TwinLife Technical Report Series

No. 03, August 2017

The sampling design and socio-demographic structure of the first wave of the TwinLife panel study:

A comparison with the Microcensus

by Volker Lang, Anita Kottwitz

volker.lang@uni-bielefeld.de

http://www.twin-life.de/en/twinlife-technical-report-series

TwinLife

Volker Lang, Anita Kottwitz

The sampling design and socio-demographic structure of the first wave of the TwinLife panel study: a comparison with the Microcensus (updated version)

TwinLife Technical Report Series No. 03
Project TwinLife “Genetic and social causes of life chances”
Bielefeld, August 2017

TwinLife Technical Report Series
General Editors: Martin Diewald, Rainer Riemann and Frank M. Spinath
ISSN 2512-403X

This publication has been funded by the German Research Foundation (DFG).

TwinLife Technical Reports are refereed scholarly papers. Submissions are reviewed by the general editors before a final decision on publication is made.

The Technical Report Series is a forum for presenting technical works (e.g., data documentation, field reports etc.) in progress. Readers should communicate comments on the manuscript directly to the author(s).

The papers can be downloaded from the project website:
http://www.twin-life.de/en/twinlife-technical-report-series

TwinLife “Genetic and social causes of life chances”
University of Bielefeld
Faculty of Sociology
PO Box 100131
D-33501 Bielefeld
Germany

Phone: +49 (0)521 106-4309
Email: martin.diewald@uni-bielefeld.de
Web: http://www.twin-life.de/en

Universität Bielefeld

UNIVERSITÄT DES SAARLANDES

The sampling design and socio-demographic structure of the first wave of the TwinLife panel study: a comparison with the Microcensus[a]

Volker Lang, Bielefeld University: volker.lang@uni-bielefeld.de
Anita Kottwitz, Bielefeld University: anita.kottwitz@uni-bielefeld.de

Abstract

The main objectives of this paper are, first, to assess differences in the socio-demographic structure between twin and multiple-child households in Germany and, second, to demonstrate the usability of the TwinLife panel study for differentiated social structural analysis. More specifically, this paper will describe the sampling design of the TwinLife study and compare the distributions of the key socio-demographic variables in TwinLife with those covered by the German Microcensus using a proxy-twin and a multiple-child household sample. The analyses show that (proxy-)twin and multiple-child households in Germany are similar with respect to most socio-demographic indicators. The sole difference is the higher prevalence of households with two children among multiple-child households. Further, the analyses demonstrate that the probability-based sampling design of TwinLife successfully counteracts the overrepresentation of monozygotic twins typically characterizing other twin samples based on self-recruitment. Most importantly, the results show that the TwinLife sample covers the full distributions including the lower and upper bounds for the core social inequality indicators – educational status, occupational status and income. This enables researchers to use the TwinLife survey for multidimensional analyses of social inequality and differences in gene expression, and thus also to study gene-environment interplay. However, the analyses also indicate that participation in the first wave of TwinLife was, to some degree, selective with regard to parental educational level, particularly in the younger cohorts. A weighting scheme addressing this selectivity is therefore suggested.

[a] We thank Martin Diewald and Kristina Krell for their very helpful comments on the manuscript of this paper. We are also grateful to Kristina Krell for her assistance in facilitating access to the Microcensus data.

1. Introduction

The TwinLife panel is the first longitudinal study of twin families in Germany based on a national probability sample (Diewald et al., 2016; Hahn et al., 2016). In contrast to many other countries (e.g., Australia: Hopper, Foley, White, & Pollaers, 2013; Denmark: Skytthe et al., 2013; Finland: Kaprio, 2013; Italy: Brescianini et al., 2013; The Netherlands: van Beijsterveldt et al., 2013; Norway: Nilsen et al., 2013; Sweden: Magnusson et al., 2013), no twin registry is available for Germany. Moreover, efforts to collect twin data for Germany have been focused on specific regions, and none of the studies to date has used probability-based sampling methods to construct a twin sample (Busjahn, 2013; Hahn, Gottschling, & Spinath, 2013; Kandler et al., 2013). Having a probability-based sample is an advantage in itself. However, a major potential asset of the TwinLife panel study is that it covers the whole range of social structural variation in Germany. This allows for a more accurate account of how genetic and socio-demographic variation mutually influence life courses, developmental trajectories, and social inequalities than studies using non-probabilistic samples which have a weak coverage of certain population groups. Further, it enables more reliable international comparisons with twin registry data or representative twin surveys (e.g., MSUTR: Burt & Klump, 2013; TEDS: Haworth, Davis, & Plomin, 2013), and also with representative national samples contained in family or household panel studies (e.g., pairfam: Huinink et al., 2011; SOEP: Wagner, Frick, & Schupp, 2007) which are not twin-based. In addition, capturing the lower and upper bounds of social structural characteristics is important for TwinLife because these extremes in particular might be relevant for analyzing related differences in gene expression (i.e., gene-environment interplay). Of course, a twin family sample does not represent all existing living arrangements in a country since childless people and one-child families are missing by definition. Thus, more specifically, the question this paper seeks to address is whether and how the socio-demographic structure of twin families participating in the TwinLife study differs from the social structure of multiple-child families in the population at large.

To assess the extent to which the socio-demographic structure of the twin family sample conforms with that of multiple-child families in general, and to demonstrate the usability of TwinLife for socio-demographic differentiated analysis, this paper contrasts the distributions of social structural indicators considered important for gene-environment interplay. The comparison is based on the first wave of the TwinLife panel and the German Microcensus Survey conducted by the Federal Statistical Office (Destatis, 2014a, 2014b; Lengerer et al., 2007. The Microcensus is a representative annual household survey covering around one percent of the German population. Specifically, we compare highest educational and occupational status, monthly net equivalent income, region and size of community of residence, and citizenship status. We also look at maternal age at childbirth as a potential reason for different social structural distributions.

2. Design and sampling strategy of the TwinLife panel study

The TwinLife study collects longitudinal data for the analysis of genetic and environmental influences on social inequality with a focus on families with monozygotic or dizygotic twin children. To exclude effects of within-twin-pair gender differences, the study includes only same-sex dizygotic twins. The base population of TwinLife consists of twins aged 5, 11, 17, and 23 to 24 at the time of the first survey. TwinLife therefore comprises four birth cohorts (see Figure 1).

