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

FDZ-DATENREPORT

Documentation on labour market data

07|2024 EN

Codebook and Documentation of the Panel Study „Labour Market and Social Security“ (PASS) Datenreport Wave 17

Marco Berg et al.

Bundesagentur für Arbeit

Codebook and Documentation of the Panel Study „Labour Market and Social Security“ (PASS) Datenreport Wave 17

Marco Berg, Ralph Cramer, Christian Dickmann, Reiner Gilberg, Birgit Jesske, Martin Kleudgen (all infas Institut für angewandte Sozialwissenschaft GmbH),

Jonas Beste, Sandra Dummert, Corinna Frodermann, Sonja Malich, Stefan Schwarz, Claudia Wenzig, Mark Trappmann, Sebastian Bähr, Nadja Bömmel, Mustafa Coban, Matthias Collischon, Stefanie Gundert, Benjamin Küfner, Jan Mackeben, Bettina Müller, Jens Stegmaier, Nils Teichler, Anja Wunder (all Institut für Arbeitsmarkt- und Berufsforschung (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

1 Introduction 10
1.1 The objectives and research questions of the panel study „Labour Market and Social Security“ 10
1.2 Instruments and interview program 11
1.2.1 Characteristics and innovations of wave 17 - Individual Questionnaire 13
1.2.2 Characteristics and innovations of wave 17 - Senior citizens questionnaire 16
1.2.3 Characteristics and innovations of wave 17 - Household questionnaire 18
1.2.4 General notes on previous waves 18
1.3 Sample and data preparation 20
1.3.1 Characteristics and innovations of wave 17 21
1.3.2 Quality control of realised interviews 22

2 Key figures 23
2.1 Sample size 25
2.2 Response rates 27
2.3 Panel participation agreements, merging data and linking with process data 28
2.4 Split-off households 29

3 Dataset structure 31

4 Generated variables 34
4.1 Coding responses to open-ended survey questions 34
4.1.1 Open-ended residual categories and open-ended items 34
4.1.2 Coding of occupation and industry 38
4.2 Harmonisation 42
4.3 Dependent Interviewing 45
4.4 Simple generated variables 48
4.5 Constructed variables 77
4.5.1 Individual Level 77
4.5.2 Household or benefit unit level 100

5 Data preparation 114
5.1 Structure checks and removing interviews 115
5.2 Filter checks and missing codes used 121
5.3 Plausibility checks 122
5.4 Retroactive changes in waves 1 to 16 124

5.5 Anonymisation ... 127
5.6 Receipt of Unemployment Benefit II ... 132
5.6.1 Concept for updating the spells of Unemployment Benefit II receipt that were ongoing in the previous wave ... 132
5.6.2 Structure of the Unemployment Benefit II spell dataset ... 133
5.6.3 Plausibility checks and corrections to the Unemployment Benefit II spell dataset ... 134
5.6.4 Updating the Unemployment Benefit II spell dataset ... 134
5.7 Employment biographies ... 135
5.7.1 Variables on the employment/inactivity status in PENDDAT ... 135
5.7.2 Income variables and working hours in the PENDDAT and in the BIO-spell dataset ... 144
5.7.3 Concept for updating the spells that were ongoing in the previous wave ... 150
5.7.4 Structure of the BIO-spell dataset ... 150
5.7.5 Plausibility checks and corrections of the spell datasets ... 152
5.7.6 Update of spell datasets ... 152
5.8 One-Euro job spell dataset (*ee_spells)* ... 154
5.8.1 Concept for updating the spells that were ongoing in the previous wave ... 155
5.8.2 Structure of the EE spell dataset ... 155
5.8.3 Plausibility checks and corrections in the EEJ spell dataset ... 155

6 Weighting Wave 17 ... 157
6.1 Design weights for the panel replenishment (municipal register sample) in wave 17 ... 157
6.2 Integration of the design weights for the panel replenishment (from municipal registers (EWO)) using the existing weights of the population sample ... 158
6.3 Design weights for the panel households in wave 17 ... 158
6.4 Design weights for the refreshment sample in wave 17 ... 159
6.5 Propensity to participate again - households ... 159
6.6 Propensity to participate - first-time interviewed split-off households ... 160
6.7 Nonresponse weighting for households from the refreshment sample of BA new inflows wave 17 ... 160
6.8 Nonresponse weighting for households from the panel replenishment (municipal register sample) wave 17 ... 161
6.9 Propensity to participate again - individuals ... 161
6.10 Integration of the weights to yield the total weight before calibration ... 162
6.11 Integration of temporary non-responses (households) ... 162
6.12 Calibration to the household weight, wave 17, cross-section ... 164
6.13 Calibration of the BA sample ... 165

6.14 Calibration of the population sample 166
6.15 Calibration of the total sample 167
6.16 Calibration of the person weight, wave 17, cross-section 168
6.17 BA sample 168
6.18 Population sample 170
6.19 Total sample 171
6.20 Estimating the BA cross-sectional weights for households and individuals not in receipt of Unemployment Benefit II 173

7 Appendix: Brief description of the dataset 175

List of Figures

Figure 1: Realised panel sample for households and individuals by survey wave 27
Figure 2: Dataset structure of PASS in wave 17 33
Figure 3: Overview of generated variables for wave 17 at the individual level 75

List of Tables

Table 1: Changes in the person questionnaire in W17 14
Table 2: Changes in the senior questionnaire in W17 17
Table 3: Changes in the household questionnaire in W17 18
Table 4: Coding responses to open-ended questions at the household level in wave 17 35
Table 5: Coding responses to open-ended questions at the individual level in wave 17 35
Table 6: Coding scheme of the additional variables used in PASS 39
Table 7: Harmonised variables in the individual dataset (*PENDDAT*) 43
Table 8: Variables in the individual dataset (*PENDDAT*) are generated across waves but not completely harmonised (*PENDDAT*) 44
Table 9: Updated information in wave 17, household questionnaire 46
Table 10: Updated information in wave 17, personal questionnaire 47
Table 11: Simple generated variables in the cross-section datasets (*HHENDDAT; PENDDAT*) for households and individuals who previously provided information on the topic 49
Table 12: Wave 17 simple generated variables in the household (*HHENDDAT*) and *KINDER* datasets (in alphabetical order) 51
Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) 53
Table 14: Wave 17 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) 64
Table 15: Simple generated variables for wave 17 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset) 67

Table 16: Wave 17 simple generated variables included in the one-euro spell dataset *(ee_spells)* (in the same order presented in the dataset) 71
Table 17: Wave 17 simple generated variables included in the person register dataset *(p_register)* (in alphabetical order) 72
Table 18: Education in years 77
Table 19: Education in years, mother 78
Table 20: Education in years, father 79
Table 21: CASMIN 80
Table 22: MCASMIN 81
Table 23: VCASMIN 82
Table 24: ISCED 97 83
Table 25: MISCED 97 84
Table 26: VISCED 97 84
Table 27: International Standard Classification of Occupations 1988 (ISCO88) 85
Table 28: International Standard Classification of Occupations 2008 (ISCO08) 86
Table 29: Classification of Occupations 1992 (KldB92) 88
Table 30: Classification of Occupations 2010 (KldB2010) 89
Table 31: Erikson, Goldthorpe and Portocarrero (EGP) Class Scheme 90
Table 32: European Socio-economic Classification (ESeC) 92
Table 33: Magnitude-Prestige Scale (MPS) 93
Table 34: Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala)-Basis ISCO-88 94
Table 35: Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala) – Basis ISCO-08 94
Table 36: International Socio-Economic Index (ISEI) – Basis ISCO-88 95
Table 37: International Socio-Economic Index (ISEI) – Basis ISCO-08 96
Table 38: Classification of Economic Activities 2003 (WZ2003) 97
Table 39: Classification of Economic Activities 2008 (WZ2008) 98
Table 40: Physiological scale of SF12v2 (SOEP-Version, NBS) 99
Table 41: Psychological scale of SF12v2 (SOEP-Version, NBS) 99
Table 42: Equivalised household income, previous OECD weighting 100
Table 43: Equivalised household income, modified OECD weighting 101
Table 44: Deprivation index, unweighted 101
Table 45: Deprivation index, weighted 103
Table 46: Household typology 105
Table 47: Wave 17 benefit unit ID 106
Table 48: Wave 17 benefit unit typology 108
Table 49: Benefit unit receiving Unemployment Benefit II on the wave 17 sampling date 109
Table 50: Benefit unit receiving Unemployment Benefit II on the wave 17 survey date 109

Table 51: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date 109
Table 52: Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date 110
Table 53: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date 110
Table 54: Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date 111
Table 55: Number of benefit units within the household 113
Table 56: Number of benefit units in the household receiving benefits on the sampling date 113
Table 57: Overview of the steps involved in preparing the data of wave 17 of PASS . 115
Table 58: Overview of the missing codes used 122
Table 59: Overview of retroactive changes to the household dataset *(HHENDDAT, KINDER)* 125
Table 60: Overview of retrospective alterations in the individual dataset *(PENDDAT)* . 126
Table 61: Overview of retroactive corrections to spell datasets *(bio_spells, alg2_spells, ee_spells)* 126
Table 62: Overview of retrospective alterations to the register datasets *(hh_register; p_register)* 127
Table 63: Overview of retrospective alterations to the weighting datasets *(hweights; pweights)* 127
Table 64: Overview of the anonymised variables in the individual dataset *(PENDDAT)* in wave 17 129
Table 65: Overview of the anonymised variables in the BIO-spell dataset *(bio_spells)* in wave 17 132
Table 66: Overview of anonymised variables in the children dataset in wave 17 *(KINDER)* 132
Table 67: Cross-sectional variables in the UB II spell dataset *(alg2_spells)* 133
Table 68: Logic of generation of *erwerb, erwerb2, nichterw, nichtew2* 137
Table 69: Revision *erwerb, erwerb2, nichterw, nichtew2* 140
Table 70: Basic assignment - Spell with higher priority beats spell with lower priority 143
Table 71: Detailed assignment for special cases 144
Table 72: Revision income variables 145
Table 73: Revision working hours variables 147
Table 74: ET-specific cross-section variables in the BIO-spell dataset (*bio_spells*) 151
Table 75: AL-specific cross-section variables in the BIO-spell dataset (*bio_spells*) 152
Table 76: Cross-sectional variables in the EE spell dataset (*ee_spells*) 155

Abstract

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. This data report documents the data preparation of the PASS wave 17 and is based upon the sixteenth wave's data report: Marco Berg, Ralph Cramer, Christian Dickmann, Vincent Gerber, Reiner Gilberg, Birgit Jesske, Martin Kleudgen (all infas Institute for Applied Social Sciences), Jonas Beste, Sandra Dummert, Corinna Frodermann, Sonja Malich, Stefan Schwarz, Claudia Wenzig, Mark Trappmann, Sebastian Bähr, Nadja Bömmel, Mustafa Coban, Matthias Collischon, Stefanie Gundert, Benjamin Küfner, Jan Mackeben, Bettina Müller, Jens Stegmaier, Nils Teichler, Anja Wunder (all Institute for Employment Research (IAB)): Codebook and documentation of the panel Study 'Labour Market and Social Security' (PASS), Datenreport wave 16, FDZ Datenreport, 12/2023 (en), Nürnberg.

Zusammenfassung

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. Dieser Datenreport dokumentiert die Aufbereitung der Welle 17 von PASS. Das Dokument basiert auf dem Datenreport der Welle 16: Marco Berg, Ralph Cramer, Christian Dickmann, Vincent Gerber, Reiner Gilberg, Birgit Jesske, Martin Kleudgen (alle infas Institut für angewandte Sozialwissenschaft GmbH), Jonas Beste, Sandra Dummert, Corinna Frodermann, Sonja Malich, Stefan Schwarz, Claudia Wenzig, Mark Trappmann, Sebastian Bähr, Nadja Bömmel, Mustafa Coban, Matthias Collischon, Stefanie Gundert, Benjamin Küfner, Jan Mackeben, Bettina Müller, Jens Stegmaier, Nils Teichler, Anja Wunder (alle Institut für Arbeitsmarkt- und Berufsforschung (IAB)): Codebuch und Dokumentation des Panel „Arbeitsmarkt und soziale Sicherung" (PASS) Band I: Datenreport Welle 16, FDZ Datenreport, 12/2023 (de), Nürnberg.

1 Introduction

1.1 The objectives and research questions of the panel study „Labour Market and Social Security“

The panel study „Labour Market and Social Security“ (PASS), established by the Institute for Employment Research (IAB), creates an empirical dataset for labour market, welfare state and poverty research and policy counseling in Germany.

This study is conducted as part of IAB research on German Social Code Book II (SGB II)[1]. The IAB must fulfill a statutory mandate to study the effects of the benefits and services provided under SGB II, which are aimed at labour-market integration and subsistence benefits. However, due to its complex sampling design, this study also enables researchers to examine additional issues. The following five core questions, which are detailed in Achatz, Hirseland and Promberger (2007), influenced the development of this study[2].

1. What are the options for regaining financial independence from Unemployment Benefit (UB) II (Arbeitslosengeld II)?
2. How does a household’s social situation change when it receives benefits?
3. How do individuals who receive benefits cope with their situations? Do recipient attitudes toward the actions required to improve their situations change over time?
4. How does contact between benefit recipients and institutions that provide basic social security take place? What actual institutional procedures are applied in practice?
5. What employment history patterns or household dynamics lead to receiving Unemployment Benefit II?

