, Panama 1, Paraguay 1, Peru 11, Uruguay 6
- 6 Australia 4, Canada 2, New Zealand 2, United States of America 30
- 7 Afghanistan 31, Bangladesh 18, Sri Lanka 8, India 18, Pakistan 5
- 8 Azerbaijan 4, Armenia 5, Cyprus 2, Georgia 3, Occupied Palestinian Territory 27, Israel 2, Jordan 2, Kuwait 5, Saudi Arabia 2
- 9 Myanmar 3, Cambodia 2, Indonesia 2, Malaysia 5, Philippines 17, Viet Nam 39, Thailand 42
- 10 Asia 1, China 26, China, Hong Kong Special Administrative Region 1, Japan 5, Republic of Korea 15, Mongolia 2, Turkmenistan 1, Uzbekistan 4
- 11 Bulgaria 13, Belarus 1, Czechoslovakia 1, Czech Republic 3, Hungary 24, Romania 18, Russian Federation 15, Slovakia 2, Ukraine 1
- 12 Albania 5, Greece 17, Italy 28, Portugal 8, Spain 10, Sinti and Roma 2
- 13 Austria 10, Belgium 1, Estonia 26, Aland Islands 1, France 9, Iceland 2, Ireland 1, Latvia 10, Lithuania 4, Luxembourg 1, Netherlands 12, Switzerland 2, United Kingdom of Great Britain and Northern Ireland 26

References

- Alba, R. (1988). "Cohorts and the dynamics of ethnic change", in: M. W. Riley (ed.), *Social Structures and Human Lives*, Newbury Park: Sage Publications, 211-228.
- Alba, R., Logan, J., Lutz, A. and Stults, B. (2002). "Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants", *Demography* 39(3), 467-484.
- Allison, P. D. (2001). *Missing Data – Quantitative Applications in the Social Sciences*, Thousand Oaks: Sage Publications.
- Barnett, M. D., Sonnert, G. and Sadler, P. M. (2012). "More like us: the effect of immigrant generation on college success in mathematics", *International Migration Review* 46(4), 891-918.
- Becker, B. (2011). "Cognitive and language skills of Turkish children in Germany: A comparison of the second and third generation and mixed generational groups", *International Migration Review* 45(2), 426-459.
- Berthoud, R. (2005). "Family formation in multicultural Britain: diversity and change", in: G. C. Loury, T. Modood and S. M. Teles (eds.), *Ethnicity, Social Mobility and Public Policy: Comparing the USA and UK*, Cambridge: Cambridge University Press, 222-253.
- Çelikaksoy, A., Nekby, L. and Rashid, S. (2010). "Assortative mating by ethnic background and education among individuals with an immigrant background in Sweden", *Zeitschrift für Familienforschung* 22(1), 65-88.
- Crul, M. and Vermeulen, H. (2003). "The second generation in Europe", *International Migration Review* 37(4), 965-986.
- Dagevos, J. (1997). "Equivalency and equal access as standards for ethnic minority policy in the Netherlands", in: Council of Europe (ed.), *Measurement and Indicators of Integration*, Strasbourg: Council of Europe Publishing, 89-100.
- Dale, A. and Ahmed, S. (2011). "Marriage and employment patterns amongst UK-raised Indian, Pakistani, and Bangladeshi women", *Ethnic and Racial Studies* 34(6), 902-924.
- Farley, R. and Alba, R. (2002). "The new second generation in the United States", *International Migration Review* 36(3), 669-701.
- González-Ferrer, A. (2006). "Who do immigrants marry? Partner choice among single immigrants in Germany", *European Sociological Review* 22(2), 171-185.
- Gresch, C. and Kristen, C. (2011). "Staatsbürgerschaft oder Migrationshintergrund? Ein Vergleich unterschiedlicher Operationalisierungsweisen am Beispiel der Bildungsbeteiligung", *Zeitschrift für Soziologie* 40(3), 208-227.
- Heath, A. and Brinbaum, Y. I. (2007). "Explaining ethnic inequalities in educational attainment", *Ethnicities* 7(3), 291-305.
- Heath, A., Rethon, C. and Kilpi, E. (2008). "The second generation in Western Europe: education, unemployment, and occupational attainment", *Annual Review of Sociology* 34, 211-235.
- Jacobs, D., Swyngedouw, M., Hanquinet, L., Vandezande, V., Andersson, R., Horta, A., . . . Statham, P. (2009). "The challenge of measuring immigrant origin and immigration-related ethnicity in Europe", *Journal of International Migration and Integration* 10(1), 67-88.