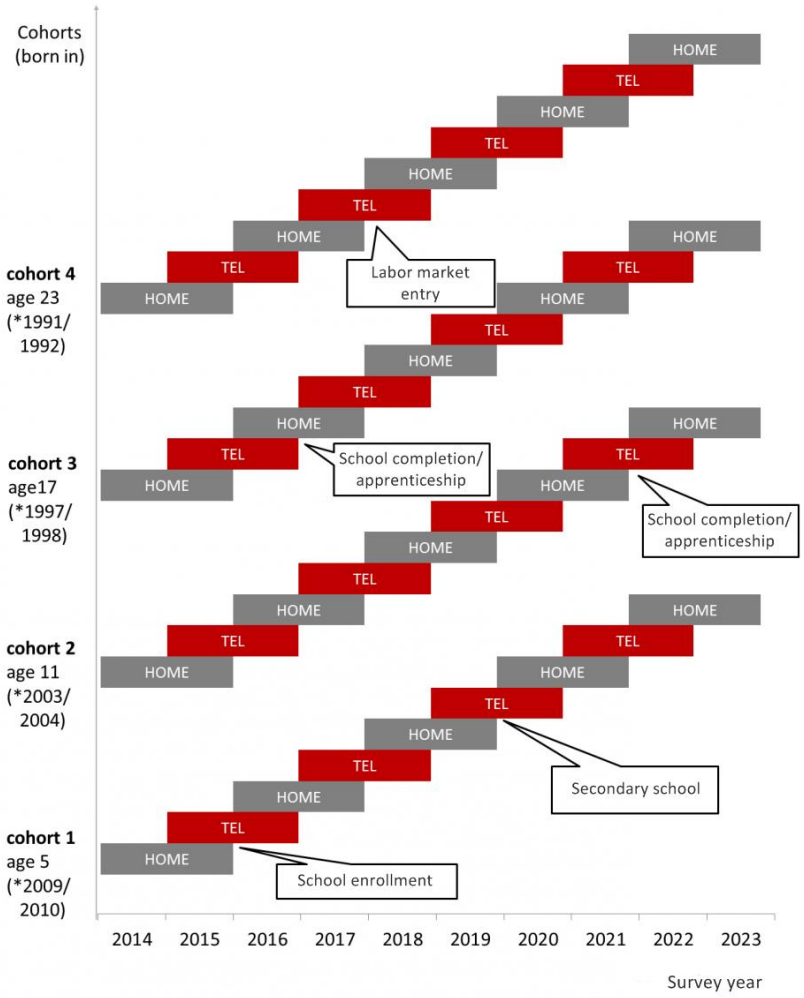

Figure 1: The cohort-sequential design and survey modes of the TwinLife panel

Due to the sampling design of TwinLife (see p. 5) the twins in these cohorts are sampled from several years of birth: the youngest twins, in cohort 1, were born in 2009 or2010, the twins in cohort 2 in 2003 or 2004, the twins in cohort 3 in 1997 or 1998, and the oldest twins, in cohort 4, between 1990 and 1993. Over the planned panel period TwinLife covers important life course transitions ranging from school entry to the labor market entry phase, and also important life stages for meeting a partner and starting a family. The TwinLife surveys are conducted annually and survey modes alternate between face-to-face interviews at home and telephone interviews.

In addition, the TwinLife study combines this cohort-sequential design with an extended twin family design (ETFD). As part of the ETFD, the biological and, if applicable, the social parents (i.e., partners of mothers and fathers also called "step-parents" in the TwinLife panel), and the sibling that is closest in age to the twins are surveyed additionally to the twins themselves (see Figure 2). Moreover, the partners of adult twins are also included. All of these family members are included in the design irrespective of whether they live in the same household as the twins or not. A family in TwinLife can therefore consist of several households, i.e., the households are nested within the families. The minimum requirement for inclusion as a valid family case in the TwinLife panel was the participation of both twins and one of the biological or social parents in the first wave.[2] A further design requirement was that the twins were raised together, i.e., lived in the same household until age 16. The family perspective of the ETFD facilitates the study of different degrees of genetic similarity which is important for detailed analysis of the manifold influences of the family environment on the development of the twins.

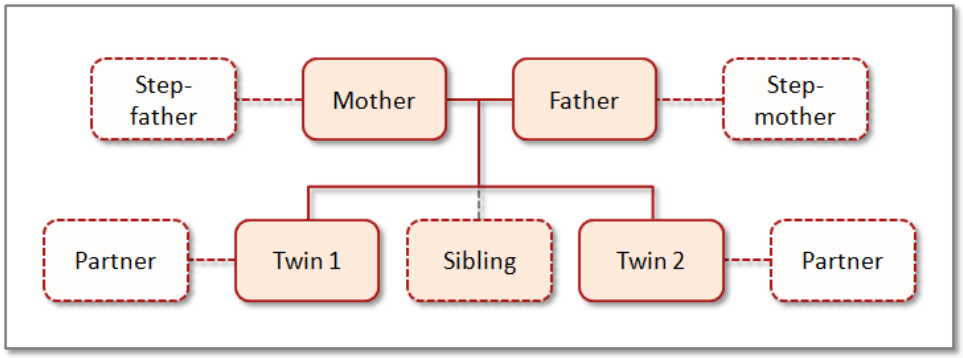

Figure 2: Extended twin family design used for the TwinLife panel

The target net sample size for wave 1 of the TwinLife panel was 1,000 twin families in each of the four birth cohorts with approximately half of the families having monozygotic and the other half having same-sex dizygotic twins. To obtain a sample with these design characteristics, a national probability-based sampling procedure was implemented in two steps (TNS Infratest, 2016): first, a sample of 500 out of approximately 11,900 communities was drawn to generate addresses where twin families matching the design requirements resided. Potential twin families in cohorts 1 to 3 were identified by locating persons of the same sex with the same or similar birthdates registered at the same address according to the current registry of residents for the respective communities. Families in cohort 4 were also selected based on previous registries of residents containing address data prior to reported house moves. Using these previous addresses, an inquiry for the current address of the persons identified as probable twins was carried out. Second, a gross sample of 13,359 addresses out of around 19,000 addresses provided by the local registry of residents was drawn; 2,736 for cohort 1, 2,697 for cohort 2, 2,823 for cohort 3, and 5,103 for cohort 4.

[2] Exceptions are orphan families where there is no parent to participate. There are four cases of this type in the net sample of the panel.

Table 1: Distribution of addresses by community size in TwinLife and population

Community size category (GKPOL)	Number of communities in TwinLife	Percent of addresses in … ... population	... TwinLife gross sample	... TwinLife net sample
5,000–19,999	362	22.6	17.4	18.9
20,000–49,999	56	26.7	10.7	11.7
50,000–99,999	42	12.1	17.0	16.3
100,000–499,999	28	18.0	22.4	22.2
500,000 or more	12	20.6	32.6	30.8
Total	500	100.0	100.0	100.0

Sources: TNS Infratest 2016 and TwinLife, wave 1 (doi: 10.4232/1.12665), own calculations

Given these gross sample sizes, it was a priori obvious that the sampling design could not be proportional. Thus, each of the cohorts 1 to 3 is composed of two years of birth and cohort 4 of four years of birth. Population statistics for twin families in Germany are not available, but it is known that there are approximately 7,000 same-sex twin births each year (about 0.01 percent of all annual births, Destatis, 2013). Consequently, a design using the gross sample sizes described above and based on a cohort composed of only one year of birth would have to cover around 40 percent of the population for cohorts 1 to 3 and 75 percent for cohort 4. Using multiple-year birth cohorts reduces this share to approximately 20 percent. A proportional implementation of this design would necessitate conducting face-to-face interviews in around 2,500 communities which is impracticable. Three subsamples of communities were therefore selected instead: first, a proportional sample of 180 communities with 10,000 or more inhabitants was drawn according to the political community size classification for Germany (GKPOL) ("base sample"). Second, a disproportional sample (with higher sampling probabilities for larger communities) of 60 communities with 50,000 or more inhabitants was selected to obtain the necessary coverage of the target population ("urban sample"). Third, an additional proportional sample of 260 communities with between 5,000 and 19,999 inhabitants was drawn ("rural sample").[3] The base sample consists of 5,575 addresses (41.7 percent of the gross sample), the urban sample of 6,558 addresses (49.1 percent of the gross sample), and the rural sample of 1,226 addresses (9.2 percent of the gross sample). This sampling design, which is disproportional overall, leads to an overrepresentation of addresses located in urban communities in the TwinLife panel in comparison to all addresses registered in communities with 5,000 or more inhabitants (see Table 1).

The gross sample of addresses described above was used for the face-to-face interviews of the TwinLife panel, wave 1. The data collection for twins born in 2009, 2003, 1997, and 1990 or 1991 was carried out between September 2014 and May 2015. For twins born in 2010, 2004, 1998, and 1992 or 1993, data collection started in September 2015 and was

[3] Communities with fewer than 5,000 inhabitants are excluded by the TwinLife design. This is because, on average, only one or two twin families over all birth cohorts studied are expected to reside in a community of this size, making conducting face-to-face twin family interviews in communities like this prohibitively expensive and at the same time particularly problematic with respect to a possible re-identification.

completed in April 2016. Table 2 shows distributions of the gross and net samples differentiated by cohort. 10.5 percent of the addresses in the gross sample were invalid contact addresses and 4.2 percent did not comply with the requirements of the design, leaving an adjusted gross sample of 11,405 cases. In cohorts 1 to 3, around 10 percent of the cases in the adjusted gross sample were permanently absent or sick during the field phase and 40 percent refused to participate. In cohort 4, the sickness rate was twice as high and half of the sample refused participation. In 1.1 percent of the cases, it was not possible to interview all the necessary family members according to the design requirements, 2.5 percent of the addresses were not used because the target sample size had already been obtained, and 1.9 percent of the cases did not participate for other reasons. This results in a net sample for wave 1 of 1,010 families in cohort 1, 1,043 families in cohort 2, 1,060 families in cohort 3, and 984 families in cohort 4, which closely matches the target sample size. The participation rate based on the adjusted gross sample is therefore over 40 percent in cohorts 1 to 3 and 23.0 percent in cohort 4. A total of 39.0 percent of the families in the net sample are part of the base sample, 51 percent are part of the urban sample, and 10.1 percent are part of the rural sample. For more information on the field process, see the report by TNS Infratest (2016).