This data report provides an overview of the seventeenth survey wave, for which 11,743 individuals in 8,862 households[3] were interviewed between February 2023 and October 2023. Due to the processing of the EWO sample, the field period was extended by four weeks until October 2023. This sample included 8,252 individuals and 6,450 households that had previously been interviewed for PASS.

[1] Social Code Book II - basic security for job-seekers (Sozialgesetzbuch (SGB) Zweites Buch (II) - Grundsicherung für Arbeitsuchende).
[2] The research questions were adapted against the background of the reform of SGB II and have taken into account the Citizen’s Income since survey wave 17.
[3] These figures include evaluable interviews only. Additionally, repeatedly interviewed households were considered even if only a household interview but no personal or senior citizen interview could be conducted.

This wave-specific data report[4] of wave 17 documents the aspects of the study. In chapter 1 an overview of the aims and research questions of the study is given with a short description of the instruments and the survey program in chapter 1.2 and the characteristics and innovations of wave 17 in chapters 1.2.1 to 1.2.3. In chapter 2 the data report provides key figures on the wave's sample and response rates. The data itself and the data preparation are the topics of the following chapters. In chapter 3 an overview of the data structure is given and in chapter 4 the generated variables are presented. Furthermore, the data preparation and the decisions taken during this process are described in chapter 5. In chapter 6 the weighting procedure is presented. In the appendix, a complete overview of all datasets of all waves of PASS is given.

Up to wave 16, the frequencies of all variables contained in the Scientific Use File were shown in separate table volumes. These volumes will no longer be published from wave 17 onwards. Frequency analyses of individual variables are available on request.

1.2 Instruments and interview program

The information in PASS is collected using separate questionnaires for the household and individual levels. First, a household interview is conducted. This interview gathers information about the entire household. The target person for this household interview[5] was selected during the contact phase preceding the interviews. Personal interviews of the household members follow the household interview. The aim is to conduct a personal interview of each individual living in the household who is 15 years of age or older. House-hold members who are 65 or older receive a shortened version of the questionnaire (the senior citizens' questionnaire)[6], which excludes questions that are irrelevant to that age group.

[4] These reports were divided into the following two components for the first time in the wave 3 documentation: a wave-specific data report (including a codebook) and a cross-wave User Guide. The PASS project team at the IAB is responsible for creating the cross-wave User Guide. As of wave 3, infas has created the documentation for the wave-specific data report, which is based on the wave 2 data report. The cross-wave User Guide documents the entire study, details the objectives and design of PASS and presents the contents and instruments of the survey. Moreover, it describes the structure of the scientific use file and the concept of the variable types and their names.

[5] The target person for the household interview should know as much as possible about general household issues, and target selection was based on the rules documented in the methods reports (Jesske & Quandt (2011); Jesske & Schulz (2012); Jesske & Schulz (2013); Jesske & Schulz (2014); Jesske & Schulz (2015); Jesske et al. (2016); Jesske et al. (2017); Jesske & Schulz (2018); Jesske et al. (2019); Jesske et al. (2020); Jesske et al. (2021); Jesske et al. (2022); Jesske et al. (2023)).

[6] Since wave 10, the survey tool has taken into account that the statutory retirement age will be gradually raised to 67 years over the next few years. The age limit for the submission of the senior citizens' questionnaire is therefore no longer rigidly 65 years, but is based on the statutory retirement age of the target person on a monthly basis.

The survey instruments and interview program for wave 17 are based on those used in wave 16. However, individual questions and modules have been revised or newly developed (see Chapter 1.3 for an overview).

The PASS survey instruments are designed to allow not only repeat interviews of individuals and households but also first-time interviews[7]. Since wave 3, dependent interviewing has been used for certain questions to update in-formation that the respondent had previously provided to avoid seam effects[8] in the repeat interviews and to increase data quality. Information about constant characteristics was generally not gathered again. Additionally, since wave 4, an integrated questionnaire for repeatedly interviewed households (HHalt) and first-time interviewed households (HHneu) has been used[9]. During the entire span of the panel, it is possible that households or persons from panel households who had suspended a wave (or even several waves), i.e. did not participate (so-called temporary dropouts), participate again. Temporary dropouts who did not participate in exactly one wave, i.e. the previous wave, are treated as panel households or panel respondents in terms of interview control. If an entire household did not participate in two consecutive waves, it will not be contacted again. At the person level, however, it is possible that respondents in panel households are not interviewed for two or more subsequent waves and then participate in the panel study again. In the person interview, this special respondent group is regarded as new respondents after the interview break, i.e. unchangeable information (e.g. social origin) is collected again and no dependent interviewing takes place.

The cross-wave PASS User Guide elaborates the individual instruments and interview program. In addition, the wave-specific questionnaires are also provided (for the scales and instruments used, see also Müller et al. 2020). The following section reviews the characteristics and innovations of wave 17. The characteristics and innovations of wave 17 affect the questions asked in the household and personal questionnaires (e.g., change of reference periods, modification of individual questions and new question modules)[10], sample and data preparation.

[7] First-time interviewed households include the following groups: (1) households from the refreshment and replenishment samples of the current wave; and (2) households that split off from households interviewed during previous waves (split-off households). (For further explanation, please see the wave 4 methods report (Jesske & Quandt, 2011).)

[8] In a panel data, the number of changes observed at the interface (seam) between interviews conducted in sequential panel waves is often considerably higher than the number of changes observed within an interview (see Jäckle 2008).

[9] In this survey, split-off households are treated like new households.

[10] Not all of the minor changes to the questionnaire (adding, modifying or deleting individual questions) are listed.

1.2.1 Characteristics and innovations of wave 17 - Individual Questionnaire

The personal questionnaire updates the employment history information gathered since wave 2[11]. From wave 4 onwards, a modified chronological retrospective surveying is included (see section 1.3.1 in Berg et al. (2011), FDZ Datenreport 08/2011).

For the personal questionnaire in wave 17, some modules and blocks of questions were newly developed and others were taken from previous waves and re-used.

[11] This information is gathered using the so-called dependent interviewing method. In dependent interviewing, information that was provided during previous interview waves is included in the interview text of the current interview to determine whether the information must be updated.

Table 1: Changes in the person questionnaire in W17

Module	Variable	Change	Explanation
Attitudes (self-efficacy)	PEO0100*	added	The module was included again from wave 14. However, a translation error was detected in the course of the fieldwork (Russian language version).
Corona	PCV0100a-PCV0100c and PCV0200 to PCV0600	deleted	The questions on the coronavirus pandemic have been completely deleted.
Employment history (Spells)	BIO0100z, BIO0100	modified	Adjustment of the mini-job limit to 520 euros.
Employment history (Spells)	ET4400, AL1410	deleted	Both questions on the identification of subsidized employment (wage cost subsidies from the job center/employment agency to the employer) were deleted.
Mini job	PET0510	modified	Adjustment of the mini-job limit to 520 euros.
Quality of employment	PQB1700	added	New question about home office added.
Attitudes (role models)	PEO0400a-d	added	The module was included again from wave 14.
Attitudes (work and family)	PEO0800a-b, PEO0900a-b, PEO1000a-b and PEO1100a-b	deleted	This module has been deleted.
Agency contacts	PX40000 to PTK1700*	modified	The content of the module has been fundamentally revised to take account of the reform of the Citizen's Income. The changes concern both conceptual adjustments (unemployment benefit II and citizen's allowance) and interviewer instructions, as well as new questions on the provider contacts that were added as a result of the reform.
Job search	PAS0920*	modified	The label of an answer code has been adjusted.

Table 1: Changes in the person questionnaire in W17 (continued)

Further training	WBH0100 to WBH0900	added	The module has been added. The background to this is the citizens' income reform, in which further training is an important component. It contains questions on continuing vocational training and the conditions under which this takes place.
Political participation	PPT0200*	modified	Minor adjustment to the question text.
Stigma awareness	PSV0300	added	The module on stigma awareness has been added. It contains items on stigma awareness in the receipt of citizens' income.
Networks	PSK0200	modified	Removed the experiment on gender-neutral language.
Internet usage	PTS1000*, PTS1010 to PTS1040, PTS2000, PTS3000*, PTS4000 and PTS5000	deleted	The module for internet usage was completely removed. The module contained items on the provision of internet-enabled devices, frequency of use of the devices, internet access, internet usage behavior and concerns with regard to data protection.
Latent functions of work	PLA0100a-PLA0100r	deleted	The module has been completely deleted.
Health	PG0001	added	A consent question was added to the health module. This allowed respondents to refuse the entire module. The input filters for the individual questions in the module were adjusted accordingly.
Health	PG0100	modified	The value „0" was added as valid information. The formulation variations on gender-neutral language, which were intended as an experiment in wave 16, were removed again.
Health	PG1205, PG1210 and PG1215*	deleted	Questions from the health focus module were deleted.
Income (pensions)	PEK1460*, PEK1470*, PEK1480*	added	New questions on the receipt of old-age pension or widow's or child-raising pension and their amount have been added.
Social assistance	PA0960	modified	In the filter, it was added that the person must have been born before 2005.

Table 1: Changes in the person questionnaire in W17 (continued)

Migration	PMI3100 to PMI3600	added	New questions on arrival in Germany, allocation of place of residence and situation in country of origin.

In the course of wave 17, errors were found in the translation of two questions, which is why all cases that had gone through the question up to this point were set to code -4 „Question mistakenly not asked“. On the one hand, this concerns question *PMI3600* in the Arabic translation of the questionnaire, in which a label was translated incorrectly. n=629 cases were therefore set to code -4. In addition, the item battery on self-efficacy *(PEO0100a-e)* in the Russian instrument (personal and senior citizens' questionnaire) is affected. The error was already present in wave 11. The values collected were converted to the special code -4 for wave 17 in the current SUF and for the previous wave 11 already in SUF wave 16.

1.2.2 Characteristics and innovations of wave 17 - Senior citizens questionnaire

Due to the gradual increase in retirement age the filter for respondents with valid information of the date of birth from wave 10 onward is carried out on a monthly basis, in order to ensure that senior citizens with age 65 and older receive the short version of the questionnaire. The age determining the transition from the individual to the senior citizens questionnaire in wave 17 is adjusted according to the standard retirement age as follows: 65 years and 11 months (for those born 1956).

Out of the list of modifications realized for the personal questionnaire the following modifications were also implied for the senior citizens questionnaire.

Table 2: Changes in the senior questionnaire in W17

Module	Variable	Change	Explanation
Attitudes (self-efficacy)	PEO0100*	added	The module was included again from wave 14. However, a translation error was detected in the course of the fieldwork (Russian language version).
Attitudes (role models)	PEO0400a-d	added	The module was included again from wave 14.
Corona	PCV0100a-PCV0100c	deleted	The questions on the coronavirus pandemic have been completely deleted.
Income (pensions)	PEK1420, PEK1470*	added	Two new questions on widow's/parent's pension have been added.
Employment	PET0500	modified	The limit for marginal employment has been adjusted from 450 euros to 520 euros.
Employment	PMJ0910, PMJ1010	added	New question on professional activity and position in mini-job.
Political participation	PPT0200*	modified	Minor adjustment of the question text.
Networks	PSK0200	modified	Removal of the experiment on gender-neutral language.
Internet usage	PTS1000*, PTS1010 to PTS1040, PTS2000, PTS3000*, PTS4000 and PTS5000	deleted	The module for internet usage has been completely deleted. The module contained items on the provision of internet-enabled devices, frequency of use of the devices, internet access, internet usage behavior and concerns with regard to data protection.
Latent functions of work	PLA0100a-PLA0100r	deleted	The module has been completely deleted.
Health	PG0001	added	A consent question was added to the health module. This allowed respondents to refuse the entire module. The input filters for the individual questions in the module were adjusted accordingly.
Health	PG0100	modified	The value „0“ was added as valid information. The wording variations on gender-neutral language, which were intended as an experiment in wave 16, were removed again.

Table 2: Changes in the senior questionnaire in W17 (continued)

Health	PG1205, PG1210 and PG1215*	deleted	Questions from the health focus module were deleted.
Migration	PMI3100 to PMI3600	added	New questions on arrival in Germany, allocation of place of residence and situation in country of origin.

The Russian translation for *PEO0100* in the senior citizens' questionnaire was also incorrect (see above), which is why all cases that had gone through the questionnaire up to this point (wave 11 & wave 17) were set to code -4 „Question mistakenly not asked".

1.2.3 Characteristics and innovations of wave 17 - Household questionnaire

In the household questionnaire, in addition to the usual adjustments of the time reference and filter controls for subsamples, there were only minor changes in wave 17, which are listed in the following overview.

Table 3: Changes in the household questionnaire in W17

Module	Variable	Change	Explanation
Housing situation	HW2200, HW2300	added	There is a new question on how the apartment was found and whether there was a choice in the event of the apartment being allocated.
Childcare Corona	HKIBLK03, HKI1200*, HKI1300*, HKI1400*	deleted	The questions on „Childcare Corona" have been completely deleted.

1.2.4 General notes on previous waves

The Scientific Use File (SUF) also contains variables of individual questions/modules from previous waves that are no longer surveyed in the current wave. Therefore, this chapter is intended to collect important information on modules/questions from earlier data reports that cannot be discussed adequately in the current data report within the scope of the

chapter on innovations and special features. In many cases, the brief description also refers to the corresponding section of the previous data report.

