Education and Employment Trajectories in NEPS-ADIAB: The Survey Data of the National Educational Panel Study Linked to Administrative Data of the Institute for Employment Research

Nadine Bachbauer[1,2,*], Clara Wolf[1,2] and Daniel Fuß[1]

[1]Leibniz Institute for Educational Trajectories, Research Data Center, 96047 Bamberg, Germany and
[2]Institute for Employment Research, Research Data Center, 90478 Nuremberg, Germany

*Corresponding author. Email: nadine.bachbauer@lifbi.de

Abstract

NEPS-ADIAB is a large-scale data product from Germany that provides a unique source for the empirical analysis of life course research questions. It is jointly administered by the Institute for Employment Research (IAB) and the Leibniz Institute for Educational Trajectories. NEPS-ADIAB includes survey data from the National Educational Panel Study (NEPS) and administrative data from the IAB, combined at the individual level using a record linkage procedure. While the NEPS data cover issues such as educational trajectories and decisions, competencies and skills, learning environments, attitudes, etc., the administrative data consist of detailed employment histories dating back to 1975, plus extensive information on establishments. The linkage of both data collections results in a considerable research potential. NEPS-ADIAB is particularly suitable for analyses of the nexus between educational and employment careers, short- and long-term returns to education, transitions from school to vocational training and working life, adult education and lifelong learning, as well as discontinuities in professional careers due to parenthood or unemployment. Contextual characteristics about establishments as well as geographic indicators can additionally be taken into account. The regularly updated and expanded NEPS-ADIAB data are accessible to the entire scientific community free of charge via on-site and remote access.

Introduction

Standard survey studies in the social sciences, which are generally based on voluntary participation, are constrained in several aspects. On the one hand, even with complex panel designs, it is hardly possible to collect all required information. To achieve a high panel stability, the respondents' burden due to long interview durations, dense wave intervals, sensitive topics and demanding tests must be kept as low as possible. On the other hand, certain types of information from surveys are subject to bias. This is especially true for sensitive issues, lengthy and strenuous instruments, and retrospective questions (Tourangeau, Groves and Redline, 2010). A promising and increasingly popular way to address these challenges is to link survey data with other sources. In this context,

administrative data have become increasingly important over the last decades (Roed and Raaum, 2003; Baur, 2011). This sort of data is either generated during administrative procedures by the principle of file regularity of modern institutions or during the documentation of production processes and results in companies. Administrative data are thus based on non-reactive measurement methods that are robust to the biases mentioned above. An additional advantage is the very large number of cases.

In this Data Brief, we would like to introduce a fairly new data product from Germany that offers unique opportunities for the analysis of educational and employment histories by combining survey and administrative data. NEPS-ADIAB brings together the extensive survey plus competency testing data from the National Educational Panel Study (NEPS) and individual employment histories including establishment information from the administrative data of the Institute for Employment Research (IAB, Nuremberg). With the NEPS, managed at the Leibniz Institute for Educational Trajectories (LIfBi, Bamberg), a large and still growing data infrastructure for the empirical study of questions in educational research, sociology, economics, and beyond exists for more than 10 years now (Blossfeld and Roßbach, 2019). The administrative data from the IAB enrich the NEPS survey data especially with regard to employment and income histories. They include detailed employment periods starting in 1975, which means a significant extension of the observation period. The administrative data also provide information on participation in labour market-related measures, the receipt of social benefits, as well as contextual indicators on the respective establishments of employment.

NEPS-ADIAB is a unique data offering that is jointly administered and made available to the scientific community by the Research Data Centres (RDC) at LIfBi and the IAB. The combination of two different data sources opens up completely new research potentials and makes it possible to empirically address research questions that could not be answered on the basis of only one of them. For example, the use of NEPS-ADIAB data can close research gaps on determinants and mechanisms of career trajectories, on monetary outcomes of educational investments, on the relationship between subjective expectations and actual labour market experiences, on the persistence of social inequality in school-to-work transitions, and on the role of competence acquisition and skill endowments. Already available analyses based on NEPS-ADIAB deal with cohort changes in educational pathways and returns to education (Zimmermann, Fitzenberger and Osikominu, 2016), wage losses due to over qualification (Kracke, Reichelt and Vicari, 2017), labour market participation and atypical employment over the life cycle (Bachmann, Felder and Tamm, 2018), labour market returns to college education with vocational qualifications (Rzepka, 2018), and the influence of early childcare on children's skill development (Kuehnle and Oberfichtner, 2020).

In the following sections, we first take a closer look at both NEPS-ADIAB data sources and present some details on data collection, main contents, and relevant limitations. Next, the process of record linkage and the resulting data structure of NEPS-ADIAB are presented with some basic descriptive statistics. A main concern of this paper is to demonstrate the diverse research potentials and thus the added value of NEPS-ADIAB by highlighting a few thematic areas. We also inform about the requirements and procedures for data access, which is free of charge, as well as about the services offered by the responsible RDCs. At the end, there is a short summary and an outlook on future developments concerning the NEPS-ADIAB data product.

Data Sources and Linkage

NEPS Survey Data

The NEPS is a nationally and internationally well-established database with a particular emphasis on educational trajectories and competence development across the whole lifespan from early childhood until late adulthood. More than 370 peer-reviewed articles with empirical findings based on it have already been published, including 15 articles in the European Sociological Review (most recently: Dräger, 2021; Hamjediers and Schmelzer, 2021; Holtmann *et al.*, 2021; Schindler and Bittmann, 2021).

Initiated and developed by an interdisciplinary network, and funded by the German Federal Ministry of Research and Education (BMBF), the NEPS was launched in August 2008 to build a research data infrastructure with Germany-wide representative samples and high-quality data for direct use by the scientific community (Fuß, von Maurice and Roßbach, 2016). In order to enable a comprehensive applicability of the database, six general topics were identified. Figure 1 provides an overview of the main theoretical approaches and some corresponding constructs of the six 'pillars' that constitute the NEPS survey programme.