Table 2: Gross and net samples of TwinLife

	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Gross sample	2,736 (100.0)	2,697 (100.0)	2,823 (100.0)	5,103 (100.0)	13,359 (100.0)
- no contact address	338 (12.4)	261 (9.7)	220 (7.8)	580 (11.4)	1,399 (10.5)
- no match with design	127 (4.6)	93 (3.4)	89 (3.2)	246 (4.8)	555 (4.2)
Adjusted gross sample	2,271 (83.0)	2,343 (86.9)	2,514 (89.1)	4,277 (83.8)	11,405 (85.4)
Adjusted gross sample	2,271 (100.0)	2,343 (100.0)	2,514 (100.0)	4,277 (100.0)	11,405 (100.0)
- absent or sick	258 (11.4)	267 (11.4)	237 (9.4)	891 (20.8)	1,653 (14.5)
- refusal	870 (38.3)	906 (38.7)	1,060 (42.2)	2,190 (51.2)	5,026 (44.1)
- family not complete	31 (1.4)	25 (1.1)	28 (1.1)	45 (1.1)	129 (1.1)
- address not used	69 (3.0)	60 (2.6)	80 (3.2)	80 (1.9)	289 (2.5)
- other reason	33 (1.5)	42 (1.8)	48 (1.9)	88 (2.1)	211 (1.9)
Net sample	1,010 (44.5)	1,043 (44.5)	1,061 (42.2)	983 (23.0)	4,097 (35.9)

Sources: TNS Infratest 2016 and TwinLife, wave 1 (doi: 10.4232/1.12665), own calculations

Table 3 displays distributions by sex and zygosity of the twin pairs over the four cohorts for the net sample of the TwinLife panel.[4] There are more dizygotic than monozygotic twin pairs in cohorts 1 to 3, and in cohort 4 the share of monozygotic twin pairs is 53.3 percent. These results indicate that the probability-based sampling design used for TwinLife was successful in counteracting the overrepresentation of monozygotic twins typically characterizing other twin samples based on self-recruitment (i.e, two-thirds monozygotic twin pairs, with overrepresentation particularly pronounced in adult samples, Lykken, McGue, & Tellegen, 1987). The findings are also in line with research showing an increase in dizygotic twining

[4] In 50 of these families, second twin pairs exist; in 38 cases these are full siblings of the other twins, in eight cases, they are half-siblings, and in three cases, step-siblings. Moreover, one of the families has full sibling triplets in addition to the twins.

rates for OECD countries, including Germany, since the 1980s (Hoekstra et al., 2008). This is primarily because dizygotic twinning is more strongly influenced by environmental factors such as the increase in maternal age at childbirth over recent decades. Overall, the distributions demonstrate that the TwinLife sample enables genetically sensitive analyses differentiated by gender and age.

Table 3: Net sample of TwinLife by sex and zygosity of twin pairs

	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Male, monozygotic	209 (20.8)	191 (18.4)	218 (20.6)	212 (21.6)	830 (20.4)
Male, dizygotic	279 (27.8)	307 (29.6)	235 (22.2)	198 (20.2)	1,019 (25.0)
Female, monozygotic	225 (22.4)	229 (22.1)	280 (26.4)	311 (31.7)	1,045 (25.6)
Female, dizygotic	291 (29.0)	309 (29.8)	326 (30.8)	259 (26.4)	1,185 (29.1)
Total	1,004 (100.0)	1,036 (100.0)	1,059 (100.0)	980 (100.0)	4,079[a] (100.0)

[a] In 11 families the multiples are triplets (five male and six female) and for seven female twin pairs no information on their zygosity is available.
Sources: TNS Infratest 2016 and TwinLife, wave 1 (doi: 10.4232/1.12665), own calculations

As described above, both twins, one sibling, their parents, and the partners of the adult twins are the target respondents for the interviews, irrespective of whether they live in the same household or not. Table 4 shows the composition of the families (upper part of Table 4) and the households (lower part of Table 4) interviewed in TwinLife, wave 1. Overall, the TwinLife net sample consists of 4,097 twin families living in 4,828 households. A total of 91.4 percent of these families are families with two parents.[5] However, the share of two-parent families decreases over the cohorts from 95.6 percent to 87.1 percent. In 62.2 percent of the families the twins have at least one sibling. Due to the correlation between children's age and parents' reproductive phase, this share increases from 54.9 percent in cohort 1 to around 65 percent in cohorts 2 to 4. The mean number of siblings per family in families with at least one sibling is 1.6, and the maximum number of siblings is ten. Overall, the distributions indicate that TwinLife facilitates studies based on the ETFD. The lower part of Table 4 illustrates the distribution of households in TwinLife across cohorts. As required by the study design, all of the twins in cohorts 1 and 2, and almost all of the twins in cohort 3 live together in one household. In more than 90 percent of the twin households in cohort 1, the twins live with two parents. This share drops to about 75 percent in cohort 3. For the young adult twins (cohort 4), the share of twin households with at least one parent is 54.1 percent. This corresponds to 43.9 percent of all households in cohort 4. A total of 76 percent of the twins from cohort 4 who had already moved out of the parental household are living without their co-twin. This represents 32.2 percent of all households in cohort 4. Further, the share of non-twin households increases from approximately 1 percent in cohorts 1 to 3 to 18.9 percent in cohort 4. Overall, these results illustrate that TwinLife captures the major

[5] In 99.1 percent of the families with a mother, these mothers are the biological mothers of the twins. The share of biological fathers is 96.6 percent. Moreover, there are more than two parents, i.e., a partner of a father or mother in addition to the biological parents, in 3.8 percent of the families.

shift in household structures resulting from the young adult twins starting to create their own families.[6]

Table 4: Family and household compositions in the net sample of TwinLife

Family composition	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Mother and father, twins	431 (42.7)	337 (32.3)	350 (33.0)	290 (29.5)	1,408 (34.4)
Mother and father, twins, sibling	534 (52.9)	644 (61.7)	591 (55.7)	566 (57.6)	2,335 (57.0)
Mother or father, twins	25 (2.5)	23 (2.2)	46 (4.3)	45 (4.6)	139 (3.4)
Mother or father, twins, sibling	20 (2.0)	39 (3.7)	74 (7.0)	78 (7.9)	211 (5.2)
No parents, (sibling)[a]	0 (0)	0 (0)	0 (0)	4 (0.4)	4 (0.1)
Total	1,010 (100)	1,043 (100)	1,061 (100)	983 (100)	4,097 (100)
Household composition	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Parents, both twins, (sibling)[b]	917 (90.3)	883 (83.4)	815 (74.1)	428 (25.9)	3,043 (63.0)
Parent, both twins, (sibling)[b]	93 (9.2)	160 (15.1)	231 (21.0)	113 (6.8)	597 (12.4)
Parent(s), one twin, (sibling)[b]	0 (0)	0 (0)	22 (2.0)	184 (11.1)	206 (4.3)
Both twins, (sibling)[b]	0 (0)	0 (0)	0 (0)	84 (5.1)	84 (1.7)
One twin, (sibling)[b]	0 (0)	0 (0)	8 (0.7)	532 (32.2)	540 (11.2)
No twins	6 (0.6)	16 (1.5)	24 (2.2)	312 (18.9)	358 (7.4)
Total	1,016 (100)	1,059 (100)	1,100 (100)	1,653 (100)	4,828 (100)