- In wave 14, changes had to be made in the already running field. This involved newly included questions that addressed the special situation of the Covid-19 pandemic. In the dataset, the variable kennungfbversW14 identifies the cases or households that were interviewed before (code 1) and after (code 2) the change. The changes in the questionnaire were active as of April 8, 2020. For more information, see Berg et al. (2021), p. 14. In principle, the use of wave 14 should be considered and interpreted in light of the Covid-19 pandemic. This is because the limited survey options may result in lower numbers of cases or declining rates for individual parameters (Berg et al. 2021, p. 16).
- The module Trust experiment (PVX0100-PVX0600) could not be collected in wave 14 due to restrictions in the CAPI field associated with the Covid-19 pandemic as of mid-March. For more information, see Berg et al. (2021), p. 13.
- The module Crowdworking (PCW0100-PCW0400) was newly developed for PASS and surveyed in waves 12 and 13. Analyses with these variables have shown considerable inconsistencies and give reason to doubt the validity of the measurement instrument. For more information, see Berg et al. (2019), p. 14.
- In wave 11, changes had to be made in the running field. These changes in the questionnaire were related to the subsample of Syrian/Iraqi households. The variable *kennungfbversW11* marks whether a case was interviewed with the original or the revised questionnaire version. For more information, see Berg et al. (2018), p.44.
- Up to wave 15, a plausibility check between *PTK0500* and *PTK0200* was carried out in the carrier contact module. In the case of implausibly high deviations (more than twice as high) between the data, the information in *PTK0500* was previously set to the special code -8 „implausible value“.
- Due to changes in the questionnaire (interviewer note) and actual changes in job center practice as a result of the corona pandemic, it was decided to no longer carry out the plausibility check retroactively from wave 14 and to deliver the variable *PTK0500* with the original respondent information (and without special code) (Berg et al. 2023, p. 159).
- A translation error (Russian language variant) was detected for *PEO0100* in fieldwork W17, which also affected wave 11. The values collected have been converted to the special code -4 both for wave 11 (already in SUF wave 16) and in the current SUF for wave 17 (see section 1.2.1).

1.3 Sample and data preparation

In wave 17, as in previous waves, a refreshment sample was drawn from the Federal Employment Agency (BA) subsample[12]. The aim is to guarantee the representativeness of the BA sample in the cross-section. For the refreshment sample, benefit units were drawn receiving UB II in July 2022 but not on the sampling date of the waves 1-16 (see Chapter 2.1 and, on the concept of the refreshment sample, Trappmann et al., 2009, page 11 ff.). All of the households that were surveyed for the first time during wave 17 can be identified via the sample indicator (**sample**).

In survey wave 10, PASS already reacted to the challenge that the composition of the target group of UB II recipients has changed. Triggered by the increased immigration of refugees to Germany since 2014, there are larger proportions of benefit recipients from Arabic-speaking countries of origin who will be surveyed in PASS over the next few years. Arabic was therefore used as an additional survey language from wave 10 onwards. This ensures that recognized refugees from the most common countries of origin (Syria and Iraq) are reached by the yearly refreshment samples and continued in the panel. While in the waves 11 to 14 refreshment samples of benefit units were drawn within the sampling points of PASS according to the usual procedure (for further details, see the methodological report of wave 12[13]), in the waves 15 and 16, an oversampling of newly arrived benefit units with persons of Syrian and Iraqi nationality into UB II was carried out so that refugees could be interviewed in sufficient numbers. Given that the UB II benefit recipients of Syrian and Iraqi nationality differ considerably from the other benefit units, they continue to be shown separately in the further descriptions and in the dataset. Households in which at least one member is of Syrian or Iraqi nationality are classified as Syrian/Iraqi households. In a minority of cases this leads to other people who live in these households but do not come from these two countries being assigned to this group. In order to be able to identify Syrian nationals in the group of persons from the subsample of Syrian and Iraqi households, the additional variable *ostaatansyr* is provided in the scientific use file from wave 11 onwards. This variable is already available retrospectively from wave 10 onwards. Due to the small case numbers, only the two categories “Syrian nationality” and “a different or no nationality” are shown.

As a result of the Russian attack on Ukraine in February 2022, Germany has seen an increased immigration of Ukrainian citizens. The immigration of people from Ukraine is also reflected in a larger proportion of people receiving benefits. For this reason, this group of people will also be considered and surveyed in the PASS in the coming years. Since wave 17, Ukrainian has therefore been used as an additional survey language in addition to Russian

[12] Wave 1 of PASS includes two subsamples: (1) a sample of households receiving UB II, which was drawn from the Federal Employment Agency (BA) process data; and (2) a general population sample, stratified by status, drawn from a database provided by the commercial provider MICROM.
[13] Jesske & Schulz (2018)

and Arabic in order to ensure that this group of people is surveyed. In wave 17, a sampling of newly added benefit communities with persons of Ukrainian nationality to SGB II was carried out in order to be able to interview a sufficient number of this group. From the Scientific Use File of wave 17 onwards, the additional variable *ostaatanukr* is supplied. The categories „Ukrainian nationality yes“ or „Ukrainian nationality no“ are shown.

In addition, the population sample was refreshed in wave 17. For this purpose, 100 new zip code areas (=sample points) were drawn. The postcode areas were drawn independently of the previously selected points. Seven of the new postcode areas were part of the previous 400 zip code areas and were not excluded. Within the municipalities, the sample was drawn from the population register. The personal addresses were drawn from the sample population in the municipalities using systematic random selection (interval sampling). A detailed description of the procedure can be found in the data report for wave 5. In the next step, the addresses provided formed the selection for a disproportionate distribution of the deployment sample. The aim was to reduce the proportion of people born in 1958 or earlier (undersampling pensioners) and to increase the proportion of non-German people (oversampling migrants). The stratified drawing of the gross addresses for the deployment sample was carried out in three steps: reduction of persons born in 1958 or earlier by half, inclusion of all non-German persons in principle and finally a random selection of German persons from the remaining gross sample.

For the 100 zip code areas, 92 registration offices were contacted and the requested addresses were supplied from all the requested municipalities.

1.3.1 Characteristics and innovations of wave 17

Waves 14 and 15 were still strongly influenced by the Covid-19 pandemic and the associated restrictions on conducting interviews, whereas the restrictions were already significantly eased in wave 16. In wave 17, there was a return to the survey procedure before the Covid-19 pandemic. In the CAPI field, interviews were only allowed to be conducted in person on site and telephone interviews only in the CATI field. In the final weeks of the fieldwork, this rule was deviated from in exceptional cases if it was not possible for an interviewer in the CAPI field to conduct an interview in person on site without further ado. A description of the survey situation can be found in the methodology report for wave 17 (Jesske et al. 2024).

The limited survey possibilities of the previous waves continued to have an impact on the individual key figures. However, the case numbers and rates again show clear trends as before the Covid-19 pandemic. In particular, the proportion of interviews conducted by telephone in the CAPI field was significantly reduced.

The data preparation was performed in close cooperation with the IAB. Basic procedures, such as updating datasets and correcting problems in the household structures, were discussed during the preparation process. Final decisions were made by the IAB.

For the Scientific-Use-File of PASS provided via the FDZ, modified identifiers (hnr, uhnr, pnr and bgnr) are created for each new version starting with wave 15. These are compatible within a version, but not between different versions.

1.3.2 Quality control of realised interviews

In the PASS study, extensive quality control measures are implemented at all levels of the survey process. For example, revision and test loops are implemented at the beginning of the questionnaire programming process to ensure an error-free instrument for field use, and extensive test steps are implemented at the end of the data processing process to guarantee the quality of the evaluation data. With various methodological experiments, PASS contributes to the optimisation of the surveys in the panel and to the improvement of the data quality[14]. In wave 14, improved quality controls were developed to detect deviating interviewer‘s behaviour. Controls focused in particular on statistical control methods such as the analysis of time stamps, longitudinal correlation analyses and other statistical indicators (see Beste et al. (2021), FDZ-Methodenreport). While the extension in wave 14 still led to the subsequent removal of interviews (see Berg et al. 2021:16ff), interviewees could already be identified in the current field of wave 17. In three cases, there were anomalies that did not lead to the removal or subsequent collection of interviews. The scope and procedure are described in detail in the methodological report of wave 17 (Jesske et al. 2024).

[14] The procedure and the various experiments are documented in the method reports of the respective waves

2 Key figures

This chapter provides a brief overview of important figures in the study, such as sample sizes (gross and net) and response rates. The panel sample is represented over the course of the previous waves. Figures are reported not only for both the original and replenishment samples but also for the complete study.

- Sample I: Subsample 1 (BA sample) refers to the sample of benefits recipients from the process data of the Federal Employment Agency.
- Sample II: Subsample 2 (MICROM sample) refers to the stratified population sample.
- Sample III: Refreshment sample 1 (BA sample) is the sample drawn from the SGB II inflow between waves 1 and 2.
- Sample IV: Refreshment sample 2 (BA sample) is the sample drawn from the SGB II inflow between waves 2 and 3.
- Sample V: Refreshment sample 3 (BA sample) is the sample drawn from the SGB II inflow between waves 3 and 4.
- Sample VI: Panel replenishment/supplement 1 (municipal register sample) is the sample drawn from the registration office inflows in 100 new postcode regions during wave 5.
- Sample VII: Panel replenishment/supplement 2 (BA sample) is the sample drawn from the SGB II inflows in 100 new postcode regions during wave 5.
- Sample VIII: Refreshment sample 4 (BA sample) is the sample drawn from the SGB II inflow between waves 4 and 5.
- Sample IX: Refreshment sample 5 (BA sample) is the sample drawn from the SGB II inflow between waves 5 and 6.
- Sample X: Refreshment sample 6 (BA sample) is the sample drawn from the SGB II inflow between waves 6 and 7.
- Sample XI: Refreshment sample 7 (BA sample) is the sample drawn from the SGB II inflow between waves 7 and 8.
- Sample XII: Refreshment sample 8 (BA sample) is the sample drawn from the SGB II inflow between waves 8 and 9.
- Sample XIII: Refreshment sample 9 (BA sample) is the sample drawn from the SGB II inflow between waves 9 and 10.
- Sample XIV: Refreshment sample 10 (BA sample Syrian/Iraqi households) is the sample drawn from the oversampling of Syrian/Iraqi households.
- Sample XV: Panel replenishment/supplement 3 (municipal register sample) is the sample drawn from the registration office inflows in the postcode regions of wave 5 (wave 11).
- Sample XVI: Refreshment sample 11 (BA sample) is the sample drawn from the SGB II inflow between waves 10 and 11.

- Sample XVII: Refreshment sample 12 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 10 and 11.
- Sample XVIII: Refreshment sample 13 (BA sample) is the sample drawn from the SGB II inflow between waves 11 and 12.
- Sample XIX: Refreshment sample 14 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 11 and 12.
- Sample XX: Refreshment sample 15 (BA sample) is the sample drawn from the SGB II inflow between waves 12 and 13.
- Sample XXI: Refreshment sample 16 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 12 and 13.
- Sample XXII: Refreshment sample 17 (BA sample) is the sample drawn from the SGB II inflow between waves 13 and 14.
- Sample XXIII: Sample XXI: Refreshment sample 18 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 13 and 14.
- Sample XXIV: Panel replenishment/supplement 4 (BA sample) is the sample drawn from the SGB II inflow of Syrian/Iraqi households (Wave 10).
- Sample XXV: Refreshment sample 19 (BA sample) is the sample drawn from the SGB II inflow between waves 14 and 15.
- Sample XXVI: Refreshment sample 20 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 14 and 15.
- Sample XXVII: Panel replenishment/supplement 5 (BA sample) is the sample drawn from the SGB II inflow of Syrian/Iraqi households (Wave 10).
- Sample XXVIII: Refreshment sample 21 (BA sample) is the sample drawn from the SGB II inflow between waves 15 and 16.
- Sample XXIX: Refreshment sample 22 (BA sample Syrian/Iraqi Households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between waves 15 and 16.
- Sample XXX: Panel replenishment/supplement 4 (municipal register sample) is the sample drawn from the registration office inflows in 100 postcode regions (wave 17).
- Sample XXXI: Refreshment sample 23 (BA sample) is the sample drawn from the SGB II inflow between wave 16 and wave 17.
- Sample XXXII: Refreshment sample 24 (BA sample Syrian/Iraqi households) is the sample drawn from the SGB II inflow of Syrian/Iraqi households between wave 16 and wave 17.
- Sample XXXIII: Refreshment sample 25 (BA sample Ukrainian households) is the sample drawn from the SGB II inflow of Ukrainian households.

2.1 Sample size

Each sample in a panel begins with the interviewed households from the first survey wave. In PASS, the gross panel sample contains the interviewed households from wave 1 and the HHneu from the refreshment samples in waves 2 to 13[15]. Only those households being interviewed for the first time that are willing to participate in the panel and are available for repeat interviews are considered[16]. Agreement to participate in the panel is only recorded during the first interview. Confirmation of these households' willingness in subsequent waves is not required. In addition to confirming willingness, access to the panel is induced during the first interview by general willingness to participate, that is, by providing an interview. Measures to ensure the best possible selection-free access to the panel as part of PASS are described in detail in the methods and field reports of waves 1 to 16[17].

Wave 1 of PASS included 12,794 household interviews, of which 12,000 households agreed to participate in the panel. These wave 1 households constitute the sample for the beginning of the first tracking survey.

