RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

# FDZ-DATENREPORT

Documentation of labour market data

## 14|2025 EN Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023

Sandra Dummert, Katrin Hohmeyer, Torsten Lietzmann, Irakli Sauer, Martina Oertel

Bundesagentur für Arbeit

# Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023

Sandra Dummert (IAB), Katrin Hohmeyer (IAB), Torsten Lietzmann (IAB), Irakli Sauer (IAB)
Martina Oertel (IAB)

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

# Contents

# List of figures



# List of tables

# Zusammenfassung

Dieser Datenreport beschreibt die Stichprobe der Integrierten Grundsicherungsbiografien (SIG) 2007 - 2023.

# Abstract

This data report describes the Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023.

# Keywords

Data manual, German administrative micro data, welfare benefit data.

# Acknowledgements

We would like to thank our colleagues at the Research Data Centre (FDZ) and the Data and IT Management Department (DIM) of the Institute for Employment Research (IAB) for their cooperation and support, in particular Alexandra Schmucker (FDZ), Andreas Ganzer (FDZ) and Tobias Graf (DIM). Text parts from DIM data descriptions and from previous FDZ data reports were used to describe the data. In addition, we would like to thank our (former) colleagues in the Research Group “Basic Income Support and the Labour Market” (FG GAMA), who contributed to the conception of earlier versions of SIG: Lena Bösel, Kerstin Bruckmeier, Helmut Rudolph and Anna-Theresa Saile.

# 1 Introduction and short description

## 1.1 Introduction

With the Sample of Integrated Welfare Benefit Biographies (Stichprobe der Integrierten Grundsicherungsbiografien, SIG) of the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB), the Research Data Centre (Forschungsdatenzentrum, FDZ) of the Federal Employment Agency (Bundesagentur für Arbeit, BA) provides data for scientific analysis on the structure and development of unemployment benefit II (UB II) receipt according to German Social Code II (Sozialgesetzbuch, SGB. The SIG consists of a 4 % random sample of all individuals who were living in a benefit unit – as defined in Section 7 SGB II – between 2007 and 2023.

The sample is based on the data of the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung, LHG). The LHG contains consolidated individual information accurate to the day, regarding individuals capable of work and eligible for UB II (Unemployment benefits in accordance with Social Code Book II) as well as the members of their benefit unit, if applicable, from the administrative procedures of the BA for this benefit as well as the data submitted from authorised municipalities.

Therefore, the SIG provides process-produced longitudinal data that can be used to analyse research questions on the dynamics of benefit receipt. In order to be able to investigate the processes of inflows into and outflows from as well as the reduction of benefit receipt over time, the data were enriched with information on periods of activation, support, registered unemployment and job search, and employment. The individuals in the sample and their employment histories form the so-called SIG core.

The eligibility for UB II receipt is defined in the household (= benefit unit) context. The main concern of the SIG is to enable research activities regarding benefit receipt considering specific situations in the household context, e.g. when children belong to the benefit unit. Therefore, information on further members of the benefit units as well as their employment histories was obtained for the individuals in the sample and added to the SIG core. More than one sampled individual may live in one benefit unit.

The major part of the employment histories of the sampled individuals and the members of the benefit units was taken from the Integrated Employment Biographies (Integrierte Erwerbsbiografien, IEB) of the IAB. The IEB contain data on all individuals who fall into one of the following categories at least once during the observation period:

- employment subject to social security (recorded from 1975 onwards)
- marginal part-time employment (recorded from 1999 onwards)
- receipt of benefits in accordance with Social Code Book III (recorded from 1975 onwards) or Social Code Book II (recorded from 2005 onwards)
- registered with the Federal Employment Agency (Bundesagentur für Arbeit - BA) or at an institution responsible for implementing SGB II as a jobseeker (recorded from 1997 onwards)
- participation in an employment or training measure (recorded from 2000 onwards)

These data, which come from different sources, are merged in the IEB and the statuses are depicted exact to the day but are not checked for plausibility in substantive terms.

While the Employee History (Beschäftigtenhistorik - BeH) is the origin of the information on employment subject to social security and marginal part-time employment, the receipt of benefits in accordance with Social Code Book III (SGB III) and Social Code Book II (SGB II) is recorded in the Benefit Recipient History (Leistungsempfängerhistorik - LeH) and the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG). The Jobseeker Histories (Arbeitsuchendenhistoriken - ASU and XASU) are the data source for the periods of registered unemployment and job search recorded by the BA or by municipal institutions responsible for implementing SGB II, while participation in employment and training measures is recorded in the Participants-in-Measures History Files (Maßnahmeteilnahmehistoriken – MTH and XMTH).

In contrast to other administrative data products of the BA's FDZ within the IAB, e.g. the Sample of the Integrated Employment Biographies (Stichprobe der Integrierten Arbeitsmarktbiografien, SIAB), in the SIG, all information on the receipt of benefits according to SGB II is drawn directly from the LHG, not from the IEB. This allows for a broader range of characteristics, especially on the structure of the benefit unit. On the one hand, the information is available as individual characteristics; on the other hand, all members of the benefit unit are included in the data as well. In addition to that, individual characteristics not available in the IEB were directly obtained from additional data sources.

The data on the sampled individuals from the LHG as well as on the members of their benefit unit can be enriched additionally – for the first time in a data product of the FDZ – with calendar month-based, comprehensive financial data from the Benefit Statistics SGB II (Leistungsstatistik SGB II, LST-S), which contain financial information on the amount of potential and actual benefit receipt, income, and sanctions on the individual level.

The SIG data were created by the FDZ and the Research Group "Basic Income Support and the Labour Market" ("Forschungsgruppe Grundsicherungsbezug und Arbeitsmarkt", FG GAMA) of the IAB. The SIG core contains a total of 905,135 sampled individuals whose employment histories are documented in a total of 24,966,840 data rows. The dataset of the members of the benefit units contains a total of 1,784,760 individuals whose employment histories are captured in a total of 50,179,541 data rows. The present data report describes the characteristics of the weakly anonymised version of the SIG, which mostly contains the original, i.e., non-coarsened data. Some characteristics, however, are classified as particularly sensitive and will only be made available upon special and reasoned request (see Chapter 1.4.1).

This data report is structured as follows: Chapter 1 contains the introduction and, among other things, information on the distinction from other administrative data products of the IAB, on data access and management as well as a short description of the data, the quantity structure, and a list of characteristics. A description of the several data sources can be found in Chapter 2. Sampling, data structure and preparation as well as the quality of the data are discussed in Chapters 3 and 4, and the individual characteristics are described in Chapter 5.

## 1.2 Population of the SIG

The basic population of the SIG includes all individuals living in benefit units and receiving UB II according to SGB II.

So-called "individuals capable of work and eligible for benefits" as well as the family and household members living together with them who form a benefit unit are eligible for benefits. According to Section 7 (1) SGB II, eligible individuals include "individuals who

1. have reached the age of 15, but not the age limit according to Section 7a,
2. are capable of work,
3. are in need of financial assistance, and
4. have their habitual residence in the Federal Republic of Germany (individuals capable of work and eligible for benefits)."

According to Section 7 (3) SGB II, a benefit unit includes

"1. the individuals capable of work and eligible for benefits,

2. the parents or one parent of an unmarried child capable of work who has not reached the age of 25 living in the household, and the partner of this parent living in the household,

3. as the partner of the individual capable of work and eligible for benefits

a) the not permanently separated spouse,

b) the not permanently separated partner in a same-sex union,

c) an individual living together in a common household with the individual capable of work and eligible for benefits, whereby, following a reasonable assessment of the circumstances, the mutual desire to bear responsibility for each other and support each other is to be assumed,

4. the unmarried children of the individuals specified in Items 1 to 3 belonging to the household, if they have not reached the age of 25, insofar as they cannot sustain a livelihood from their own income or assets."

In legal terms, unmarried children under 25 years of age with their own income covering their requirements are no members of the benefit unit of their parents (according to Section 7 (3) SGB II). BA statistics, however, regard minor children as individuals not eligible for benefits in a benefit unit "in the sense of comprehensive social reporting". This group of people is sub-summarised under "members of the benefit unit" in this data report as well[1].

## 1.3 Distinction from other administrative data products of the IAB

The SIG is the first solely administrative longitudinal dataset on UB II receipt provided to external researchers by the FDZ. The purpose of the SIG is the analysis of benefit receipt histories on the

[1] Moreover, children whose benefit unit solely receives Benefits for Education and Participation (Leistungen für Bildung und Teilhabe, BuT) have a special status. Their parent(s) is/are no members of the benefit unit in this case.

individual level as well as the analysis of the structure and dynamics of benefit receipt on the aggregate level. The SIG provides three unique features in this context, distinguishing it from the other data products of the IAB and serving the specific research purpose.

First, the SIG contains a larger case number of individuals receiving UB II benefits as compared to the sample of the Integrated Employment Biographies (Stichprobe der Integrierten Arbeitsmarktbiografien, SIAB) and the PASS-ADIAB (PASS survey data linked with the administrative data of the IAB). This allows for analyses of individual groups within the very heterogeneous group of UB II recipients, such as single parents or employed individuals receiving benefits, as well as regional analyses to be performed. However, due to unavailability of the necessary high level of detail in the type of measure in the SIG it is not suitable for measure evaluation.

Second, the SIG is the only administrative dataset of the FDZ that contains a benefit unit context. For the first time, the SIG data include information on all members of the benefit unit and their employment histories for external researchers. This allows for a differentiation of problematic situations in the household context required for the analyses of the benefit receipt, such as the size of the household, the labour participation of the members of the benefit unit, or the care of children. In addition to that, the representation of comprehensive employment histories of all members of the benefit unit helps to investigate the benefit receipt in the context of labour market-related life situations, such as the labour market transition of children receiving benefits.

Third, the data from the Benefits Statistics According to SGB II (Leistungsstatistik nach dem SGB II, LST-S) that have not been available as micro data for research outside the IAB up to now can be requested as an extension module for the SIG. It will therefore be possible for the first time to analyse the potential and actual benefit receipt, income, and sanctions on the benefit unit level. This expands the analysis options by the investigation of requirements and analyses on the income structure.

## 1.4 Data use

### 1.4.1 Data access and data management

The SIG data in the weakly anonymous version may be analysed in the context of a research visit at the FDZ, via remote desktop access and subsequent remote data access.

In order to be able to use the data, it is first necessary to submit an application to the FDZ. The FDZ decides on the approval of the research project on behalf of and, if necessary, in coordination with the Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales – BMAS). When approval has been granted, a data use agreement is concluded with the researcher's scientific institution. Details on applying for the dataset and possibilities for data processing are available on the FDZ website.

The SIG data, which include both German and English labels,[2] have a modular structure and are stored in several files. One module, which is henceforth called "SIG-Core", contains identifiers (individual IDs, benefit unit IDs and establishment IDs), the personal variables, information on benefit receipt, registered unemployment and job search activity, participation in an employment or

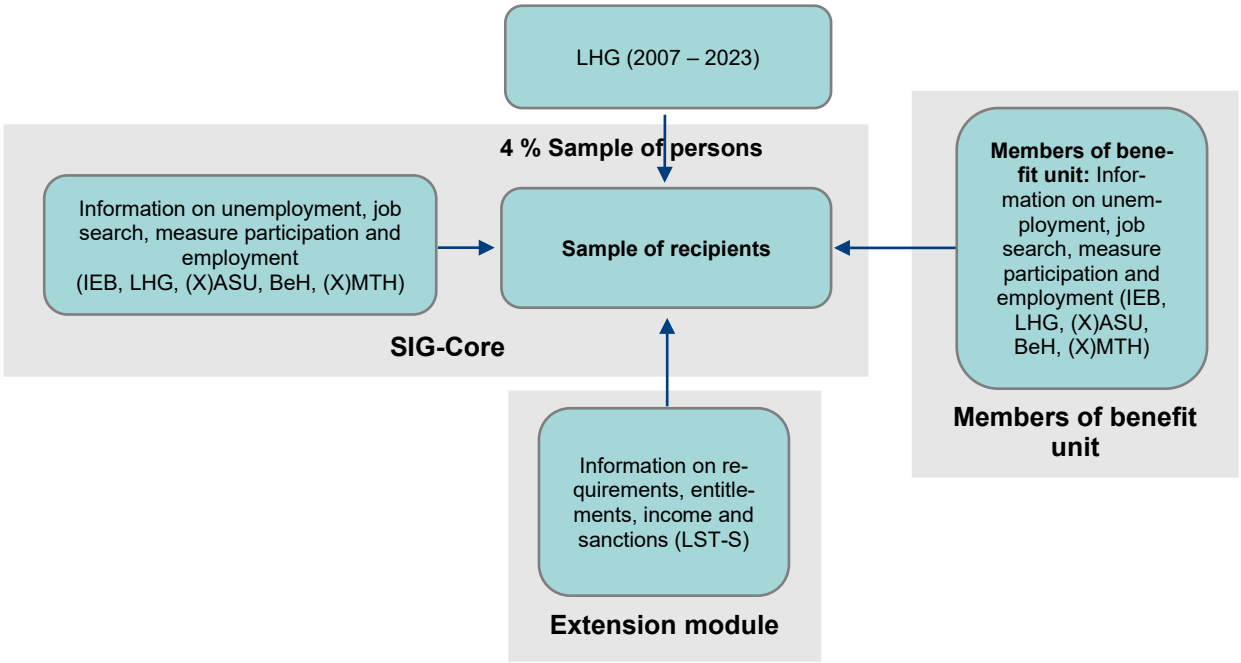
[2] With the Stata command *label language en* / *label language de* labels can be switched to English or German.

training measure and employment, variables regarding the place of residence of the sample persons, and technical variables. The personal characteristics as well as the information on the employment history of the members of the benefit unit are stored in a separate file (members of the benefit unit).

Upon reasoned request, the Benefits Statistics According to SGB II (Leistungsstatistik nach dem SGB II, LST-S) – Benefits Statistics – can also be provided as an extension module. The LST-S contains individual-level information on the potential and actual benefit receipt, income and sanction.

In terms of organisation, the modules are kept separate, underlining the data structure and saving storage space (cf. Figure 1). Linking the LST-S with the SIG core is possible on a monthly level using the month and the benefit unit ID. A description and an instruction are outlined in Chapter 2.6.

**Figure 1: Storage of SIG data**




Source: own illustration

Certain original characteristics facilitating the de-anonymisation of individuals and establishments are only shared if it is necessary for the purpose of the analysis, and only if a reason has been explicitly stated in the data access application. These particularly sensitive characteristics in terms of data protection laws include:

**Modules SIG-Core and members of the benefit unit:**

- month of birth (gebmon)
- nationality (nation_lhg, nation_beh, nation_leh, nation_x_asu, nation_x_mth)
- occupational sub-group - current employment (beruf2010_4_aktT)
- occupational sub-group - most recent completed training (beruf2010_4_ausb)
- occupational sub-group - sought-after job (beruf2010_4_gesucht)

- place of residence: district (Kreis) (wo_kreis_lhg, wo_kreis_beh, wo_kreis_leh, wo_kreis_x_asu)
- occupational sub-group – last employment (KldB 2010) before 2007 (beruf2010_4_vor2007)
- measure type – group (mass)
- place of work: district (Kreis) (ao_kreis)
- economic activity 08 - sub-class of economic activity (five-digit code) (w08_5_kons)

Please note that the sensitive characteristics are typically included in the data already in coarsened form. The sensitive characteristics will be made available only if the information contained in the coarsened characteristic is not enough to perform the desired research.

The attributes of the LST-S are sensitive in their entirety due to their level of detail and the associated de-anonymisation potential and may be provided upon reasoned request only.

### 1.4.2 File names in SIG 0723

| | |
|---|---|
| SIG-Core: | sig_kern_0723_v1.dta |
| Module benefit unit members: | sig_bg_0723_v1.dta |
| LST-S: | sig_0723_v1_lsts_v1.dta |

## 1.5 Changes as compared to SIG version 0720

### 1.5.1 Set of variables

Compared to version SIG 0720, SIG 0723 includes new variables on daily earnings that support the addition of one-off special payments and offer an imputation of the censored values to the right of the upper earnings threshold[3]. The underlying data work is described in Drechsler et al. (2023) and Drechsler/Ludsteck (2025). Further information on the new variables can be found under “Daily wage (incl. one-off special payment) (entgelt_tag_bonus)”, “Daily wage (imputed) (entgelt_tag_imp)” and “Daily wage (incl. one-off special payment) - second employment (mehrfbesch_bonus_tentg)”.

In addition, the variable “Integration forecast (ipo)” has been added to SIG 0723 compared to version SIG 0720. The integration forecast has been used in the placement process to categorize BA clients since the end of 2016.

Furthermore, the sensitive variable “Measure type – group (mass)” is now included. This provides more detailed information on participation in specific measures of active labor market policy. The variable is in contrast to the coarsened version (measure type – category (mass_gr)) only available upon request and in well-founded cases.

In addition, the variable “Vocational training (imputed) (ausbildung_imp_beh)” was also newly included. This variable is filled in the source BeH.

Table 1 provides an overview of the changes in the set of variables across all sources.

[3] The imputation procedure for wages contained in the scripts by Stüber et al. (2023) can now be replaced by the newly offered variables.