[a] Orphan families; three with at least one sibling and one with no sibling.
[b] Living in a household either with or without at least one sibling.
Source: TwinLife, wave 1 (doi: 10.4232/1.12665), own calculations

3. Data, methods, and indicators for comparing TwinLife with the German Microcensus

To evaluate the social structural coverage of the net sample of TwinLife, we compare it with the 2013 Microcensus (Destatis, 2014a, 2014b; Lengerer et al., 2007), which is a German household survey based on a nationally representative sample of one percent.[7] While the sampling strategy of TwinLife is focused on families defined by the ETFD, the sampling design of the Microcensus is based on households, specifically persons living together at the same address sampled from the population register (Lengerer, Bohr, & Janßen, 2005). Hence, to obtain comparable groups, we have to implement a two-step procedure: first, we define a household sample based on the net household sample of the TwinLife panel, which consists of the twins and at least one parent. Second, we select two different household samples from the Microcensus similar to the cohort and person composition of the TwinLife panel households as defined in the first step.

There are 3,640 (out of 4,828) TwinLife households in which the twins live together with at least one parent ("parent-twin sample," see Table 4). These households comply with our household definition for comparison with the Microcensus. This sample of households

[6] 43.4 percent of the twins in cohort 4 have a partner and 30.7 percent of these twins live in a household with their partners.
[7] The 2013 Microcensus provides the most recent data currently available and thus most accurately reflects the population of 2015—the year in which the majority of the families in the TwinLife panel was sampled.

comprises almost all twin families and households in which at least one twin resides. For cohort 4 it covers 73.8 percent of all twin families and 54.1 percent of all households with twins. The Microcensus survey does not collect information on whether the children living in the household are twins or not. Thus, we need to construct a suitable comparison sample to match the cohort and person composition of the TwinLife parent-twin sample described above without this information. First, we select one-family households with one or two parents and at least two children under the age of 25 of which at least one child—the "anchor child"—belongs to the same birth cohorts as in TwinLife ("multiple-child sample"). Second, we construct a sample of proxy-twin households, i.e., one-family households in which two children of the same sex are born in the same year and live with at least one of their parents ("proxy-twin sample"). In view of the approximately 7,000 same-sex twin births each year (see Destatis, 2013), we can expect to find around 70 proxy twins in the 2013 Microcensus for each year of birth from circa 2000 and declining numbers for the years prior to 2000 based on the following assumptions: 1) a household sample of one percent from the population approximates a population sample of one percent; 2) there are only rare cases,[8] other than twin births, of same-sex children in a household being born in the same year; 3) most twin children live together and with at least one parent. To gain a proxy-twin sample of sufficient size for socio-demographic differentiated analyses, we use six-year birth cohorts: 2007–2012 (cohort 1), 2001–2006 (cohort 2), 1995–2000 (cohort 3), and 1989–1994 (cohort 4). Moreover, to match the TwinLife sampling design, households in communities with fewer than 5,000 inhabitants are excluded. These represent about 16 percent of the households in both Microcensus comparison samples.

Table 5 shows the household structures in the TwinLife parent-twin sample in contrast to the two Microcensus comparison samples. The number of children living in a household with both parents differs in the Microcensus multiple-child sample compared to the TwinLife parent-twin and the Microcensus proxy-twin samples. While there are around 58.9 percent of two-children households with both parents in the former sample, this share is approximately 40 percent in the latter two. This difference is plausible since potential parents often plan to have two children (Ruckdeschel, 2007) but if the second birth is a twin birth, they have three children. The share of single-parent households is about 16 percent in all three samples. Overall, these results indicate that the main difference in the composition of twin and non-twin multiple-child households is the higher prevalence of two-children households in the latter group. In addition, the findings confirm that the probability-based sampling procedure used for TwinLife was appropriate in this regard since the household structures in the TwinLife parent-twin and the Microcensus proxy-twin samples are similar.

[8] There are rarely any women who give birth to two children within the same calendar year. However, the Microcensus does not differentiate between biological and step-children. Thus, there might be a negligible number of cases which are spuriously considered as twin families. These might be foster or blended families with same-sex children born in the same year.

Table 5: Household structures in the TwinLife and Microcensus comparison samples

TwinLife parent-twin	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Couples, twin(s)	428 (42.4)	355 (34.0)	401 (38.3)	259 (47.9)	1443 (39.6)
Couples, twin(s), sibling	489 (48.4)	528 (50.6)	414 (39.6)	169 (31.2)	1600 (44.0)
Single parent, twin(s)	50 (5.0)	80 (7.7)	149 (14.2)	76 (14)	355 (9.8)
Single parent, twin(s), sibling	43 (4.3)	80 (7.7)	82 (7.8)	37 (6.8)	242 (6.6)
Total	1,010 (100.0)	1,043 (100.0)	1,046 (100.0)	541 (100.0)	3,640 (100.0)
Microcensus multiple-child	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Couples, 2 children	3,680 (61.1)	3,523 (55.6)	3,531 (55.7)	3,558 (63.9)	14,292 (58.9)
Couples, 3 or more children	1,713 (28.5)	1,774 (28.0)	1,544 (24.3)	948 (17.0)	5,979 (24.6)
Single parent, 2 children	426 (7.1)	732 (11.5)	958 (15.1)	924 (16.6)	3,040 (12.5)
Single parent, 3+ children	199 (3.3)	310 (4.9)	309 (4.9)	142 (2.5)	960 (4.0)
Total	6,018 (100.0)	6,339 (100.0)	6,342 (100.0)	5,572 (100.0)	24,271 (100)
Microcensus proxy-twin	Cohort 1 (%)	Cohort 2 (%)	Cohort 3 (%)	Cohort 4 (%)	Total (%)
Couples, 2 children	139 (46.8)	82 (28.3)	99 (33.2)	70 (45.5)	390 (37.5)
Couples, 3 or more children	122 (41.1)	149 (51.4)	139 (46.6)	48 (31.2)	458 (44.1)
Single parent, 2 children	20 (6.7)	34 (11.7)	30 (10.1)	27 (17.5)	111 (10.7)
Single parent, 3+ children	16 (5.4)	25 (8.6)	30 (10.1)	9 (5.8)	80 (7.7)
Total	297 (100.0)	290 (100.0)	298 (100.0)	154 (100.0)	1,039 (100.0)

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

With respect to the social structural indicators used for the analysis, we compare the region of residence and the size of the communities where the household is located, German citizenship status on the household level, highest educational and occupational status of parents in the household, and also monthly net equivalent household income in euros. To assess the potential use of the TwinLife study for multidimensional analysis of social structural (dis-)advantage, we also look at the bivariate distribution of highest educational status in the household by monthly net equivalent household income. Moreover, we contrast maternal age at childbirth as a potential reason for social structural differences between the samples. Region of residence is defined as eastern and western German federal states. The size of the community where the household is located is categorized based on the German community size classification (GKPOL). German citizenship is used as a proxy for migration background since the definitions of migration background used in TwinLife and the Microcensus are not directly comparable. We assign German citizenship status on the household level if both parents have German citizenship. For the highest educational and occupational status, the individual-level information on the parents is used to calculate the maximum status on the household level. The highest educational status within the household is based on the International Standard Classification of Education (ISCED) 1997 (Schneider, 2008) and the highest occupational status is based on the International Socio-Economic Index (ISEI) (Ganzeboom, Graaf, & Treiman, 1992). The ISCED is coded as an ordered categorical variable with “no educational degree” (1) as the lowest and “Ph.D. degree” (6) as the highest category. The ISEI is coded as a metric score with a

range between 12 and 89 and is generated based on the International Standard Classification of Occupations (ISCO) 2008. Information on monthly net income is surveyed on the household level. To make the household incomes comparable across different household structures, an equivalence weight according to the new OECD scheme (OECD, 2011) is applied. In addition, income is adjusted for inflation by dividing the nominal income by the Consumer Price Index for Germany using 2015 as the base year.