The panel concept in PASS assumes that new or split-off households emerge as individuals move out of panel households, which are considered separate households as soon as a household interview is conducted. This design results in a higher number of households compared to the original sample. Details about the procedures for the PASS panel concept can be found under „split-off households". In addition to the expansion of the panel, loss of households can occur due to panel mortality. Households in which all respondents passed away or moved abroad are removed from the gross panel in subsequent waves. Moreover, panel losses may occur if no household interview could be conducted for a household for two consecutive waves.

This situation arose for the first time at the end of wave 3 and affected the gross panel in waves 4 to 16[18]. The gross sample used for wave 17 included 9,159 panel households. In addition, there are households surveyed for the first time from the BA access samples

[15] The interviews with a part of so-called pure senior citizen households were discontinued before wave 10. Half of the PASS households, in which only persons over the age of 67 lived (pure senior citizen households) were selected randomly and removed. In total this affected 420 households (see also Datenreport wave 10 in Berg et al. (2017))

[16] Willingness to participate in the panel is confirmed by the household reference person and is thus valid for all household members. Households that were willing to participate in the panel have allowed their addresses to be stored for the purposes of this study's repeat interviews.

[17] see Hartmann et al. (2008); Büngeler et al. (2009); Büngeler et al. (2010); Jesske & Quandt (2011); Jesske & Schulz (2012); Jesske & Schulz (2013); Jesske & Schulz (2014); Jesske & Schulz (2015); Jesske et al. (2016); Jesske et al. (2017); Jesske & Schulz (2018); Jesske et al. (2019); Jesske et al. (2020); Jesske & Schulz (2021); Jesske et al. (2022); Jesske et al. (2023)

[18] The survey institute change also influenced the panel gross in wave 4 because transmitting participant addresses from the IAB to infas required the target person's permission. For details on this procedure and its results, please refer to the methods report for wave 4 (Jesske & Quandt, 2011).

(n=6,341, of which 310 Syrian/Iraqi households and 3,271 Ukrainian households) and the EWO access sample (13,325 households) as well as the split households that were newly formed in waves 16[19] (n=174) or 17 (n=236) .[20]

The case numbers for the gross sample size of the panel households in the respective survey waves and subsamples[21] are reported in → **Table A1**. In wave 17, at least one interview could be conducted for 6,507 households in the panel sample. In addition, there are 1,311 first-time interviewed households from the EWO replenishment sample, of which 1,204 declared their willingness to participate in the panel, as well as 349 from the usual BA access sample, of which 330 declared their willingness to participate in the panel. A total of 62 households come from the access sample of Syrian/Iraqi households, 56 of which were willing to participate in the panel. There were 633 households from the first-time access sample of Ukrainian households, 596 of which were willing to participate in the panel. Also first-time interviewed households of wave 17 are 82 split-off households, which result from the subsamples of the previous waves 1 to 16 and are included in the households of the panel sample.

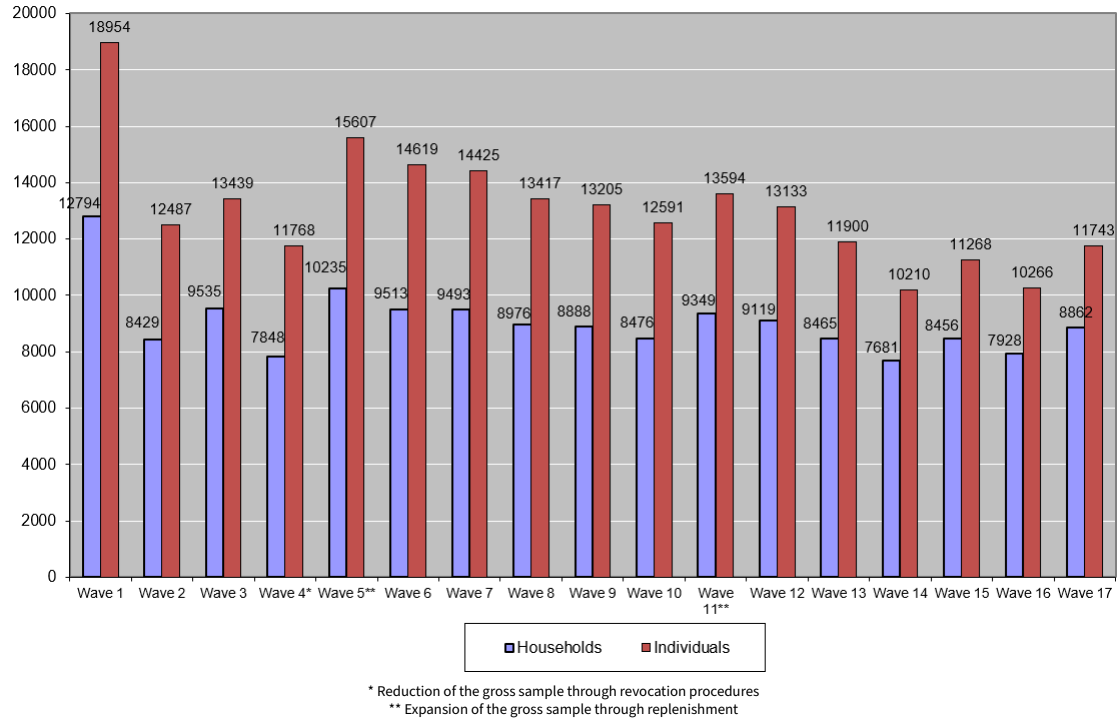
The 8,862 household interviews conducted in wave 17 correspond to 11,743 personal interviews. → **Table A2** lists the distribution of respondents across subsamples and survey waves.

For respondents without sufficient German language skills, interviews were offered in Turkish and Russian in wave 1 to 9. To also interview Syrian and Iraqi households, Arabic was added as an interview language from wave 10 onwards. Since wave 10 interviews in Turkish were not offered anymore. In wave 17, Ukrainian was also included as an additional foreign language for the first time, so that Ukrainian households could also be surveyed. → **Table A3** indicates how many households or persons were interviewed in these additional survey languages.

For the overall data pool of the realised panel sample, the following figure outlines households and individuals over the survey waves.

[19] Split-off households which could not be interviewed in the wave before, were considered like temporary drop outs and should be interviewed again in the following wave. Cases which could not be realized in the following wave were considered like final drop outs.
[20] Case numbers for the gross sample see Methodenbericht wave 17 (Jesske et al. 2024).
[21] The case numbers contain all cases of the register file. Deviations to the method data are possible because of subsequent data checks and cleaning procedures.

Figure 1: Realised panel sample for households and individuals by survey wave



2.2 Response rates

The response rate is calculated according to the standards of the American Association for Public Opinion Research (AAPOR) (AAPOR, 2011). The response rate (RR1) is reported, which includes all cases of unknown eligibility in the denominator and therefore provides the minimum value of all response rates[22]. The response rate at the household level is calculated from the share of usable household interviews as a proportion of the total usable household interviews and non-neutral nonresponses. Only households in which all members have passed away or moved abroad permanently are considered cases of neutral nonresponse. Households are considered usable if at least one complete household interview is available. New households are considered usable if both the household interview and at least one complete personal interview are available. → **Table A4** shows the response rates at the household level for wave 17.

[22] This issue is addressed in very different ways in Germany. Frequently, a large number of individuals or households that were not interviewed are considered ineligible and are removed from the denominator when the response rate is calculated. When a sample is drawn from registers, neither a household that is not living at the expected address nor a household that claims not to belong to the target group may be considered to have provided a neutral nonresponse. Moreover, the population of PASS is not restricted to German-speaking respondents or individuals who can be interviewed; therefore, the nonresponse reasons „does not speak German“or „respondent is sick/unable to be interviewed“cannot be considered cases of neutral nonresponse.

In a household survey, one can distinguish between the response rates at the household level and within the household.

The response rate within households indicates the average proportion of household members aged 15 or older within non valuable households for whom a complete personal interview is available.

The average response rates within interviewed households are shown in → **Table A5**

In addition to the between- and within-household response rates, → **Table A6** provides the repeat interview rate at the individual level. This value is the proportion of individuals willing to participate in the panel with whom an interview could be conducted in the subsequent wave.

2.3 Panel participation agreements, merging data and linking with process data

Respondent consent is always required to store addresses for repeat interviews in a subsequent wave and to merge survey data with the process data obtained from the Federal Employment Agency.

Panel participation agreement was explained in detail in Chapter 2.1. HHneu[23] consent to participate in the panel is illustrated in → **Table A7**

The consent to participate in the panel is recorded following the first personal interview in a new household during each wave. The information provided by that individual is assumed to apply to the household. That is, if the individual consents to participate in the panel, the household is considered willing to participate in the panel and if the individual does not agree to participate in the panel, the household is considered unwilling to participate in the panel (see also Chapter 2.1)[24].

[23] All households in wave 1 are HHneu. Subsequently, only households from the refreshment samples and split-off households participating for the first time are considered HHneu. Therefore, since wave 2, households interviewed for the first time have been in the minority - the majority of household interviews conducted in these waves were conducted previously.

[24] One individual confirms household willingness to participate in the panel. The information available on the household level was integrated into the individual dataset (*PENDDAT*) during data preparation. The individual respondents in the household were assigned the corresponding information available for that household. The same procedure was applied during wave 2. In wave 1; however, consent was recorded after each individual and senior citizen interview; therefore, data could vary within a household. Households with at least one individual willing to participate in the panel were considered willing to participate in the panel. As part of updating address information after the first personal interview in re-interviewed households, it was explained that an interview would be conducted again the following year. If the respondent did not explicitly object to this notification, the household was considered to agree to participate in the panel and the panel variable in the individual dataset (*PENDDAT*) was updated accordingly.

In contrast, permission to merge process data from the Federal Employment Agency with the survey data was obtained for each respondent who was interviewed using the personal questionnaire. This question does not apply to individuals aged 65 and over because it is not included in the senior citizens questionnaire. Consent to merging of these data is not obtained again in each wave[25].

→ **Table A8** provides an overview of obtained consent to merge data in each wave. Only interviews in which consent to merge data was requested in that wave as part of the personal questionnaire are listed.

2.4 Split-off households

PASS is designed as a dynamic panel. Individuals who join or are born into the household are interviewed if they are at least 15 years old. Individuals who move out of sample households for one year or more should continue to be interviewed; however, these individuals are considered new, split-off households. These split-off households also become sample households in PASS. All individuals 15 years of age or more living in these households become target persons for personal interviews. If part of this split-off house-hold in turn splits off in subsequent waves, then this new split-off household also becomes a PASS sample household regardless of whether that new household contains anyone from the original sample (see infinite degree contagion model, Rendtel & Harms 2009, 267). However, individuals who have moved abroad are removed from the survey because they no longer belong to this population and research questions specific to SGB II no longer apply. Individuals who leave the household for less than one year continue to be considered household members.

There are 1,938 split-off households from waves 1 to 17, of which 591 could be interviewed during wave 17, including 40 newly split-off households from wave 17 and 42 Split households surveyed for the first time that could be identified in wave 16. Please refer to the methods report for wave 17 for further information about split-off households (Jesske et al. 2024).

The interviewed split-off households can be identified in the datasets by comparing the current household number (*hnr*) with the original household number (*uhnr*), which differs in

[25] Due to filtering modifications, there were cases in which permission to merge data was raised again in waves 2 and 3 if the respondent had not previously agreed to that during the previous waves. Since wave 6 respondents who refused to give permission to merge data in the previous wave are asked for permission once again. The question is not raised again if the respondent refuses to give permission a second time.

these cases. The original household number (*uhnr*) contains the household number of the panel household from which the new household has separated. Split-off households assume the sample indicator (*sample*), sampling year (*jahrsamp*), primary sampling unit (*psu*) and stratification (*strpsu*) of their original household.

3 Dataset structure

The usual structure for editing a panel dataset - for example, the German Socio-Economic Panel (GSOEP) or the british longitudinal study „The UK Household Longitudinal Study“ (Understanding Society)“ - involves storing individual and household information in annual individual datasets. If required, these individual datasets can be supplemented with specific datasets, which might have a cross-wave data structure, such as register or spell data.

This data structure allows the information to be stored using relatively little storage space. The variables for each year can be identified immediately when examining the datasets. Identifying the merged additional information via key variables, such as household or personal identification numbers, is also quite simple. However, this common panel data structure increases the difficulty of working with these datasets. If analyses are conducted not only cross-sectionally but also longitudinally, then first, all of the relevant variables from each wave dataset must be integrated into a common dataset and care must be taken to ensure that the constructs are comparable for each year. For typical longitudinal analyses, the cross-wave dataset created in this way then must be reshaped into the so-called long format. Unlike the wide format, which contains a data matrix with one row per observation unit (e.g., the household or individual) and several datasets for each survey wave, in the long format, all of the waves assigned to an observation unit are arranged below one another. Rather than arranging information in wave-specific variables in the same row, in long format, the information is assigned to the same variable in each case in wave-specific rows for the observation units.