**Table 1: Added, deleted, renamed and altered variables in the new SIG version 0723**

| Variable | Explanation |
|---|---|
| Daily wage (incl. one-off special payment) (entgelt_tag_bonus) | New variable |
| Daily wage (imputed) (entgelt_tag_imp) | New variable |
| Daily wage (incl. one-off special payment) - second employment (mehrfbesch_bonus_tentg) | New variable |
| Integration forecast (ipo) | New variable |
| Measure type - group (mass) | New variable, more detailed differentiation of measure type by IAB |
| Measure type – category (mass_gr) | Renamed, old variable name „massn“ |
| Vocational training (imputed)” (ausbildung_imp_beh) | New variable |

### 1.5.2 Documentation

In this data report, the descriptions for the following variables have been revised or extended:

- Vocational training (ausbildung_beh, ausbildung_x_asu)
- Vocational training (imputed) (ausbildung_imp_beh)
- Daily wage (entgelt_tag)
- Daily wage (incl. one-off special payment) (entgelt_tag_bonus)
- Daily wage (imputed) (entgelt_tag_imp)
- Daily wage (incl. one-off special payment) - second employment (mehrfbesch_bonus_tentg)
- Measure type – category (mass_gr)
- Measure type - group (mass)
- Integration forecast (ipo)

Section “1.7 List of characteristics” has been deleted. It has been replaced by detailed overviews of the individual characteristics (labels, counts, missing values) in the online appendix to the data report.

## 1.6 Outline

**Table 2: Outline**

| | |
|---|---|
| Topics/ groups of variables | **Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG):**<br>Data on individuals in receipt of basic social security benefits in accordance with Social Code Book II (SGB II) (Types of institution: cooperation of employment agencies and municipalities/joint facilities, separated responsibilities/municipalities exercising their duties separately, authorised municipalities)<br>**Benefit Recipient History (Leistungsempfängerhistorik - LeH):**<br>Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefit, unemployment assistance and maintenance allowance.<br>**Jobseeker History (Arbeitsuchendenhistorik - ASU):**<br>Information on job search activities that are recorded in BA procedures<br>**Jobseeker History from XSozial-BA-SGB II (Arbeitsuchendenhistorik aus XSozial-BA-SGB II - XASU):**<br>Information on job search activity reported to the BA by authorised municipalities via the transmission standard XSozial-BA-SGB II |

| | |
|---|---|
| | **Participants-in-Measures History File (Maßnahmeteilnahmehistorik - MTH):**<br>Information on participation in employment and training measures (not including measures of authorised municipalities)<br>**Participants-in-Measures History File from XSozial-BA-SGB II (Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II - XMTH):**<br>Information on participation in employment and training measures reported to the BA by authorised municipalities via the transmission standard XSozial-BA-SGB II<br>**Employee History (Beschäftigtenhistorik - BeH):**<br>Annual notifications and end-of-employment notifications submitted to the social security agencies for employees covered by social security and employees in marginal part-time employment.<br>**Benefits Statistics According to SGB II (Leistungsstatistik nach dem SGB II LST-S)**<br>Information on the need for assistance, income, sanctions, and claims of individuals eligible for benefits and other groups of people in the context of a benefit unit. |
| Data units | Benefit recipients and the members of their benefit unit, employees subject to social security contributions (as of 1999, also individuals in marginal employment), job seekers, participants in measures, employing companies |
| Number of cases | 905,135 individuals (SIG-Core)<br>24,966,840 non-overlapping observations (after episode splitting) (SIG-Core)<br>1,784,760 individuals (Members of the benefit unit)<br>50,179,541 non-overlapping observations (after episode splitting) (Members of the benefit unit) |
| Period covered | 01.01.2007 – 31.12.2023 |
| Time reference | exact to the day |
| Regional structure | German federal states (Bundesländer), districts (Kreise) |
| Date of territorial allocation | Territorial allocation updated as of 31.12.2023 |
| Data collection method | 4% random sample of persons drawn directly from the LHG |
| Institutions involved | Social security agencies, Federal Employment Agency (Bundesagentur für Arbeit), municipal institutions |
| Frequency of data collection | Continuous |
| File format and size | Stata; SIG-Core: 4,7 GB; Benefit unit members: 9,4 GB; LST-S: 6,5 GB |
| File architecture | The data are stored in three files. One contains individual-level information of the sample persons and the other individual-level information of the benefit unit members. A further file with the Benefits Statistics According to SGB II LST-S can be provided following a justified application. |
| Data access | Remote desktop, on-site use or remote data access |
| Degree of anonymisation | Weakly anonymous |
| Sensitive variables | **SIG-Core and module benefit unit members:** Month of birth (gebmon), nationality (nation_lhg, nation_beh, nation_leh, nation_x_asu, nation_x_mth), measure type – group (mass) occupational sub-group current employment (beruf2010_4_aktT), occupational sub-group of the last completed training (beruf2010_4_ausb), occupational sub-group of the employment searched for (beruf2010_4_gesucht), place of residence: district (Kreis) (wo_kreis_lhg, wo_kreis_beh, wo_kreis_leh, wo_kreis_x_asu), occupational sub-group – last employment (KldB 2010) before 2007 (beruf2010_4_vor2007), place of work: district (Kreis) (ao_kreis), economic activity 08 - sub-class of economic activity (five-digit code) (w08_5_kons)<br>**LST-S:** completely sensitive |
| Citation of data and data documentation | Data:<br>„The data basis of this paper is the weakly anonymous Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023. The data access was provided via remote |

| | |
|---|---|
| | desktop/on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and remote execution." DOI: 10.5164/IAB.SIG0723.de.en.v1<br><br>Data documentation:<br>Dummert, Sandra; Hohmeyer, Katrin; Lietzmann, Torsten; Sauer, Irakli; Oertel, Martina (2025): Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023. FDZ-Datenreport, 14/2025 (en), Nürnberg. DOI: 10.5164/IAB.FDZD.2514 .en.v1 |
| Dataset version | Sample of Integrated Welfare Benefit Biographies (SIG) 2007 - 2023 (SIG 0723); DOI: 10.5164/IAB.SIG0723.de.en.v1 |

## 1.7 Volume structure

**Table 3: Volume structure**

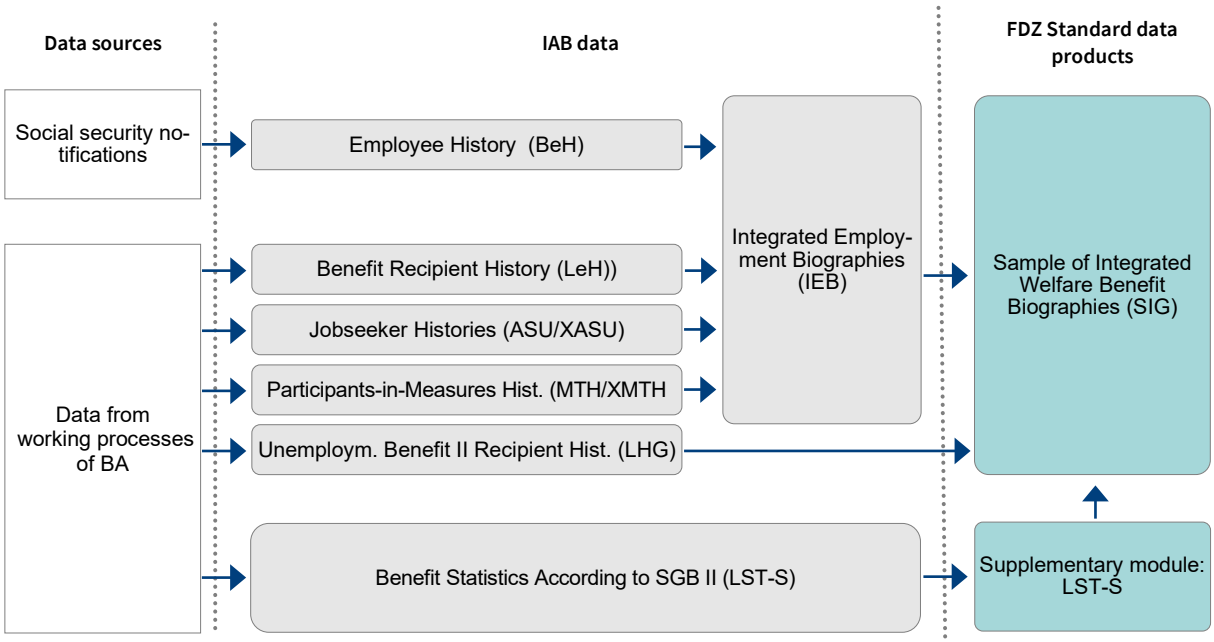
| **Number of cases** | **SIG-Core** | **Benefit unit members** |
|---|---|---|
| LHG | 13,577,419 | 31,887,733 |
| LeH | 1,224,743 | 1,887,418 |
| ASU | 13,233,142 | 25,652,579 |
| XASU | 2,209,447 | 5,086,523 |
| MTH | 1,764,780 | 3,329,212 |
| XMTH | 451,848 | 960,224 |
| BeH | 12,267,868 | 21,599,699 |
| **Total number of observations*** | **24,966,840** | **50,179,541** |
| **Individuals** | **905,135** | **1,784,760** |

* The total number of observations does not equal the sum of observations over all sources due to the data structure with spells from potentially several sources.

# 2 Data sources

Sampling for the SIG is based on the data of the Unemployment Benefit II Recipient Histories (Leistungshistoriken Grundsicherung, LHG). Information on the receipt of benefits according to SGB II is also taken from the LHG, while the other administrative individual data were mostly drawn from the Integrated Employment Biographies (Integrierte Erwerbsbiografien, IEB) of the IAB. They unite data from different data sources, each of which may contain information from different administrative procedures. In addition, some supplementary variables from these data sources, which are not part of the IEB, are incorporated into the administrative individual data. Figure 2 illustrates the data flows that lead to the SIG. These data sources are briefly described below.

**Figure 2: Data sources of the SIG**




**Note: The IAB sampling procedure for LST-S is based on the UB II Income and Benefit Statistics of the Federal Employment Agency (BA).**

Before the data are merged from the above-mentioned sources to become the IEB, they undergo source-specific data adjustments (see also the following sections). The following adjustment processes are applied to the entire IEB:

- Spells where the age is below 13 at the end or over 75 at the beginning are deleted.
- Spells where the end date is smaller than the start date are deleted.
- Inconsistent information on the gender or date of birth within one account is adjusted.
- Spells without information on the date of birth or gender after the adjustment are deleted.

Additional adjustments, such as completing apparently missing notifications, strike adjustments, etc., are not performed.

## 2.1 Unemployment Benefit II Recipient History (LHG)

The Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG) contains information about individuals who are eligible for benefits and capable of work and about the members of their benefit unit (Bedarfsgemeinschaft) in accordance with § 7 SGB II (see chapter 1.2). With the SIG data it is possible for the first time for researchers outside the IAB to link individuals with SGB-II benefits at the individual level. The receipt of benefits in accordance with SGB II covers both basic social security benefits (e.g., Unemployment Benefit II) and supplements to unemployment benefit or additional benefits. The LHG does not contain any information about the benefit rates. However, such information is part of the benefit statistics according to SGB II (LST-S) and is available upon justified request.

Unlike the benefits in the sphere of Social Code Book III, the Federal Employment Agency (BA) is not the sole institution responsible for administering the benefits. The data therefore distinguish between the three possible types of institution responsible for implementing SGB II:

- Cooperation of employment agencies and municipalities (Arbeitsgemeinschaften – ARGE) until the end of 2010 / joint facilities (gemeinsame Einrichtungen) since 2011), in which the BA and the municipality deal with tasks jointly,
- separated responsibilities (getrennte Trägerschaft - gt) / municipalities exercising their duties separately (until 2011) – here the tasks are divided between the BA and the municipality,
- authorised municipalities, which are also called opting local authorities or opting municipalities according to the initial experimental clause of Section 6a - here the local authority is responsible for all tasks in the sphere of SGB II.

The data of the “Unemployment Benefit II Recipient History drawn from A2LL” (LHG) come from different reporting procedures. As a rule, the IT procedure A2LL was used in all ARGE cooperation projects until 2010, and in joint facilities from 2011 onwards. In 2014 the IT procedure A2LL was substituted by the IT procedure ALLEGRO. In this process each benefit unit was assigned a new benefit unit ID.[4] Authorised municipalities use various IT procedures of their own and transmit their data to the BA by means of the XSozial-BA-SGB II standard once a month. Both of the procedures are used by municipalities with separated responsibilities. The different data standards affect the scope and quality of the data supplied.

The earliest available data in the LHG are from 1 January 2005. However, the data source is incomplete until the beginning of 2007. For this reason, the year 2007 was chosen as the left margin for the sampling of SIG.

## 2.2 Benefit Recipient History (LeH)

The Benefit Recipient History (Leistungsempfänger-Historik - LeH) of the IAB covers periods during which individuals receive earnings replacement benefits from the Federal Employment Agency (sphere of Social Code Book III). The benefits comprise unemployment benefit, unemployment as-

[4] LHG spells belonging to the same benefit unit cannot be identified across the two different IT procedures. On the individual level this is possible because the individual ID is stable across the whole timespan.

sistance, maintenance allowance, and contributions paid by the BA to private health or care insurance while benefits are being drawn. Furthermore, information on suspension and interruption periods is also included:

- A suspension period is a period during which a person who is entitled to benefits does not receive them for reasons he or she is responsible for. A typical example is a short suspension period at the beginning of a benefit claim, if the person registered as a jobseeker too late.
- An interruption period is a period without receipt of benefits for reasons the person entitled to benefits is not responsible for. This might be, for example, a compulsory absence from the place of residence. A period of interruption does not reduce the duration of entitlement, i.e. the person does not receive any benefit during the period of interruption, but the regular end of entitlement to benefits is extended accordingly.

Benefits in the context of Social Code Book II (e.g., unemployment benefit II) are not included in the data. Since entitlement to bene-fits depends on meeting certain legal requirements, periods of unemployment in which the requirements are not met (e.g., no eligibility for unemployment assistance, or non-completion of the qualifying period for unemployment benefit) are not captured in the Benefit Recipient History.

The following LeH adjustment processes relevant for the data product are performed:

- Observations without a valid start date are excluded.
- Observations with no end date or an invalid end date are excluded, since in these cases it cannot be assumed that a benefit payment was made at all.
- Information on unemployment assistance and alimonies is removed for the SIG.

## 2.3 Jobseeker Histories (ASU/XASU)

Data about jobseekers are stored in the Jobseeker Histories (Arbeitsuchendenhistoriken – ASU/XASU). The ASU data source contains information on jobseekers who are registered with employment agencies, and from 2005 onwards also includes ARGE cooperation projects and separated responsibilities for the implementation of SGB II. The XASU data source, on the other hand, contains the data of jobseekers in receipt of Unemployment Benefit II (ALG-II) from authorised municipalities from 2005 onwards. These data are reported in accordance with the X-Sozial-BA-SGB II standard. The earliest available data in the ASU are from 1 January 1997 and in the XASU from 1 January 2005, respectively. Besides unemployed and non-unemployed job seekers, information on other registered individuals capable of work is also included, especially the members of benefit units who are not available for job search according to Section 10 SGB II (reliably recorded in both data sources as of 2011).

The following (X)ASU adjustment processes relevant for the data product are performed:

- There is no consolidation of the ASU observations on individual level. Therefore, over-laps between ASU observations might occur.
- Individual-related variables that are only available for the (X)ASU sources always refer to the beginning of the spell.
- A new ASU spell is generated as soon as a change of status occurs (e.g., from seeking work to unemployed). This also applies if the type of institution (employment agency, cooperation of

employment agency and municipality, joint facility, authorised municipalities, separated responsibilities) changes.

- The XASU contains non-overlapping time periods for individuals. If one of the following variables changes, in each case a new data spell is generated for the XASU:
  - change of job search status
  - change of availability
  - change of SGB II institution (due to notification procedure)
  - change of place of residence

## 2.4 Participants-In-Measures History Files (MTH/XMTH)

The Participants-In-Measures History Files (Maßnahmeteilnahmehistoriken - MTH) contain information that can be assigned to different legal spheres. First, they contain active labour market policy measures in accordance with Social Code Book III and participation in such measures. Second, the MTH contain measures in the legal sphere of Social Code Book II if these are recorded in BA administrative procedures. Measures implemented by authorised municipalities or opting local authorities are recorded in the XMTH from 2005 onwards.

The following MTH adjustment processes relevant for the data product are performed:

- Observations generated more than a year after the end of the measure are deleted if another observation exists that was generated within the year after completion of the measure.
- Only the most recent record of an individual case of participation in a measure is used.
- Only cases of participation in measures that are classified as “actually took place” are included in the IEB. Cases of participation that did not take place or have not yet taken place are deleted. Cases of participation are also classed as not having taken place when a deletion date is set during the participation in a measure.
- Certain types of measure are not included. These include services to support careers advice and job placement, mobility assistance and pure rehabilitation measures.

The following (X)MTH adjustment processes relevant for the data product are performed:

- For XMTH, the particular challenge is to identify and handle multiple notifications of participation that are caused by technical or organisational issues. The causes of these duplicate notifications are, on the one hand, a missing identification number for participation in measures until April 2009 or improper handling of the subsequently introduced promotion ID and, on the other hand, a change of provider numbers in connection with mergers or other reorganisations of institutions responsible for implementing SGB II due to the reorganisation of the employment agencies in 2012/2013. In addition, provider-specific problems arose with version or product changes of the municipal software or with version changes of the XSozial standard. At times, this results in a considerable number of reports of participations in measures of the same type, which partly or completely overlap in time per participant. Real duplicates are identified and sorted out because they are highly likely technical duplicates. Hidden duplicates can also be identified and sorted out. In principle, the rule always applies that the most up-to-date information is retained.

- Overlapping and immediately adjacent notifications of the same type of measure (from XSozial) are combined to one observation. Measure-specific characteristics of the combined periods are set to system missings.
- Subsequently, the remaining characteristics are compiled and calculated. For this, consolidated individual data from XSozial-histories and LHG are used.
- Finally, the following measures are excluded:
  - one-off benefits (like UBV/Mobi-/Vermittlungsbudget /LES)
  - specific rehabilitation measures

## 2.5 Employee History (BeH)

The source of data regarding employment is the Employee History (Beschäftigtenhistorik - BeH) of the IAB. The data basis is the integrated notification procedure for health, pension and unemployment insurance, which came into effect as of 1 January 1973 (and was extended to cover East Germany as of 1 January 1991) and is known by the abbreviation DEÜV (previously DE-VO / DÜVO) (for further details see: Bender et al. 1996, p. 4 et seq.; Wermter /Cramer 1988). Under this procedure, employers are required to submit notifications to the responsible social security agencies concerning all of their employees covered by social security at least once a year. The BeH covers all white- and blue-collar workers as well as apprentices as long as they are not exempt from social security contributions. This means that civil servants, self-employed persons and regular students[5] (see Cramer 1985) are not recorded in the BeH in principle. Since the notification procedure was changed on 1 January 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until 1 April 1999). The data are recorded by the health insurance companies, collected and edited by the Federal Employment Agency (BA) and subsequently integrated into a history file by the IAB.