There is some data missing for several indicators in the TwinLife parent-twin sample.[9] To account for these missing values, we set up a multiple imputation model on the household level using the information on mothers, fathers, and households.[10] We impute 20 values for each missing observation. Specifically, we do multiple imputations with chained equations (van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006), a method which iterates over a sequence of univariate imputation models for each variable. For the univariate imputation models, we use predictive mean matching with ten nearest neighbors in the case of continuous variables and logistic or ordered logistic regressions in the case of categorical variables.[11] We assess the influence of the imputation procedure on the distributions of the social structural indicators compared. Here, we find slight increases in the lower categories of the indicators (typically about 2 percent) and converse declines in the upper categories. However, there are only minor differences between imputed and non-imputed estimates. Therefore, in the following chapter, we refrain from presenting non-imputed in addition to imputed results for reasons of clarity.

[9] Information is missing on ISCED for 4.5 percent of the mothers and 22.9 percent of the fathers, on ISEI for 8.1 percent of the mothers and 26 percent of the fathers, on German citizenship status for 4 percent of the mothers and 22.6 percent of the fathers, and on monthly net household income for 12.2 percent of the households.

[10] As predictors, we use all imputed variables, information on the years of birth, migration background, cognitive test scores, monthly gross income, weekly working hours of mothers and fathers, and also information on household structure and community size.

[11] The values presented in the descriptions are calculated as the mean of the imputations in the case of continuous variables and as the mode of the imputations in the case of categorical variables.

4. The distributions of social structural indicators in the TwinLife and Microcensus samples

Table 6 shows the shares of households located in eastern compared to western German federal states in the TwinLife and Microcensus samples. Overall, approximately 15 percent of the households in all samples are in eastern German federal states. In line with the declining East-West migration within Germany over the last two decades, the shares are slightly higher in younger cohorts and lower in the older cohorts. Consequently, there are no indications of different regional distributions between (proxy-)twin and all multiple-child households or between the TwinLife and Microcensus samples.

Table 6: Households by region

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Eastern Germany (in %)	18.0	17.6	15.5	12.4	16.3
Western Germany (in %)	82.0	82.4	84.5	87.6	83.7
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Eastern Germany (in %)	17.5	16.1	11.7	9.5	13.8
Western Germany (in %)	82.5	83.9	88.3	90.5	86.2
Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Eastern Germany (in %)	19.5	16.9	15.8	9.7	16.3
Western Germany (in %)	80.5	83.1	84.2	90.3	83.7

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Table 7 reports shares of households by community size in the samples compared. Around two-thirds of the TwinLife households are located in communities with 50,000 or more inhabitants while this share is approximately 40 percent in the Microcensus samples. This difference is mainly attributable to the oversampling of urban communities in TwinLife which was implemented to obtain the necessary coverage of the twin family target population (see Section 2). However, if we exclude the oversampled urban population from the analysis, the distributions of the TwinLife and Microcensus samples are roughly comparable. The group of TwinLife households in communities with 500,000 or more inhabitants is around four percentage points larger than the Microcensus samples, and the share of households in communities with 100,000 to 499,999 inhabitants is approximately six percentage points smaller in the TwinLife sample than in the Microcensus samples. Regarding the Microcensus proxy-twin and multi-child samples, there are no considerable differences in shares of households by community size between the samples.

Table 7: Households by community size

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
5,000–19,999 (in %)	18.4	18.9	19.2	21.8	19.3
20,000–49,999 (in %)	10.5	13.3	10.9	14.1	12.0
50,000–99,999 (in %)	18.0	16.1	15.2	16.1	16.4
100,000–499,999 (in %)	21.9	21.1	22.6	20.5	21.7
> 500,000 (in %)	31.2	30.6	32.1	27.5	30.7
TwinLife, without urban sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
5,000–19,999 (in %)	38.4	37.4	39.0	37.0	38.1
20,000–49,999 (in %)	20.1	25.0	21.1	23.1	22.3
50,000–99,999 (in %)	10.1	9.3	8.8	10.5	9.5
100,000–499,999 (in %)	11.7	11.0	11.0	7.5	10.6
> 500,000 (in %)	19.7	17.3	20.1	22.0	19.5
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
5,000–19,999 (in %)	31.3	33.9	35.5	35.6	34.1
20,000–49,999 (in %)	22.1	23.8	23.5	24.1	23.4
50,000–99,999 (in %)	10.6	10.1	11.0	11.4	10.8
100,000–499,999 (in %)	17.4	16.2	15.9	15.6	16.3
> 500,000 (in %)	18.7	15.9	14.1	13.3	15.5
Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
5,000–19,999 (in %)	26.6	33.8	35.6	31.8	32.0
20,000–49,999 (in %)	22.9	22.1	20.1	23.4	21.9
50,000–99,999 (in %)	11.5	11.0	12.1	13.6	11.8
100,000–499,999 (in %)	18.2	17.9	15.1	18.8	17.3
> 500,000 (in %)	20.9	15.2	17.1	12.3	16.9

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Table 8 contrasts the shares of households with German citizenship across the samples. Overall, this share is 84.7 percent in the TwinLife sample while the corresponding shares are around 80 percent in the Microcensus samples. Moreover, the share is constant across cohorts in the TwinLife sample while it declines in the Microcensus samples from about 85 percent in the older cohorts to about 75 percent in the younger cohorts. Consequently, there are around five to ten percentage points more households with German citizenship in the TwinLife sample for cohorts 1 and 2. The shares of households with German citizenship in the Microcensus proxy-twin and multiple-child samples are similar.

Table 8: Households by German citizenship

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
German citizenship (in %)	85.5	85.0	84.1	83.7	84.7
No German citizenship (in %)	14.5	15.0	15.9	16.3	15.3
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
German citizenship (in %)	74.1	77.9	82.3	81.7	79.0
No German citizenship (in %)	25.9	22.1	17.7	18.3	21.1
Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
German citizenship (in %)	75.8	76.9	85.9	85.7	80.5
No German citizenship (in %)	24.2	23.1	14.1	14.3	19.5

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Table 9 describes the distributions of highest educational level in the households for the TwinLife parent-twin and the Microcensus multiple-child samples based on the ISCED.[12] We observe that the TwinLife sample covers the full distribution of educational levels. The lower tail (ISCED 1 and 2) encompasses around 5 percent of the cases. The results also indicate that there are more households with a university education (ISCED 5a and 6) and fewer with medium or low education (ISCED 1 to 3) in TwinLife than the Microcensus multiple-child sample, particularly in the younger cohorts.

Table 9: Highest educational level (based on ISCED) in household

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
1 (in %)	0.6	0.8	1.1	1.5	0.9
2 (in %)	4.9	4.0	3.6	5.2	4.3
3a, b, c (in %)	25.2	27.0	33.5	37.9	30.0
4a, b (in %)	9.8	7.2	8.0	8.1	8.3
5b (in %)	10.8	13.2	12.8	16.5	12.9
5a (in %)	41.4	42.7	36.0	27.9	38.2
6 (in %)	7.4	5.1	5.0	3.0	5.4
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
1 (in %)	3.4	3.7	3.6	4.2	3.7
2 (in %)	9.9	8.6	7.9	10.2	9.1
3a, b, c (in %)	34.1	36.8	39.3	43.5	38.3
4a, b (in %)	9.4	9.2	9.2	7.8	8.9
5b (in %)	11.5	12.7	14.7	14.0	13.2
5a (in %)	27.8	25.5	22.7	18.4	23.7
6 (in %)	4.0	3.4	2.6	2.0	3.0

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

[12] Cell-specific case numbers in the Microcensus proxy-twin sample (see Table 5) are too small to present detailed distributions for highest ISCED and ISEI in households or for net equivalent monthly household income.