Reshaping the data into long format has both advantages and disadvantages. The decisive advantage of this variant is that this data structure is required for many longitudinal analyses (such as event history analyses). It is no longer necessary to invest additional time and effort creating a cross-wave file. The switch from long format to wide format is also quite easy to perform. Stata, for example, provides an option to switch between formats with little effort using the „reshape“command. Until a few years ago, the central argument against using this type of data structure was the significantly larger storage space required because even variables recorded in only one or a small number of survey waves require a complete column across all of the waves in the dataset. In addition, these long files become quite large with the increasing duration of the panel because all annual waves are appended, which significantly increases the storage space required and time needed to perform individual operations. The current wide availability of fast processors and large storage capacities even on simple desktop computers render this objection irrelevant. Another disadvantage occurs when merging additional data sources. Unlike datasets prepared in wide format, an additional variable is now required to identify an observation

clearly. This variable may be a wave identifier in the household or individual datasets or the spell number in the spell datasets, which are also available in long format. Furthermore, it is not immediately apparent which variables were included in each wave because all variables are present in the dataset. These variables are assigned a special code (-9) to identify waves during which they were not surveyed.

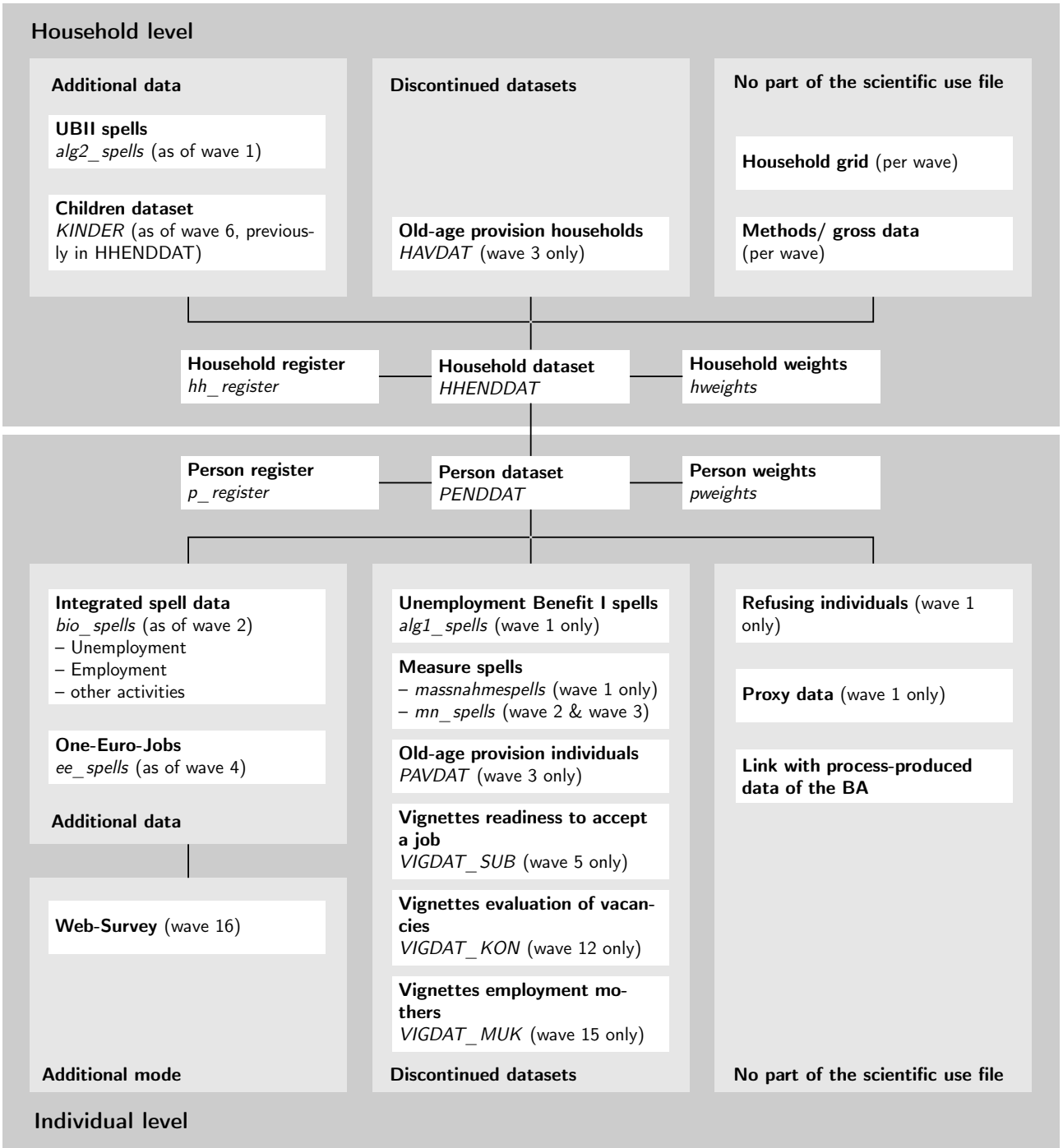
When the advantages and disadvantages of long format are weighed, the advantages of the long format clearly outweigh the disadvantages. Accordingly, household and individual PASS datasets (*HHENDDAT; PENDDAT*), corresponding weighting data (*hweights; pweights*) and a new dataset since wave 6 on children (*KINDER*) were prepared in long format.

At the household level, the scientific use file contains the data on household receipt of Unemployment Benefit II in spell form (*alg2_spells*). Since wave 4, the individual level has contained an integrated biographic spell dataset (*bio_spells*), that integrates and replaces the previous spell datasets *et_spells, al_spells* und *lu_spells*. Furthermore, a one Euro spell dataset (*ee_spells*) is available for waves 4 to 17. The household and person registers (*hh_register; p_register*) are available in wide format. During wave 5, the scientific use file was extended at the individual level by one dataset for the vignette module (*VIGDAT_SUB*) and was complemented by a dataset on resident children (*KINDER*), which includes household information. Vignette modules were again collected in waves 12 and 15. The data set was named *VIGDAT_KON* in wave 12 and *VIGDAT_MUK* in wave 15. The vignette modules collected in the waves 5, 12, and 15 differ in content. Therefore, the *VIGDAT* datasets from the different survey waves should not be merged. For further information on the structure of each dataset, please refer to the PASS User Guide (Fuchs 2013).

In order to make an entry into the self-administered survey in web mode, a web survey was carried out for the first time as part of the „Panel Labour Market and Social Security“ (PASS) following wave 16 in 2022. In particular, the aim was to test optimal procedures for conducting web surveys in the PASS sample (Dickmann et al., 2023). The dataset is not part of the Scientific Use File, but can be obtained from the RDC.

In the labeling of all datasets of the Scientific Use File, umlauts and special characters are resolved in order to ensure a uniform display of the label texts regardless of the character set used by the individual users.

Figure 2: Dataset structure of PASS in wave 17



1

4 Generated variables

4.1 Coding responses to open-ended survey questions

4.1.1 Open-ended residual categories and open-ended items

Some items of the survey were gathered as closed items with an open residual category or as open-ended items. In such cases, additional variables were usually generated, which differed from the original variable only insofar as the information from the open-ended responses could not be coded to the corresponding categories. Moreover, in some cases, new categories were created based on the information obtained from open-ended questions. The name of these additional variables frequently differs from that of the original variable in the last digit only, where "0" is replaced by "1." The items on country of birth, nationality and parent/grandparent country of residence before migration were anonymised and assigned variable names[26]. The following two tables provide an overview of the open-ended survey questions that were coded for wave 17[27].

[26] *ogebland* (country of birth); *ostaatan* (nationality); *ostaatansyr* (syrian nationality); *ostaatanukr* (ukrainian nationality); *ozulanda* to *ozulandf* (parent/grandparent country of residence before migration).

[27] Variables for which information was obtained via open-ended questions and coded in the previous waves but not in the current wave are not listed (with the exception of the spell dataset for Unemployment Benefit II). Observations in waves without obtaining information on these variables were coded -9 (item not asked in wave) and documented in the survey wave data report.

Table 4: Coding responses to open-ended questions at the household level in wave 17

Regular Variable name	Coded to variable	Dataset	Name
HD1100a-o	*HD1101a-o*	*HHENDDAT*	Other Employment status of HH members, proxy information, if necessary
HW0880a-i	*HW0881a-j*	*HHENDDAT*	Other reason for moving out, not listed
AL20550a-h	*AL20551a-h*	*alg2_spells*	Other reasons for the beginning of UB II receipt
AL21300a-h to *AL22100a-h*	*AL21301a-h* *AL21401a-h* *AL21501a-h* *AL21601a-h* *AL21701a-h* *AL21801a-h* *AL21851a-h* *AL21901a-h* *AL22001a-h* *AL22101a-h* *AL22102a-h* *AL22103a-h*	*alg2_spells*	Other reason for benefit cut, not listed
AL22200a-h	*AL22201a-h*	*alg2_spells*	Other reason for discontinuation of receipt of UB II, not listed

Table 5: Coding responses to open-ended questions at the individual level in wave 17

Regular Variable name	Coded to variable	Dataset	Name
PB0230 (Code 6)	*PB0231*	*PENDDAT*	Other German school qualification, not listed (update)
PB0230 (Code 7)	*PB0231*	*PENDDAT*	Other foreign school qualification, not listed (update)
PB0400 (Code 9)	*PB0401*	*PENDDAT*	Other German school qualification, not listed (first survey or not reported in previous wave)
PB0400 (Code 10)	*PB0401*	*PENDDAT*	Other foreign school qualification, not listed (first survey or not reported in previous wave)
PB1000	*PB1001*	*PENDDAT*	Other foreign school qualification, not listed (first survey or not reported in previous wave)

Table 5: Coding responses to open-ended questions at the individual level in wave 17 (continued)

PB1300a-j (Item I)	*PB1301a-j*	*PENDDAT*	Other German training qualifications not contained in the list (first survey or no statement in the previous wave)
PB1300a-j (Item J)	*PB1301a-j*	*PENDDAT*	Other foreign training qualifications not contained in the list (first survey or no statement in the previous wave)
PB1600	*PB1601*	*PENDDAT*	Other qualification to which the foreign qualification corresponds, not listed
AL0600	*AL0601*	*bio_spells*	Other reason for no longer being registered as unemployed, not listed
BIO0100	*BIO0101*	*bio_spells*	Other type of activity, not listed
ET2400	*ET2401*	*bio_spells*	Other source to get notice of a job
ET2420	*ET2421*	*bio_spells*	Other social network as source to get notice of a job
EE0300a-h	*EE0301a-h*	*ee_spells*	Other reason for not participating in a one-euro job
EE1000a-e	*EE1001a-e*	*ee_spells*	Other reason why one-euro job was terminated prematurely
EE1600	*EE1600z*	*ee_spells*	Other field of the one-euro job
PTK0320b-g	*PTK0321b-g*	*PENDDAT*	Other reasons not contained in the list regarding why no job was searched
PTK1700a-m	*PTK1701a-m*	*PENDDAT*	Other support from job-center
PTK3200a-j	*PTK3201a-j*	*PENDDAT*	Other topics of discussion with the job center
PAS0900a-g	*PAS0901a-g* *PAS0901i*	*PENDDAT*	Other places where target pers. obtained information about job vacancies, not listed
PAS0920a-l	*PAS0921a-l*	*PENDDAT*	Other social network used for job search
PER0200a-e	*PER0201a-e*	*PENDDAT*	Other reasons for employment in retirement
PER0400	*PER0401*	*PENDDAT*	Other way to be informed of the vacancy
PER1000a-e	*PER1001a-e*	*PENDDAT*	Other reasons for the planned employment in retirement
PER1300a-g	*PER1301a-g*	*PENDDAT*	Other reasons against employment in retirement

Table 5: Coding responses to open-ended questions at the individual level in wave 17 (continued)

PSK0700a-e	*PSK0701a-e*	*PENDDAT*	Other reasons for voluntary engagement
PSK1100a-e	*PSK1101a-e*	*PENDDAT*	Other reasons for the planned voluntary engagement
PSK1200a-f	*PSK1201a-f*	*PENDDAT*	Other reasons against voluntary engagement
PAS0950a-i	*PAS0951a-i*	*PENDDAT*	Other form of disability/impairment
PG1310	*PG1311*	*PENDDAT*	Statutory health insurance fund
PG1320	*PG1321*	*PENDDAT*	Private health insurance fund
PG1330	*PG1331*	*PENDDAT*	Private health insurance fund and eligible for additional allowances for public
PG1340	*PG1341*	*PENDDAT*	Free provision of health services for civil servants
PG1350	*PG1351*	*PENDDAT*	Insured differently
PMI0200	*ogebland*	*PENDDAT*	Other country of birth, not listed
PMI0500	*ostaatan*	*PENDDAT*	Other nationality, not listed
PMI1000a-f	*ozulanda-f*	*PENDDAT*	Other country of birth, not listed country from which parent/grandparent migrated
PMI1700	*PMI1701*	*PENDDAT*	Legal basis of the entry into Germany
PMI3000	*PMI3001*	*PENDDAT*	Other reason not to apply for recognition of a vocational qualification obtained abroad in Germany
PSH0200 (Code 9)	*PSH0201*	*PENDDAT*	Other German school qualification of mother, not listed
PSH0200 (Code 10)	*PSH0201*	*PENDDAT*	Other foreign school qualification of mother, not listed
PSH0300a-i (Code 7)	*PSH0301a-i*	*PENDDAT*	Other German vocational qualification of mother, not listed
PSH0300a-i (Code 8)	*PSH0301a-i*	*PENDDAT*	Other foreign vocational qualification of mother, not listed
PSH0500 (Code 9)	*PSH0501*	*PENDDAT*	Other German school qualification of father, not listed
PSH0500 (Code 10)	*PSH0501*	*PENDDAT*	Other foreign school qualification of father, not listed

Table 5: Coding responses to open-ended questions at the individual level in wave 17 (continued)

PSH0600a-i (Code 7)	*PSH0601a-i*	*PENDDAT*	Other German vocational qualification of father, not listed
PSH0600a-i (Code 8)	*PSH0601a-i*	*PENDDAT*	Other foreign vocational qualification of father, not listed

4.1.2 Coding of occupation and industry

Occupations are coded in accordance with ISCO (ISCO-88/ISCO-08) and the German Classification of Occupations (KldB) (1992/2010), and industries in accordance with the German Classification of Economic Activities (WZ) (2003/2008). The coding of occupations requires specific knowledge which is taught to the coders in training courses. The training courses use standardised training materials. The first training session for new coders comprises a presentation in which the basic rules of coding and the ISCO/KldB coding are taught, as well as the coding and discussion of selected test cases with various levels of difficulty. The training course lasts one and a half days.