The following BeH adjustment processes relevant for the data product are performed:

- To capture a person group that is as constant as possible over time, some person groups for which data are not available throughout the entire observation period are excluded. From the reporting year 2011 onwards, the BeH data originate from newly designed source data. As a result, a number of person groups have been introduced or reactivated as they are classified by the BA statistics as being subject to social security contributions. The person groups 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 and 203 - 205 are therefore contained from that time onwards as well as the two groups 109 and 209, which indicate people in marginal part-time employment. Groups that are not included are, for example, people in short-term employment, i.e. person groups 110, 202 and 210.
- Person groups 123, 124 and 127 have been newly introduced in 2011.
- For data protection reasons, the person groups 107, 111, 113, 114, 127 and 204 are combined to form the person group “other workers” (599).
- From the reporting year 2012 onwards apprentices were included as the new person groups 121 and 122.

[5] Students may still appear in the BeH if, for example, they had a marginal part-time job parallel to their degree course.

- Observations with no details on earnings, and the value 101 for the person group variable, and the value 50 for the reason for notification (annual notification) are not incorporated into the IEB.
- Gender and date of birth are taken from the Data Warehouse (DWH) of the BA. This information is harmonised across data sources.

## 2.6 Benefit Statistics According to SGB II (LST-S)

The Benefit Statistics – Sampling of the IAB (LST-S) contains information on potential and actual benefit receipt, income, sanctions of individuals eligible for benefits and other groups of people in the context of a benefit unit[6]. The data are obtained from the statistics of the Federal Employment Agency (Bundesagentur für Arbeit, BA) from A2LL, ALLEGRO and zPDV / STEP, as well as via the transmission standard XSozial-BA-SGB-II (XSozial) and consolidated across all sources. The "LST-S" process generates datasets with information accurate to the month. The characteristics range concentrates on amounts of money. For members of a benefit unit excluded from benefit receipt, there are generally no financial data available.

The financial data of the LST-S provide information on the potential and actual benefit receipt, income and sanctions. Simply put, it can be said that the potential benefit receipt of one individual is identified first. This need is then reduced by existing income from employment (also self-employment), other transfer benefits as well as capital and assets, if applicable. The benefit eligibility resulting from this calculation frequently corresponds to the actual benefit receipt, but may deviate if sanctions are applied by the job centre which reduce the claim.

For the SIG data product, children without benefit eligibility, i.e. with their own income covering their requirements, are excluded due to the unreliable data quality. The smallest temporal unit is a month. Subsequently, for reasons of data economy, the information of several consecutive months of a person is aggregated or added up whenever all selected financial information of these months is identical.

Linking the LST-S with the SIG core is possible in various ways and depends on the research question. The preferred strategy is to link both datasets on the monthly level. However, this step is computing- and memory-intensive, which is why an adequate amount of time must be scheduled. Moreover, it is strongly recommended to first adjust both datasets – the SIG core and the LST-S – for information not required for the analyses, i.e., limit the datasets to those groups of people, periods of time, and variables actually required for the analyses. This reduces the runtime and prevents the process from failing because of the capacity limit of the FDZ user servers (30 GB).

When linking the datasets, we recommend to first break down both the LHG spells of the SIG core and the LST-S spells to the monthly level, then merge them, and finally aggregate them again to the original periods of time of the SIG core spells.

The following Stata code (see Box 1) shows a variant of how the LST-S can simply be transformed into a monthly representation. As part of this representation, all spells aggregated over large periods of time without content-related changes are separated again and assigned to their respective months.

[6] Statistics of the Federal Employment Agency: statistics on basic benefits for job seekers according to SGB II, Nuremberg, April 2018.

**Box 1: Example code for Stata 18**

```
* Extension of the LST-S-data set according to the respective spell length:
expand spell_length
* Allocation of the aggregated financial information by month:
qui d bedarf* zanspr* zbeink* veink* sankt*, varl
foreach v in `r(varlist)' {
	dis "`v'"
	gen z_`v'=`v'/spell_length
	drop `v'
	rename z_`v' `v'
}
* Generation of a monthly identifier:
bys persnr bg_id beg_mon end_mon: gen count = _n-1
by persnr bg_id beg_mon end_mon: gen int mon_id = beg_mon if _n==1
by persnr bg_id beg_mon end_mon: replace mon_id = mon_id[1] + count if
mon_id==.
format mon_id %tm
drop count beg_mon end_mon
```

# 3 Data preparation and sampling procedure

## 3.1 Sampling

The SIG is a 4 % random sample[7] of all individuals with a record in the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung, LHG). Therefore, the SIG is representative for the group of people from the LHG starting from 2007.

## 3.2 Structure and preparation of the SIG

The SIG constitutes the most comprehensive data basis for the receipt of benefits according to SGB II up to now available through the FDZ. Accordingly, the data focus on the information on periods when benefits were received and on the corresponding characteristics of the individuals and benefit units. In order to be able to map the histories of individuals receiving benefits as broadly as possible, the corresponding biographical information was added from other administrative data products on unemployment and job search, employment, participation in measures of active labour market policy, and receipt of benefits according to SGB III (unemployment benefit I). Due to the institutional structure of the basic benefits system, these additional spells can intersect with

[7] As part of data preparation and anonymising, a few cases will be lost, so strictly speaking, the SIG contains slightly less than 4 % of all potential cases.

the benefit receipt (e.g. employed individuals receiving benefit, “Aufstocker”) as well as occur before or after benefits receipt.

The basic principle of the SIG structure is that there is only one entry (spell) per person and time period. On the one hand, this kind of structure is useful in substantial terms, since there is a dominant source in the data, i.e. the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung, LHG) with the periods of receipt within UB II, which occurs at least once for each individual due to the sample of individuals receiving benefits. Information from other sources is then assigned to this basic information. On the other hand, due to this structure the user does not have to prepare the data across several spells and sources. Moreover, this structure is more space-saving than the provision of all parallel pieces of information in the various individual sources. The final data structure thus provides non-overlapping information on benefit receipt and employment histories accurate to the day on a uniform timeline as, e.g., it is required for duration or sequence pattern analyses.

Since this structure does not correspond to the one in the raw data, the data had to be prepared. This is first done on the basis of the individual data sources and then for the combined data. To do so, decisions had to be made regarding what information is classified as relevant or as a priority and is thus maintained. This procedure will be described in the following sections.

### 3.2.1 SGB II anonymisation

In order to reduce the risk of de-anonymisation, only the year of birth is available in the SIG by default. The month of birth can only be requested as a sensitive variable if there is a justified need for it. However, in the LHG and (X)ASU there is the risk that the exact date of birth may still be obvious due to the chronological structure of the observations. Observations might end systematically on certain birthdays and/or the day before, or start again on the birthday.

In order to prevent an indirect determination of the exact date of birth, the following procedure is applied. Spells separated at sensitive dates of birth in the SGB II context, i.e. if the date of birth triggers a split, receive the 15th day of the respective month as a new spell start date. Accordingly, the end date of the previous spell is set to the 14th day of the month. In rare cases, another – full – spell lies between this date and the 15th day of the month, which is deleted in the course of anonymisation. Parallel episodes of other members of the benefit unit are adjusted accordingly.

Besides the exact date of birth, no other information is lost due to these anonymisation steps. However, this correction results in a slight distortion of the spell duration.

### 3.2.2 Unemployment Benefit II Recipient History (LHG)

- The LHG information is further processed as follows:
- In the LHG, there may be double notifications at the personal level (see Chapter 4.2), which are adjusted for as follows:
    - Double notifications existing in transitional periods of changes of reporting systems are adjusted for using a filter table.
    - In case of parallel spells of an individual who is part of two benefit units at the same time, the following adjustments are made:

  - the spell with the more important position in the benefit unit is kept (main person or partner);
  - if the position in the benefit unit is identical, the one with higher work capability level is kept;
  - if in both spells the information on the work capability level is identical, then the one based on the notification from the BA or a joint facility is kept;
  - if this is identical, too, then the one with the lower benefit unit ID will be kept.
- In case of overlapping ends of two spells, the start of the later spell is set to the day after the end of the previous spell.
- Benefit episodes lasting only a single day are deleted. These are usually one-time benefits granted. Individuals having only one benefit episode which lasts only one day are therefore not included.
- Implementation of a monthly concept: Start and end of benefit episodes are set to the beginning or end of a month, and gaps within one calendar month are closed, since eligibility for benefit receipt applies to the whole month.
- Successive spells are joined, if this is possible without losing information, and data are censored at the beginning and the end of the SIG observation window.

### 3.2.3 Benefit Recipient History (LeH)

From the LeH, only information on the receipt of unemployment benefit is used for the SIG, and no information, for example, on unemployment assistance. In addition to that, only the information on the amount of benefits received and on the reason for deregistration is included.

The episodes – as is the case with the other sources – are adjusted so that only one information is available simultaneously from this source. To do so, the following individual steps are performed:

- One-day episodes are deleted.
- In case of parallel episodes with the same characteristics, only one is kept.
- In case of parallel episodes with different information, the one with the higher amount of benefit receipt and a lower (but valid) reason for deregistration is kept.
- LeH episodes without gaps are joined, if this is possible without losing information, and data are censored at the beginning and the end of the SIG observation window.

### 3.2.4 Job Seeker Histories (ASU und XASU)

The ASU and XASU are processed together. First, the highest school education and professional training level before the observation window of the SIG is saved. Then the following steps are performed to adjust for parallel episodes:

- In case of parallel episodes originating from different sources, the one from the ASU is kept, since better data quality can be assumed here.

- In case of parallel episodes from the same source, the one with the respective lower employment status value is kept (i.e. unemployment trumps job search or non-availability).
- If one of the parallel episodes does not contain information on important variables, but the other one does, only the latter is kept.
- Of the remaining parallel episodes, the one with the longer (anonymized) duration in the original dataset is used.
- Of the remaining parallel episodes beyond that, a random one is kept.
- Subsequently, episodes with a maximum gap of 7 days which do not show any changes in the central variables are joined, and the data are censored at the beginning and the end of the SIG observation window.

### 3.2.5 Participants-in-Measures History (MTH) and Participants-in-Measures History from XSozial (XMTH)

- The MTH and XMTH are processed individually but analogously. The preparation is done in the following steps:
- One-day episodes are removed. These are, for example, moving allowances or application trainings.
- The measures are categorised and then prioritised as follows according to their assumed intensity: (1) subsidized employment, (2) direct job creation measures, (3) further training, (4) occupational choice & vocational training, (5) activation & professional integration, (6) discretionary and other support. In case of parallel episodes, the one with the measure from the category rated as more important is kept. In case of multiple parallel episodes of the same measure category the episode without any missing values is kept in preference. Now there is only a maximum of one episode per partial source and period of time.
- Subsequent episodes without gaps are joined if this is possible without losing information.
- Episodes are censored at the beginning and the end of the SIG observation window.

Afterwards, the two sources are merged and prepared together, so that there is only information for each period of time. In case of parallel episodes in the MTH and XMTH, the one from the MTH is kept, since better data quality can be assumed here. If one episode is included in another one, the internal episode is deleted and the longer one is kept. In case of overlapping episodes, the data from the XMTH episode are corrected so that the MTH parts are kept and the XMTH parts are deleted. Then successive episodes are joined, if this is possible without losing information.

### 3.2.6 Employee History (BeH)

Only BeH spells with a start date as of 1 January 1993 are included in the data preparation, since data from pension insurance providers for Eastern Germany have been available in evaluable quality since 1993. First, the episodes of full pensioners (ieb_erw_stat= 119, 149) and pensioners with assumed employment (120) are removed.

The target data structure, which permits only one entry per period of time and source per individual, requires the main employment to be identified in case of several parallel or overlapping em-

ployment episodes. In case of parallel employment episodes, the one that is subject to social security contributions or the one with a higher daily wage is kept as the main employment. In addition to that, some central information on possible second (parallel) employment relationships subject to social security contributions and marginal employment relationships is saved in separate variables.

In the next step, missing annual notifications (for the process, see FDZ methods report 6/2007) are be added if the BeH notification ends on 31 December, and one year later, there is again a notification on 1 January, and the establishment ID, employment status, and position in the job are identical.

Next, episodes that have a maximum gap of seven days and where occupation, wage, job position, employment status, economic sector, and secondary employment do not change, are merged. Very short episodes of up to seven days are then removed if they have no other adjacent spells within seven days before and after.

For analyses, it may be desired to have information on employment before the observation period of the SIG starting in 2007. Therefore, some information on work experience and the last employment are saved between 1 January 1993 and the beginning of the SIG observation window. The episodes are joined, to the extent that this is possible without loss of information.

### 3.2.7 Benefit Statistics According to SGB II (LST-S)

The LST-S financial data of several consecutive months are aggregated or added up for the SIG whenever all selected financial information of these months is identical. The smallest temporal unit in the end product is thus one month, and every piece of information is valid for at least one month, not for shorter periods of time.

In order to reduce the risk of de-anonymisation, especially of exact dates of birth, various steps are conducted in the LST-S: First, some characteristics are provided in dichotomous form only to minimise the risk associated with the combination of rather unusual financial flows and their amounts. This information is not lost, though, but rather allows for the identification of certain additional requirement cases within a given period of time. These are cases of pregnancy, disability, or special, cost-intensive dietary requirements due to medical necessity (see Section 21 SGB II). Moreover, this kind of dichotomous indicator can be used to identify the granting of loans instead of bonuses – but not their amounts.

Furthermore, all benefit information on income, potential and actual benefit receipt, and sanctions is distorted by adding random integers between -10 and +10 and are subsequently rounded up to the nearest multiple of 10.

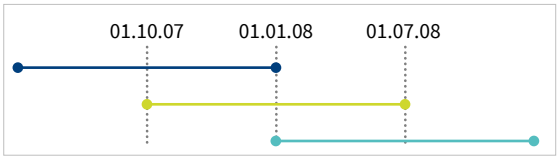
## 3.3 Episode splitting

The administrative individual data are available as “split” episodes. If observations overlap within an account, they are replaced by artificial observations with new dates so that completely parallel episodes and non-overlapping episodes are created. This increases the number of observations (see Figure 3). In the SIG, this episode splitting is performed as follows: First, as in the original sources the individual observations are provided as split episodes, if individual characteristics

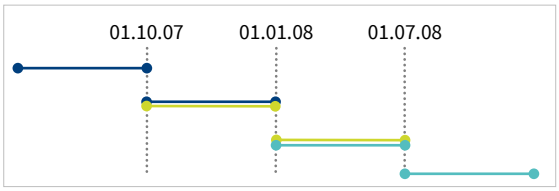
change, the spells are separated in the respective source as of the date the new characteristic manifestation occurs. In the preparation of the separate sources (see above), spells are aggregated again if the split was due to a characteristic that is not included in the SIG. Moreover, short interruptions of a maximum of seven days (BeH, (X)ASU) are closed. In the LHG, spells of basic benefit receipt starting or ending during a given month are set to the respective start or end of the month.

In a second step, the individual sources are joined and split to reach the desired data structure of one spell per individual and period of time. If there are changes within one source, the information on the other sources will also be split. For example, if the unemployment status of an individual changes from unemployed to job-seeking ((X)ASU source) during a consecutive period of benefit receipt within basic income support (LHG source), the spell of this individual is split at this point. This results in two parallel spells for which all relevant information is then transferred to one spell, and the second spell is deleted. The information stemming from the different data sources is neither consolidated nor further substantially checked for plausibility. This kind of end (or start) generated by splitting does not necessarily mean that a consecutive episode actually ends (or starts) with regard to each of the sources. For the identification of new starting (or ending) episodes of benefit receipt, unemployment, participation in measures, or employment, it must be verified in each case whether the respective status also exists prior to that (afterwards). For which sources the respective spell contains information can be seen in the "quelle-*"-variables.

**Figure 3: Episode splitting**




| persnr | begorig | endorig | quelle_lhg | quelle_asu | quelle_beh |
|---|---|---|---|---|---|
| 4711 | 01.01.07 | 31.12.07 | 1 | | |
| 4711 | 01.10.07 | 30.06.08 | | 1 | |
| 4711 | 01.01.08 | 31.12.08 | | | 1 |




| persnr | begepi | endepi | quelle_lhg | quelle_asu | quelle_beh |
|---|---|---|---|---|---|
| 4711 | 01.01.07 | 30.09.07 | 1 | | |
| 4711 | 01.10.07 | 31.12.07 | 1 | 1 | |
| 4711 | 01.01.08 | 30.06.08 | | 1 | 1 |
| 4711 | 01.07.08 | 31.12.08 | | | 1 |

## 3.4 Cleanup of inconsistencies

Various - necessary - adjustments in the episode dates may in some constellations lead to certain variables within a benefit unit being affected by slight inconsistencies. This concerns in particular the type of benefit unit as well as the correct identification of newborns.

In both cases, the inconsistencies are eliminated by means of small adjustments. For the 1st day of each episode (of a person within a benefit unit), a new benefit unit type variable is created, which is used in the few cases of a deviation of the original variable. Prior to this – with respect to the 1st day as the cut-off date – episodes of newborns are removed if due to data preparation or anonymization their date of birth is after the 1st day or even after the end of an episode.