To analyze potential reasons for these differences, Table 10 shows the shares of university educated households compared to all other households. Overall, the share of university educated households is 43.5 percent in the TwinLife sample while it is around 30 percent in the Microcensus samples. In cohort 4 the difference in the share of university educated households is around ten percentage points between the samples while it is around 15 percentage points in cohorts 2 and 3 and approximately 20 percentage points in cohort 1. The latter difference declines to around 15 percentage points if we restrict the samples to households with German citizenship to account for the higher shares of these households in TwinLife (see Table 8).[13] The shares of households with a university education in the Microcensus proxy-twin and multiple-child samples are approximately the same.

Table 10: Households by university education (based on ISCED)

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	51.3	52.3	59.1	69.2	46.5
ISCED 5a or 6 (in %)	48.7	47.7	40.9	30.8	43.5
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	68.2	71.1	74.7	79.6	73.3
ISCED 5a or 6 (in %)	31.8	28.9	25.3	20.4	26.7
Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	43.6	74.4	76.4	77.8	72.4
ISCED 5a or 6 (in %)	36.4	25.6	23.6	22.2	27.6
TwinLife, without urban sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	52.9	57.1	60.3	69.5	58.9
ISCED 5a or 6 (in %)	47.1	42.9	39.7	30.5	41.1
TwinLife, only German citizenship	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	48.7	48.8	55.8	66.1	53.3
ISCED 5a or 6 (in %)	51.3	51.2	44.2	33.9	46.7
Microcensus multiple-child sample, only German citizenship	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISCED 1, 2, 3, 4, or 5b (in %)	65.1	65.8	72.8	77.3	70.8
ISCED 5a or 6 (in %)	34.9	34.2	27.2	22.7	29.2

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Table 11 presents the distributions of highest occupational status in the households for the TwinLife and Microcensus samples based on the ISEI. The TwinLife sample encompasses the full occupational distribution. Approximately 5 percent of the households are non-employed, around 15 percent have a low ISEI (12–29 points), approximately 50 percent have a medium ISEI (30–69 points), and around 30 percent have a high ISEI (70–89 points). Overall, the mean of the highest household ISEI is 52 points in TwinLife while it is around 46 points in the Microcensus samples. Differentiated by cohort, the means are approximately

[13] Excluding the urban sample to address the oversampling of households like this in TwinLife (see Section 2) does not change the differences observed between TwinLife and the Microcensus samples to a relevant degree, however.

the same in cohort 4 while there are mean differences of around 3, 6, and 9 points in cohorts 3, 2, and 1, respectively; similarly, the difference in the share of households with a high ISEI changes from zero in cohort 4 to around 9 percentage points in cohort 1. Restricting the samples to households with German citizenship or excluding the TwinLife urban sample to account for discrepancies in these aspects cannot explain the differences in the younger cohorts. The highest household ISEI mean values are similar in the Microcensus proxy-twin and multiple-child samples.

Table 11: Highest occupational status (based on ISEI) in households

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISEI score (mean)	55.6	53.3	51.4	46.0	52.3
ISEI score (in %):					
12–29	13.9	15.6	17.8	19.6	16.4
30–49	18.1	17.9	19.2	23.5	19.2
50– 69	25.5	27.0	28.8	28.3	27.3
70– 89	38.3	34.4	29.1	20.5	31.9
Non-employed	4.3	5.0	5.2	8.1	5.3

Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISEI score (mean)	45.8	46.6	47.1	43.9	45.9
ISEI score (in %):					
12– 29	22.1	22.0	22.0	26.3	23.0
30– 49	18.6	19.3	21.6	22.6	20.5
50– 69	18.6	22.1	24.1	23.1	22.0
70– 89	29.1	27.3	25.4	20.5	25.7
Non-employed	11.6	9.3	6.7	7.5	8.9

Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
ISEI score (mean)	48.7	44.3	45.7	46.3	46.3

Mean ISEI score in subsamples	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
TwinLife, without urban sample	54.6	52.9	51.1	47.0	51.8
TwinLife, only German citizenship	58.4	57.0	55.1	49.3	55.7
Microcensus multiple-child sample, only German citizenship	50.0	50.5	50.1	47.3	49.5

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Table 12 reports the distributions of monthly net equivalent household incomes for the TwinLife and Microcensus samples. The TwinLife sample also covers the full income distribution. Across all cohorts, around 20 percent of the households have an adjusted income of less than €1,000 per month, around 53 percent have between €1,000 and €2,000 per month, around 20 percent have between €2,000 and €3,000 per month, and

approximately 7 percent have more than €3,000 per month. These shares are roughly comparable to the Microcensus samples where the share of households with less than €1,000 per month is slightly higher and the share with between €2,000 and €3,000 per month is marginally lower. Overall, the median monthly net equivalent household income in the TwinLife sample is €1,528 while it is around €150 less in the Microcensus samples. Differentiated by cohort, these differences between monthly median incomes are approximately €100 in cohorts 3 and 4 and around €200 in cohorts 1 and 2. Similar to the distributions of highest occupational status in the households (see Table 11), restricting the TwinLife and Microcensus samples to households with German citizenship or excluding the TwinLife urban sample does not account for the differences observed. The household income medians are similar in the Microcensus proxy-twin and multiple-child samples.

Table 12: Monthly net equivalent household income

TwinLife parent-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Household income in € (median)	1,618	1,574	1,403	1,610	1,528
Household income in € (in %):					
< €1,000	18.3	18.6	25.5	17.4	20.3
€1,000 to < €2,000	50.9	52.4	54.4	51.9	52.5
€2,000 to < €3,000	23.3	20.9	15.1	23.8	20.3
≥ €3,000	7.5	8.2	5.0	6.8	6.9
Microcensus multiple-child sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Household income in € (median)	1,324	1,373	1,376	1,537	1,375
Household income in € (in %):					
< €1,000	26.6	23.9	25.8	19.8	24.1
€1,000 to < €2,000	50.8	50.1	55.0	56.6	53.0
€2,000 to < €3,000	16.2	18.3	14.2	18.7	16.8
≥ €3,000	6.4	7.7	4.9	5.0	6.0
Microcensus proxy-twin sample	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Household income in € (median)	1,433	1,285	1,303	1,537	1,373
Median income in € in subsamples	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
TwinLife, without urban sample	1,574	1,549	1,405	1,612	1,520
TwinLife, only German citizenship	1,670	1,670	1,499	1,733	1,664
Microcensus multiple-child sample, only German citizenship	1,469	1,478	1,433	1,601	1,495

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

The monthly net equivalent household income distributions shown in Table 12 are further differentiated by the highest educational status in the households based on ISCED in order to assess the TwinLife studies potential for multidimensional analysis of social structural

(dis-)advantage (Table 13). The parts of this two-dimensional social structural distribution covered in the Microcensus multiple-child sample are also represented in the TwinLife parent-twin sample indicating that the latter can be used for related multidimensional analysis. Further, the distributions are also roughly comparable; the shares of households with a university education (ISCED 5a or 6) and an adjusted income of between €1,000 and €3,000 are slightly larger in the TwinLife parent-twin sample, while those with medium education (ISCED 3) and an income of between €1,000 and €2,000 and also those with low education (ISCED 1 or 2) and an income of less than €1,000 are slightly lower.

