Schmollers Jahrbuch 131 (2011), 625 – 634
Duncker & Humblot, Berlin

European Data Watch

This section offers descriptions as well as discussions of data sources that are of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@leuphana.de⟩. Past "European Data Watch" articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: http://www. ratswd.de.

ALWA – New Life Course Data for Germany

By Corinna Kleinert, Britta Matthes, Manfred Antoni, Katrin Drasch, Michael Ruland, and Annette Trahms

1. Introduction

In 2007 and 2008 the Institute for Employment Research (IAB) carried out a new retrospective life course study named ALWA.[1] This survey was designed to analyse the interplay between educational processes and basic cognitive skills in various phases of employment trajectories (for details, see Kleinert et al., 2008). Since 2010, the study's life course data is available at the Research Data Centre (FDZ) of the Federal Employment Agency.

[1] The acronym ALWA is derived from the study's German name 'Arbeiten und Lernen im Wandel' which means 'Working and Learning in a Changing World'.

The ALWA data provides seamless longitudinal data on educational and employment histories, places of living, partners and children. Thus it allows examining various dynamic processes in life courses, e.g. school-to-work transitions, employment mobility or family formation. Furthermore, different aspects of life course data can be combined to examine the interplay between different spheres of life. Finally, the change of life courses in different cohorts can be studied, because ALWA includes data on persons born within a range of more than 30 years. Possible fields of research are, for example, long-term returns of educational activities in youth, decisions to participate in further education, the embeddedness of education in employment histories and other life domains, spatial mobility, or the interplay between family and employment careers. Beyond that, the dataset allows answering many other research questions in sociology, economics, demography and related fields.

This overview summarizes the basic information necessary for analysing the ALWA data and complements existing texts (Kleinert et al., 2008; Antoni et al., 2010; Matthes/Trahms, 2010). First, it provides information on the design of the study, the sample, survey modes and the questionnaire. Second, different aspects of survey implementation, such as field work, editing, coding and weighting, are described. Third, the article explains the data structure of the scientific use file and how to access the data. The article concludes with a short outlook on the next project steps.

2. Survey Design

2.1 Sampling Process

The target population of the IAB ALWA study consists of persons living in private households in Germany who were born between 1956 and 1988. Thus, it includes all German inhabitants in these birth cohorts irrespective of their first language, their nationality or their employment status.

The sample was drawn in two stages: first, a representative sample of municipalities (Gemeinden) was selected from municipality data made available by the Federal and State (Länder) Statistical Offices. The municipality sample (primary sampling units or PSUs) was drawn using the probability proportional to size technique, i.e. the selection probability of the municipality was proportional to the extrapolated resident population in the target group. Thus, ALWA is a self-weighting sample consisting of 281 sample points, which represent 250 municipalities.[2]

[2] Owing to the comparatively large number of PSUs, the expected design effect is very small due to cluster sampling, i.e. intra-cluster correlation coefficients (ICC) are assumed to be small on average.

At the second stage, the addresses of the target persons (secondary sampling units or SSUs) were drawn from the registers of the residents' registration offices in the 250 selected municipalities using a systematic random sampling procedure (interval sampling). Taking a random address as starting point, the subsequent addresses were systematically sampled at set intervals. Only registered residents of the birth cohorts from 1956 to 1988 with a primary residence in the selected municipalities on the sampling date were included in the selection.

Finally, for the entire address sample telephone numbers were searched using up-to-date and older editions of digital telephone directories. In this way it was possible to identify telephone numbers for 52.6 percent of the addresses (see Table 2 for details).

2.2 Study Design and Methods

The main survey consisted of two interviews. In the first part, data on individuals' life courses and demographic information was collected in computer assisted telephone interviews (CATI). After this step, literacy and numeracy tests were conducted with those respondents who had agreed to participate in these tests by phone. These tests were administered by paper and pencil (PAPI) interviews in the respondents' homes. As the data on cognitive skills is not included in the scientific use file yet, this overview concentrates on describing the CATI survey.

We chose the CATI mode for collecting life course data due to cost reasons, feasibility, but also because computer-assisted interviews have proven to be particularly suitable for recording complex life histories. The questionnaire can be tailored flexibly to adapt to the large heterogeneity in life courses, filters are routed automatically and data checks are implemented directly into the survey. The interviewers can refer to information provided earlier during the interview in order to assist the respondents' memory (Matthes et al., 2007). These tools minimise the work involved in editing complex life course data.

For Turkish and Russian speaking target persons who did not speak German well enough to answer the survey shorter versions of the German questionnaire were programmed in these two languages and conducted by native speaking interviewers. These instruments contained the same cross-sectional questions as the German language version, but only rough cross-sectional indicators of their previous life courses.