## 3.5 Missing values

In the data, missing values are coded as follows:

**Table 4: Missing values**

| Term | Value | Description |
|---|---|---|
| No (valid) details available | .z | Values of a variable that are not systematically missing, i.e., the variable is available in principle for the data source, but no details are available for the value considered or cannot be interpreted reasonably. |
| Systematically not available | .n | A variable is not available in principle for a data source or is not available for a certain period. |
| Value implausible (LST-S only) | .p | The plausibility of the benefit data transferred from the LST-S is questioned as the result of a BA statistics check in coordination with the providers of basic benefit and thus set to the “missing” value of .p for the SIG. |

# 4 Data quality and problems

## 4.1 Introduction of SGB II and subsequent institutional changes

With the introduction of the SGB II on 1 January 2005, the responsibilities for the support of job seekers were redesigned, so that tasks in the sphere of the SGB II can be taken over by different types of providers:

- Usually, the Arbeitsgemeinschaften (ARGE) between BA and the district took over the responsibility according to the SGB II. These were replaced in 2011 by the joint facilities (gE) - also known as "Job Centers".
- In addition, until the end of 2011 it was possible for the BA and the districts to perform the tasks assigned to them in their own (separate) responsibility (gT and gAw).
- Since 2005, it has also been possible for authorised municipalities (zkT; also: opting municipalities) to take over the responsibilities. Initially, 69 administrative districts or independent towns took sole responsibility for the basic provision for job seekers until 31.12.2010. Since the constitutional amendment in 2012, the option has been extended to 110 providers.

While the ARGEn/gE maintain benefit and case management via the BA procedures and transfer the recorded data to the statistics department of the BA, the municipal providing institutions responsible for implementing SGB II each use their own software systems. The transfer to the statistics department of the BA takes place via the XSozial standard. A division of responsibilities existed for separate providers. Any given provider was only responsible for the collection and delivery of the data falling within its area of responsibility. For this reason, there were reduced data requirements for the municipal side of the separate providers. Transmission via the XSozial standard takes place once a month at a fixed time window. In the following week, there is the possibility of repeating failed deliveries on a second reporting day. Over the years, numerous quality assurance instruments were developed. Nevertheless, version changes of the standard or the software used on site regularly increase the risk of data quality problems.

As part of the reorganisation of SGB II institutions on 1 January of 2011 and of 2014, various ARGEn/gE and separate providers were transferred to municipal providers or converted from zkT to gE. This change of responsibility and the associated change of the software used locally also lead to some breaks in the data of the SGB-II sources. Detailed information on the resulting quality problems in the data can be found in the individual chapters on the data sources.

## 4.2 Unemployment Benefit II Recipient History (LHG)

- Longitudinal analyses for individuals comprise uncertainties, since monthly underrecording with regard to benefit receipt due to data delivery issues or failures at the providers' level is possible. This means that for individuals living in or moving to such regions, any non-observable benefit receipt can be attributed to delivery issues, or there was actually no receipt at all.
- Cases of underrecording occur at times. These generally last one month and occur mainly in the authorised municipalities.
- The gap variable can identify whether such gap may have existed prior to or after a spell.
- Underrecording and overrecording occur in connection with changes in the type of institution responsible for implementing SGB II:
  - In the context of the reform of the territories covered by the institutions, which came into force on 1 January 2011, cases of underreporting occurred in the districts covered by the employment agencies of Dessau-Roßlau, Halberstadt, Halle and Sangerhausen.
  - Double notifications due to the territorial reforms in 2009/2011 and the changes in the form of the institutions as of 1 January 2012 and of 1 January 2013 are already corrected as far as possible in the IEB. Nonetheless double notifications may still occur.
- In the following job centres there are inaccuracies with regard to the allocation of benefit cases:
  - between Emden and Norden between September and December 2009
  - between Döbeln and Mittelsachsen from October to December 2012
  - between Tirschenreuth and Wunsiedel from November 2012 to March 2013
- Some individuals for whom a (X)LHG spell exists are excluded entirely or partly from benefit receipt according to SGB II, for instance because they take part in a subsidised training programme, receive an old-age pension, live in an in-patient facility or a residential institution or receive insurance payments aimed at prevention of need in assistance. This affects on average 3 to 5 percent of all cases. In XSozial this person group is sometimes underrecorded by some institutions. Exclusion from benefits cannot be identified in the SIG.
- Due to the reporting logic, information from the XSozial transmission standard can only be updated monthly.
- In the official performance statistics of the BA, reporting gaps are supplemented by a statistical estimation procedure at an aggregated regional level. No supplementary data sets are provided in the LHG. A comparison of the IEB with the performance statistics is therefore only possible to a limited extent.

- A peculiarity of basic benefit data is that although these are individual data, the type and amount of benefit receipt are not always determined individually. Instead, applicable law stipulates that individuals within a benefit unit are willing to "bear responsibility for each other and support each other" (Section 7 (3) SGB II). This means that changes in the benefit unit may affect the entitlement to benefits of all individuals who belong to the respective basic benefit case.

  Multiple benefit unit numbers may be assigned to one benefit unit, which was adjusted for as part of data preparation (see Chapter 3.2.2). The possible causes are following:

  - Relocation to an area of a different provider of basic benefits. Since the providers are legally distinct from one another, the cases are managed partly using the common IT process A2LL, but must be kept separate in terms of function. Therefore, the provider number is part of the benefit unit number (the first five digits).
  - Provider number changes due to reorganisation of the provider or the assigned employment agency result in new customer and benefit unit IDs without any relocation.
  - Technical reason in the IT process, e.g. in case of data migration with the issue of a new number, or correction of a wrong notification. Switching from A2LL to ALLEGRO will also result in a change of the benefit unit number.
  - Inflows or outflows of benefit units over time or changes of benefit units of individuals are not reliably identifiable. Studying the sampling of benefit units at a given point in time is mostly unproblematic with the benefit unit ID.
  - If children live only temporarily in the household, e.g., because they regularly live with one of their separated parents on a rotating basis, they can theoretically belong to two benefit units at the same time.

## 4.3 IEB

### 4.3.1 Gaps in employment histories

The IEB contain information on various employment statuses. However, the administrative data comprise not every type of employment. Additionally, there is no comprehensive information on other biographical conditions (e.g. with regards to educational periods or parenting). This leads to individual gaps in the IEB. For evaluation purposes, however, it might well be interesting to have information on these gaps in the IEB (e.g. to create control groups, for life course analyses, etc.). For some of these gaps, the "notification reason/termination reason/reason for SGB II discontinuation/deregistration reason" characteristic of the predeceasing episode (if any such episode exists) may provide information on the kind of gap (e.g. if it contains information that the termination of an unemployment episode was due to taking up self-employment).

### 4.3.2 Job seeker histories (ASU/XASU)

#### 4.3.2.1 ASU

- With the introduction of SGB II on 1 January 2005, jobseekers are no longer fully covered by BA procedures. From this date, the ASU only covers persons who are supported by the BA in the

sphere of the SGB III (employment promotion) or by ARGEn, gE or gT in the sphere of the SGB II (basic security).

- For the placement staff it is not always possible to allocate to the legal sphere immediately, since it is frequently only clear which institution is primarily responsible after a certain time due to a possible entitlement to SGB II benefits.
- For some individuals for whom an authorised municipality has been responsible since 2005, parallel "artificial" ASU episodes were created by the Federal Employment Agency. These can be identified via the variable estatvor (transfer to an authorised municipality).
- The coArb procedure, which was used until June 2006, supported only the placement of unemployed persons and jobseekers. Some data were also collected about individuals who were only seeking advice, but these data are incomplete. The careers advice data were collected in a separate system. In VerBIS the attributes of the job-search status were extended to include ‘seeking advice’ and individuals ‘without status’. The latter group includes individuals eligible for Unemployment Benefit II who are only available for job placement to a limited degree. The recording of this group in VerBIS is only regarded as largely complete since January 2008.
- A change of the institution responsible for implementing SGB II or a change of place of residence does not lead to a new ASU observation, the value of the variable at the start of an episode is continued. The longer the observation becomes, the greater the risk is that the institution responsible or the place of residence is no longer correct.
- Since spring 2011, the jobseeker data from BA procedures and XSozial have been consolidated in the integrated unemployment statistics. This may lead to larger deviations. At the current margin, however, the stock data are identical.

#### 4.3.2.2 XASU

- Since the XASU also contains LHG episodes as a basis for the notifications for non-job seekers according to Section 10 SGB II, individual months are affected by delivery gaps. Besides, there is underrecording due to technically induced wrong cancellations that cannot be completely traced back with the de-cancellation processes applied.
- Overreporting may also occur at certain points. A large number of technical problems can be responsible for this, including systematically missing deregistrations, incorrect reversals of cancellations and parallel job search notifications in several districts.
- A variety of variables sometimes have only a very low degree of completeness for the XASU. Variables which are affected by this include 'School leaving qualification', 'Reason of notification' as well as 'Employment status prior to job search'. Although the degree of completeness of these variables improves over time, some of them are still unsatisfactory. The 'Occupation - current/most recent' variable is not available in the XASU for almost the entire period available.
- For a number of institutions (districts), the proportion of registered recipients of unemployment benefit II who are also registered jobseekers is implausibly large at times or continuously in the IEB. One possible reason for this could be an incorrect determination of the status 'not unemployed but seeking work' by these institutions.

- The institution-related and time-dependent plausibility of the XASU data should be examined before use, taking the research question into account.
- Differences in consolidation rules, time references and regional assignments may result in differences to the published BA statistics.

### 4.3.3 Participants-In-Measures History Files (MTH/XMTH)

#### 4.3.3.1 MTH

- The MTH only contains notifications that are recorded in BA procedures. The use of these procedures in cooperations of employment agencies and municipalities/separated responsibilities/municipalities exercising their duties separately increases continuously between 2005 and 2007. The notifications for these institutions are complete from March 2007 onwards. Measures that are reported by authorised municipalities via the XSozial standard are contained in the XMTH.
- Because of the reorganisation of the institutions responsible for implementing SGB-II in 2011 to 2014, a split of the documentation of participations in measures in the MTH and in the XMTH may occur when there is a change in the reporting procedure.[8] This might result in a split or a duplication of the spell of the measure (see Section 2.4).
- In the case of notifications regarding the bridging allowance (Überbrückungsgeld) the maximum permissible duration of six months is sometimes exceeded. In most cases this can be explained by a default setting in the input mask of the data recording system.
- The MTH is supplemented by applicant characteristics (e.g., vocational training) from other BA procedures. For these variables the administrative procedure was switched from coArb to VerBIS in 2006. The same quality limitations as for the ASU apply here.

#### 4.3.3.2 XMTH

- About 13 % of all providers of basic benefits reported almost no municipal integration benefits (formerly supporting benefits) between 2005 and the beginning of 2017, which are included, among other things, in the summarised “Other support” category. Many other Jobcenters reported only occasionally and/or only selected types. This leads to under-recording with regards to “Other support”.
- Overregistration of participations: Total stocks are considered to be stable from the end of 2008. However, the majority of the providers still register notification profiles which are temporarily conspicuous and which raise doubts about the quality. Only the introduction of the XSozial promotion-ID in 2009 and the meanwhile several years of experience of all participants stabilise the reporting process and the subsequent data processing. Also the new zkT, introduced in 2012, report inconspicuously for the most part, so that the scope and duration of presumed overrecording decrease noticeably.
- The reorganisation of SGB II providers in 2011-2014 (see Section 4.1) lead to a split in the documentation of participation in measures in MTH and XMTH if the reporting procedure was

[8] Further information concerning the territory structure of the institutions responsible for implementing Social Code Book II and relevant changes is available at https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Regionale-Gliederungen/Gebietsstruktur-Traeger-Grundsicherung-Nav.html

changed. This could result in a split of the measure spell, but also in duplications (see Section 2.4).

- The figures or individuals obtained from the monthly cutoff counts in the XMTH differ in several ways from the published BA statistics due to extensive consolidation work.

### 4.3.4 Employee History (BeH)

- Information on vocational training, the occupation/activity performed and the occupational status is transmitted by means of notifications made by the employer in accordance with the Data Collection and Transmission Regulation (DEÜV) (see Section 2.5) using a so-called occupation code. The new occupation code 2010 was adopted for notifications with an end date later than 30 November 2011 (for further details, see Bertat et al., 2013). The decision to switch to the new occupation code was made by the central organisations of the social security agencies as a number of facts could no longer be recorded in a way that was up-to-date and realistic using the occupation code 2003. As the notifications made by employers in accordance with DEÜV only enter the Employee History (BeH), the change of the occupation code only affects observations from this source. The measurement of the following characteristics previously reported using the occupation code 2003 is affected by this change: working hours, occupation, occupational status and school and vocational qualification level. In addition, since the switch to the new occupation code, details are also available about whether an employment relationship is fixed-term and whether a person is employed by a temporary work agency to be hired out to other firms. The most important consequence is the switch to a new occupational classification. Instead of the previous Classification of Occupations 1988 (Klassifikation der Berufe 1988 (KldB 1988)), the more highly differentiated KldB 2010 is reported with the new occupation code.[9]
- The introduction of the new occupation code in 2011 led to a number of problems. For example, during the transition period granted to employers in the social security notification procedure,[10] there was a temporary increase in the number of missing details. Analyses of the BA statistics (Bertat et al., 2013, p. 10) show that in 20 to 30 percent of cases no information was available in the new or converted variables “occupation/activity performed”, “working time” and “vocational education and training” after the switch. This situation began to improve significantly in the first half of 2013. In order to improve the quality of the “working time” variable in the transition period, Ludsteck/Thomsen (2016) developed an imputation procedure to replace the missing values by imputed values. The imputed data are included since SIG 0720. No imputation is performed regarding the gaps in the other variables.
- Within the employment notification procedure, a certain time lag is unavoidable. Although changes in employment relationships have to be reported immediately, and existing employment relationships have to be confirmed annually by April (until the annual notification 2012) or mid-February (from the annual notification 2013 onwards) of the following year, some notifications actually arrive years later. The History File of the IAB is not updated continuously, however, but at certain intervals. This is done using files of employment notifications for one

[9] See Paulus/Matthes, 2013, for details regarding the Classification of Occupations 2010.

[10] The test programs used in the notification procedure permitted missing details in the occupation code 2010 until the end of May 2012.

particular year which were submitted 36, 18, 12 or 6 months after the end of the reporting year (e.g. the 18-month file for 2013 can be created in July 2015 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-month file shows a 100 % degree of completeness by definition.

- In the underlying IEB, the year 2014 is the last year with a degree of completeness of BeH observations of 100%.[11] For the years 2015 and 2016, the 18-month files were used, and the observations for 2017 originate from a 6-month file. It can therefore be assumed that employment notifications for 2015 and 2016 are slightly underreported, and that those for 2017 are underreported to a slightly higher degree. However, this should not reduce the ability to analyse the data at individual level. The missing notifications occur more frequently in a few establishments, however. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.

## 4.4 Benefit Statistics according to SGB II (LST-S)

Evaluations on provider or district level are partly limited due to temporary delivery failures. On the one hand, the two attributes “Previous notification gap” and “Subsequent notification gap” must be considered, which show general underrecording. Additionally, the statistic of the BA observes the aggregate values of a notifying provider. If these are implausible with regard to the subject, all episodes of the provider are labelled accordingly. In the SIG, these periods of time are set to the missing value .p and therefore provide the precondition for project-specific imputations.

The significance for regional and temporal comparisons of the costs of accommodation has been considerably limited since mid-2016, since benefits for accommodation of refugees living in collective accommodation facilities were partly claimed collectively for a possibly longer period of time in the past and were not always assigned to the corresponding entitlement months by the notifying provider, but were allocated to one month only. This also affects the totals of pecuniary claims and total standard benefits (Federal Employment Agency, 2017).

In addition, the time periods from LHG and LST-S do not always match completely. Thus, it can be the case that due to necessary alterations of the data, e.g. for anonymization purposes, an LHG episode either occurs before or ends after the corresponding LST-S spell. For an approximate quantification of the phenomenon: For almost 9 percent of observations (at the monthly level) with information from the LST-S, there is no LHG equivalent in the current SIG version, i.e., no LHG spell that includes at least one day of the respective month. Conversely, for 3 percent of the observations with LHG information (at the monthly level), no LST-S equivalent exists for the respective month.

Amounts of exactly 0.00 euros may be specified not only if there is actually no need for assistance, but also for other reasons. This is the case if the type of need for assistance was not yet or no longer valid at the time of notification or if (technical) restrictions or poor data quality prevent capturing. Cases with a total need for assistance of 0 euros are excluded from the LST-S file for the SIG.

[11] Due to a redesign of the data basis of the BeH, at the time of the preparation of the BeH as an exception only the 30-month files was available instead of the 36-month files. However, analyses with earlier data versions have shown that, as a rule, the 18-month file already has a degree of completeness of around 99%. This means that in the next 18 months there will only be extremely few follow-up or correction notifications.

The validity of characteristics is first determined by legal validity; legally, the subjects are only possible within certain temporal limits. However, the data are not consolidated by their legal validity. For example, individual cases of unemployment benefit II bonuses are reported even after 2010. The LST-S financial data may also be of varying quality in individual cases, regardless of the provider-specific plausibility. Above all, after introducing new benefit types it must be expected that the quality will stabilise only over time. The characteristics are not consolidated in this regard, but correspond to the notification flow. Their quality thus pretty much affects the quality of the total values published in the SIG.

The amount characteristics on “Education and Participation” (Bildung und Teilhabe, BuT) have special quality limitations and are thus included in the SIG as indicators for the need for assistance only. They have only been published since mid-2015 with retroactive effect as of 2013. During this period of time, the number of notifying providers was rising continually, but at the time of the first publication, there were still some providers that did not report BuT cases. The data are not suitable for a regional comparison, since additional comprehensive information on regional offers would be necessary to do so (e.g. travel costs allowances from the federal states, municipal contribution to lunchtime meals, etc.) (for further information on the general quality, see Federal Employment Agency (2015)).