Table 13: Highest educational level (using ISCED) by net equivalent income in households

Monthly net equivalent household income	**TwinLife parent-twin sample**				**Microcensus multiple-child sample**			
	Highest educational level (based on ISCED) in household in % (cell percentages)							
Cohort 1	1, 2	3a, b, c	4a, b, 5b	5a, 6	1, 2	3a, b, c	4a, b, 5b	5a, 6
< €1,000	3.8	7.8	2.2	4.6	10.0	11.6	2.8	2.2
€1,000 to < €2,000	1.7	16.0	14.1	19.1	3.1	20.9	13.8	13.0
€2,000 to < €3,000	0	1.1	4.0	18.2	0.1[a]	1.8[a]	3.4	11.2
≥ €3,000	0	0.2	0.4	6.9			0.6	5.4
Cohort 2	1, 2	3a, b, c	4a, b, 5b	5a, 6	1, 2	3a, b, c	4a, b, 5b	5a, 6
< €1,000	2.6	8.6	3.3	4.1	8.3	10.7	3.1	1.7
€1,000 to < €2,000	2.2	16.4	13.7	20.0	4.1	23.1	13.4	9.6
€2,000 to < €3,000	0	1.7	2.8	16.4	0.2[a]	2.9[a]	4.2	11.6
≥ €3,000	0	0.3	0.7	7.2			1.1	6.0
Cohort 3	1, 2	3a, b, c	4a, b, 5b	5a, 6	1, 2	3a, b, c	4a, b, 5b	5a, 6
< 1,000	3.7	12.5	3.7	5.5	7.9	12.7	3.2	2.0
€1,000 to < €2,000	1.0	18.9	14.4	20.1	3.4	24.2	16.2	11.2
€2,000 to < €3,000	0	1.7	2.3	11.1	0.2[a]	2.5[a]	3.4	8.5
≥ €3,000	0	0.3	0.4	4.3			0.9	3.6
Cohort 4	1, 2	3a, b, c	4a, b, 5b	5a, 6	1, 2	3a, b, c	4a, b, 5b	5a, 6
< €1,000	2.8	7.4	3.0	4.3	7.3	8.9	2.3	1.2
€1,000 to < €2,000	3.1	21.8	15.0	12.0	6.8	28.0	13.1	8.8
€2,000 to < €3,000	0.7	7.2	6.3	9.6	0.5[a]	6.8[a]	4.9	7.2
≥ €3,000	0	1.5	0.4	5.0			1.1	3.2

[a] Due to size, the cells €2,000 to < €3,000 and ≥ €3,000 are aggregated for ISCED 1, 2 and 3a, b, c.
Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

Finally, Table 14 displays the mean values of maternal age at the birth of the twins or the anchor child for the TwinLife and Microcensus samples. This value is approximately 31 years in all samples. It increases from around 30 years in cohort 4 to approximately 32 years in cohort 1 which is accompanied by an increase in the share of mothers aged 35 or older at childbirth (from around 15 to 30 percent). The changes are less pronounced in the Microcensus multiple-child sample. Overall, there are no indications of relevant differences

in maternal age at childbirth which could be responsible for the social structural differences observed between the samples.

Table 14: Maternal age at birth of twins or anchor child

TwinLife parent-twin	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Maternal age (mean in years)	32.2	31.9	30.7	29.8	31.3
Maternal age (age groups in %):					
Age < 25	6.7	7.4	9.9	12.4	8.7
Age 25–34	60.8	61.6	70.8	73.9	65.8
Age ≥ 35	32.5	31.0	19.3	13.7	25.6
Microcensus multiple-child	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Maternal age (mean in years)	31.1	31.4	30.2	30.5	30.8
Maternal age (age groups in %):					
Age < 25	8.6	5.8	5.4	12.7	8.0
Age 25–34	59.1	59.8	66.0	65.0	62.4
Age ≥ 35	32.3	34.4	28.6	22.3	29.6
Microcensus proxy-twin	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Maternal age (mean in years)	32.4	31.1	30.0	29.6	30.9

Sources: TwinLife, wave 1 (doi: 10.4232/1.12665) and RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations

5. Discussion and conclusion

The present paper contrasted the distributions of key socio structural variables in the first wave of the TwinLife study with those in covered by the Microcensus using a proxy-twin and a multiple-child household sample. First, in terms of the differences between the two Microcensus samples, the analyses show that (proxy-)twin and multiple-child households in Germany have comparable distributions regarding many socio-demographic indicators such as region of residence, community size, citizenship status, educational and occupational status, and also income – but also with respect to maternal age at birth of anchor children. The only difference we found between (proxy-)twin and multiple-child households is the higher prevalence of two-children households in the latter; a difference that can be explained by parents often planning to have two children (Ruckdeschel, 2007). The absence of relevant differences in the distributions of social background indicators is important for the TwinLife study, since it would otherwise be impossible to capture the full range of social structural variation using a twin-based sampling design. Moreover, this is also beneficial for comparative studies using twin registry data in other countries and for generalizing inferences of social structural influences based on twin studies to the corresponding population at large. If different outcomes in twin and other multiple-child families are not a

consequence of different social structural distributions, these varying outcomes indicate different inequality-generating processes for twin and non-twin families. Therefore, if one has reason to assume that there are no different inequality-generating processes for an outcome, the results of twin studies can be generalized to other multiple-child families.

Further, our analyses demonstrate that the probability-based sampling design of TwinLife was successful in counteracting the overrepresentation of monozygotic twins typical of other twin samples based on self-recruitment (Lykken et al., 1987). The regional distributions of households in the TwinLife and Microcensus samples are similar while there are larger shares of urban households in TwinLife due to the oversampling of populous communities that was necessary to achieve the target sample sizes. The share of households with migration background (indicated by no German citizenship) is approximately five to ten percentage points smaller in the younger cohorts of the TwinLife compared to the Microcensus samples. This is a typical difference that can be addressed using specialized sampling strategies (Brücker et al., 2014; Schupp & Wagner, 1995). TwinLife did not have funding for an additional migration sample, however. Looking at the social structural distributions in greater detail, we found around 15 percentage points more university educated households in the TwinLife sample, particularly in the younger cohorts. Regarding the highest occupational status based on the ISEI, the analyses showed that mean values were around 8 percentage points higher in the TwinLife sample for the younger cohorts, while the means for the older cohorts are similar to the Microcensus samples. For the monthly net equivalent household income, we found that median values were around €200 higher for the younger TwinLife cohorts and that the corresponding values were around €100 higher in the older cohorts. Additional analyses showed that the oversampling of urban communities in TwinLife cannot account for these differences. The smaller share of households with no German citizenship in TwinLife can explain a quarter of the difference in the shares of university educated households between the samples for cohort 1.

These results indicate that participation in TwinLife was, to some degree, selective with respect to parental education, particularly in the younger cohorts.[14] Such selectivity is largely unproblematic if analyses are multidimensional, i.e., if they are conditional on or differentiated by several socio-demographic indicators. In addition, to assess the robustness of results, we suggest conducting additional analyses using a cohort-specific weighting scheme based on the distribution of highest education by German citizenship in the Microcensus multiple-child sample (see Annex I). Using such a weighting scheme for TwinLife is justified by the social structural similarity between (proxy-)twin and multiple-child households in Germany found in this study. Overall, for the core social structural

[14] A potential reason for this selectivity is the demanding questionnaire program for the first wave of TwinLife, particularly for the children aged around 5 at the time of the survey in cohort 1. To ensure panel stability, plans had already been made to reduce the survey program for future TwinLife waves prior to the first wave but the program will be further reduced given the results of this study. The expected overall workload on the family level for the second wave was reduced from around 180 minutes in the first wave to around 120 minutes.

indicators – educational and occupational status and income, the results show that the TwinLife sample covers the full distributions including the lower and upper bounds. TwinLife can therefore be used for multidimensional analysis of social inequality and for studying related differences in gene expression – gene-environment interplay – based on the Extended Twin Family Design.

References

Brescianini, S., Fagnani, C., Toccaceli, V., Medda, E., Nisticò, L., D'Ippolito, C.,. . . Stazi, M. A. (2013). An Update on the Italian Twin Register: Advances in Cohort Recruitment, Project Building and Network Development. *Twin Research and Human Genetics*, *16*(1), 190–196.

Brücker, H., Kroh, M., Bartsch, S., Goebel, J., Kühne, S., Liebau, E.,. . . Schupp, J. (2014). *The new IAB-SOEP migration sample: An introduction into the methodology and the contents*. *SOEP survey papers Series C, Data documentations: Vol. 216*. Berlin: DIW. Retrieved from http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0216.pdf

Burt, S. A., & Klump, K. L. (2013). The Michigan State University Twin Registry (MSUTR): An Update. *Twin Research and Human Genetics*, *16*(1), 344–350.