If coders have not done any occupation coding for more than six months, the coding rules are refreshed at the start of a new project and all the coders' results are compared. To this end at least 500 randomised cases are coded by all the participants and the discrepancies are analysed. With this procedure individual coders' systematic errors can be detected and discussed before the coding process.

In the course of the project, regular quality checks are conducted in addition to the training in order to assure quality. During the coding process the coders receive individual feedback about any discrepancies arising. To this end, cases in which a suggested code was rejected are listed for all the coders. If systematic errors emerge, they are discussed with the respective coder.

The coding of occupations and industries involves the following process steps:

1. Preparation of the coding materials
   For coding occupations, not only the responses to the open-ended questions about the respondent's occupation from the interview should be used but also additional variables. Before the coding begins, the main staff responsible for the coding agree with those working in data preparation regarding what additional information is available in the survey questions and will be given to the coders together with the open-ended responses regarding occupation.

In PASS the following additional variables are generated from the information reported and are given to the coding staff as a coding list in Excel format together with the open responses on the occupation:

Table 6: Coding scheme of the additional variables used in PASS

Abbreviation	Title
StiB_g	Basic classification of the occupational status
ang	White-collar worker
arb	Blue-collar worker
bea	Civil servant or judge
selbst_f	Self-employed in an independent profession
selbst_H/DL	Self-employed in trade or craft, commerce, industry, services
landw	Self-employed farmer
mith_f	Family member working for a self-employed relative
sol	Professional soldier
k.A.	Details refused
wn	Don’t know
StiB_f	Detailed classification of the occupational status
xxHektar	Farmer with xx hectare
xxMitarbeiter	Self-employed or academic independent profession with xx employees
40	Civil servant, simple administrative duties
41	Civil servant, mid-level administrative duties
42	Civil servant carrying out senior administrative duties
43	Civil servant, executive duties
45	Enlisted personnel, other than non-commissioned officer
46	Enlisted personnel, non-commissioned officer
47	Commissioned officer, captain or lower rank
48	Commissioned officer, major or higher rank
51	Employee, simple duties
52	Employee, under close supervision
53	Employee, carrying out responsible tasks independently

Table 6: Coding scheme of the additional variables used in PASS (continued)

54	Employee, wide managerial responsibilities
60	Unskilled worker
61	Semi-skilled worker
62	Skilled worker
63	Foreman
64	Master craftsman, site foreman
k.A.	Details refused
wn	Don’t know
Aufs,x	Supervising responsibility, number of supervised employees
Aufs,x	Supervising responsibility, number of supervised employees
k.Aufs	No supervising responsibility
Schul	Highest school qualification
(fa)Abi, Eos12	General/subject-specific upper secondary school
Fabi	Upper secondary school
Real, Pos.10	Intermediate secondary school
Haupt, Pos.8/9	Lower secondary school
Sonder	School incorporating physically or mentally disabled children
and	Other degree
Ausl	Foreign degree
kAB	No degree
Schüler	Still pupil in a general-education school
k.A.	Details refused
wn	Don’t know
Aus	Vocational Qualification (multiple entries possible)
Anlern/Tfach.	Training as a semi-skilled worker
Le	Apprenticeship, vocational training
Ges	School for health care professionals
BerAk	Professional college
BeruFab	Full-time vocational school
Meist/Tech	Master craftsman qualification, a technician qualification

Table 6: Coding scheme of the additional variables used in PASS (continued)

Dipl (FH), BA (Uni,FH)	Diploma (University of Applied Sciences) or Bachelor (University, University of Applied Sciences)
Dipl (Uni), BA + MA (Uni)	Diploma and such(University) or Bachelor/Master (University, University of Applied Sciences)
Prom/Hab	Doctorate or post-doctoral lecturing qualification
Schüler	Student in a general-education school
and	Other degree
Ausl	Foreign degree
kAB	No vocational qualification
k.A.	Details refused
wn	Don't know
ÖD	Public service
ÖD	Employed in public service
nÖD	Not employed in public service

Besides the coding list, the coding materials also include further information, such as rules for as-signing codes when the variable attributes are not clear, which are provided in the form of a continuously growing collection of cases. This list is continually filled with the occupational codes im-plemented in the institute. The internet can also be used for researching occupations (e.g. berufenet provided by the Federal Employment Agency; the classification server of the Federal Statistical Office, ILO, Statistics Austria for ISCO-08).

At the start of a project, if necessary, the general coding rules are adapted or special rules are drawn up for the particular specific project, depending on the data provided or rules from previous waves of the project. These adapted coding rules are documented and passed on to the coders.

The content of the columns in the coding lists is standardised across all projects and is designed to document permanently not only the final result but also all the steps described in the following. The lists document not only the codes of the individual coding steps and the coders' coding numbers but also, where applicable, comments regarding difficulties occurring in the coding process.

2. **First coding**

   Initial coding is carried out with the help of automatic precoding software (supervised learning). The data is imported into the electronic coding system and pre-coded there. With the help of the electronic pre-coding software, up to 40 percent of the open data can already be clearly assigned to a code, i.e. in 40 percent of cases the automatic

pre-coding assigns exactly one code. All other data is pre-coded with two suggested codes.
The prerequisites for a high hit rate are good data quality and the collection of all relevant additional information.

3. **Second coding**
   For all entries, a critical review of the cases from the automatic pre-coding for plausibility of content is carried out as part of a second coding. The second coder sees the result/suggested results of the first coding and can accept it (one suggestion) or select it (two suggestions) and correct it. With the help of the automatic pre-coding software, a hit rate of 70 percent of correct codes can now be achieved, which are no longer corrected by a coder in the second coding.
4. **Third coding**
   Differences in the codes assigned in the first and second coding steps are clarified by a third coder. Problem cases are discussed and decided in discussion groups. If the third coder clearly agrees with one of the two assigned codes because the other code is clearly incorrect, the correct code is entered in the decision column. If the third coder is unable to decide between the two codes or suggests another code, then this is marked. This case is then to be discussed in the meeting concerning problem cases. In addition a comment column can be used to justify a decision.
5. **Discussion of problem cases**
   The coders meet regularly to discuss problem cases and to make decisions regarding codes.
6. **Last check**
   Finally, the main staff responsible for the coding process check the correctness of the codes and compliance with the most important coding rules.

4.2 Harmonisation

The survey instruments for some variables changed across waves. In particular, the integration of the module “employment biography” in wave 2 provided critical information on employment status, current main employment, status of economic inactivity and receipt of UB I in a different way than in wave 1. Since then, information has been collected not only for the date of the interview but also for particular periods.

To facilitate cross-wave analyses in such cases, variables are generated for important indicators, which are harmonised across waves. Harmonisation creates a special group within the generated variables (see Section 4.4) that is used to standardise indicators collected in different ways retrospectively.

Changes between the waves can affect the entire survey concept, categories and interviewed groups. Harmonised variables thus consider different source variables that result from changed survey concepts, categories or interviewed groups. This was an effort to standardise them across waves as much as possible before variables were generated.

Thus far, the simple classification for occupational status (*stibkz*) has been harmonised; however, the need harmonisation is expected to increase with the duration of the panel.

Table 7: Harmonised variables in the individual dataset (*PENDDAT*)

Variable	Subject	Name
stibkz	Employment	Current occupational status, simple classification, harmonised (anonymised)

Although explicitly harmonised variables also consider changes in categories and interviewed group across waves - in addition to changes in the survey concept - a second type of variable does not explicitly consider changes in the interviewed groups. These variables are generated for all waves but may contain information for different groups of respondents in each wave. These differences result from revisions to the filtering processes performed between waves and affect the source variables of generated variables.

Accordingly, cross-wave variables of this type apply in addition to harmonisations and standardise individual aspects across waves. In contrast to the harmonised variables, they are generated for each wave for all groups for which the corresponding source variables were collected. Thus, they can easily be used to evaluate the cross-section of a specific wave. However, in the longitudinal section, these differences must be considered before statements about changes between the waves can be made.

Before working with cross-wave but not harmonised variables, it should be verified whether differences in the interviewed groups might cause problems in the evaluations, and it should be determined whether standardisation is necessary[28]. Subsequent cross-wave variables are different for the group for which they are generated.

[28] For example, in wave 1, the groups of respondents that were questioned about their employment were different from those questioned in the waves that followed. Accordingly, the respective groups that provided information about occupational status, occupational activities, working hours, fixed-term employment, etc., varied.

Table 8: Variables in the individual dataset (*PENDDAT*) are generated across waves but not completely harmonised (*PENDDAT*)

Variable	Subject	Name
isco88	Employment	Intern. Standard Classification of Occupations 88, current employment, gen.
kldb1992	Employment	Classification of occupations 1992, current employment, gen.
azhpt2	Employment	Current actual working hrs. main employment (without marginal employment, incl. cat. info.), gen.
azges2	Employment	Current total actual working hrs. (without marginal employment , incl. cat. info.), gen.
befrist	Employment	Current activity: limited contract? Generated (all waves)
mps	Employment	Magnitude Prestige Scale, current employment, gen.
siops1	Employment	Standard Intern. Occupational Prestige Scale (Basis ISCO88), current employment, gen.
isei1	Employment	International Socio-Economic Index (Basis ISCO88), current employment, gen.
egp	Employment	Class scheme acc. to Erikson, Goldthorpe and Portocarrero (EGP), current occupation, gen.
esec	Employment	European Socio-economic Classification (ESeC), current occupation, gen.
stib	Employment	Occupational status, code number, current employment, gen.
netges	Employment	Current total net income (without marginal employment, incl. cat. info.), gen.
alg1abez	Benefit receipt	Current receipt of UB I, gen.
aktmassn	Participation in measures	Current participation in a programme funded/promoted by the employment agency, gen.
isco08minj	Minijob	ISCO 08, current minijob, gen.
kldb2010minj	Minijob	Classification of occupations 2010, current minijob, gen.

4.3 Dependent Interviewing

At various times in both the household and personal interviews, information was gathered via dependent interviewing, i.e., interviews that were dependent on the responses provided during a previous wave. In this approach, data from the previous interview are used to control the filter questions or are integrated directly into the question text of the current interview.

Two main goals were pursued, utilising information from previous waves[29]. First, changes that occurred since the previous wave were recorded, depending on the information available from the previous wave. At those points, information from previous waves was used to control the filter. Second, the respondent should have received information. In places where changes since the previous wave were to be collected, the interview date of the previous wave was included in the question text to clarify the definition of the reporting period[30]. In other places, especially where spell information was updated[31], the previous response was integrated into the question text to remind the respondent and prevent incorrect changes in status. Such changes are artifacts of the open-ended survey question arising out of inaccurate memories or imprecise information.

If information from a single wave in the dataset is reviewed, information is incomplete for some respondents due to dependent interviewing, which only represents the changes between survey dates. For respondents who are interviewed for the first time about a certain topic, complete information might be information available for that wave[32].

During data preparation, the recorded changes are combined with information from the previous wave to create variables and datasets with complete information. The spells in the existing spell datasets are then updated. In the cross-section datasets (*HHENDDAT, PENDDAT*), however, generated variables are created in which the information from the previous wave is combined with the reported changes.

The following two tables provide a brief overview of the relevant updates to the questionnaires and indicate the variables for which updated information was obtained.

[29] For example, individuals were only asked about their highest school qualification once. Only qualifications obtained since the previous interview were reported in subsequent waves.

[30] For example, if only new school qualifications were to be reported, the following question was asked: ”Have you obtained a general school qualification since our last interview on [interview date of previous wave]?”

[31] Examples include updates of UB II receipts since the previous wave in the household interview or employment or unemployment updates in the individual interview.

[32] Individuals who were asked about their school qualifications for the first time reported their highest school qualification. Therefore, complete information on the highest school qualification is available for this wave in the recorded variable. In the subsequent wave, only newly obtained school qualifications are recorded. For example, if a school qualification is recorded, it is not clear whether it represents the individual’s highest school qualification. In that sense, the information obtained in the subsequent wave is incomplete in its reported variables.

Cases for which generated variables were updated or continued are listed in Chapter 4.4 of this data report.