Trainees whose vocational training is eligible for support according to the Federal Training Assistance Act (Bundesausbildungsförderungsgesetz, BAföG) or as part of Vocational Training Assistance (Berufsausbildungsbeihilfe, BAB) were actually excluded from benefit receipt according to SGB II up to and including August 2016. However, under certain conditions, they may claim benefits according to SGB II if their requirements cannot be covered otherwise. Trainees may thus receive payments for additional requirements, cost of accommodation bonuses, loans for standard requirements and contributions for health and nursing care insurance as well as benefits for rental arrears. None of the benefits granted to trainees are considered unemployment benefit II. The statistical representation does not differentiate between the individual types of benefits, but shows the total value for benefits for trainees.

No financial data are available for individuals completely excluded from benefit eligibility. When average amounts are calculated per benefit unit, this must be considered.

The figures of pecuniary claims are obtained differently depending on the source procedure: From ALLEGRO, consolidated notifications of requirement-specific pecuniary claims are submitted, and for XSozial data, the figures are calculated during statistics preparation from requirement, income, and sanctions notifications. A2LL data are calculations directly from the notification flow of payment instructions. That is why negative figures are possible in this context.

No imputed datasets for delivery failures are available in the LST-S, only the original notifications. Therefore, the aggregate data extrapolated to population differ from the published data of the statistic of the BA in months with delivery failures.

# 5 Description of variables

Frequency counts and overviews of the individual values and labels of the variables can be found in separate files under https://fdz.iab.de/en.aspx. Further detailed overviews of the individual variables (labels, counts, missing values) can be found in the online appendix to the data report.

## 5.1 Identifiers

### 5.1.1 Unsystematic individual ID (persnr)

| Category | Description |
|---|---|
| Variable label | Unsystematic individual ID |
| Variable name | persnr |
| Category | Identifiers |
| Origin | BeH, LeH, LHG, ASU, XASU, MTH, XMTH, LST-S |
| Data type | Numerical |
| Detailed description | The artificial individual ID indicates which observations belong to the same person. Artificial means that it is not possible to infer any of the person's characteristics or any original identifiers from this individual ID.<br>As there is no uniform individual identifier in the different data sources, the allocation of the information from different data sources (e.g., employment and benefits) to individuals is not always unambiguous. In such cases, implausible employment histories may arise.<br>The formation of the individual identifier which spans all data sources is based on a heuristic developed by the BA. |

### 5.1.2 Unsystematic establishment ID (betnr_sig)

| Category | Description |
|---|---|
| Variable label | Unsystematic establishment ID |
| Variable name | betnr_sig |
| Category | Identifiers |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The establishment ID indicates which observations belong to the same establishment. To make the data anonymous, the original establishment numbers allocated by the Federal Employment Agency are replaced by randomly generated, yet unambiguous establishment IDs (further information on the allocation of establishment numbers by the BA can be found in Bender et al. (1996: p. 15 et seq. and pp. 27-30) as well as directly on the website of the establishment number service of the BA at https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige). The establishment number and year specification can be used to merge individual and establishment information.<br>For the establishment number, the following should be taken into account in general:<br>If the company has only one office, or if the company has only one office in one municipality, this office is the establishment and is given an establishment number.<br>If the company has **several** branch offices in **one** municipality, these establishment premises / workplaces must be merged into a single establishment under one establishment number, if they belong to the same economic class. If they do **not** belong to the same economic class, each branch office is regarded as a separate establishment and is given its own establishment number.<br>If the company has **several** branch offices in **several** municipalities, each of these branch offices is an establishment and is given its own establishment number.<br>In this context, the following definitions with regards to the allocation of establishment numbers as part of the notification procedure for social security must be taken into account:<br>An **establishment** is a regionally and economically delimited unit in which employees work and which is allocated an establishment number according to the above-mentioned principles. |

| Category | Description |
|---|---|
| | A **workplace** is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles.<br>A **company** as a term combines establishment premises and workplaces belonging to the same employer.<br>An **employer** is any natural person or legal entity that employs at least one employee subject to social security contributions or in marginal part-time employment.<br>Establishment and establishment premises are synonyms; branch office is a synonym for subsidiary, district office, out-sourced office, workplace etc. if it is not an establishment. |
| Notes on quality | The establishment ID is only missing in a very small number of cases. These observations are notifications for the person group “205” (earnings notifications for casual workers). As establishment variables (place of work, economic activity, establishment size etc.) are merged via the establishment ID, they are missing in these observations. |

### 5.1.3 Unsystematic benefit unit ID (bg_id)

| Category | Description |
|---|---|
| Variable label | Unsystematic benefit unit ID |
| Variable name | bg_id |
| Category | Identifiers |
| Origin | LHG, LST-S |
| Data type | Numerical |
| Detailed description | The artificial benefit unit number shows the datasets belonging to the same benefit unit. Artificial means that this benefit unit number is not indicative of properties of the individuals/benefit unit or original identifiers. |
| Notes on quality | Analysing benefit units at a given point in time is mostly unproblematic with the benefit unit number. Inflows or outflows of benefit units over time or changes of benefit units of individuals are not reliably identifiable. |

## 5.2 Generated technical variables

### 5.2.1 Source (quelle_lhg, quelle_asu, quelle_xasu, quelle_leh, quelle_mth, quelle_xmth, quelle_beh)

| Category | Description |
|---|---|
| Variable label | Source |
| Variable name | quelle_lhg, quelle_asu, quelle_xasu, quelle_leh, quelle_mth, quelle_xmth, quelle_beh |
| Category | Generated technical variables |
| Origin | LHG, ASU, XASU, LeH, MTH, XMTH, BeH |
| Data type | Numerical |
| Detailed description | These characteristics specify the data sources from which information is included in the spell. |

### 5.2.2 Previous reporting gap (lck_links)

| Category | Description |
|---|---|
| Variable label | Previous reporting gap |
| Variable name | lck_links |
| Category | Generated technical variables |
| Origin | LHG; SIG calculation |
| Data type | Numerical |

| Category | Description |
|---|---|
| Detailed description | This characteristic indicates whether the reporting provider of basic benefits had data submission problems prior to the start date of the spell. This resulted in a notification gap in the existing data. The start date may therefore be incorrect. |

### 5.2.3 Subsequent reporting gap (lck_rechts)

| Category | Description |
|---|---|
| Variable label | Subsequent reporting gap |
| Variable name | lck_rechts |
| Category | Generated technical variables |
| Origin | LHG; SIG calculation |
| Data type | Numerical |
| Detailed description | This characteristic indicates whether the reporting provider of basic benefits had data submission problems after the end date of the spell. This resulted in a notification gap in the existing data. The end date may therefore be incorrect. |

### 5.2.4 Spell left-censored 1.1.2007 (zensierung_links_lhg, zensierung_links_x_asu, zensierung_links_leh, zensierung_links_x_mth, zensierung_links_beh)

| Category | Description |
|---|---|
| Variable label | Spell left-censored 1.1.2007 |
| Variable name | zensierung_links_lhg, zensierung_links_x_asu, zensierung_links_leh, zensierung_links_x_mth, zensierung_links_beh |
| Category | Generated technical variables |
| Origin | SIG calculation |
| Data type | Numerical |
| Detailed description | This characteristic indicates whether a spell was left-censored effective 1 January 2007 (according to source). |

### 5.2.5 Spell length (spell_length)

| Category | Description |
|---|---|
| Variable label | Spell length |
| Variable name | spell_length |
| Category | Generated technical variables |
| Origin | LST-S |
| Data type | Numerical |
| Detailed description | This variable indicates the number of months the current LST-S spell comprises. It can be used to reverse the data structure back to the monthly representation (as suggested in Section 2.6). As part of this representation, all spells aggregated over large periods of time without content-related changes are separated again and assigned to their respective months. |

## 5.3 Period of validity

### 5.3.1 Episode start date (begepi)

| Category | Description |
|---|---|
| Variable label | Episode start date |
| Variable name | begepi |

| Category | Description |
|---|---|
| Category | Generated period of validity |
| Origin | SIG calculation |
| Data type | Date |
| Detailed description | The start date of the split episode is always equal to or greater than the start date of the original observation (see also the comments on episode splitting in Section 3.3). |

### 5.3.2 Episode end date (endepi)

| Category | Description |
|---|---|
| Variable label | Episode end date |
| Variable name | endepi |
| Category | Generated period of validity |
| Origin | SIG calculation |
| Data type | Date |
| Detailed description | The end date of the split episode is always equal to or smaller than the end date of the original observation (see also the comments on episode splitting in Section 3.3). |

### 5.3.3 Start date of spell (beg_mon)

| Category | Description |
|---|---|
| Variable label | Start date of spell |
| Variable name | beg_mon |
| Category | Generated period of validity |
| Origin | LST-S; SIG calculation |
| Data type | Date (month) |
| Detailed description | This characteristic determines the start of an episode from the LST-S in accordance with the summarising SIG-internal calculation for months with unchanged benefits statistics on the benefit unit level. |

### 5.3.4 End date of spell (end_mon)

| Category | Description |
|---|---|
| Variable label | End date of spell |
| Variable name | end_mon |
| Category | Generated period of validity |
| Origin | LST-S; SIG calculation |
| Data type | Date (month) |
| Detailed description | This characteristic determines the end of an episode from the LST-S in accordance with the summarising SIG-internal calculation for months with unchanged benefits statistics on the benefit unit level. |

## 5.4 Personal information

### 5.4.1 Sex (frau)

| Category | Description |
|---|---|
| Variable label | Sex |
| Variable name | frau |

| Category | Description |
|---|---|
| Category | Personal variable |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | Gender dummy (0 - man, 1 - woman). The gender information is constant within one individual account. |
| Notes on quality | The information on gender was overwritten from the information on benefit receipt (LHG) to all spells of a person. |

### 5.4.2 Year of birth (gebjahr)

| Category | Description |
|---|---|
| Variable label | Year of birth |
| Variable name | gebjahr |
| Category | Personal variable |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | The year of birth is constant within one individual account. |
| Notes on quality | In the original data, it may happen that the date of birth changes between the data sources. This is corrected during the data preparation process. The information from the social security number is given highest priority here. |

### 5.4.3 Month of birth (gebmon)

| Category | Description |
|---|---|
| Variable label | Month of birth |
| Variable name | gebmon |
| Category | Personal variable |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | The month of birth is constant within one individual account.<br>One can use the variables Year of birth (gebjahr) and Month of birth (gebmon) to generate a variable in the date format JJJJmM (e.g., 1984m6) with the following syntax in Stata:<br>`gen int gebdat = ym(gebjahr, gebmon)`<br>`format gebdat %tm` |
| Notes on quality | In the original data, it may happen that the date of birth changes between the data sources. This is corrected during the data preparation process. The information from the social security number is given highest priority here. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (gebjahr) is provided. |

### 5.4.4 Nationality (nation_lhg, nation_x_asu, nation_leh, nation_x_mth, nation_beh)

| Category | Description |
|---|---|
| Variable label | Nationality |
| Variable name | nation_lhg, nation_x_asu, nation_leh, nation_x_mth, nation_beh |
| Category | Personal variable |
| Origin | LHG, ASU, XASU, LeH, MTH, XMTH, BeH |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | The variable contains the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2019). |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (nation_gr) is provided. |

### 5.4.5 Nationality, grouped (nation_gr_lhg, nation_ gr_x_asu, nation_ gr_leh, nation_gr_x_mth, nation_ gr_beh)

| Category | Description |
|---|---|
| Variable label | Nationality, grouped |
| Variable name | nation_gr_lhg, nation_ gr_x_asu, nation_ gr_leh, nation_gr_x_mth, nation_ gr_beh |
| Category | Personal variable |
| Origin | LHG, ASU, XASU, LeH, MTH, XMTH, BeH |
| Data type | Numerical |
| Detailed description | The variable contains a grouped version of the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2019). |

### 5.4.6 Marital status (famst_lhg, famst_x_asu, famst_leh)

| Category | Description |
|---|---|
| Variable label | Marital status |
| Variable name | famst_lhg, famst_x_asu, famst_leh |
| Category | Personal variable |
| Origin | LHG, ASU, XASU, LeH |
| Data type | Numerical |
| Detailed description | This variable describes the marital status. It is not filled for employment episodes.<br>**1) LeH**<br>In the LeH, the variable has only two values (0 - not married, 1 - married).<br>**2) LHG, ASU, XASU**<br>In the sources LHG, ASU and XASU a distinction is made between six values (values 11-16).<br>The information from the different sources was not compared. |
| Notes on quality | **1) LeH**<br>From year 2014 onwards, the data quality of this variable should be regarded as questionable. For example, the proportion of married persons drops significantly between 2013 and 2017.<br>**2) LHG, XASU, XMTH**<br>Due to deviating reporting standards in XSozial, the data from the zkT should be considered less valid until December 2009. |

### 5.4.7 Age youngest child categorized (alter_juk_kat)

| Category | Description |
|---|---|
| Variable label | Age youngest child categorized |
| Variable name | alter_juk_kat |
| Category | Personal variable |
| Origin | LHG |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | This characteristic specifies the age of the youngest child (below the age of 15) in the following categories:<br>0 no child under the age of 15<br>1 0 to 2 years<br>2 3 to 6 years<br>3 7 to 14 years |

### 5.4.8 Vocational training (ausbildung_x_asu, ausbildung_beh)

| Category | Description |
|---|---|
| Variable label | Vocational training |
| Variable name | ausbildung_x_asu, ausbildung_beh |
| Category | Personal variable |
| Origin | ASU, XASU, BeH |
| Data type | Numerical |
| Detailed description | This variable contains the vocational training qualification. A classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ. It must be taken into account that this variable has a different meaning depending on the data source:<br><br>**1) BeH**<br>For observations obtained from the BeH, the variable contains the vocational education reported by the employers as part of the employment notification procedure. The following values exist:<br><br>1 Without vocational training<br>2 In-company voc. training/traineeship/external (on-school) voc. training<br>11 University of applied sciences without further specifications<br>12 University without further specifications<br><br>In notifications that rely on the new occupation code (see Section 4.3.4) it is no longer possible to identify graduates of universities of applied sciences clearly, as the new occupation code no longer has a separate category for this vocational qualification. They are assigned to category 12.<br><br>**2) ASU**<br>For these observations the vocational education completed most recently is reported. The following values exist in spells with a start date until 30 June 2006:<br><br>1 Without vocational training<br>2 In-company voc. training/traineeship/External (on-school) voc. training<br>3 Technical school (voc. training)<br>4 Technical school (advanced voc. training)<br>5 University of applied sciences (FH)<br>6 University<br><br>In 2006, the IT procedure from which the jobseeker data originate was switched from coArb to VerBIS. Many variables, such as training, were reported with different levels of differentiation in the two systems. This means that in spells from ASU and MTH which have a start date from 1 July 2006 onwards, the following differentiated categories are available:<br><br>7 Voc. training not accepted in Germany<br>8 University degree not accepted in Germany<br>9 In-company voc. training/traineeship / In-school voc. training<br>10 Other exams<br>11 University of applied sciences without further specifications<br>12 University without further specifications<br>13 Doctorate<br>14 Bachelor (BA)<br>15 Bachelor (FH)<br>16 Bachelor (University)<br>17 Master (FH)<br>18 Master (University)<br>19 Diploma (BA)<br>20 Diploma (FH)<br>21 Diploma (University) |

| Category | Description |
|---|---|
| | 23 Undergraduate studies<br>24 Secondary/additional studies<br>25 Other (promotion-) advanced training for graduates<br>26 Dual study programme, integrating vocational training<br>27 Dual study programme, integrating practical experience<br><br>**3)XASU**<br>For spells that originate from these sources, the vocational education completed most recently is reported. The following values exist:<br><br>1 Without vocational training<br>2 In-company voc. training/traineeship/External (on-school) voc. training<br>3 Technical school (voc. training)<br>4 Technical school (advanced voc. training)<br>5 University of applied sciences (FH)<br>6 University<br>7 Voc. training not accepted in Germany<br>8 University degree not accepted in Germany<br><br>For analyses that cover a longer period of time, the values can be aggregated as follows:<br><br>1, 22 to 1<br>7 to 2<br>8 to 3<br>2, 3, 4, 9 to 4<br>5, 11, 14, 15, 17, 19, 20 to 5<br>6, 10, 12, 13, 16, 18, 21, 26, 27 to 6<br>23, 24, 25 to 7<br><br>The aggregated categories take the following labels:<br>1 Without (recognised) vocational training<br>2 Vocational training not accepted in Germany<br>3 University degree not accepted in Germany<br>4 In-company / school-based training<br>5 University of applied sciences<br>6 University<br>6 Other exams<br>7 Undergraduate studies<br>7 Secondary/additional studies<br>7 Other (promotion-) advanced training for graduates |
| Notes on quality | **1) BeH**<br>“Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee obtained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until he/she joins a new firm. It can generally be assumed that when a person is employed in a firm for a longer period, the personal data that they reported when they joined the firm is simply carried forward” (own translation of Meinken / Koch 2004, p. 63).<br><br>The share of missing values increases almost continuously over time. Due to the introduction of the new occupation code in 2011, the share even temporarily strongly increased to around 51%. Since 2014, however, the proportion of missing values in the BeH has levelled off at around 40%.<br><br>Missing values occur particularly frequently in the following groups: marginal part-time employees, part-time workers, foreign employees and workers from Eastern German. The reason for this is that the variable is not relevant for social security contributions (see Meinken/Koch, 2004, p. 63).<br><br>The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments.<br><br>For the variable “Vocational training (imputed) (ausbildung_imp_beh)”, a method was applied to correct missing values or inconsistent changes of the training variable. However, this variable is only filled in the source BeH and has different categories than the variable ausbildung.<br><br>**2) ASU**<br>As a result of the switchover from coArb to VerBIS it is not possible to distinguish correctly between “no completed vocational training” and “no information available” in the ASU data source between 2006 and 2008. A missing value in this period therefore does not necessarily mean that the person has no vocational training or that there is no information available on vocational education and training, but may also mean that it was not possible to apply the relevant data generation procedure. In the source MTH, older categories partially still occur even after 1 July 2006. |

| Category | Description |
|---|---|
| | **3) XASU**<br>The degree of completeness in the XASU is generally low. |