Apart from the main survey, the ALWA study consisted of three additional components to control data quality: in 2006, a pilot study was conducted to test feasibility, interview length, and problems in questionnaire design. In order to get evidence for the sample's selectivity bias the main study was accompanied by two short surveys with basic personal information. 375 persons for whom

no telephone number could be found completed a written questionnaire. A telephone interview was conducted with nearly 1,800 interview non-respondents (for results, see infas 2011).

2.3 Questionnaire Design

The major part of the questionnaire consisted of a retrospective longitudinal section in which the entire previous educational and working history as well as the residence and household history of the interviewee were collected. The question wording and the sequence of the instruments in these areas are related to existing instruments and experiences from life course studies conducted at the Max-Planck-Institute for Human Development in Berlin (MPIB).

In one important aspect the ALWA questionnaire design differs from the MPIB life course studies, however: in the later studies, the educational and employment history has been surveyed as a quasi-continuous life course with employment following training, interrupted only by spells of unemployment or non-employment. Due to the increasing complexity of educational and employment histories, in ALWA all the different aspects of the educational and working life (schooling, vocational preparation, training, military service, employment, unemployment, and inactivity) were designed as independent longitudinal modules, i.e. the respondents had to go through their lives again in each of these modules.

The problem of anchoring that arises in this modularized form of data collection was addressed by two strategies: in every module the interviewers were given tools that enabled the respondents to anchor their information about a certain episode with other events they had already reported. After data collection, the time consistency of dating was checked for overall coherence in cooperation with the respondents and corrected if necessary in special data revision modules, one for the residence history and a second one for the education and employment history (Drasch / Matthes, 2012).

In sum, the questionnaire consists of six blocks of questions, split into 14 modules based on their content (see table 1). The introductory cross-sectional module contains basic socio-demographic information such as sex, date of birth, country of birth, date of immigration and questions on nationality.

For the residential history all the municipalities in which the respondents had lived since their birth were recorded. In addition to changes in the place of residence that were due to relocation, also secondary residences were collected. After the residence history, the collected episode data was checked in the first data revision module.

In the first part of the education, training and employment history all educational activities were recorded. In the schooling module, data on all periods of

general schooling, whether ending successfully with a school-leaving certificate or not, were collected. Afterwards, the respondents were asked if they had taken part in vocational preparation schemes and, if so, from when and until when they did so. Subsequently, all initial and further vocational training and academic courses that the interviewees had ever begun were recorded, e.g. apprenticeships, college and university studies, but also postgraduate studies and formal further training episodes leading to a certificate of any kind.

To complete our information on the educational histories of our respondents, also shorter and less salient further training activities were collected. Here, ALWA relied on context related memory strategies (Dürnberger et al., 2011): questions about undertaking courses were integrated into the various longitudinal modules. Thus, respondents were asked if they participated in courses within every reported episode of e.g. military service, employment, unemployment, parental leave or inactivity.

Table 1

Main contents of the ALWA questionnaire

Question block	Module	Time reference
Cross section 1	Basic personal information, immigration, nationality	Cross-sectional
Residential history	Residential history	Longitudinal
	Residential history data revision	
Education, training and employment history	Schooling	Longitudinal
	Vocational preparation	Longitudinal
	Vocational training / studying	Longitudinal
	Military / civilian service	Longitudinal
	Periods of employment	Longitudinal
	Periods of unemployment	Longitudinal
Family history	Partners in household	Longitudinal
	Children	Longitudinal
	Periods of parental leave	Longitudinal
Data revision	Gap activities	Longitudinal
Cross section 2	Additional personal information, parental background, household	Cross-sectional

In the second part of the education, training and employment history, all periods of military, alternative civilian and national service, and voluntary social, ecological and European service were recorded. In the third part of this module, information on all episodes of employment and unemployment was collected. First, all 'regular' jobs were recorded. Follow-up questions were asked about training-related employment, e.g. as student worker, trainee or intern, about

marginal employment and about activities performed part-time besides another job. After the employment history, times of unemployment were recorded, irrespective of whether the respondents were officially registered as unemployed or not.

The next part of the questionnaire contains information on partners, children, and parental leave. Partner information was limited to all partners who had ever lived in the households of the ALWA respondents. Children include the respondent's own, foster and adopted children as well as other children who had ever lived in the respondents' households. In addition, the respondents were asked whether they currently are or had been entitled to parental leave and, if so, whether this was taken.

After the life course data was collected, it was automatically merged and displayed as biographic calendar. In the second data revision module, the interviewers checked it together with the respondents for missing dates, gaps, and internal consistency (for details, see Drasch / Matthes, 2012).