Besides, the NEPS also addresses current topics related to education, such as the increasing digitization in professional life or the consequences of the Corona

Competence Development	Learning Environments
- *domain-general*: cognitive functions (DGCF) - *domain-specific*: reading, listening, math, science - *meta-competencies*: ICT literacy, metacognition - *stage-specific abilities*: orthography, English etc. - *questionnaire*: self-regulation, social competence	- *formal*: educational institutions (e.g. school) - *non-formal/informal*: family, peers, work place - *characteristics*: structural features, excitation and support potentials - *transitions*: between learning environments
Personality & Motivation	**Educational Decisions**
- *temperament and personality*: IBQ, Big Five - *self-concept*: global self-esteem, self-efficacy - *motivation*: learning & achievement motivation - *personal goals*: locus of control, goal pursuit - *interests*: topic related & general orientation - *social & parenting behavior*: interpersonal competence (ICQ), strengths & difficulties (SDQ)	- *rational choice*: costs & benefits, probability of success, aspirations, information & awareness - *social capital*: reference groups, networks - *cultural capital*: objectified, institutionalized, embodied cultural capital, reading habits - *socio-demographics*: educational and occupational trajectories, basic characteristics
Migration Background	**Returns to Education**
- *migration biography*: origin, nationality, residence - *language*: use, training, competence - *ethnic networks*: religion & religiosity, identity, cultural practices, traditionalism - *specific environments*: ethnic composition, attitudes of others, perceived discrimination	- *economic*: labour income, property, job stability - *expectations*: later job and income - *non-economic*: well-being, satisfaction, health status and behavior, couple formation and starting a family, social and political participation, deviant behavior

Figure 1. Main topics of the NEPS survey programme.

pandemic on the individual and the family, as well as on school and work.

A basic principle of NEPS is the notion of education and competence development as lifelong processes. The six thematic key areas are designed to be prevalent in all phases of the life course; however, primary attention is paid to eight educational stages or critical transitions. Both dimensions combined represent the conceptual framework of the study, as shown in Figure 2. In practice, this means that stage-specifically adapted questions are asked for all topics, e.g. the cost-benefit considerations for the upcoming educational decision.

In order to collect and provide data on all stages within a reasonable period of time, a multicohort sequence design was implemented with six initial starting cohorts (SCs) ranging from newborns to adults. The sampling was done between 2009 and 2012 on either an individual or an institutional basis. Individual samples originate from population registration offices, while institutional samples are based on school records or lists of people studying at universities and universities of applied sciences. Table 1 lists the cohorts sorted by average age of the target subjects. At the beginning of the panel, more than 70,000 respondents participated in the NEPS. Depending on the expected heterogeneity of

educational transitions and in order to ensure sufficiently large comparison groups over time, the initial sample sizes are distributed differently across the cohorts. The close-meshed panel design, which also includes context persons such as parents, educators, and teachers for information on relevant learning environments in the younger cohorts, results in a data portfolio with scientific use files containing data from at least 8 up to 15 survey waves (August 2021).

The starting cohorts shown in normal font colour in Table 1 are those that are already linked to the IAB administrative data and are thus part of the NEPS-ADIAB data package. These four cohorts cover all stages with corresponding transitions:

Under the title 'Education from the very beginning', the *newborns' cohort* (SC1) focuses on early childhood development processes as of the age of 6–8 months. As the central source of information, mothers were and are interviewed at regular intervals; fathers are surveyed only in very few cases. The main interest here is on the skills and abilities of children in the first years of life, on parenting behaviour, on support by settings of child care and education within and outside the family, as well as on the framework conditions for different care arrangements and their effects.
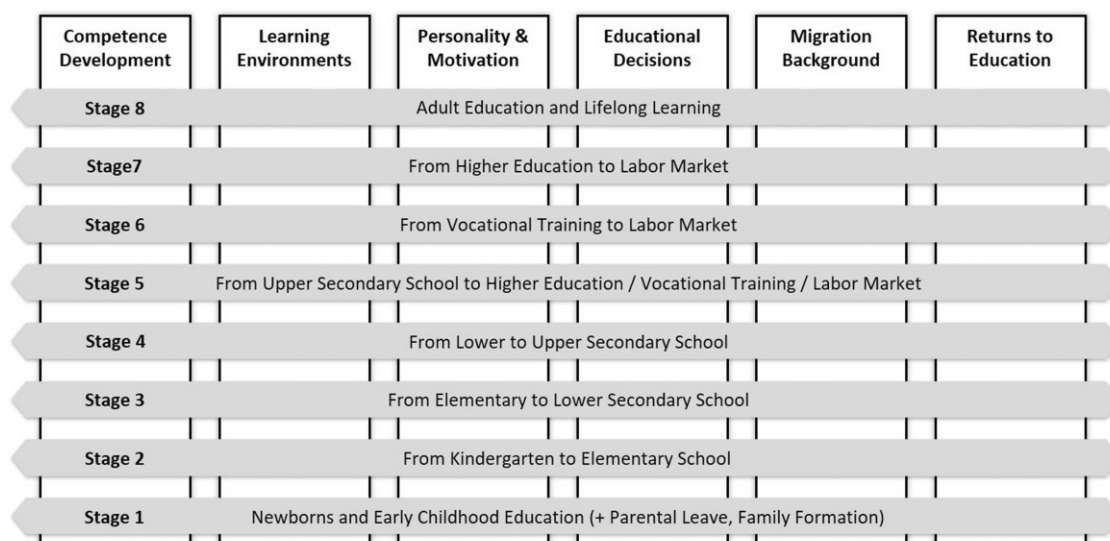
Figure 2. Conceptual framework of the NEPS.