### 5.4.9 Vocational training (imputed) (ausbildung_imp_beh)

| Category | Description |
|---|---|
| Variable label | Vocational training (imputed) |
| Variable name | ausbildung_imp_beh |
| Category | Personal variables |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable Vocational training (imputed) is a supplement to the variable Vocational training (ausbildung_beh) and contains additional and harmonized information on the vocational training of employees for BeH spells. The variable thus offers a solution to the problems concerning the variable Vocational training (ausbildung_beh)described in Section 5.4.8. The imputation procedure is described in Thomsen et al (2018), which is based on the work of Fitzenberger et al. (2006).<br>As the variable only uses the training information from BeH notifications and because the educational categories of the old and the new occupation codes had to be harmonised for the variable Vocational training (ausbildung_beh) the variable ausbildung_imp_beh has other categories than the variable Vocational training (ausbildung_beh).<br>A classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/). |
| Notes on quality | Despite the imputation procedure, the share of missing data increases over time, reaching about 11% in 2023. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. |

### 5.4.10 School leaving qualification (schule_x_asu, schule_beh)

| Category | Description |
|---|---|
| Variable label | School leaving qualification |
| Variable name | schule_x_asu, schule_beh |
| Category | Personal variable |
| Origin | ASU, XASU, BeH |
| Data type | Numerical |
| Detailed description | This variable contains the school leaving qualification. Classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ. Different values are possible depending on the source.<br>**1) BeH**<br>With the switch to the new occupation code (see Section 4.3.4) the possible values of the variable change. The values from the old occupation code are:<br>5 Grade-/lower school certificate, intermediate school or equivalent qualification<br>8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling<br>9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling<br>With the new occupation code, the values are:<br>1 No school leaving certificate<br>4 Lower secondary school certificate/ grade school certificate<br>6 Intermediate school leaving certificate<br>8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling.<br>The differentiation options in lower secondary education were thus expanded (1, 4, 6 instead of 5 previously), while those in upper secondary education were reduced (8 and 9 now combined under 8). |

| Category | Description |
|---|---|
| | **2) ASU, XASU**<br>The following values are possible for these data sources:<br>1 No school leaving certificate<br>4 Lower secondary school certificate/ grade school certificate<br>6 Intermediate school leaving certificate<br>7 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college<br>9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling<br>They are valid at the beginning of the period of job-search or participation in a measure. In the case of people seeking an apprenticeship position, the variable may also contain the school qualification they are working towards in the XASU data source. |
| Notes on quality | **1) BeH**<br>The degree of completeness in the BeH has been decreasing continuously over time and seems to have levelled off at under 2/3 in recent years. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments.<br>**2) ASU, XASU**<br>In the XASU, the degree of completeness increases continuously and has levelled off at over 75% since 2018 (XASU). In the ASU, the degree of completeness is generally high. |

### 5.4.11 Vocational training until 2006 (ausbildung_bis2006)

| Category | Description |
|---|---|
| Variable label | Vocational training until 2006 |
| Variable name | ausbildung_bis2006 |
| Category | Personal variable |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | This characteristic indicates the highest educational level of a given individual until 2006. |

### 5.4.12 School leaving qualification until 2006 (schule_bis2006)

| Category | Description |
|---|---|
| Variable label | School leaving qualification until 2006 |
| Variable name | schule_bis2006 |
| Category | Personal variable |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | This characteristic indicates the highest school education level of a given individual until 2006. |

## 5.5 Information on benefit receipt

### 5.5.1 Daily benefit (leistung_tag)

| Category | Description |
|---|---|
| Variable label | Daily benefit |
| Variable name | leistung_tag |
| Category | Information on benefit receipt |
| Origin | LeH |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | The variable shows the daily benefit rate, converted into euros. It must be taken into account that the daily benefit rate applies to calendar days. Since 1 January 2005, a daily benefit rate reported as 0 euros can be put down to benefit suspension periods or interruptions of benefit payments. If a reason for an end of benefit is reported for an observation with a daily benefit rate equal to 0, then it is a notification of interruption of benefit payments. In the case of observations that reflect a period of benefit suspension, the entitlement is the same as before the start of the benefit suspension period. |

### 5.5.2 Position in the benefit unit (bg_rolle)

| Category | Description |
|---|---|
| Variable label | Position in the benefit unit |
| Variable name | bg_rolle |
| Category | Information on benefit receipt |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | This characteristic describes the relationships of the individuals in a benefit unit with one another (partner, children). The following values are possible:<br>1 Main person/partner<br>2 Underage unmarried child<br>3 Adult, unmarried person under 25 year |

### 5.5.3 Ability to work (erwf)

| Category | Description |
|---|---|
| Variable label | Ability to work |
| Variable name | erwf |
| Category | Information on benefit receipt |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | This characteristic was calculated from the specifications on the SGB II status. The following values are possible:<br>0 No<br>1 Yes |
| Notes on quality | The SGB II status does not take into consideration whether there is any exclusion reason. To identify individuals eligible for unemployment benefit II or social benefits, the SGB II exclusion reason (IAB) {sgbii_ausschl} characteristic must therefore be considered additionally. Furthermore, the SGB II status of the LHG does not indicate whether there is individual benefit eligibility. Minor children without benefit eligibility cannot be identified in this characteristic, either.<br>A higher-than-average number of entries shows 15- to 17-year-olds as not capable of work, which is implausible. |

### 5.5.4 Type of benefit unit (bg_typ)

| Category | Description |
|---|---|
| Variable label | Type of benefit unit |
| Variable name | bg_typ |
| Category | Information on benefit receipt |
| Origin | LHG |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | This characteristic contains the calculated benefit unit type of the benefit unit member according to the IAB definition. The following values are possible:<br>1 Single<br>2 Single with adult(s) under 25 years of age<br>3 Single parent with child(s) under 18 years of age<br>4 Single parent with child(s) under 18 years of age and adult(s) under 25 years of age<br>5 Couple without child<br>6 Couple without child and with adult(s) under 25 years of age<br>7 Couple with child(s) under 18 years of age<br>8 Couple with child(s) under 18 years of age and adult(s) under 25 years of age<br>9 Other |

### 5.5.5 Reason of determination (grund_end)

| Category | Description |
|---|---|
| Variable label | Reason of determination |
| Variable name | grund_end |
| Category | Information on benefit receipt |
| Origin | LeH |
| Data type | Numerical |
| Detailed description | This characteristic states the reason for the end of unemployment benefit receipt. The LeH does not include information on the reasons for the start of benefit receipt, since the LeH is populated with the notifications from the employment agencies to the health insurance providers on completed benefit receipts durations. The values of this variable are only filled for the validity period provided by the technical or legal regulations. If values occur outside their validity period, they are converted to substitute values {value range 1173-1181} because their meaning cannot then be clarified with certainty. |

### 5.5.6 SGB-II reason for exclusion (sgbii_ausschl)

| Category | Description |
|---|---|
| Variable label | SGB-II reason for exclusion |
| Variable name | sgbii_ausschl |
| Category | Information on benefit receipt |
| Origin | LHG |
| Data type | Numerical |
| Detailed description | The SGB II exclusion reason specifies the person-related reason for a benefit unit member to be excluded from the receipt of benefits (fully or partially). |
| Special note | No financial data are included for individuals fully excluded from the receipt of benefits (values 1–10). When average amounts are calculated per benefit unit member, this must be considered accordingly. |

### 5.5.7 Income from (dependent) employment to be taken into account (zbeink_besch)

| Category | Description |
|---|---|
| Variable label | Income from (dependent) employment to be taken into account |
| Variable name | zbeink_besch |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | The variable indicates the sum of a person's income from dependent employment to be taken into account. |

| Category | Description |
|---|---|
| Special note | For individuals fully excluded from the receipt of benefits, no information on income is available, which means that underrecording may exist regarding the total income of the benefit unit.<br>Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.8 Income from self-employment to be taken into account (zbeink_selbst)

| Category | Description |
|---|---|
| Variable label | Income from self-employment to be taken into account |
| Variable name | zbeink_selbst |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | The variable indicates the sum of a person's income from self-employment to be taken into account. |
| Special note | For individuals fully excluded from the receipt of benefits, no information on income is available, which means that underrecording may exist regarding the total income of the benefit unit.<br>Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.9 Total disposable income (veink_sum)

| Category | Description |
|---|---|
| Variable label | Total disposable income |
| Variable name | veink_sum |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the total of the available income of the individual. In this context, the following applies – as well as to the subsequent characteristics on available income:<br>Total of the entire income of one individual<br>- privileged income<br>= income to consider (gross income or income from self-employment)<br>= taxes and social insurance contributions (or operating costs for self-employed individuals)<br>= available income (net income or profit for self-employed individuals)<br>- statutory tax-exempt amounts<br>= allowable income<br>The allowable incomes of all members of a benefit unit are added up and then assigned again to the members according to the requirement share method. This results in the allowable income of a given individual. This allowable income then reduces the requirements for benefits to secure a livelihood for the individuals eligible for benefits. For the SIG data, these individual components are aggregated back to the benefit unit level.<br>For the SIG data product, only the actually available income is taken into consideration, since this is significant for the reduction of the need for assistance, and in case of self-employment it specifies the actual profit of the enterprise.<br>Even though the subsequent characteristics are sub-categories of the total income available and are thus included in its total values, they are not exhaustive. Therefore, the specifically calculated total from these sub-categories may deviate from the “Total income available” variable, not only because of rounding. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on income is available, which means that underrecording may exist regarding the total income of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.10 Disposable income from employment (veink_besch)

| Category | Description |
|---|---|
| Variable label | Disposable income from employment |
| Variable name | veink_besch |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from employment of the individual. This includes not just the income from employment subject to social security contributions and marginal employment, but also other income from employment such as pocket money for voluntary services, income from volunteering, etc. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.11 Disposable income from self-employment (veink_selbst)

| Category | Description |
|---|---|
| Variable label | Disposable income from self-employment |
| Variable name | veink_selbst |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from self-employment of the individual. It is not possible to identify all self-employed individuals in SGB II, since only a positive balance after the deduction of operating costs from the income is represented in the available income. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.12 Disposable income from child benefit (veink_kindg)

| Category | Description |
|---|---|
| Variable label | Disposable income from child benefit |
| Variable name | veink_kindg |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from child allowance payments of the individual. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.13 Disposable income from alimony (veink_unterh)

| Category | Description |
|---|---|
| Variable label | Disposable income from alimony |
| Variable name | veink_unterh |
| Category | Information on benefit receipt |
| Origin | LST |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from maintenance allowance payments of the individual. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.14 Disposable income from sickness benefit (veink_krankg)

| Category | Description |
|---|---|
| Variable label | Disposable income from sickness benefit |
| Variable name | veink_krankg |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from sickness allowance payments of the individual. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.15 Disposable income from unemployment insurance (ALG I) (veink_algI)

| Category | Description |
|---|---|
| Variable label | Disposable income from unemployment insurance (ALG I) |
| Variable name | veink_algI |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from unemployment benefit I payments (SGB III) of the individual. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.16 Disposable income from pension payments (veink_rente)

| Category | Description |
|---|---|
| Variable label | Disposable income from pension payments |
| Variable name | veink_rente |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from benefits from pension insurance prospective entitlements of the individual. This includes all types of pensions (e.g. also orphan's and widow's pensions). |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.17 Disposable income from capital und property (veink_kub)

| Category | Description |
|---|---|
| Variable label | Disposable income from capital und property |
| Variable name | veink_kub |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of available income from capital as well as from rent and lease of the individual. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.18 Total needs without education & social participation (bedarf_sum)

| Category | Description |
|---|---|
| Variable label | Total needs without education & social participation |
| Variable name | bedarf_sum |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the total requirement of an individual without requirements from education and participation. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.19 Need from running expenses for accommodation (bedarf_lkdu)

| Category | Description |
|---|---|
| Variable label | Need from running expenses for accommodation |
| Variable name | bedarf_lkdu |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the total requirement of an individual from running costs of accommodation. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.20 Need from onetime expenses for accommodation (bedarf_ekdu)

| Category | Description |
|---|---|
| Variable label | Need from onetime expenses for accommodation |

| Category | Description |
|---|---|
| Variable name | bedarf_ekdu |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the total requirement of an individual from one-time costs of accommodation. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.21 Additional need from pregnancy (d) (d_schwanger)

| Category | Description |
|---|---|
| Variable label | Additional need from pregnancy (d) |
| Variable name | d_schwanger |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This dichotomous characteristic indicates whether an individual has any additional requirements due to pregnancy. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.22 Additional need from disability (d) (d_behind)

| Category | Description |
|---|---|
| Variable label | Additional need from disability (d) |
| Variable name | d_behind |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This dichotomous characteristic indicates whether an individual has any additional requirements due to disability. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.23 Additional need from special nutrition (d) (d_ernaehr)

| Category | Description |
|---|---|
| Variable label | Additional need from special nutrition (d) |
| Variable name | d_ernaehr |

| Category | Description |
|---|---|
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This dichotomous characteristic indicates whether an individual has any additional requirements due to special, cost-intensive dietary requirements due to medical necessity.<br>Notifications from A2LL also contain additional requirements according to Section 21 (6) (hardship case) due to a workaround. |
| Data quality | For individuals fully excluded from the receipt of benefits, no information on requirements is available, which means that underrecording may exist regarding the total requirements of the benefit unit. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.24 Need for loan (d) (d_darleh)

| Category | Description |
|---|---|
| Variable label | Need for loan (d) |
| Variable name | d_darleh |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This dichotomous characteristic indicates whether (individual or all) requirements of an individual were granted as a loan. The designated use may be manifold, from a loan for maintaining an owner-occupied house or rental arrears to multi-day school trips and orthopaedic shoes. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.25 Total pecuniary claim without education & social participation (zanspr_sum)

| Category | Description |
|---|---|
| Variable label | Total pecuniary claim without education & social participation |
| Variable name | zanspr_sum |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the total pecuniary claim of an individual without claims to social security contributions and benefits for education and participation, and after deduction of sanction payments, if any. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.26 Pecuniary claim from long-term unemployment benefits (ALG II) (zanspr_algII)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from long-term unemployment benefits (ALG II) |
| Variable name | zanspr_algII |
| Category | Information on benefit receipt |
| Origin | LST |

| Category | Description |
|---|---|
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from unemployment benefit II. It corresponds to the total of standard benefits, additional requirements, costs of accommodation, and unemployment benefit II bonuses, after the deduction of sanction payments, if any, for members of a benefit unit capable of work. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.27 Pecuniary claim from social welfare benefits (zanspr_sozg)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from social welfare benefits |
| Variable name | zanspr_sozg |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from social benefits. It corresponds to the total of standard benefits, additional requirements, and costs of accommodation, after the deduction of sanction payments, if any, for members of a benefit unit not capable of work. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.28 Pecuniary claim from basic maintenance (zanspr_regel)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from basic maintenance |
| Variable name | zanspr_regel |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from standard requirements, after the deduction of sanction payments, if any. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.29 Pecuniary claim from unemployment benefits supplement (zanspr_algIIzu)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from unemployment benefits supplement |
| Variable name | zanspr_algIIzu |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from unemployment benefit II bonuses, after the deduction of sanction payments, if any. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.30 Pecuniary claim from additional needs (zanspr_mbedarf)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from additional needs |
| Variable name | zanspr_mbedarf |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from additional requirements, after the deduction of sanction payments, if any. There is no differentiation according to the type of additional requirements. This is done for the requirements only. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.31 Pecuniary claim from running expenses for accommodation (zanspr_lkdu)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from running expenses for accommodation |
| Variable name | zanspr_lkdu |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from recognised running costs of accommodation, and after the deduction of sanction payments, if any. This includes both direct lease costs and operating and heating costs. Related subsequent payments also come under running costs of accommodation. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.32 Pecuniary claim from onetime expenses for accommodation (zanspr_ekdu)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from onetime expenses for accommodation |
| Variable name | zanspr_ekdu |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the amount of the pecuniary claim from recognised one-time costs of accommodation, after the deduction of sanction payments, if any. This includes, for example, rental arrears, housing costs, and maintenance costs for owner-occupied houses. Subsequent payments of operating and heating costs, however, come under running costs of accommodation. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.33 Pecuniary claim from social insurance (zanspr_sozvers)

| Category | Description |
|---|---|
| Variable label | Pecuniary claim from social insurance |
| Variable name | zanspr_sozvers |
| Category | Information on benefit receipt |

| Category | Description |
|---|---|
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the total of contributions to statutory health/nursing care/pension insurance. These contributions cannot be reduced by sanctions. Bonuses for insurance contributions according to Section 26 SGB II are included in the total. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

### 5.5.34 Total amount of sanctions (sanktionen_sum)

| Category | Description |
|---|---|
| Variable label | Total amount of sanctions |
| Variable name | sanktionen_sum |
| Category | Information on benefit receipt |
| Origin | LST |
| Data type | Numerical |
| Detailed description | This characteristic indicates the total of deductions from standard benefits, unemployment benefit bonus, additional requirements, and costs of accommodation. |
| Anonymisation | Due to its special sensitivity in terms of data protection legislation, this characteristic is provided only upon request and only in reasonable cases. |

## 5.6 Information on job search and participation in employment and training measures

### 5.6.1 Occupational group - last completed training (KldB 2010), 3-digit (beruf2010_3_ausb)

| Category | Description |
|---|---|
| Variable label | Occupational group - last completed training (KldB 2010), 3-digit |
| Variable name | beruf2010_3_ausb |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | This characteristic contains the applicant's last successfully completed vocational training (since 2000) at the occupational group level. |