The survey's final cross-sectional module dealt with specific information about the respondents themselves, such as the importance of different life domains, leisure activities, cultural capital, (foreign) language skills, languages spoken in the household, and religion. Information on informal learning was collected for a two-year period before the interview. Additionally, information on the respondents' family background and on their household was collected. Additional cross-sectional information was collected directly in some of the longitudinal modules, for example subjective indicators of school performance and current reading and maths skills, desired occupations after school, or information on future career plans and observed occupational success.

3. Implementation

3.1 Field Phase

The main survey started in August 2007 and was conducted by infas (Institut für angewandte Sozialforschung, Bonn). All target persons for whom a telephone number had been obtained received a personal letter in advance, which also contained information on the selection of addresses, data protection legislation and the voluntary nature of participation. As an incentive for participation in the CATI interview, a lottery with several prizes (notebooks, travel vouchers, iPods) was announced.

The appointed CATI interviewers were experienced in conducting studies with retrospective longitudinal designs. They underwent extensive interviewer training and completed several practical exercises to familiarize themselves with the instrument. To ensure standardized information on the project and its implementation an interviewer manual was developed. During the whole field

phase the interviewers were closely supervised and received a continuous follow-up training.

The response rate of the CATI interviews was 45.9 percent with reference to the sample used in the field (addresses with a valid telephone number). Specially trained interviewers contacted target persons who initially had refused to participate in the CATI interview a second time in order to convince them to take part. The average duration of the interview conducted in German, including the interviewer questions, was 62 minutes. The foreign-language interviews lasted 20 minutes on average. The dropout rate during the interview was very low and amounted to around 0.6 per cent. A summary of the sample realization can be found in table 2.

Table 2

Sample realization

	abs. n.	per cent
Address sample	133,451	100.0
Telephone number sample	70,183	52.6
Field sample	22,656	100.0
Non-contact/out of range	3,820	16.8
Non-response	8,432	37.2
Realized interviews	10,404	45.9
in German	10,177	44.9
in Russian	150	0.3
in Turkish	77	0.7

3.2 Data Editing and Coding

The tool that was developed for editing the ALWA data, PatchTales, is based on the TrueTales survey instrument (Matthes et al., 2007). The program was intended to point out gaps in the data, problematic transitions between episodes and errors in the start and end dates of episodes, and to enable changing the data immediately. The ALWA data were checked for inconsistencies and, if necessary, corrected in a three-step data editing process performed by a specifically trained team of data editors. In the first step, the editors identified interviews that did not require any corrections. In the second step, easy to solve cases were edited, for example cases in which a period of training had been recorded as a period of employment. In the third step, the editors worked in pairs to deal with the remaining complex cases with, for example, gaps in the data. The focus of editing was on correct allocation of events to the corresponding modules and on temporal consistency of the life course. During the editing process all open-ended responses were checked to see whether they could be recoded using the codes available in the survey.

All information on the respondent's occupation and economic sector were collected open-ended and coded after data collection. The respondents' desired occupations, the occupations in which they were trained, and all past and ongoing occupational activities were coded into the two most common classification schemes, the Classification of Occupations of the Federal Employment Agency (KldB88) and the International Standard Classification of Occupations 1988 (ISCO88). Information about the economic sector was coded in accordance with the German Classification of Economic Activities (WZ 2008).

3.3 Selection Bias and Weighting

In order to estimate the amount of selection bias in our study and potential sources of data loss, we compared the ALWA population with the same birth cohorts in the German Mikrozensus 2007, a one-per cent sample of all households in Germany. The results in Table 3 show that ALWA has several sources of bias: similar to other surveys, lower educated persons, in particular persons with lower schooling, and migrants are underrepresented – an attribute which can be explained by the lower readiness of these groups to participate in complex and time-consuming surveys. In ALWA, this problem was possibly aggravated by the study's focus on education and employment. Besides, but less severely, older persons are overrepresented. This source of bias may be due to their higher availability – older persons are less mobile and have land line telephones more often than younger persons.

Since the ALWA sample is self-weighted, our data does not contain a design weight, but only a calibration weight that adjusts the marginal distributions of the sample to those of the German Mikrozensus 2007. Four variables were used to compute the weight factor: federal state (Bundesland), district local community size (measured in BIK categories), immigration status, and level of education (a combined variable for schooling and training / studying), interacted with sex and age groups (for details see infas, 2011).

Table 3

Basic population distributions in Mikrozensus 2007 and ALWA (in %)

	MZ 2007	ALWA CATI
45 years and above	21.6	31.5
Immigrants	16.3	10.3
School dropouts / lower secondary schooling	30.9	18.7
Without training	26.2	19.0

Sources: ALWA 2007, Mikrozensus 2007.