Table 1. Panel characteristics of the NEPS starting cohorts (August 2021)

NEPS Starting Cohort	Panel start			Panel status		Context
	Year	Age	N	Waves	N	Additional surveys
Newborns SC1	2012	6–8 months	3,481	8	2,070	– Parents (usu. Mothers) – Educators/Teachers – Institution Heads
Kindergarten[a] SC2	2011 (2013)	4–5 years	2,996 (+6,341)	9	4,098	– Parents – Educators/Teachers – Institution Heads
Grade 5[a] SC3	2010 (2012)	10–11 years	6,112 (+2,205)	10	3,846	– Parents – Teachers – Institution Heads
Grade 9 SC4	2010	14–16 years	16,425	11	6,272	– Parents – Teachers – Institution Heads
First-year students SC5	2010	18–29 years	17,910	15	6,531	
Adults[a] SC6	2007 (2009) (2011)	21–63 years	6,855 (+5,077) (+5,208)	12	7,052	

[a]The years and case numbers in parentheses refer to sample refreshments during the panel progress.

The *ninth graders' cohort* (SC4) addresses the topic of 'School and vocational training: Educational pathways of students in grade 9 and higher' by following adolescents on their various paths into and through upper secondary education, into the vocational training system, into higher education, and into the labour

market. Because of the diversity of transitions and in view of the high mobility of people in this age group, the cohort started with a very large initial sample.

Within the *first-year students' cohort* (SC5), young adults are accompanied 'From university to the labour market', considering the entire range of study subjects at public and private universities as well as universities of applied sciences in Germany. The cohort features an oversampling of students to the teaching profession. In addition to the study progress from the first semester to graduation, including subject changes and dropouts, the focus in this cohort is again on the transition to professional life and the first career steps.

Finally, the *adults' cohort* (SC6) is designed to study processes of 'Adult education and lifelong learning' at all stages of a working career. The broad age range allows for research on educational activities, employment trajectories, and skill development throughout the lives of working-age adults up to the transition into retirement.

IAB Administrative Data

The administrative data originate from operative processes of the German Federal Employment Agency (BA) and from social security notifications. There are different data sources that start at different points in time due to changes in statutory regulations. For most of the sources, however, information is available since 1975 for West Germany and since 1992 for East Germany (Antoni and Schmucker, 2019; Müller and Möller, 2019).

Administrative data from the social security notifications result from the legal obligation of employers in Germany to annually report the fixed-period wages of their employees, with the duration of the reporting period to be specified in calendar days. In the case of varying wages within a year, this may lead to certain inaccuracies in the provision of daily income information. Nevertheless, the validity of the earnings information in the administrative data is much higher compared to survey data and the number of missing values is lower, as both memory deficits and refusals due to the sensitivity of such details restrict the data quality when (retrospectively) requesting income information from the respondents. Further details on the employees, such as education and nationality, are also part of the mandatory notification. Since April 1999, marginally part-time employees and employees with earnings below the marginal part-time earnings threshold—these persons were not relevant for the social security notifications until then—have also been included in the data. Wages above the upper earnings limit for statutory pension insurance are truncated. Truncated means that the respective

threshold replaces wages above it. Social security notifications only exist for individuals with an employment subject to social security contributions. Soldiers, self-employed persons, and civil servants are therefore not part of the administrative data.

An important element of the social security notifications comprises a set of indicators about the respective establishments (company, business). These indicators refer to the size of the establishment, the distribution of male and female employees, the number of persons with a certain employment status, qualification, income level, etc., but also to worker fluctuation (inflows and outflows) and the founding or closing of establishments.

The Federal Employment Agency is also obliged to provide administrative information from its operative processes including compulsory unemployment insurance, calculation of unemployment benefits and corresponding entitlement periods, registered job search, placement offers, and active labour market measures.

The data from both the social insurance notifications and the working processes of the Federal Employment Agency are finally combined in the so-called Integrated Employment Biographies (IEB). Together with the Establishment History Panel (BHP), it constitutes the administrative data part of NEPS-ADIAB. Figure 3 illustrates this compilation; more details on each of the data sources are available in the NEPS-ADIAB data reports (e.g., Bachbauer and Wolf, 2020).

Linkage Procedure

In Germany, linking different data sources at the individual level generally requires so-called informed consent (Antoni and Schmucker, 2019). The NEPS participants have been instructed about our intention of data enrichment and asked whether they agree with the linkage of their survey data with the administrative data of the IAB or not. In case, there was no consent; yet, the corresponding question was asked repeatedly in order to also reach temporary dropouts from the panel and to possibly convince those respondents who had not indicated their agreement so far. All survey participants of the four NEPS starting cohorts who finally gave their consent became part of the sample for the data linkage approach. Consenters who left the panel permanently before the linkage was started and whose contact details were no longer available due to the deletion obligation had to be omitted from the procedure.

These personal contact details are the key characteristics for the linkage, since there is no common identifier for a unique assignment of the administrative data to the survey data. First and last name, date of birth, gender
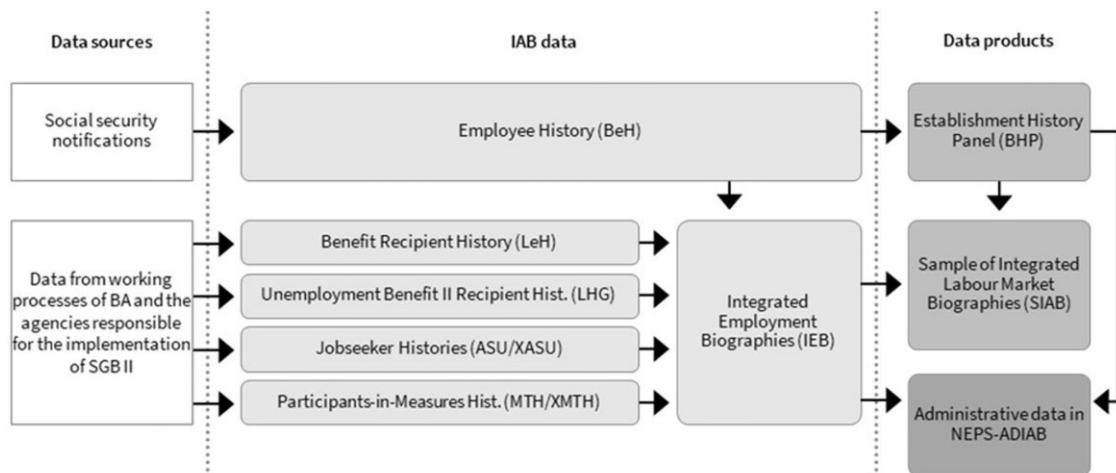
Figure 3. Administrative data sources of the IAB data in NEPS-ADIAB.