### 5.6.2 Occupational sub-group - last completed training (KldB 2010), 4-digit (KldB 2010) (beruf2010_4_ausb)

| Category | Description |
|---|---|
| Variable label | Occupational sub-group - last completed training (KldB 2010), 4-digit |
| Variable name | beruf2010_4_ausb |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |

| Category | Description |
|---|---|
| Detailed description | This characteristic contains the applicant's last successfully completed vocational training (since 2000) at the occupational sub-group level. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (beruf2010_3_ausb) is provided. |

### 5.6.3 Occupational group - sought-after (KldB 2010), 3-digit (beruf2010_3_gesucht)

| Category | Description |
|---|---|
| Variable label | Occupational group - sought-after (KldB 2010), 3-digit |
| Variable name | beruf2010_3_gesucht |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | The variable gives the occupation at the beginning of the episode at the level of the occupational group, which is the mainly sought after by the customer. |

### 5.6.4 Occupational sub-group - sought-after (KldB 2010), 4-digit (beruf2010_4_gesucht)

| Category | Description |
|---|---|
| Variable label | Occupational sub-group - sought-after (KldB 2010), 4-digit |
| Variable name | beruf2010_4_gesucht |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | The variable gives the occupation at the beginning of the episode at the level of the occupational sub-group, which is the mainly sought after by the customer. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (beruf2010_3_gesucht) is provided. |

### 5.6.5 Job search status (asustat)

| Category | Description |
|---|---|
| Variable label | Job search status |
| Variable name | asustat |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | This variable displays the job placement status. It distinguishes between "unemployed jobseeker", "not unemployed jobseeker", "unemployed - unable to work for up to 42 days", "seeking advice" and "not seeking work". Applicants who only want advice from the BA are considered as “seeking advice”. In addition, there are rehabilitants and, before 2008, persons aged 58 or older who are not fully available for placement. As of 1 August 2016, this may also include persons who no longer require support, but who are nevertheless still under the job center’s care.<br>"Not seeking work" mainly subsumes persons of whom activation or placement can-not be reasonably expected according to § 10 SGB II. Similarly, persons with an incapacity to work of more than 42 days who continue to receive ALG II are listed in the system under this status. |

| Category | Description |
|---|---|
| | In addition, the status "ALG-I supplementing" exists for XASU cases as of 1.1.2017. According to this, a job search status has been reported by the ZkT, but according to the current legal situation, the responsibility for the job placement lies with the employment agency. A status notification from the ASU is therefore expected, which is then also decisive.<br>Since the reporting logic would make it possible to re-identify the exact date of birth in many cases, the original dates were changed by the anonymization procedure described in Section 3.2.1. |
| Notes on quality | The categories "seeking advice" and "not seeking work" have only existed since the introduction of Verbis (see Section 4.3.2). Due to late notifications, however, entries before 2006 can also be found. The characteristic "not seeking work" is considered to be under-recorded before 2008. In the XASU, the quality of status determination is partly limited for some institutions, especially in the first years. |

### 5.6.6 Employment status prior to job search (estatvor)

| Category | Description |
|---|---|
| Variable label | Employment status prior to job search |
| Variable name | estatvor |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | For ASU/XASU observations, the ‘employment status’ variable reports the job search status. If a data record is artificially split within the ASU/XASU (e.g. in the case of a change of legal status between SGB III and SGB II), the entry reason "DS generated by splitting" is assigned. However, this does not apply to episode splitting within SIG. The specification therefore does not necessarily refer to the start date of the episode, but to the start of the original period. |

### 5.6.7 Employment status after job search (estatnach)

| Category | Description |
|---|---|
| Variable label | Employment status after job search |
| Variable name | estatnach |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU |
| Data type | Numerical |
| Detailed description | The variable contains the person’s status after leaving unemployment. Longer periods of illness can be identified via this variable.<br>The values are classified as follows:<br>• 1000s: measure (assisted employment)<br>• 2000s: non-assisted employment<br>• 3000s: training etc.<br>• 4000s: self-employment<br>• 5000s: exclusion<br>• 6000s: other |
| Notes on quality | The relative frequencies vary over time depending on the values of the variable. Particular caution is therefore required in analyses. |

### 5.6.8 Integration forecast (ipo)

| Category | Description |
|---|---|
| Variable label | Integration forecast |
| Variable name | ipo |
| Category | information on employment, benefit receipt and job search |

| Category | Description |
|---|---|
| Origin | ASU |
| Data type | numerical |
| Detailed description | The variable reports the integration forecasts assessed by the placement, counseling and integration specialist (or the pilot in SGB III) during the initial contact. The integration forecast is the result of the assessment of the client's proximity to the market. "Close to market" applies if integration is expected within six months, "not close to market" if integration is only expected to succeed after more than six months. Identifiable needs for action that can be remedied within six months do not rule out market proximity.<br>The integration prognosis "integrated, but in need" is only valid for applicants from the SGB II.<br>If there is only a desire for placement in training, no integration forecast is to be determined.<br>This variable will be available from the end of 2016 and is the successor to the client profile (profil) variable. For the period before that, the information from the client profile is recoded into the integration forecast as follows:<br>The old values 2 (market profile), 3 (empowerment profile) and 4 (incentive profile) become the new value 1 (Close to market), the old values 5 development profile), 6 (stability profile) and 7 (benefit profile) become the new value 2 (Not close to market), the old value 8 (assignment unnecessary) becomes 3 (Assignment unnecessary) and 9 (integrated, but in need) becomes 4 (Integrated, but in need) and 1 (still undetermined) becomes .z (No entry).. |
| Notes on quality | The variable was introduced in 2016 but was recoded back to 2006 using other variables. The quality has been assessed as reliable by the BA statistics department since 2010. |

### 5.6.9 Reason for end of previous employment (art_kuend)

| Category | Description |
|---|---|
| Variable label | Reason for end of previous employment |
| Variable name | art_kuend |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU |
| Data type | Numerical |
| Detailed description | This variable is not filled for employment episodes.<br>This variable describes how the last employment or training relationship was terminated before a period of job search. It can therefore be used to identify job-to-job placements. |
| Notes on quality | The share of missing values is particularly high in this variable, averaging around 65%. However, this is mainly due to the fact that the variable can only contain information for persons who are looking for work and who were previously either employed subject to social security contributions or marginally employed. |

### 5.6.10 Working hours of job application (arbzeit)

| Category | Description |
|---|---|
| Variable label | Working hours of job application |
| Variable name | arbzeit |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU |
| Data type | Numerical |
| Detailed description | During the placement procedure, jobseekers indicate how many working hours the job they are seeking should have. |

### 5.6.11 Reason for termination (grund_abm)

| Category | Description |
|---|---|
| Variable label | Reason for termination |
| Variable name | grund_abm |

| Category | Description |
|---|---|
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Numerical |
| Detailed description | **1) ASU**<br>In the case of ASU observations, the variable contains the deregistration reason. In the case of a change of legal sphere, the observation is split artificially and 'generated by data splitting' is entered as the reason for deregistration. In order to depict the reasons for deregistration correctly it is also necessary to take into account the variable 'status after job search'.<br><br>The number of values of the variable was reduced from 26 April 2003 onwards. For analyses over long periods of time, the old values can be recoded to the currently valid ones using the rules below:<br><br>229 to 260, 230 to 260, 231 to 261, 232 to 260, 233 to 260, 234 to 260,2 35 to 260, 236 to 261, 237 to 266, 238 to 266, 239 to 271, 240 to 269, 242 to 265, 243 to 270, 244 to 274, 245 to 277, 246 to 267, 247 to 267, 248 to 278, 249 to 269, 250 to 275, 251 to 274, 252 to 276, 253 to 268, 254 to 278.<br><br>**2) XASU**<br>In the XASU, overlap-free episodes are mapped in relation to an opting municipality (zkT). When a number of characteristics change, a new observation is generated for the XASU in each case, of which the following are part of the SIG:<br>▪ Change of job search status<br>▪ Change of availability<br>▪ Change of SGB II provider<br>▪ Change of place of residence<br>If the episode reported by authorised municipalities is artificially split for other reasons (see Episode start date), "DS generated by splitting" is entered as the deregistration reason. |

### 5.6.12 Start date left-censored unemployment spell (begin_alo)

| Category | Description |
|---|---|
| Variable label | Start date left-censored unemployment spell |
| Variable name | begin_alo |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Datum |
| Detailed description | The start date of the left-censored unemployment spell indicates the date the unemployment episode started prior to 1 January 2007. The left margin is 1 January 1997.<br>This technical date may not be compared to the start of an actually continuous period of unemployment, since, among others, short sick periods do not interrupt unemployment in legal terms, or unemployment episodes may border on each other for technical reasons. |

### 5.6.13 Start date left-censored job search spell (begin_asu)

| Category | Description |
|---|---|
| Variable label | Start date left-censored job search spell |
| Variable name | begin_asu |
| Category | Information on job search and participation in employment and training measures |
| Origin | ASU, XASU |
| Data type | Datum |
| Detailed description | The start date of the left-censored job search spell indicates the date the job search episode started prior to 1 January 2007. The left margin is 1 January 1997.<br>This technical date may not be compared to the start of an actually continuous period of job search, since (X)ASU episodes may border on each other for technical reasons. |

### 5.6.14 Measure category (mass_gr)

| Category | Description |
|---|---|
| Variable label | Measure category |
| Variable name | mass_gr |
| Category | Information on job search and participation in employment and training measures |
| Origin | MTH, XMTH |
| Data type | Numerical |
| Detailed description | This characteristic shows the measure type category. This is the highest level in the hierarchy of measure type classifications of the Federal Employment Agency statistics. |

### 5.6.15 Measure type – group (mass)

| Category | Description |
|---|---|
| Variable label | Measure type – group |
| Variable name | mass |
| Category | Information on employment, benefit receipt and job search |
| Origin | MTH, XMTH |
| Data type | numerical |
| Detailed description | The variable indicates the measure type group. This is the second highest level in the hierarchy of the measure type classifications of the BA. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the measure category is provided (mass_gr). |

## 5.7 Information on Employment

### 5.7.1 Daily wage (entgelt_tag)

| Category | Description |
|---|---|
| Variable label | Daily wage |
| Variable name | entgelt_tag |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | In BeH observations, this variable shows the employee's gross daily wage. It is calculated from the fixed-period wages reported by the employer and the duration of the (unsplit) original notification period in calendar days. The daily wage is shown in euros.<br><br>Until 1998, employers in principle only reported the earnings which were subject to social security contributions. Earnings below the marginal part-time income threshold were not reported. Earnings exceeding the upper earnings limit for statutory pension insurance are only reported up to this limit. There are two upper earnings limits in the statutory pension insurance scheme. The earnings limit of the miners' pension insurance is generally higher than the earnings limit of the pension insurance for wage and salary earners. However, it is not possible to differentiate between these two insurance providers in the data.<br><br>Since the inclusion of marginal part-time employees in the employment notification procedure on 1 April 1999, earnings below the marginal part-time income threshold have also been recorded; the upper earnings limit still applies as the upper ceiling. In some cases, however, the reported earnings nonetheless exceed the upper earnings limit. Generally, this can probably be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual, employment interruption or end of employment notifications. In this case, it is irrelevant whether the upper earnings limit in the statutory pension insurance which is decisive for the notification period is exceeded as a result of |

| Category | Description |
|---|---|
| | this addition. However, such earnings notifications could also be due to incorrect details in the employment period. (The earnings information, however, may be considered less error-prone due to its insurance relevance.)<br><br>The marginal part-time income threshold and the upper earnings limit for statutory pension insurance differ from year to year as well as between eastern and western Germany (the decisive factor is the location of the establishment). An overview of these limits and thresholds can be found under (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/).<br><br>A daily wage reported as 0 euros can be put down to “employment interruption notifications”. During these periods, the employment relation-ship continues to exist in legal terms, but without pay. This is the case for periods of illness after the end of continued payment of wages, for periods of maternity leave and for sabbaticals.<br><br>The daily wage is shown with two decimal places. All values greater than 0 and smaller than 0.01 were rounded up to 0.01. This makes it possible to identify the above-mentioned employment interruption notifications with the condition daily wage = 0. |
| Notes on quality | From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off wage) increases sharply (see Section 5.7.11). It is likely that special payments which were reported with the annual declarations before 2013 are now reported separately. It is therefore advisable to add these variable one-time payments to the corresponding wages for simultaneous employment episodes within the same establishment when analysing wages over time. |

### 5.7.2 Daily wage (incl. one-off special payment) (entgelt_tag_bonus)

| Category | Description |
|---|---|
| Variable label | Daily wage (incl. one-off special payment) |
| Variable name | entgelt_tag_bonus |
| Category | Information on employment, benefit receipt and job search |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable Daily wage (incl. one-off special payment) is a supplement to the variable “Daily wage (entgelt_tag)” and contains harmonised information on daily wage supplemented by one-off payments/special payments for BeH spells. Values above 97.5% of the mode of daily wages rounded to whole euro amounts for full-time male employees (per year and separately for East and West Germany) are truncated. This value is lower than the official upper earnings limit. The procedure is described in chapter 4.1 of Drechsler et al (2023). |
| Notes on quality | The variable is not available for all observations from the BeH source. On the one hand, episodes with notification on one-off payments/special payments (154) are not taken into account, on the other hand, Drechsler et al. (2023) clean up multiple employment for the same period in the same establishment in advance, not using it to generate the variable. These episodes are still included in the data product presented here. |

### 5.7.3 Daily wage (imputed) (entgelt_tag_imp)

| Category | Description |
|---|---|
| Variable label | Daily wage (imputed) |
| Variable name | entgelt_tag_imp |
| Category | Information on employment, benefit receipt and job search |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable ‘Daily wage’ (imputed) is an additional variable to the variable “Daily wage (entgelt_tag)” and contains imputed daily wages for BeH notifications of full-time employees aged between 14 and 64 (excluding trainees) if these are censored in the notification data because they exceed the upper earnings limit. The variable is based on the variable “Daily wage (incl. one-off special payment) (entgelt_tag_bonus)”. This means that the daily wages of this subpopulation that do not exceed the upper earning limits are taken from “Daily wage (incl. one-off special payment) (entgelt_tag_bonus)”.<br><br>Details of the imputation procedure are described in Drechsler et al. (2023) and Drechsler/Ludsteck (2025). |

| Category | Description |
|---|---|
| Notes on quality | Drechsler et al. (2023, Chapter 5) warn against using the variable in analyses without further test steps. In particular, it should always be checked how high the proportion of censored wages is in the respective model and whether all regressors of the model have also found their way into the imputation model either directly or through sufficiently good proxies. It is therefore recommended that you familiarise yourself with the assumptions and limitations of the procedure before using it. |

### 5.7.4 Occupation - current employment (KldB 1988) (beruf_aktT)

| Category | Description |
|---|---|
| Variable label | Occupation – current job (KldB 1988) |
| Variable name | beruf_aktT |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes applies to one employee, the employer is required to select the job title that best defines the main activity performed (see BA 2005, p. V).<br><br>For this, the employer encodes the employee's job in accordance with the "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles" (published by the Federal Employment Agency, Nuremberg, 1988), which contains approx. 25,000 job titles. The occupational classification consists of a 3-digit code and comprises about 330 values.<br><br>Employment notifications with an end date later than 30 November 2011 are reported using the new occupation code 2010 (KldB2010) (see Section 4.3.4). These values are transcoded to the KldB1988 via a priority switch. This results in inaccuracies. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. |

### 5.7.5 Occupational group - current employment (KldB 2010), 3-digit (beruf2010_3_aktT)

| Category | Description |
|---|---|
| Variable label | Occupational group - current employment (KldB 2010), 3-digit |
| Variable name | beruf2010_3_aktT |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes applies for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br><br>For this the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational group is recorded by the first three digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau).<br><br>Employment notifications with an end date earlier than 30 November 2011 are re-ported using the old occupation code 1988 (KldB 1988) (see Section 4.3.4). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data. |

| Category | Description |
|---|---|
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. |

### 5.7.6 Occupational sub-group - current employment (KldB 2010), 4-digit (beruf2010_4_aktT)

| Category | Description |
|---|---|
| Variable label | Occupational sub-group - current employment (KldB 2010), 4-digit |
| Variable name | beruf2010_4_aktT |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes applies for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br><br>For this the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational group is recorded by the first three digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau).<br><br>Employment notifications with an end date earlier than 30 November 2011 are re-ported using the old occupation code 1988 (KldB 1988) (see Section 4.3.4). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (beruf2010_3_akt) is provided. |

### 5.7.7 Level of requirement - current employment (KldB 2010) (niveau_aktT)

| Category | Description |
|---|---|
| Variable label | Level of requirement - current employment (KldB 2010) |
| Variable name | niveau_aktT |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes applies for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br><br>For this, the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau).<br><br>Employment notifications with an end date earlier than 30 November 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 4.3.4). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data. |

| Category | Description |
|---|---|
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. |

### 5.7.8 Part-time (teilzeit)

| Category | Description |
|---|---|
| Variable label | Part-time |
| Variable name | teilzeit |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable “Part-time” (teilzeit) distinguishes between full-time and part-time employees. The decisive factor is the ratio between the contracted hours and the usual working hours in the establishment. For part-time employees the variable only records whether their working hours exceed a certain limit or not. Since 1988 this limit is 18 hours per week. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. In order to reduce this problem, the working hours were imputed at the IAB for the period in question. Further information about the procedure can be found in Ludsteck/Thomsen (2016). No imputation is performed here.<br>The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments.<br>For a more detailed discussion of the problem and a possible corrective approach, see Fitzenberger/Seidlitz (2020). |