Table 9: Updated information in wave 17, household questionnaire

Construct	Q.No.	Note	Update in var.
Housing situation		Form of accommodation, type of tenancy and type of hostel/home/hall of residence updated during the interview	*HHENDDAT*: *HW0200a* to *HW0400*
Household structure		Household size updated during the interview	*HHENDDAT*: *HA0100*
		Sex of the individuals in the household corrected during the interview, if necessary	*HHENDDAT*: *HD0100a* to *HD0100o*
		Age of the individuals in the household updated during the interview	*HHENDDAT*: *HD0200a* to *HD0200o*
		Family relationships updated during the interview	not provided in the SUF
Size of dwelling in sqm	HW1000	Updated in generated variable	*HHENDDAT*: *wohnfl*
Receipt of Unemployment Benefit II	Module "Unemployment Benefit II"	Updated in Unemployment Benefit II spell dataset	*alg2_spells*: Variables of the Unemployment Benefit II spell dataset
		Information on the HH's current receipt of Unemployment Benefit II	*HHENDDAT*: *alg2abez*
		Information on the benefit units's Unemployment Benefit II receipt	*p_register*: *bgbezs17; bgbezb17*

Table 10: Updated information in wave 17, personal questionnaire

Construct	Q.No.	Note	Update in var.
Highest general school qualification	PB0220-PB1100	Updated in generated variable	*PENDDAT*: *schul1* (without responses to open-ended questions) *schul2* (responses to open-ended questions)
Year in which highest school qual. was gained	PB0410	Updated in generated variable	*PENDDAT*: *schulabj*
Vocational qualification	PB1200-PB1600	Highest vocational qualification, updated in generated variable	*PENDDAT*: *beruf1* (without responses to open-ended questions) *beruf2* (responses to open-ended questions)
Year of vocational qualification	PB1310a-k	Updated in generated variable	*berabj*
Periods of updated activities in the BIO spell dataset	BIO0600z1, BIO0600z2, BIO0400z, BIO0500z	Updated in the BIO spell dataset for attached spells	*bio_spells*: *BIO0400, BIO0500, BIO0600*
		Information about finished activities	*bio_spells*: *ET2300, ET2700*
		Information on current employment, updated in generated variables	*PENDDAT*: *isco88; isco08; kldb1992; kldb2010; stib; stibkz; azhpt1; azhpt2; azges1; azges2; befrist; mps; siops1; siops2; isei1; isei2; egp; esec; branche1; branche2*
		Information on current economic inactivity/employment status, updated in generated variables	*PENDDAT*: *etakt; alakt; statakt*
Periods of receipt of Unemployment Bene-fit I in updated unemployment spells		Information on current receipt of Unemployment Benefit I	*bio_spells*: *AL0700, AL0800, AL0900, AL1000, AL1100, AL1200*

Table 10: Updated information in wave 17, personal questionnaire(continued)

		Information on finished periods of unemployment Information on the current receipt of unemployment benefit I	*bio_spells*: *AL0600, AL0601* *PENDDAT*: *alg1abez*
Periods of updated activities in the EE spell dataset			*ee_spells*: *EE0800a, EE0800b*
Information regarding premature end in the EE spell dataset			*ee_spells*: *EE0900, EE1000a-EE1000e, EE1001a-EE1001e*

A distinction must be drawn between characteristics for which previously collected information is updated with information on changes between the survey dates and so-called constant characteristics that are not expected to change over time. Therefore, these characteristics are recorded only once in PASS, but in some cases, corrections are possible. Because information on these characteristics is usually only available for the surveyed variables during the first interview, they are subsequently provided in the form of generated variables (see Chapter 4.4, User Guide PASS Wave 6).

4.4 Simple generated variables

Simple generated variables include variables for which different items in a construct are surveyed separately for technical reasons and then aggregated. Alternatively, information from the current wave is combined with information from the previous wave (see Chapter 4.3), such as the highest educational qualification (see Chapter 4.3). Important information can also be obtained by merging partial datasets (e.g., indicators for current receipt of UB I or II).

The simple generated variables for households and individuals who are interviewed on a topic for the first time can always be generated based on information from the current wave. Households and individuals who provided information on a topic during a previous wave can be differentiated in the cross-section datasets (*HHENDDAT; PENDDAT*) to indicate the origin of the variables necessary for variable generation. The three different types of simple generated variables are provided in the following table.

Table 11: Simple generated variables in the cross-section datasets (*HHENDDAT; PENDDAT*) for households and individuals who previously provided information on the topic

Type	Generation based on source data from wave of the first survey of the topic for HH/individ.	Generation based on source data from current wave	Description
constant (uv)	yes	no	Information gathered in the first survey is generally adopted in the subsequent wave, unless input errors were corrected in the current wave. Example: *zpsex* (sex)
continued (fs)	yes	yes	Information that was current in the previous wave is combined with information of the current wave and updated, if necessary. Example: *schul1* (highest school qualification)
independent (new)	no	yes	The variable is newly generated from the data of the current wave in each wave, regardless of the information from the previous wave. Example: *hhincome* (net income of household)

Explanations that are more detailed must be provided on the type “unveränderlich (uv)” simple generated variables for *PENDDAT*. A first-time survey of a topic with an individual does not always take place during the first wave in which the individual provides an interview. Two groups of individuals are considered first-time interview respondents even if they provide a repeat interview.

The first group is individuals moving back into a household. Individuals who move out of their previous household to form a split-off household (see Chapter 2.4) take their preload information with them. Thus, they can be treated correctly as either first-time interviews or repeated interviews. However, if an individual returns from a split-off household into a panel household in which he/she lived during a previous wave, the preload of this individual is not transferred from the split-off household to the original household. Individuals returning home are treated as first-time interviewees. This situation has occurred since wave 3. The first move-outs of HHalt occurred during wave 2, and returns may occur by wave 3.

An individual preload for dependent interviewing is created for an individual (see Chapter 4.3) only if he/she provided an interview during one of the two preceding waves. The

context for this rule is that there is a point in time until which an individual is expected to remember the response in spell form. Individuals who last provided a personal or senior citizen interview during the third wave or earlier had passed this point. To reduce respondent stress and protect the validity of the information provided, which is presumably severely threatened beyond this limit, individuals whose reference date for information about spell results is before the relevant date are treated as first-time respondents[33]. This situation first occurred in wave 4 because that wave was the first time that a previous personal interview could have taken place more than two waves previously.

The information on which these generated variables are based is collected again for these two groups (e.g., in the module "social origin") because they are treated as first-time interviews. Data preparation treats this survey information identically to the information from individuals engaged in actual first-time interviews within the PASS framework. These generated variables, e.g., the status of the mother and father, are thus based on information from the current wave. No transfer of information from previous waves takes place, and there is no attempt to make the data fit plausibly with previous information. We assume that the information provided by the target person, which is processed to become generated variables, is consistent with previous information in a repeated survey. However, deviations from previously obtained information in the previous waves cannot be generally excluded. Individuals included in either group are flagged in *PENDDAT* by the variable *altbefr* as first-time respondents (code "0" or "-9" for wave 1).

These simple generated variables are provided in the following six tables. The tables include short descriptions of each variable. Furthermore, the source variables to generate the variable are indicated[34]. For the cross-section datasets (*HHENDDAT; PENDDAT*), additional information identifies the type based on the simple generation from the previous table (uv; fs; new). This division is not used for spell datasets because there are no wave-specific observations. Instead, variables are newly generated at the spell level if the spell was newly included in the wave or was updated with information obtained in the current wave. In addition, register datasets follow a different logic, and no further differentiation was made.

[33] Excluding previously granted consent to the merging of data. This preload information is generated regardless of when the previous personal interview was provided to avoid individuals negating question *RegP0100* and de facto withdrawing their consent. The option to with-draw consent to the merging of data remains unaffected by this decision.

[34] The data report documents how the variables in the cross-section datasets (*HHENDDAT; PENDDAT*) were generated for observations in previous waves. The documentation for specific waves also describes the generation of wave-specific variables in the register datasets. The generated variables in the spell datasets were always generated in the updated datasets. If a spell was not updated, the generated variables remain unchanged (with the exception that a special code was used in the censoring indicator if the spell could not be continued for technical reasons). If a spell was updated, then the most current information was used, i.e. the variables provided with information from the current wave or cross-section variables in the spells relevant for the current wave.

Table 12: Wave 17 simple generated variables in the household (*HHENDDAT*) and *KINDER* datasets (in alphabetical order)

Variable	Label and description	Source var. for gen. var wave 17
alg2abez	*Current receipt of UB II of the HH, generated*: Indicator for the household's current receipt of Unemployment Benefit II	*zensiert; AL20300; AL20400; AL20500 (alg2_spells)* information on further receipts of Unemployment Benefit II *(AL22700); hintjahr (HHENDDAT)*
anzgeschw	*Number of siblings in the household*: Indicator of an individual's number of siblings Parenthood and sibling status are surveyed separately. Individuals may share one parent but not call themselves siblings. Therefore in some cases, *anzgeschw* is not equivalent to sibling status, which can be generated through the parent indicator variable in *p_register*.	Information to relations in the household *household grid*
bik	*BIK region size classes (GKBIK10), generated*: The information on region size was generated by infas by converting the postcode from the address to *GKBIK10* (new).	Supplied by survey institute
blneualt	*Western German States or Eastern German States, generated*: Divides the German states into the western states of the former FRG (excluding Berlin) and the eastern states of the former GDR (with Berlin). Infas determined the state based on the postcodes the address data (new).	*bundesld* Information generated and supplied by the survey institute on the federal state in which the household is resident at the survey date.
butaber	*Eligibility for education package at point of interview*: This variable indicates that a household is eligible to draw benefits from the education and participation package if he draw one of the benefits like UB II, children's allowance, housing or social benefit since January of the year before the actual year of the survey (new).	*AL20200; AL20400; AL20500 (alg2_spells); HA0250a-b; HW1800; HW1950; HEK0100; HEK0115; HEK1630; HEK1645 (HHENDDAT)*
hhinckat	*Categorised household income per month (in EUR), gen.*: Categorised information on the household's income aggregated from several survey items into one variable (new)	*HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hhincome	*Household income per month (in EUR) incl. categorised information, gen.*: This generated variable integrates information from categorised and openended survey questions on net household income (new).	*HEK0600; HEK0700; HEK0800; HEK0900; HEK1000; HEK1100 (HHENDDAT)*
hintdat	*Date of household interview*: This generated variable indicates the date on which the household interview was conducted in the format YYMMDD (new)	*hintjahr; hintmon; hinttag (HHENDDAT)*

Table 12: Wave 17 simple generated variables in the household (*HHENDDAT*) and *KINDER* datasets (in alphabetical order) (continued)

hintnum	*Interviewer in household interviews*: The artificial identifier indicates the interviewer who conducted the interview. This information is consistent between *PENDDAT* and *HHENDDAT* as well as across waves. A definite characteristic of the label always identifies the same interviewer (new).	information that is generated and supplied by the survey institute
kindu4	*Control variable: child under the age of 4 in the HH*: A variable indicating that at least one individual in the household is under the age of four in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged four is actually the child of another individual living in the household (new).	*HD0200a - HD0200o (HHENDDAT)*
kindu13	*Control variable child under the age of 13 in the HH*: A variable indicating that at least one individual in the household is under the age of 13 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 13 is actually the child of another individual living in the household (new).	*HD0200a - HD0200o (HHENDDAT)*
kindu15	*Control variable: child under the age of 15 in the HH*: A variable indicating that at least one individual in the household is under the age of 15 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 15 is actually the child of another individual living in the household. If the response to the open-ended question on age was missing, the categorical follow-up question about the age groups was also used to generate the variable (new).	*HD0200a - HD0200o;* categorical follow-up question about age group (in cases of no response in *HD0200 (HHENDDAT)*)
kindu18	*Control variable: child under the age of 18 in the HH*: A variable indicating that at least one individual in the household is under the age of 18 in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 18 is actually the child of another individual living in the household (new).	*HD0200a - HD0200o;* categorical follow-up question about age group (for KA in HD0200) (HHENDDAT)

Table 12: Wave 17 simple generated variables in the household (*HHENDDAT*) and *KINDER* datasets (in alphabetical order) (continued)

kindu25	*Control variable: child under the age of 18 or pupils under the age of 25 in the HH.*: A variable indicating whether at least one individual in the household is under the age of 18 or that at least one individual is between the age of 18 and 25 and pupil. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual of the age group is actually the child of another individual living in the household. If the response to the open-ended question on age was missing, the categorical follow-up question about the age groups was used to generate the variable as well (new).	*HD0200a - HD0200o;* categorical follow-up question about age group (for KA in HD0200); HD1100a-o (HHENDDAT) (in cases of no response in *HD0200*); *HD1100a-o (HHENDDAT)*
kind5u15	*Control variable: child from 5 to under the age of 15 in the HH*: A variable indicating that at least one individual in the household is 5 to under 15 years old in the wave. As the generated variable is based only on the age details in the household dataset, it is irrelevant whether this individual aged 5 to under 15 is actually the child of another individual living in the household (new).	*HD0200a - HD0200o*
wohnfl	*Living space in sqm, gen.*: Information on the size of the living space in the household's current dwelling. In the case of re-interviewed households, the size of the living space was only asked as of the second wave if the household had moved house or if the house/apartment had changed since the previous wave (fs).	For first survey: *HW1000 (HHENDDAT)* For repeated survey:: wohnfl from previous wave; *HW1000; (HHENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order)