and address (postal code, city, street, house number) were used to identify the NEPS respondents in the central address files of the Federal Employment Agency via record linkage methods. Prior to the actual linkage, a so-called pre-processing has been implemented, which includes certain routines of correction and standardization of the key characteristics in order to maximize the matching probability between both data sources. These routines are based on established procedures that have been adjusted by the German Record Linkage Center to the specifics of the German context (Antoni and Schnell, 2019).

The actual linkage was conducted in successive steps. At first, an exact matching was carried out, i.e. all entries of the two data sources were compared and those with exactly the same information on all available key characteristics were linked. In the following steps, only minor deviations in one of the particularly error-prone items such as certain address fields or parts of the birth date were tolerated. For the remaining non-matches, a probabilistic linkage technique was applied. It involves the calculation of weighted similarity measures based on the key characteristics and the aggregation into a quality index. The higher this score, the better the match. Of all compared pairs of records that exceeded a specified threshold of this quality index, the administrative record with the highest score was selected as the match for the respective NEPS respondent. Finally, the records with scores below the threshold were checked and, if possible, manually assigned. The complete linkage procedure is also described in the NEPS-ADIAB data reports (e.g., Bachbauer and Wolf, 2020).

Data Structure and Descriptive Statistics

At the end of the linkage procedure, the survey data of 56 per cent (SC4) to 74 per cent (SC6) of the NEPS respondents could be successfully enriched with administrative data. The exact case numbers are shown in the last line of Table 2. The table also illustrates the consecutive linkage steps and their respective success. The striking difference between the adults' cohort and the other cohorts is already evident in the response to the consent question in the survey. While 93 per cent of the SC6 sample agreed with the linkage, the consent rates in SC1, SC2 and SC4 range between 66 and 69 per cent. The reasons for this discrepancy cannot be discussed here. It should be noted, however, that the wording of the consent question was identical for all respondents. The further dropouts in the linkage procedure occurred due to the non-availability of contact details for NEPS panel leavers ('available consenters'; minus 5–11 per cent points), the non-linkability of data because of insufficient similarity scores or completely missing entry in the administrative data ('linked consenters'; minus 1–7 per cent points), and the non-existence of employment information in the IEB ('successfully linked consenters'; minus 1–4 per cent points).

This raises the question of possible biases in the course of the linkage process. Although an analysis of differences between the initial NEPS samples and the NEPS-ADIAB samples of successfully linked individuals reveals some significant deviations with respect to average age as well as gender and education distribution, these deviations do not show any systematic pattern. The descriptive sample characteristics in Table 3 show only very moderate differences (which may nevertheless

Table 2. Sample sizes, consent, and linkage rates

	SC1	SC4	SC5	SC6
NEPS respondents	3,481 (100%)	16,425 (100%)	17,910 (100%)	17,140 (100%)
Consenters	2,285 (66%)	11,134 (68%)	12,373 (69%)	15,974 (93%)
Available consenters	2,127 (61%)	10,016 (61%)	11,340 (63%)	14,065 (82%)
Linked consenters	2,052 (59%)	9,640 (59%)	11,141 (62%)	12,846 (75%)
Successfully linked consenters[a]	2,004 (58%)	9,169 (56%)	10,468 (58%)	12,621 (74%)

[a] Case numbers and analyses in this chapter are based on the data versions with the following DOI: 10.5164/IAB.NEPS-SC1-ADIAB7518.de.en.v1, 10.5164/IAB.NEPS-SC4-ADIAB7517.de.en.v1, 10.5164/IAB.NEPS-SC5-ADIAB7518.de.en.v1, and 10.5164/IAB.NEPS-SC6-ADIAB7518.de.en.v1.

Table 3. Comparison of age, gender, and education distribution (full sample|linked sample).

	SC1	SC4	SC5	SC6
Mean age in years[a]	36.9|37.3**	20.3|20.3	28.2|28.1**	48.7|47.9***
Percentage of female respondents[b]	97.4|98.4**	49.5|49.8	60.5|60.2	50.5|50.3
Percentage of low educated[c]	12.3|9.0**	22.8|21.0***	n. a.	24.7|24.6

Significance level: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

[a] The average age refers to the (month-specific) age of the respondents at the time of the linkage wave.

[b] Proportion of female respondents in the total sample; in SC1, mothers were explicitly invited to participate in the panel.

[c] The proportion of low educated includes respondents without a degree and with a lower secondary school leaving certificate. For the significance analysis, the four education categories (0) no graduation, (1) low degree, (2) middle degree, and (3) high degree were used. The categorisation is based on the self-reported educational level at the respective linkage wave.

Table 4. Survey waves and time ranges in NEPS-ADIAB (August 2021).

	SC1	SC4	SC5	SC6
Number of NEPS waves NEPS-ADIAB[a]	7	10	14	10
NEPS survey period covered	2012–2018	2010–2017	2010–2018	2007–2018
Administrative data period covered	1980–2017	2008–2017	1975–2017	1975–2017

[a] Since autumn 2021, updated versions with administrative data and NEPS survey data up to 2019 are available for SC1 and SC5.

lead to significant results due to the large sample sizes; see also the Figures A1–A3 in the Appendix). A serious selectivity of the NEPS-ADIAB samples cannot be deduced from these rather general analyses.