### 5.7.9 Employment status (beschstat)

| Category | Description |
|---|---|
| Variable label | Employment status |
| Variable name | beschstat |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable corresponds to the person group recorded in the new notification procedure (DEÜV) from 1 January 1999 onwards. It indicates contribution- or benefit-related particularities of the employment relationship.<br>If multiple codes apply to an employment notification, the smallest must be indicated by the reporting employer. The majority of these cases are employment relationships subject to social security contributions without any distinctive characteristics, which are recorded under code number 101. Accordingly, it is possible that these employment relationships are slightly overestimated.<br>The notification procedure stipulates those changes in the employment status - e.g., when an apprentice is taken on by his/her training company after completing his/her vocational training - must be indicated by a new notification.<br>Since 1 April 1999, employees in marginal part-time employment have also been recorded in the DEÜV notification procedure. This person group can be distinguished via the values 109 and 209. For employees in marginal part-time employment, no data prior to the introduction of the notification obligation in 1999 could be collected. |
| Notes on quality | The person group can be contained in employment notifications that refer to the years prior to 1999 but were not received until 1999 or later. For notifications which were received before 1999, an attempt is made to allocate the notifications to the person groups on the basis of certain rules and with the aid of the vocational education and training', 'occupational status and working hours' and 'occupation' variables as well as other information. In many cases, however, conclusive allocations are not possible. |

### 5.7.10 Transition zone (gleitz)

| Category | Description |
|---|---|
| Variable label | Transition zone |
| Variable name | gleitz |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates whether the employment notification relates to employment in the low-wage sector, within the so-called transition zone (in German “Gleitzone”, since July 2019 “Uebergangsbereich”), (so-called midi jobs). Employee only have to pay a reduced overall social security contribution. As employees with earnings in the transition zone can voluntarily pay the “regular” social security contribution, not all employees with corresponding earnings are automatically classified as being in the transition zone.<br><br>The transition zone is based on gross pay and changes over time:<br><br>01.04.2003 to 31.12.2012: € 400.01 to € 800.00<br>01.01.2013 to 30.06.2019: € 450.01 to € 850.00<br>From 01.07.2019: € 450.01 to €1,300.00<br><br>Different lower limits exist for trainees and employees in voluntary services. In the case of several employment relationships, assignment to the transition zone is based on the total gross pay.<br><br>The corresponding legislation has been in force since 1 April 2003. |

### 5.7.11 Temporary agency work (leih)

| Category | Description |
|---|---|
| Variable label | Temporary agency work |
| Variable name | leih |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable reports whether the person’s employment is a temporary job via an employment agency. The employee is recorded with the establishment that provides the social security notifications and this is the establishment which leases out the employee (not the establishment that the employee physically works in). The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than 30 November 2011. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is already down to about 5% and decreases further below 1% until 2023. |

### 5.7.12 Fixed-term contract (befrist)

| Category | Description |
|---|---|
| Variable label | Fixed-term contract |
| Variable name | befrist |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable reports whether the person’s employment relationship is fixed-term or permanent. The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than 30.11.2011. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is already down to about 5% and decreases further below 1% until 2023. |

### 5.7.13 Reason of notification (grund_abg)

| Category | Description |
|---|---|
| Variable label | Reason of notification |
| Variable name | grund_abg |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable indicates the reason why the employer submitted the employment notification in question to the social security agencies. However, not all of the possible reasons for submitting a notification that may occur in the context of the notification procedure are available in the IEB. For instance, the IEB only includes notifications that have information on earnings (i.e., annual, employment interruption and end of employment notifications), while initial registrations are not contained as they contain no information on earnings. However, this does not involve a loss of information, as the details from a registration are transmitted again with the following annual, employment interruption or end of employment notification.<br><br>The reasons for submitting employment notifications are encoded according to the regulations of the notification procedure, which has been in effect since 1 January 1999 (in accordance with DEÜV) with a leading "1". To obtain the original coding, the value "100" must be subtracted from the coding in the variable. |
| Notes on quality | From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off payment) increases sharply. As long as an employment relationship exists, special payments that are paid out by March of the following year can be included into the usual notifications (mostly annual notifications) for the previous year. A separate notification with a reason for deregistration of 54 is then not required. Until 2012, the annual notifications could be submitted until mid-April; since 2013, they must now be submitted by mid-February at the latest. Special payments made in February and March must now be reported separately. |

### 5.7.14 Second employment (mehrfbesch)

| Category | Description |
|---|---|
| Variable label | Second employment yes/no |
| Variable name | mehrfbesch |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This characteristic indicates whether an individual has a second simultaneous employment. |

### 5.7.15 Daily wage – second employment (mehrfbesch_tentg)

| Category | Description |
|---|---|
| Variable label | Daily wage – second employment |
| Variable name | mehrfbesch_tentg |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This characteristic contains the daily wage of the second simultaneous employment (see also Chapter 5.7.1). |

### 5.7.16 Daily wage (incl. one-off special payment) - second employment (mehrfbesch_bonus_tentg)

| Category | Description |
|---|---|
| Variable label | Daily wage (incl. one-off special payment) – second employment |
| Variable name | mehrfbesch_bonus_tentg |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This characteristic contains the daily wage (incl. one-off special payment) of the second simultaneous employment (see also Chapter 5.7.2). |

### 5.7.17 Employment status – second employment (mehrfbesch_beschstat)

| Category | Description |
|---|---|
| Variable label | Employment status - second employment |
| Variable name | mehrfbesch_beschstat |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This characteristic contains the employment status of the second simultaneous employment (see also Chapter 5.7.9). |

### 5.7.18 Part-time – second employment (mehrfbesch_teilzeit)

| Category | Description |
|---|---|
| Variable label | Part-time – second employment |
| Variable name | mehrfbesch_teilzeit |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | The variable “Part-time” (teilzeit) distinguishes between full-time and part-time for the second simultaneous employment (see also Chapter 5.7.8). |

### 5.7.19 Employment experience before 2007 (exp_vor2007)

| Category | Description |
|---|---|
| Variable label | Employment experience before 2007 (in days) |
| Variable name | exp_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the reported work experience of an individual between 1 January 1993 and 31 December 2006 (in days). |

### 5.7.20 Time since last employment before 2007 (time_vor2007)

| Category | Description |
|---|---|
| Variable label | Time since last employment before 2007 (in days) |

| Category | Description |
|---|---|
| Variable name | time_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the amount of time elapsed since the last reported employment episode of a person between 1 January 1993 and 31 December 2006 (in days). |

### 5.7.21 Daily wage before 2007 (tentgelt_vor2007)

| Category | Description |
|---|---|
| Variable label | Daily wage before 2007 |
| Variable name | tentgelt_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the daily wage of the last reported primary employment episode of a person between 1 January 1993 and 31 December 2006. |

### 5.7.22 Occupation - last employment (KldB 1988) before 2007 (beruf1988_vor2007)

| Category | Description |
|---|---|
| Variable label | Occupation - last employment (KldB 1988) before 2007 |
| Variable name | beruf1988_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the occupation (KldB 1988) of the last reported primary employment episode of a person between 1 January 1993 and 31 December 2006. |

### 5.7.23 Occupational group - last employment (KldB 2010) before 2007, 3-digit (beruf2010_3_vor2007)

| Category | Description |
|---|---|
| Variable label | Occupational group - last employment (KldB 2010) before 2007, 3-digit |
| Variable name | beruf2010_3_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the occupational group (KldB 2010, 3-digit) of the last reported primary employment episode of a person between 1 January 1993 and 31 December 2006. |

### 5.7.24 Occupational sub-group - last employment (KldB 2010) before 2007, 4-digit (beruf2010_4_vor2007)

| Category | Description |
|---|---|
| Variable label | Occupational sub-group - last employment (KldB 2010) before 2007, 4-digit |
| Variable name | beruf2010_4_vor2007 |

| Category | Description |
|---|---|
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the occupational group (KldB 2010, 4-digit) of the last reported primary employment episode of a person between 1 January 1993 and 31 December 2006. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (beruf2010_3_vor2007) is provided. |

### 5.7.25 Part-time before 2007 (teilzeit_vor2007)

| Category | Description |
|---|---|
| Variable label | Part-time before 2007 |
| Variable name | teilzeit_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates whether a person was exclusively reported as part-time employed between 1 January 1993 and 31 December 2006 (last primary employment episode). |

### 5.7.26 Employment status before 2007 (beschstat_vor2007)

| Category | Description |
|---|---|
| Variable label | Employment status before 2007 |
| Variable name | beschstat_vor2007 |
| Category | Information on employment |
| Origin | BeH |
| Data type | Numerical |
| Detailed description | This variable indicates the employment status of the last reported primary employment episode of a person between 1 January 1993 and 31 December 2006. |

## 5.8 Location data

### 5.8.1 Place of residence – district (wo_kreis_lhg, wo_kreis_x_asu, wo_kreis_leh, wo_kreis_beh)

| Category | Description |
|---|---|
| Variable label | Place of residence - district |
| Variable name | wo_kreis_lhg, wo_kreis_x_asu, wo_kreis_leh, wo_kreis_beh |
| Category | Location data |
| Origin | LHG, ASU, XASU, LeH, BeH |
| Data type | Numerical |
| Hierarchy | federal state \| district |

| Category | Description |
|---|---|
| Detailed description | The variable indicates the district (urban district or rural district) in which the social security contributor lives. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), and positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3.<br><br>For the LHG and XASU sources, the place of residence applies to the whole period of the original observation. For the BeH, ASU, LeH, MTH and XMTH, the variable contains the place of residence at the beginning of the original period of time. This means that the longer the spell lasts, the higher the risk that the place of residence will become obsolete and that the information given for later dates will be incorrect. In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded with reference to the territorial allocation of 31 December 2023 for all sources, i.e. in all calendar years, a place of residence is assigned to a district in accordance with the boundaries that the district had on 31 December 2023. As the district boundaries have changed over time, cases would occur in which the district code changes without the individual concerned having relocated if the territorial allocations of the districts were not updated. |
| Notes on quality | There are inaccuracies in the information provided in the BeH for some employees with regard to where they live. The reporting requirement does not clarify which residence - main or secondary residence with predominant residence - is to be reported by the employer. In the employment history, this can lead to the impression of "long-distance commuters" between the registered main residence and the place of work, even though the person is in fact employed at the secondary residence, i.e. does not actually commute.<br>In the year 2015, the BeH data show a reduction in the change of residence of approx. 10-15% compared to the usual level of the surrounding years, with regional differences occurring. It has not yet been possible to determine the exact reason for this deviation. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (wo_bula) is provided. |

### 5.8.2 Place of residence – federal state (wo_bula_lhg, wo_bula_x_asu, wo_bula_leh, wo_bula_beh)

| Category | Description |
|---|---|
| Variable label | Place of residence – federal state |
| Variable name | wo_bula_lhg, wo_bula_x_asu, wo_bula_leh, wo_bula_beh |
| Category | Location data |
| Origin | LHG, ASU, XASU, LeH, BeH |
| Data type | Numerical |
| Hierarchy | federal state \| district |
| Detailed description | This variable is an aggregation of the “district” variable to the 16 German federal states. Further information on the district of the place of residence can be found in Section 5.8.1. |

### 5.8.3 Place of work – district (ao_kreis)

| Category | Description |
|---|---|
| Variable label | Place of work – district |
| Variable name | ao_kreis |
| Category | Location data |
| Origin | BeH |
| Data type | Numerical |
| Hierarchy | federal state \| district |

| Category | Description |
|---|---|
| Detailed description | The variable indicates the district (urban district or rural district) in which the employee's establishment is located. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomen-clature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3.<br><br>In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded to the territorial allocation of 31 December 2023, i.e. in all calendar years, a place of work is assigned to a district in accordance with the boundaries that the district had on 31 December 2023. As the district boundaries have changed over time, cases would occur in which the district code of the location of the establishment would change without the establishment concerned having relocated, if the territorial allocations were not updated. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the federal state (ao_bula) is shown as regional information. |

### 5.8.4 Place of work – federal state (ao_bula)

| Category | Description |
|---|---|
| Variable label | Place of work – federal state |
| Variable name | ao_bula |
| Category | Location data |
| Origin | BeH |
| Data type | Numerical |
| Hierarchy | federal state \| district |
| Detailed description | The variable indicates the federal state in which the establishment is located. This variable is generated from the district code (ao_kreis). The first two positions of the district code indicate the federal state (NUTS 1). |

## 5.9 Establishment variables

### 5.9.1 Economic subgroup 08 (completed by extrapolation/imputation) (w08_5_kons)

| Category | Description |
|---|---|
| Variable label | Economic subgroup 08 (completed by extrapolation/imputation) |
| Variable name | w08_5_kons |
| Category | Establishment variables |
| Origin | BeH |
| Data type | Numerical |
| Hierarchy | section (1-digit code) \| division (2-digit code) \| group (3-digit code) \| class (4-digit code) \| sub-class (5-digit code) of economic activity |
| Detailed description | This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available from 2008 onwards. From 2008 up to and including 2017, the variable contains the original values from w08_3. Before 2008, the information is either written back or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity based on the economic activity classification WZ08. A detailed description can be found in Eberle et al. (2011). |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, the economic activity is only shown as the 3-digit code (w08_3_kons). |

### 5.9.2 Economic group 08 (completed by extrapolation/imputation) (w08_3_kons)

| Category | Description |
|---|---|
| Variable label | Economic group 08 (completed by extrapolation/imputation) |
| Variable name | w08_3_kons |
| Category | Establishment variables |
| Origin | BeH |
| Data type | Numerical |
| Hierarchy | section (1-digit code) \| division (2-digit code) \| group (3-digit code) |
| Detailed description | This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available from 2008 onwards. From 2008 up to and including 2017, the variable contains the original values from w08_3. Before 2008, the information is either written back or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity based on the economic activity classification WZ08. A detailed description can be found in Eberle et al. (2011). |

# References


Bender, Stefan; Hilzendegen, Jürgen; Rohwer, Götz; Rudolph, Helmut (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.

Bertat, Thomas; Dundler, Agnes; Grimm, Christopher; Kiewitt, Jochen; Schomaker, Christine; Schridde, Henning; Zemann, Christian (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, URL: https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erbebungsinhalte-Arbeitszeita-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf (letzter Zugriff: 27.11.2019).

Bundesagentur für Arbeit (Hrsg.) (2005): Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. Ausgabe Januar 2005, Nürnberg.

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg. URL: https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf (letzter Zugriff: 27.11.2019).

Federal Employment Agency (2015): Bundesagentur für Arbeit – Statistik (2015): Statistik zu Bildung und Teilhabe nach dem SGB II. Methodenbericht. Nürnberg, September 2015. URL: https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Grundsicherung-Arbeitsuchende-SGBII/Generische-Publikationen/Methodenbericht-Bildung-und-Teilhabe.pdf (letzter Zugriff: 22.11.2019).

Bundesagentur für Arbeit – Statistik (2017): Aktuelle Hinweise zur Interpretation der Kosten der Unterkunft in der Grundsicherungsstatistik. Kurzinfo. Juni 2017 (erweiterte Fassung).

Bundesanstalt für Arbeit (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html (letzter Zugriff: 27.11.2019).

Cramer, Ulrich (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv, Jg. 69, S. 56-68.

Drechsler, Jörg; Ludsteck, Johannes (2025): Imputation strategies for rightcensored wages in longitudinal datasets. In: Journal for Labour Market Research, 59(19). DOI: https://doi.org/10.1186/s12651-025-00410-4.

Drechsler, Jörg; Ludsteck, Johannes; Moczall, Andreas (2023): Imputation der rechtszensierten Tagesentgelte für die BeH. FDZ-Methodenreport 05/2023 (de), Nuremberg, DOI:10.5164/IAB.FDZM.2305.de.v1

Eberle, Johanna; Jacobebbinghaus, Peter; Ludsteck, Johannes; Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes * Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport, 05/2011 (en), Nuremberg.

Fitzenberger, Bernd; Seidlitz, Arnim (2020): The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. Journal for Labour Market Research 01/2020.

Fitzenberger, Bernd; Osikominu, Aderonke; Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 126(3), pp. 405-436.

Ludsteck, Johannes; Thomsen, Ulrich (2016): Imputation of the Working Time Information for the Employment Register Data. FDZ-Methodenreport, 01/2016 (en), Nürnberg.

Meinken, Holger; Koch, Iris (2004): BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.

Paulus, Wiebke; Matthes, Britta (2013): Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport, 08/2013 (de), Nürnberg.

Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ08.html (Zugriff: 21.04.2016)

Statistisches Bundesamt (2019): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5%20 (Zugriff: 05.12.2019).

Stüber, Heiko; Dauth, Wolfgang; Eppelsheimer, Johann (2023): A guide to preparing the sample of integrated labour market biographies (SIAB, version 7519 v1) for scientific analysis. In: Journal for Labour Market Research, 57(7). https://doi.org/10.1186/s12651-023-00335-w.

Thomsen, Ulrich; Ludsteck, Johannes; Schmucker, Alexandra (2018): Skilled or unskilled - Improving the information on qualification for employee data in the IAB Employee Biography. FDZ-Methodenreport, 09/2018 (en), Nuremberg.

Wermter, Winfried; Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 4/88, S. 468-482.

# Appendix

**Frequency tables**

Frequency tables and overviews of the individual values and labels of the variables can be found in separate files at https://fdz.iab.de/en/startseite-en/.

# Imprint

**FDZ-Datenreport 14|2025**

**Date of publication**
18 December 2025

**Publisher**
Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
D-90478 Nuremberg



**Download**
https://doku.iab.de/fdz/reporte/2025/DR_14-25_EN.pdf

**Documentation version**
SIG0723_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2514.en.v1

**Dataset version**
SIG 0723 v1, DOI: 10.5164/IAB.SIG0723.de.en.v1

**All publications in the series "FDZ-Datenreport" can be downloaded from**
https://fdz.iab.de/en/research/publications/fdz-datenreport-series/

**Website**
https://fdz.iab.de

---

**Corresponding author**
Sandra Dummert
Phone: 0911 179-7005
Email: sandra.dummert@iab.de