- Kalmijn, M. and van Tubergen, F. (2006). "Ethnic intermarriage in the Netherlands: confirmations and refutations of accepted insights", *European Journal of Population* 22(4), 371-397.
- Kalter, F., Heath, A. F., Hewstone, M., Jonsson, J. O., Kalmijn, M., Kogan, I. and Tubergen, F. v. (2014a). *Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) – Full Version*. Data file for on-site use. GESIS Data Archive, Cologne, ZA5353 Data file Version 1.1.0, doi:10.4232/cils4eu.5353.1.1.0 (forthcoming).
- Kalter, F., Heath, A. F., Hewstone, M., Jonsson, J. O., Kalmijn, M., Kogan, I. and Tubergen, F. v. (2014b). *Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) – Reduced version*. Reduced data file for download and off-site use. GESIS Data Archive, Cologne, ZA5656 Data file Version 1.1.0, doi:10.4232/cils4eu.5656.1.1.0 (forthcoming).
- Kalter, F. and Schroedter, J. H. (2010). "Transnational marriage among former labour migrants in Germany", *Zeitschrift für Familienforschung* 22(1), 11-36.
- Kogan, I. (2010). "Introduction to the special issue on minority ethnic groups' marriage patterns in Europe", *Zeitschrift für Familienforschung* 22(1), 3-10.
- Muttarak, R. (2010). "Explaining trends and patterns of immigrants' partner choice in Britain", *Zeitschrift für Familienforschung* 22(1), 37-64.
- Oropesa, R. S. and Landale, N. S. (1997). "In search of the new second generation: alternative strategies for identifying second generation children and understanding their acquisition of English", *Sociological Perspectives* 40(3), 429-455.
- Parameshwaran, M. (2013). "Explaining intergenerational variations in English language acquisition and ethnic language attrition", *Ethnic and Racial Studies* 37(1), 27-45.
- Perlmann, J. and Waldinger, R. (1997). "Second generation decline? Children of immigrants, past and present – a reconsideration", *International Migration Review* 31(4), 893-922.
- Ramakrishnan, S. K. (2004). "Second-generation immigrants? The '2.5 Generation' in the United States", *Social Science Quarterly* 85(2), 380-399.
- Rumbaut, R. (1997). "Paradoxes (and orthodoxies) of assimilation", *Sociological Perspectives* 40(3), 483-511.
- Rumbaut, R. (2004). "Ages, life stages, and generational cohorts: decomposing the immigrant first and second generations in the United States", *International Migration Review* 38(3), 1160-1205.
- Segeritz, M., Walter, O. and Stanat, P. (2010). „Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland. Evidenz für segmentierte Assimilation?“, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62(1), 113-138.
- Smith, T. W. (1984). "The subjectivity of ethnicity", in: C. F. Turner and E. Martin (eds.), *Surveying Subjective Phenomena, part 2*, New York: Russell Sage, 117-129.
- Van Tubergen, F. and Maas, I. (2007). "Ethnic intermarriage among immigrants in the Netherlands: an analysis of population data", *Social Science Research* 36(3), 1065-1086.
- Warner, W. L. and Srole, L. (1945). *The Social Systems of American Ethnic Groups*, New Haven: Yale University Press.
- Waters, M. C. and Jimenez, T. R. (2005). "Assessing immigrant assimilation: new empirical and theoretical challenge", *Annual Review of Sociology* 31, 105-125.

Appendix

Table A.1

Rule	Correct	Total	%
1 Top-down missing-replacement if descendant is foreign-born	26,792	27,985	95.7
from child to parents	3,719	4,033	92.2
from child to mother	1,872	2,036	91.9
from child to father	1,847	1,997	92.5
from parents to grandparents	23,073	23,952	96.3
from mother to maternal grandparents	11,373	11,804	96.4
from mother to maternal grandmother	5,703	5,921	96.3
from mother to maternal grandfather	5,670	5,883	96.4
from father to paternal grandparents	11,700	12,148	96.3
from father to paternal grandmother	5,857	6,088	96.2
from father to paternal grandfather	5,843	6,060	96.4
2 Complement grandparent-dyads	32,796	34,738	94.4
maternal grandparents	16,602	17,699	93.8
paternal grandparents	16,194	17,039	95.0
3 Bottom-up missing replacement if ancestors are born in the survey country	64,946	66,139	98.2
from grandparents to parents	41,262	42,141	97.9
from maternal grandparents to mother	21,225	21,656	98.0
from maternal grandmother to mother	10,721	10,939	98.0
from maternal grandfather to mother	10,504	10,717	98.0
from paternal grandparents to father	20,037	20,485	97.8
from paternal grandmother to father	10,121	10,352	97.8
from paternal grandfather to father	9,916	10,133	97.9
from parents to child	23,684	23,998	98.7
from mother to child	12,076	12,240	98.7
from father to child	11,608	11,758	98.7
4 Replace missing triad by complete triad	12,836	16,511	77.7
5 Replace one parent with other parent	15,998	18,176	88.0
6 Replace completely missing grandparent-dyad information with known grandparent-dyad	13,631	16,716	81.5

Note: Comparisons are based on dyadic comparisons. This means that we are comparing the fit of the predictive power between the child and the mother as well as between the child and the father, or between the mother and the maternal grandmother as well as between the mother and the paternal grandfather etc. This dyadic procedure comes together with an increase in sample size, as the level of analysis is not the single respondent, but rather the dyads.