Regarding the data scope of NEPS-ADIAB, administrative information is available for different time periods depending on the age of the individuals. Only for the two cohorts of adults (SC6) and university students (SC5)—some of them were 50 years and older when they began their studies—the maximum of retrospective data from 1975 onwards is reached. For the other two cohorts, mothers of newborn children (SC1) and ninth graders (SC4) at the start of the panel, the employment histories extend less far into the past because of the

biographical situation of the respondents. Table 4 also shows that the number of NEPS survey waves included in NEPS-ADIAB varies between the cohorts. On the one hand, these differences are due to different starting times of the surveys. On the other hand, the ninth graders and first-year students were followed more intensively, i.e. in shorter survey intervals, than the respondents of the two other cohorts.

The next three charts refer once again to the administrative data part of NEPS-ADIAB and provide an exemplary insight into the richness of available information. First, Figure 4 shows the percentage shares of the maximum lengths of (un-)employment histories. For the younger cohorts (SC4 and SC5), the shorter periods up
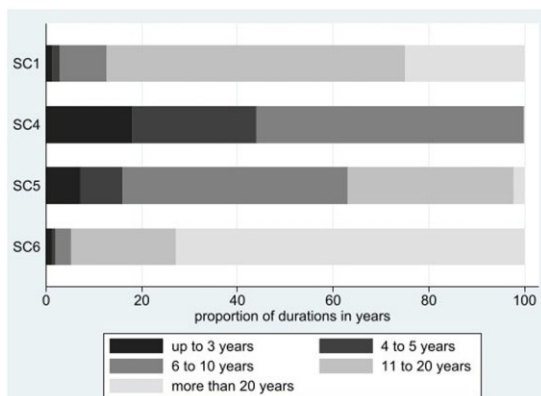
course research in sociology, economics and beyond. While the NEPS focuses on a detailed documentation of the educational career of individuals, the administrative data primarily contain information on employment histories, establishments, and income development. Additional survey information is available on family biographies (including childcare) and on the origin history of persons with a migration background. This opens up a wide field for researching the interdependencies between biographical events and for comparing types of career trajectories.

NEPS-ADIAB integrates subjective measures such as attitudes, expectations, and assessments from the NEPS survey with information obtained from administrative processes such as income, receipt of social benefits or participation in certain measures of active labour market support. A special feature concerns the competencies of respondents, which are objectively measured in several domains through standardized tests within the NEPS.

NEPS-ADIAB is not limited to the micro perspective of individuals. The administrative data also include the establishment level, which means detailed information on the respective companies in which people were or are employed. Furthermore, various geographical indicators are accessible in the data packages, allowing for spatial analyses in addition to the consideration of structural establishment characteristics.

In the following, we would like to illustrate the rich research potentials of NEPS-ADIAB data with some cohort-specific notes. The description is of course not exhaustive, other topics can also be empirically explored with this database. It should be noted, however, that the use of the administrative data presupposes a reference to labour market and occupational research.

Long-Term Employment Histories

Figures 5 and 6 show that the linked data package for the adults' cohort (NEPS-SC6-ADIAB) contains more than 650,000 episodes, and for three quarters of the (successfully linked) individuals there is at least one data point going back 20 or more years. This means that long-term (un)employment histories can be traced and analysed with regard to income development, job interruptions or changes, and so on. In light of the flexibilisation of the labour market and shifts in the concept of a normal biography organized around a working career, the research focus here is increasingly on the diversity of trajectories and on atypical paths. Supplemented by further biographical information from the NEPS, there are numerous approaches for dealing with issues such as the persistence of educational inequalities in later working

life or the reconciliation of family and private life with the demands of a professional career.

The long observation period and the broad age range of the respondents in the adults' cohort (born between 1944 and 1986) make it possible to distinguish cohort, period, and age effects. In this context, for example, questions about the consequences of increasing digitization of work processes or changes in job profiles, as well as about the impacts of the financial crisis at the beginning of the millennium and, in perspective, the Corona pandemic could be explored—both at the macro level and the micro level. Also worth mentioning at this point are the important research efforts focussing on gender-specific issues such as the gender pay gap.

Of major relevance in NEPS-ADIAB, however, are the various interrelationships between education and working life. In addition to long-term returns to schooling and education, the emphasis with respect to adults is on lifelong learning. Detailed survey information from the NEPS, in combination with the administrative episodes and establishment data, offers a wide scope of analysis perspectives. These perspectives relate to individual and structural predictors of participation in formal, non-formal, and informal learning activities, but also to the effects of adult education and skills endowment, for example, on income, job security, and job satisfaction.

Entry into the Labour Market

The main concern of the NEPS cohorts 4 and 5 is the transition of young adults into working life. In the first case, the panel started with ninth graders and followed them on the different paths into the labour market. In the latter case, first-year students were followed on their educational trajectories at universities and universities of applied sciences until they entered the labour market. Even though longer employment histories are not yet available for large parts of these two cohorts, they still provide a great potential for research. Especially the first steps of a career and the first years of proving oneself on the labour market are a critical phase with far-reaching consequences for later life. Thus, the NEPS-SC4-ADIAB and NEPS-SC5-ADIAB data can be utilized to answer key questions about more short-term returns to education, i.e., the role of educational trajectories, educational attainments, and competencies for a successful transition into working life.

Such analyses, on the one hand, might refer to the initial entry into the labour market: which education-related factors facilitate this step, and which hinder it? How are entry-level salaries distributed? What is the

significance of other characteristics such as gender, origin, personality traits, the availability of social capital, spatial mobility, etc.? Which impact do experiences have that are gained in the context of vocational preparation measures or in vocational training? On the other hand, it appears worth taking a closer look at the first time after completing school or vocational training: how do previously expressed salary expectations and career aspirations match up with the actual situation, and how do individuals react to mismatches? Which competencies and skills are particularly important when meeting the first career challenges? Under what conditions do educational inequalities increase or can be compensated at the beginning of work life? How do early phases of unemployment affect the further transition process?