4. Data Structure and Access

As the survey was conducted in modular form, the data in the scientific use file is organized in one cross-sectional and eleven longitudinal datasets (Matthes / Trahms 2010). The data of the cross-sectional modules at the beginning and the end as well as the cross-sectional data in the longitudinal modules are recorded in a single cross-sectional file. Here, also a few variables with methodological information are included, e.g. a variable that marks life-course inconsistencies that could not be solved by editing, a variable containing information on language of the interview as well as the calibration weight. In this dataset, one observation represents one respondent and includes the non-German-language interviews.

Each of the eleven longitudinal datasets stands for one life course domain in the questionnaire. In these datasets, the information is structured in the long format, i.e. one observation represents one episode in the respondent's life course. For every spell a person and module specific spell number as well as a start and end date is provided. By using this information the different life course modules can be linked for multidimensional life course analysis.

In order to make the data of the ALWA study easily available to the scientific community while at the same time guaranteeing the respondents' privacy rights, the data were anonymized. Certain variables in the original data are not available in this file, e.g. address data or open-ended answers, and others have been aggregated or classified: regional information was aggregated into federal states, nationality and country of origin were aggregated into clusters of countries, and occupational information was classified to 3-digit-codes.

The ALWA scientific use file can be obtained via the Research Data Centre (FDZ) of the German Federal Employment Agency at the IAB. Further information on requesting a scientific use file is available on the homepage of the FDZ, which also provides related survey documentation, e.g. a detailed description of the dataset and frequency tables (http://fdz.iab.de/en/FDZ_Individual_Data/ALWA.aspx).

5. Life after ALWA: the National Educational Panel Study (NEPS)

From the very beginning, the ALWA survey was planned and conducted as a panel study. 93 percent of the ALWA respondents interviewed in German and 82 percent interviewed in other languages agreed to participate in subsequent interviews.

Instead of realizing additional ALWA panel waves, those respondents were integrated in the 'National Educational Panel Study' (NEPS), in which their life courses are constantly updated (for details, see Allmendinger et al. 2011; www.

bildungspanel.de). Together with additional birth cohorts ranging up to persons born in 1944, they constitute the survey population of the NEPS sub-study 'Adult Education and Life-Long Learning'. As one of eight sub-studies, each representing different educational stages in the life course, the survey sets a stronger focus than ALWA on educational activities in adulthood. In cooperation with the Social Science Research Center Berlin (WZB) and the University of Bamberg, two additional panel waves have been conducted so far, the first in 2009/2010, the second one year later. Further panel waves are planned in yearly intervals. The University of Bamberg provides access to all the NEPS data.

References

Allmendinger, J./*Kleinert,* C./*Antoni,* M./*Drasch,* K./*Janik,* F./*Leuze,* K./*Matthes,* B./*Pollak,* R./*Ruland,* M. (2011): Adult Education and Lifelong Learning, in: Blossfeld, H.-P./Roßbach, H.-G./von Maurice, J. (eds): Education as a Lifelong Process. The German National Educational Panel Study (NEPS), Zeitschrift für Erziehungswissenschaft, Sonderband 14, 283–290.

Antoni, M./*Drasch,* K./*Kleinert,* C./*Matthes,* B./*Ruland,* M./*Trahms,* A. (2010): Arbeiten und Lernen im Wandel, Teil I: Überblick über die Studie, FDZ Methodenreport 05/2010.

Drasch, K./*Matthes,* B. (2011): Improving Retrospective Life Course Data by Combining Modularized Self-Reports and Event History Calendars: Experiences From a Large Scale Survey, Quality & Quantity 46, online first, published Sept. 13, 2011.

Dürnberger, A./*Drasch,* K./*Matthes,* B. (2011): Kontextgestützte Abfrage in Retrospektiverhebungen: Ein kognitiver Pretest zu Erinnerungsprozessen bei Weiterbildungsereignissen, Methoden, Daten, Analysen 5 (1), 3–35.

infas (2011): Arbeiten und Lernen im Wandel, Teil III: Methodenbericht. FDZ Methodenreport 10/2011.

Kleinert, C./*Matthes,* B./*Jacob,* M. (2008): Die Befragung „Arbeiten und Lernen im Wandel". Theoretischer Hintergrund und Konzeption, IAB Forschungsbericht, 05/2008.

Matthes, B./*Reimer,* M./*Künster,* R. (2007): Techniken und Werkzeuge zur Unterstützung der Erinnerungsarbeit bei der computergestützten Erhebung retrospektiver Längsschnittdaten, Methoden, Daten, Analysen 1 (1), 69–92.

Matthes, B./*Trahms,* A. (2010): Arbeiten und Lernen im Wandel, Teil II: Codebuch, FDZ Datenreport 02/2010.