Busjahn, A. (2013). The Berlin Twin Registry for Health Research. *Twin Research and Human Genetics*, *16*(1), 163–166.

Destatis. (2013). *Natürliche Bevölkerungsbewegung*. Wiesbaden.

Destatis. (2014a). *Bevölkerung und Erwerbstätigkeit – Haushalte und Familien: Ergebnisse des Mikrozensus 2013*. Wiesbaden.

Destatis. (2014b). *Mikrozensus 2013: Qualitätsbericht*. Wiesbaden.

Diewald, M., Riemann, R., Spinath, F. M., Gottschling, J., Hahn, E., Kornadt, A. E.,. . . Peters, A.-L. (2016). TwinLife: GESIS Data Archive. ZA6701 (doi:10.4232/1.12665).

Ganzeboom, H. B. G., Graaf, P. M. d., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social science research : a quarterly journal of social science methodology and quantitative research*.

Hahn, E., Gottschling, J., Bleidorn, W., Kandler, C., Spengler, M., Kornadt, A. E.,. . . Spinath, F. M. (2016). What Drives the Development of Social Inequality Over the Life Course?: The German TwinLife Study. *Twin Research and Human Genetics*, *19*(6), 659–672.

Hahn, E., Gottschling, J., & Spinath, F. M. (2013). Current Twin Studies in Germany: Report on CoSMoS, SOEP, and ChronoS. *Twin Research and Human Genetics*, *16*(1), 173–178.

Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twin Research and Human Genetics*, *16*(1), 117–125.

Hoekstra, C., Zhao, Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I., & Montgomery, G. W. (2008). Dizygotic Twinning. *Human Reproduction Update*, *14*(1), 37–47.

Hopper, J. L., Foley, D. L., White, P. A., & Pollaers, V. (2013). Australian Twin Registry: 30 years of progress. *Twin Research and Human Genetics*, *16*(1), 34–42.

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., & Feldhaus, M. (2011). Panel analysis of intimate relationships and family dynamics (pairfam): Conceptual framework and design. *Journal of Family Research*, *23*(1), 77–101.

Kandler, C., Riemann, R., Spinath, F. M., Bleidorn, W., Thiel, W., & Angleitner, A. (2013). The Bielefeld Longitudinal Study of Adult Twins (BiLSAT). *Twin Research and Human Genetics*, *16*(1), 167–172.

Kaprio, J. (2013). The Finnish Twin Cohort Study: An Update. *Twin Research and Human Genetics*, *16*(1), 157–162.

Lengerer, A., Bohr, J., & Janßen, A. (2005). *Households, families and ways of life in the microcensus: concepts and stylizations* (ZUMA-Arbeitsbericht). Mannheim.

Lengerer, A., Janßen, A., & Bohr, J. (2007). Possibilities for family research with the German Microcensus. *Journal of Family Research*, *19*(2), 186–209.

Lykken, D. T., McGue, M., & Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behavior Genetics*, *17*(4), 343–362.

Magnusson, P. K. E., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H.,. . . Lichtenstein, P. (2013). The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments. *Twin Research and Human Genetics*, *16*(1).

Nilsen, T. S., Peggy Knudsen, G., Gervin, K., Brandt, I., Røysamb, E., Tambs, K.,. . . Harris, J. R. (2013). The Norwegian Twin Registry from a Public Health Perspective: A Research Update. *Twin Research and Human Genetics*, *16*(1), 285–295.

OECD. (2011). What Are Equivalence Scales?: OECD Project on Income Distribution and Poverty. Retrieved from http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf, 25 April 2017

Ruckdeschel, K. (2007). Fertility Intentions of Childless Persons. *Journal of Family Research*, *19*(2), 210–230.

Schneider, S. L. (2008). Applying the ISCED-97 to the German Educational Qualifications. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries* (pp. 76–102). Mannheim.

Schupp, J., & Wagner, G. G. (1995). Die Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung*, *64*(1), 16–25.

Skytthe, A., Christiansen, L., Ohm Kyvik, K., Bødker, F. L., Hvidberg, L., Petersen, I.,. . . Christensen, K. (2013). The Danish Twin Registry: Linking Surveys, National Registers, and Biological Information. *Twin Research and Human Genetics*, *16*(1), 104–111.

TNS Infratest. (2016). *TwinLife: First Funding Period: Methodology Report*. Munich.

van Beijsterveldt, C., Groen-Blokhuis, M., Jan Hottenga, J., Franic, S., Hudziak, J. J., Lamb, D.,. . . Boomsma, D. I. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal Twin and Family Studies in Over 70,000 Children. *Twin Research and Human Genetics*, *16*(1), 252–267.

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully Conditional Specification in Multivariate Imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements. *Schmollers Jahrbuch*, *127*(1), 139–170.

Annex I

Instructions for constructing a weighting scheme matching the cohort specific highest ISCED by German citizenship distribution of parents on the household level for TwinLife analysis samples with the Microcensus multiple-child sample:

The aim of the proposed weighting scheme is to address the selectivity of the TwinLife sample regarding parental education and German citizenship status, particularly in the younger cohorts. We advise using it as a robustness check, i.e., to assess discrepancies in the results between analyses conducted with and without the weighting scheme. Comparable results in both analyses indicate that the conclusions drawn are not influenced by the selectivity.

We construct weights specific to each of the four TwinLife cohorts. First, for a cohort-specific weighting scheme like this, we need to calculate the shares of observations in the TwinLife analysis sample used by highest ISCED and German citizenship of the parents on the household level for each cohort using the categorization presented in Table 15. This share is given by the number of observations in a specific highest ISCED by German citizenship cell (J) for a specific cohort divided by the total number of observations in the analysis sample (N) for a specific cohort. Second, we need to divide the cell-specific correction factors (C) presented in Table 15 by the cohort-specific shares calculated for the analysis sample. The correction factors in Table 15 are based on the cohort-specific shares of observations in the Microcensus multiple-child sample by highest ISCED and German citizenship. Hence, the cohort-specific weights (W) assigned to each observation in the analysis sample depending on highest parental ISCED and parental German citizenship on the household level are given conducting the following calculation:

$$W = C/(J/N) = C \times N/J \quad (1)$$

The resulting weighted analysis sample has the same number of observations as the sample without weights in each cohort but its cohort-specific highest ISCED by German citizenship distribution matches the one in the Microcensus multiple-child sample.

If the distributions of parental background indicators for a specific analysis sample based on TwinLife do not differ significantly between the household- and the family-level of aggregation this weighting scheme can also be implemented on the family level.

If you use this weighting scheme for your research, please cite this paper.

Table 15: Factors for a selectivity correction weighting scheme based on the Microcensus[a]

	Highest educational level (using ISCED) in household			
Cohort 1	1, 2	3a, b, c	4a, b, 5b	5a, 6
German citizenship	0.05735661	0.24804655	0.17722361	0.25835412
No German citizenship	0.07547797	0.09293433	0.03142145	0.05918537
Cohort 2	1, 2	3a, b, c	4a, b, 5b	5a, 6
German citizenship	0.05561700	0.28282509	0.18991942	0.25075051
No German citizenship	0.06794122	0.08547954	0.02938853	0.03807869
Cohort 3	1, 2	3a, b, c	4a, b, 5b	5a, 6
German citizenship	0.05466035	0.32669826	0.21800948	0.22353871
No German citizenship	0.06082149	0.06650869	0.02085308	0.02890995
Cohort 4	1, 2	3a, b, c	4a, b, 5b	5a, 6
German citizenship	0.06781795	0.36643281	0.19769743	0.18546501
No German citizenship	0.07609282	0.06817773	0.01978773	0.01852851

[a] The correction factors in the table are not the weights. Please read Annex I for instructions on how to construct weights using these correction factors.

Source: RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2013, own calculations