For individuals with university education, questions arise first and foremost about the consequences of the course of study and the choice of subject: to what extent do the bachelor's and master's degrees, introduced as part of the Bologna reform, differ in terms of entry into the labour market? For which fields of study does a degree pay off particularly well at the beginning of working life? How do work-related experiences before or during the study period or a more practice-oriented study at a university of applied sciences play a role in the transition? Under what conditions do changing majors or dropping out have a positive or negative effect on the start of a career? To what extent do students' subject-related and non-subject-related competencies correspond to the requirements of the workplace, with which consequences?

Maternal Employment and Early Childhood Development

The sampling of NEPS cohort 1 was based on newborns who were 6–8 months old at the time when the panel started (2012/13). The respective interviewees were the mothers (only in very rare cases the fathers), who have been surveyed at least once a year. Accordingly, NEPS-SC1-ADIAB is a data source that provides comprehensive information on care arrangements in the first years of the child's life and repeated measures of early childhood development, as well as dense longitudinal information on mothers' employment histories from the administrative data. The empirical material covers maternal employment biographies ranging from several years before birth through parental leave to several years after birth. This gives a unique opportunity for explicitly considering the previous employment history of mothers for examining questions about the timing of births, the

availability and choice of care options after childbirth, the decision for a parental leave model.

In addition, the data source promises new empirical evidence on the complex relationships between maternal education and employment biographies on the one hand and early childhood development on the other. In the tension between demands of the labour market and the importance of early childhood competence development for later educational success, controversial debates about the influence of early (full-time) maternal employment on the child's development in cognitive, socio-emotional and health terms are going on. The same applies to questions about the prospects and preconditions to compensate for any negative effects by providing high-quality childcare. If one changes the perspective and takes the transmission from the child's development to the mother's further employment biography into account, other questions come to the fore: what effects do temperament, health, and cognitive characteristics of a newborn child have on the duration of absence from work and the type of re-entry into employment? How do the care situation and child development affect the mother's further career?

The study of maternal employment histories, including the duration of parental leave, the extent of career re-entry, the possible disadvantages of career breaks ('motherhood wage penalty'), and so on, is not limited to the linked data of NEPS Starting Cohort 1. Mothers with relevant events can also be identified in the other linked NEPS cohorts. An interesting approach for cohort-based analyses on changing societal conditions is offered by the reforms of German family policy since the mid-2000s. These reforms are intended to enable women to return to the labour market more quickly after the birth of their children and to achieve a better reconciliation of family and professional career. Another aspect concerns the change in occupations and activities that women do today, which are generally much more complex and subject to faster obsolescence than they were a few decades ago.

Establishment Context and Work Environment

Across the cohorts, the information contained in the administrative data on establishments can be used for context analyses or macro-micro questions. This information covers, in particular, the number of employees in the establishment, the structure of the workforce, the employee fluctuation, but also the classification by economic activities and the location of the establishment. It is only available for the time periods in which at least one of the linked persons was employed in the respective establishment. Accordingly, the data are not suited for longitudinal

analyses at the establishment level. However, they can be very well applied to address questions about the role of contextual factors, e.g., regarding the influence of establishment size on individual wage levels or wage growth ('employer size-wage effect') or the significance of staff composition in terms of gender, age and type of employment on personal career paths and job satisfaction. Circumstances of job change, job loss and re-employment (e.g. after parental leave) can be analysed not only at the individual level, but also with an explicit consideration of the establishment level. For this and other purposes, additional information is available on the founding and closure of establishments. Further research potential arises from the location data provided at the federal state and district level. These geographical indicators allow, for example, the exploration of regionally determined differences when it comes to the transition of young adults into the labour market.

Data Access

NEPS-ADIAB is a data product jointly managed by the LIfBi in Bamberg and the IAB in Nuremberg. The offering is administered by the Research Data Centers (RDC) at both institutions. Access to the data is free of charge for all members of the scientific community. It is provided by the IAB and bound to the conclusion of two agreements—one with each institution. In a first step, an application form has to be submitted to the RDC at the IAB stating that (i) the planned research project is related to labour market or occupational research and (ii) the linked NEPS-ADIAB data are suited best to answer the research questions, whereas the unlinked data sources are not sufficient. The second step requires a separate data use agreement with the RDC at LIfBi for the NEPS survey data. Here, too, a form must be completed with personal contact data of the researchers involved, a title and a brief description of the project, and its duration. The NEPS data use agreement allows for downloading all Scientific Use Files from the NEPS

study independently from NEPS-ADIAB data access. Once both documents are approved, the third step is to issue the final contract between the applicant researcher's institution and the IAB. Figure 7 graphically depicts this process, at the end of which NEPS-ADIAB can be analysed via on-site guest stay at one of currently 20 locations in different countries or via remote data processing.

Due to the high density of information and the sensitivity of personal data in NEPS-ADIAB, special precautions must be taken to ensure data protection and data security. For the NEPS part, this means that certain characteristics such as geographic or country-level information (e.g. federal state, native language) are sometimes only available in an aggregated form in order to avoid any potential re-identification of survey participants. Also, the use of NEPS-ADIAB is only possible in restricted access environments. The two options available—on-site and remote—differ in handling, but not with regard to the content or scope of the data. On-site guest access gives researchers full access to and insight into the data at specifically equipped workstations. Remote data access allows researchers to submit their analysis scripts via web interface of the remote execution software JoSuA (Job Submission Application) and to receive the output files after verification of compliance with data protection regulations. The data files are provided in Stata format, with variable and value labels that can be switched between German and English.

The bilingualism is not limited to the datasets, but also applies to the service and advice offered around NEPS-ADIAB. Comprehensive instructions are available to users both for data access and for working with the data. The two RDCs put much effort into making the data as suitable as possible for analyses despite the given complexity. In addition to the standard documentation with data reports, codebooks and survey instruments, test data for the administrative part are provided. They contain synthetically generated information, whereby the structure and metadata are identical to the original
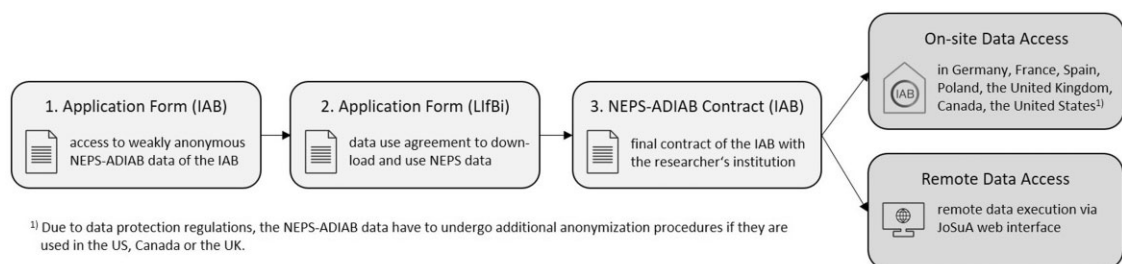

Figure 7. Application process for access to NEPS-ADIAB.

data files. While these test data are not appropriate for substantive research, they do make it possible to familiarize oneself with the structure of the administrative data and to prepare scripts for the on-site or remote processing of NEPS-ADIAB. Another service are training courses organized by LIfBi at least once a year, in which introductory presentations are offered to get a deeper understanding of the data. The newsletters of both RDCs regularly inform about relevant dates and recent developments. Individual requests concerning the data or data access can be addressed at any time by e-mail or telephone to the colleagues at both RDCs (LIfBi = NEPS survey and competence data; IAB = administrative data and NEPS-ADIAB). The RDC at LIfBi also operates an online discussion forum where questions and comments on NEPS-ADIAB can be posted publicly. All necessary application forms, data manuals and reports, tools such as the so-called 'NEPSplorer' for browsing the entire NEPS survey programme, services such as the just mentioned 'NEPSforum' or an NEPS-ADIAB video tutorial, and the contact details can be found on the respective websites:

- *RDC at LIfBi*: https://www.neps-data.de/Data-Center
- *RDC at IAB*: https://fdz.iab.de/en/FDZ_Individual_Data/NEPS-ADIAB.aspx

Summary and Outlook

With NEPS-ADIAB, the scientific community benefits from a data package that links two different sources with different strengths at the micro level of individuals—survey and competency data from the NEPS on the one hand and administrative data based on processes of the Federal Employment Agency and social security notifications on the other. Given the wide range of topics covered and the long observation periods, it offers multiple research opportunities at the intersection of educational trajectories and employment histories. The use of the data via the secure infrastructure of the RDC at the IAB is subject to certain preconditions and is by no means trivial; however, the responsible RDCs assist with extensive documentation materials and individual consulting services. Each of the currently available four scientific use files within the NEPS-ADIAB portfolio is clearly identified with a DOI ('digital object identifier'), which in turn constitutes an essential part of the contractually required data citation for any publication based on NEPS-ADIAB.

Free access to NEPS-ADIAB is guaranteed until 2026 through a cooperation between the LIfBi and the IAB. Within this period, regular data actualizations and new linkages are planned. At the beginning of 2022, updated NEPS-ADIAB data will be available for all currently linked cohorts. This will extend the administrative data horizon up to and including 2019 and the NEPS survey data up to and including 2020. With the subsequent round of updates in 2023, Corona-related questions can then be addressed in detail, as the time of the pandemic (up to and including 2021) will be represented in the employment histories in addition to information from the supplementary NEPS Corona modules collected since 2020. The intended expansion of NEPS-ADIAB refers to two cohorts of lower secondary school students whose educational trajectories have been or will be followed up to the transition to upper secondary school and beyond. The first is the cohort of fifth graders (SC3), which started in 2010 and whose participants had already gained initial work experience at the time of the linkage. The second is Starting Cohort 8, which will enter the field in 2022 and where parents are also surveyed as relevant context persons and scheduled for data linkage.

Finally, a data-related user conference is envisaged for 2023 at LIfBi for scientists to present and discuss their findings from analyses based on NEPS-ADIAB. The announcement will be published in due time via the NEPS website and the above mentioned newsletters. The exchange with researchers is not at least a key element for us to continuously improve the quality and user-friendliness of the database.

Acknowledgements

References

Antoni, M. and Schmucker, A. (2019). The research data centre of the German Federal Employment Agency at the Institute for Employment Research (RDC-IAB)—linked microdata for labour market research. *International Journal of Population Data Science*, 4, 1–10.

Antoni, M. and Schnell, R. (2019). The past, present and future of the German Record Linkage Center (GRLC). *Jahrbücher für Nationalökonomie und Statistik*, 239, 319–331.

Bachbauer, N. and Wolf, C. (2020). *NEPS-SC5 Survey Data Linked to Administrative Data of the IAB (NEPS-SC5-ADIAB) (FDZ-Datenreport No. 10-2020 en)*. Nuremberg: Institute for Employment Research.

Bachmann, R., Felder, R. and Tamm, M. (2018). *Labour Market Participation and Atypical Employment over the Life Cycle: A Cohort Analysis for Germany*. IZA Discussion Paper No. 12010. Bonn: IZA – Institute of Labor Economics.

Baur, N. (2011). Mixing process-generated data in market sociology. *Quality & Quantity*, **45**, 1233–1251.

Blossfeld, H.-P. and Roßbach, H.-G. (Eds.) (2019). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*, Edition ZfE, 2nd edn. Wiesbaden: Springer VS.

Dräger, J. (2021). The role of parental wealth in children's educational pathways in Germany. *European Sociological Review, jcab027, 10.1093/esr/jcab027*.

Fuß, D., von Maurice, J. and Roßbach, H.-G. (2016). A unique research data infrastructure for educational research and beyond: the National Educational Panel Study. *Journal of Economics and Statistics*, **236**, 517–528.

Hamjediers, M. and Schmelzer, P. (2021). Marriage, the risk of overeducation, and selection into both: evidence from Germany. *European Sociological Review, jcab030, 10.1093/esr/jcab030*.

Holtmann, A. C. *et al.* (2021). Improving formal qualifications or firm linkages: what supports successful school-to-work transitions among low-achieving school leavers in Germany? *European Sociological Review*, **37**, 218–237.

Kracke, N., Reichelt, M. and Vicari, B. (2017). *Wage Losses Due to Overqualification: The Role of Formal Degrees and Occupational Skills*. IAB-Discussion Paper 10. Nuremberg: Institute for Employment Research.

Kuehnle, D. and Oberfichtner, M. (2020). Does starting universal childcare earlier influence children's skill development? *Demography*, **57**, 61–98.

Müller, D. and Möller, J. (2019). Giving the international scientific community access to German labour market data: a success story. In Crato, N. and Paruolo, P. (Eds.), *Data-Driven Policy Impact Evaluation*. Cham: Springer, pp. 101–117.

Roed, K. and Raaum, O. (2003). Administrative registers—unexplored reservoirs of scientific knowledge? *Economic Journal*, **113**, 258–281.

Rzepka, S. (2018). Labour market returns to college education with vocational qualifications. *Education Economics*, **26**, 411–431.

Schindler, S. and Bittmann, F. (2021). Diversion or inclusion? Alternative routes to higher education eligibility and inequality in educational attainment in Germany. *European Sociological Review, jcab025, 10.1093/esr/jcab025*.

Tourangeau, R., Groves, R. M. and Redline, C. D. (2010). Sensitive topics and reluctant respondents. Demonstrating a link between nonresponse bias and measurement error. *Public Opinion Quarterly*, **74**, 413–432.

Zimmermann, M., Fitzenberger, B. and Osikominu, A. (2016). *Cohort changes in educational pathways and returns to education, VFS Annual Conference 2016 (Augsburg): Demographic Change*. Verein für Socialpolitik: German Economic Association.

Nadine Bachbauer is a Research Assistant at the Research Data Center of the Leibniz Institute for Educational Trajectories (LIfBi, Bamberg) and at the Research Data Center of the Institute for Employment Research (IAB, Nuremberg). She works primarily on the NEPS-ADIAB project and on TAEPS (Teacher in Adult Education: A Panel Study). She is interested in methodological issues such as nonresponse analysis and research on data quality.

Clara Wolf is a Research Assistant at the Research Data Center of the Leibniz Institute for Educational Trajectories (LIfBi, Bamberg) and the Research Data Center of the Institute for Employment Research (IAB, Nuremberg). She joined the NEPS-ADIAB project in 2019 and is particularly interested in methodological issues.

Daniel Fuß is head of the Research Data Center of the Leibniz Institute for Educational Trajectories (LIfBi, Bamberg). His work focuses on the coordination of LIfBi's research data infrastructure, including the services offered to data users. He is also responsible for several projects, including two projects within the Consortium for the Social, Behavioural, Educational, and Economic Sciences (KonsortSWD).
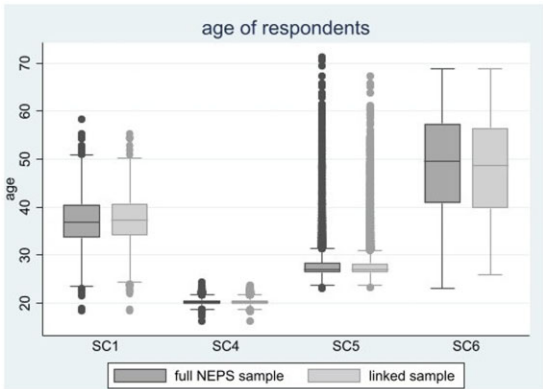
Appendix

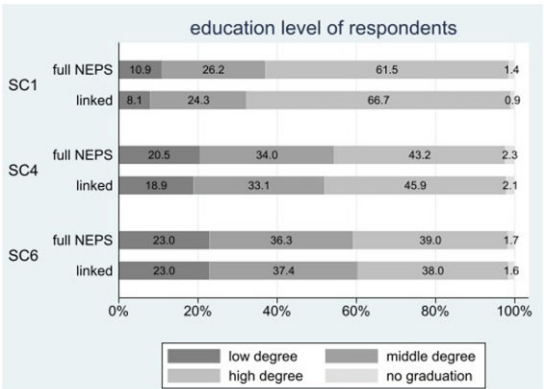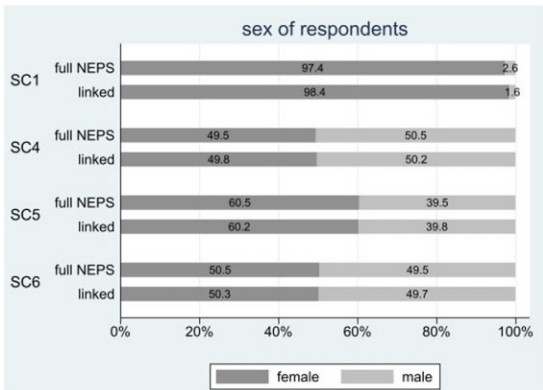**Figure A1.** Age compared between NEPS samples and linked NEPS-ADIAB samples.

Figure A2. Gender compared between NEPS samples and linked NEPS-ADIAB samples.

Figure A3. Education compared between NEPS samples and linked NEPS-ADIAB samples.