Variable	Label and description	Source var. for gen. var wave 17
akt1euro	*Current part. in one-euro job, generated*: Indicator: respondent is participating in a one-euro job program at the time of the interview (new).	*zensiert (ee_spells)*
alakt	*Currently reported as unemployed, generated (as of wave 2))*: Indicator: the TP was unemployed at the date of the personal interview of that wave (new).	*zensiert; spintegr; BIO0101 (bio_spells)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

alg1abez	*Current receipt of UB I, generated*: Indicator: respondent is receiving Unemployment Benefit I at the interview date. In wave 17, the periods since January 2021 during which the respondent was unemployed were surveyed. For each spell, additional questions about whether and when the respondent received UB I (new).	*AL0700; AL1000; AL1100; AL1200 (bio_spells)*
apartner	*Control variable: unmarried partner living in HH*: Indicator: respondent has a cohabitee or partner whose status is not specified in the household (new).	Information on relationships between household members *(Haushaltsgrid); PD0500 - PD0800 (PENDDAT)*
azhpt1	*Current contractual working hrs. main employment (without marginal employment), gen* : Weekly contractual working hours provide the respondent's primary employment at the time of the interview. Generated from open-ended questions about working hours (new).	*ET2015 (bio_spells)*
azhpt2	*Act. effective working time main employment (without minijobs, incl. cat. statements), gen.*: Weekly effective working time of the main job that the respondent performed at the moment of the interview, which is generated using from open-ended questions about working hours and a categorical follow-up question in which irregular working hours were reported (new).	*ET2115; ET2215 (bio_spells)*
azges1	*Current contractual working hrs. (without marginal employment), gen.*: Weekly contractual working hours for all positions held by the respondent at the time of the interview. Generated from open-ended questions about working hours.	*ET2015 (bio_spells)*
azges2	*Current total actual working hrs. (without marginal employment, incl. cat. info.), gen.* : Actual weekly working hours for all positions held by the respondent at the time of the interview. Generated from responses to open-ended questions on working hours and a categorical follow-up question in which irregular working hours were reported (new).	*ET2115; ET2215 (bio_spells)*
befrist	*Current employment: limited contract? Generated (all waves)*: Indicator: the employment position held by the respondent at the interview date is on a limited contract (new).	*PET2510a; PET2510b (PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

begjeewt	*Start year of first employment, generated*: The first year during which the respondent was employed in a regular position. To generate this variable, information about the first regular position was combined with information from the employment spells if the respondent had previously reported his/her first regular employment since January 2021 (uv).	For first survey: *bjahr (bio_spells); PET3200b (PENDDAT)* After first survey: *begjeewt* from previous wave *(PENDDAT)*
begjminj	*Start year of current mini-job, generated*: Year, since which participant is employed in current (main) mini-job (new)	*PMJ0800b*
begmeewt	*Start month of first employment, generated*: The month during which the respondent first held regular employment (generated, see *begjeewt*) (uv).	For first survey: *bmonat (bio_spells); PET3200a (PENDDAT)*; After first survey: *begmeewt* from previous wave *(PENDDAT)*
begmminj	*Start month of current mini-job, generated*: Month, since which participant is employed in current (main) mini-job (new).	*PMJ0800a*
berabj	*Year of the highest vocational qualification*: The year in which the respondent obtained his/her highest vocational qualification at the interview date (fs). *Note: The year in which the reported vocational qualifications reported in wave 1 but asked in wave 2.*	For first survey: *PB1310aj-kj (PENDDAT)* For repeated survey: *berabj* from previous wave *PB1310aj-kj (PENDDAT)*
beruf1	*Highest vocational qual., excluded foreign qual. and open info., generated*: Identifies the highest vocational qualification obtained by the interview date by ranking the vocational qualifications cited by the respondents, excluded information from open-ended questions (fs).	For first survey: *PB0100; PB0200; PB0300; PB1200b; PB1200c; PB1300a-j; (PENDDAT)* For repeated survey: *beruf1* from previous wave; *PB0100; PB0200; PB1200a; PB1300a-j (PENDDAT)*
beruf2	*Highest vocational qual., incl. foreign qual and open info., generated*: Defined as in beruf1 with the following differences: 1. Inclusion of responses to open-ended questions; 2. Inclusion of foreign qualifications; and 3. Degrees are not distinguished by type of institution (e.g., university or other institution of higher education) but by level (Bachelor's degree; Master's degree; Ph.D.) (fs).	For first survey: *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)* For repeated survey: *beruf2* from previous wave; *PB0200; PB1301a-j; PB1500a; PB1500b; PB1500c; PB1601 (PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

brges	*Current total gross income (without marginal employment, incl. cat. info.), gen.*: Contains the cumulative information on gross income from all employment (above the mini-job earnings limit). Generated from the answers provided in open-ended questions on gross income and categorical follow-up question when the "don't know" or "details refused" answers were provided to open-ended questions (new).	*ET2812; ET2912; ET3012; ET3112; ET3212; ET3312 (bio_spells)*
brutto	*Gross income from the current main employment incl. categorised information, generated*: A generated variable integrating information from categorised and open-ended survey questions on gross income (new).	*ET2812; ET2912; ET3012; ET3112; ET3212; ET3312 (bio_spells)*
bruttokat	*Categorised gross income from the current main employment, generated* : This variable aggregates the categorised information on gross income for a specific variable, which combines several items on income categories (new).	*ET2812; ET2912; ET3012; ET3112; ET3212; ET3312 (bio_spells)*
emonlewt	*Time when last employment ended (month)*: Month in which the respondent was most recently employed. To generate this variable, see *ejhrlewt* (fs) .	For first survey: *PET1200a (PENDDAT); ejahr; emonat (bio_spells)* For repeated survey: *ejhrlewt* from previous wave *(PENDDAT); ejahr; emonat (bio_spells)*
ejhrlewt	*Time when last employment ended (year)*: Year, in which the respondent was most recently employed. To generate this variable, information from the employment spells was combined with information on the last employment if the respondent had been out of work since January 2021 (fs).	For first survey: *PET1200b (PENDDAT); ejahr; emonat (bio_spells)* For repeated survey: *ejhrlewt* from previous wave *(PENDDAT) ejahr; emonat (bio_spells)*
ekin1517	*Control variable: own child aged between 15 and 17 in the household.*: A variable indicating whether the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 15 and 17 in the household (new).	Information on relationships between household members *(household grid)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

ekind	*Control variable: own child in HH*: A variable indicating whether the respondent has a natural child, a stepchild/adopted child or a child of non-specified status of any age in the household (new). It can occur in rare household constellations that according to *ekind*, an individual has children living in the household, but their *pnr* does not appear in the pointers *zmhh* and *zvhh* of *p_register*. This can occur in case of same-sex relationships with children or if both the current and the former partner live in the household.	Information on relationships between household members *(household grid)*
ekin614	*Control variable: own child aged between 6 and 14 in the household*: A variable indicating whether the respondent has a natural child, a stepchild/adopted child or a child of non-specified status aged between 6 and 14 in the household (new).	Information on relationships between household members *(household grid)*
ekinu15	*Control variable: own child under the age of 15 in HH*: A variable indicating whether the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 15 in the household (new).	Information on relationships between household members *(household grid)*
ekinu18	*Control variable: own child under the age of 18 in HH*: A variable indicating whether the respondent has a natural child, a stepchild/adopted child or a child of non-specified status under the age of 18 in the household (new).	Information on relationships between household members *(household grid)*
epartner	*Control variable: spouse or registered partner in HH* : A variable indicating whether the respondent has a spouse or a same-sex registered partner in the household (new).	Information on relationships between household members *(household grid)*
etakt	*Currently employed (above the mini-job earnings limit), gen. (as of wave 2)*: A variable indicating whether the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. employment earning above the mini-job earnings limit) (new).	*zensiert, spintegr, BIO0101 (bio_spells)*
famstand	*Marital status, gen.*: Generation of a marital status variable integrating information from the personal questionnaire and the control variable *epartner*; generated from the household dataset (new).	*epartner; PD0500; PD0700 (PENDDAT)*
gebhalbj	*Half-year of birth, gen.*: A variable indicating whether the date of birth is in the first or second half of the year of birth (new).	Information on month of birth

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

kennungfb-versW17	*Version identifier of the P-questionnaire wave 17*: In the current field of wave 17, a translation error in the arabic version of question *PMI3600* was corrected. The identifier marks whether a case was interviewed with the original or the revised version of the questionnaire. A detailed description of the changes in the questionnaire can be found in chapter 1.2.1 (new).	Information generated and supplied by the survey institute
kindzges	*Total number of own children (living in and outside the household), gen.*: Total number of the respondent's children including the children living in his/her household and the children living outside the household (new).	Information on relationships between household members *(household grid) PD0900; PD1000; PD1100 (PENDDAT)*
kindzihh	*Number of own children in the household, gen.*: Variable generated on the basis of the responses in the household questionnaire concerning the number of children that an individual in the household has (total number of individuals in the household (half) matrix who count as children of the respondent plus the number of individuals in the household (half) matrix for whom the respondent is classified as being a parent) (new). *Note: When using this variable it should be borne in mind that it relates to each individual person. This means that a child who lives in a household together with his/her parents is counted as a "child in the household" for both the father and the mother. Aggregating this variable across the household members will therefore not produce any meaningful results.*	Information on relationships between household members *(household grid)*
mberuf1	*Highest vocational qualification attained by the mother, incl. mother in the HH, excl. information from open-ended survey questions, gen.*: In wave 1, the question about the mother's vocational qualification was asked only if the mother was not living in the survey household. If she was living in the household, this information was obtained from her personal interview.	For first survey: *PSH0300a-i (PENDDAT)* After first survey: *mberuf1* aus Vorwelle *(PENDDAT)*
mberuf2	*Highest vocational qualification attained by the mother, incl. mother in the HH, incl. information from open-ended survey questions, gen.*: Defined as in *mberuf1* except that responses to open-ended questions were also considered to generate *mberuf2* (uv).	For first survey: *PSH0301a-i (PENDDAT)* After first survey: *mberuf2* from previous wave *(PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

mhh	*Control variable: mother living in HH*: A variable indicating whether the respondent's biological mother, stepmother, adoptive mother or mother of non-specified status lives in the household (new).	Information on relationships between household members *(household grid)*
migration	*Respondent's migration background, generated*: The following four categories were included in a generated variable for migration background: no migration background; personal migration (first generation); migration of at least one parent but no personal migration (second generation); migration of at least one grandparent but not the respondent or either parent (third generation) (uv). *Note: The concept for generating this variable has been revised as of wave 2. Previously, only the information on whether the respondent was born in Germany and which ancestor moved to Germany was collected. Now, information on whether an ancestor was born outside Germany and if applicable, which ancestor, is included. To guarantee consistency across waves, the variable for wave 1 was regenerated.*	For first survey: *PMI0100; PMI0700; PMI0800a-f; PMI0900a-f (PENDDAT)* After first survey: *migration* from previous wave *(PENDDAT)*
mschul1	*Highest general school qualification attained by the mother, incl. mother in HH, excl. information from open-ended questions, gen.*: In wave 1, the mother's highest academic qualification was inquired about only if the mother was not living within the survey household. If she was living in the household, this information was obtained from her personal interview (uv). As of wave 2, the mother's highest academic qualification has been asked of all newly interviewed individuals regardless of whether the mother was living in the survey household.	For first survey: *PSH0200 (PENDDAT)* After first survey: *mschul1* from previous wave *(PENDDAT)*
mschul2	*Highest general school qualification attained by the mother, incl. mother in HH, incl. information from open-ended questions, gen.*: Same as *mschul1* apart from the fact that responses to open-ended questions were also taken into account for the generation of *mschul2* (uv).	For first survey: *PSH0201 (PENDDAT)* After first survey: *mschul2* from previous wave *(PENDDAT)*
mstib	*Mother's occupational status, code number, gen.*: The detailed occupational status of the mother was generated from the individual variables (uv).	For first survey: *PSH0320; PSH0330; PSH0340; PSH0360; PSH0370; PSH0380 (PENDDAT)* After first survey: *mstib (PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

netges	*Current total net income (without marginal employment, incl. cat. info.), gen.*: This variable contains the accumulated information on net income from all employment positions (above the mini-job earnings limit), which is generated from the answers to open-ended questions on net income and a categorical follow-up question when respondents provided "don't know" or "details refused" answers to open-ended questions (new).	*ET3412; ET3512; ET3612; ET3712; ET3812; ET3912 (bio_spells)*
netto	*Net income of the current main employment incl. categorised information, gen.*: A generated variable integrating information from categorised and open-ended survey questions on net income (new).	*ET3412; ET3512; ET3612; ET3712; ET3812; ET3912 (bio_spells)*
nettokat	*Categorised net income from the current main employment, gen.*: This variable aggregates the categorised information on net income for a specific variable, which combines several items on income categories (new).	*ET3412; ET3512; ET3612; ET3712; ET3812; ET3912 (bio_spells)*
ostaatansyr	*Nationality syr./iraq. HH, incl. open info., categories (anon.)*: From wave 10 onwards, this identifier can be used to determine for the sub-samples of Syrian and Iraqi households whether a person has the Syrian nationality or another nationality. A separate designation of persons with Iraqi nationality had to be omitted due to a small number of cases (new).	Non-anonymous variant of *ostaatan, sample (PENDDAT)*
ostaatanukr	*Nationality ukrain. HH, incl. open info., categories (anon.)*: From wave 17 onwards, this identifier can be used for the subsamples of Ukrainian households to determine whether a person has Ukrainian or another nationality (new).	Non-anonymous variant of *ostaatan, sample (PENDDAT)*
palter	*Age (from PD0100), gen.*: The respondent's age is generated from the date of birth and date of the current personal interview (new).	*PD0100; pintjahr, pintmon, pinttag (PENDDAT)*
panel	*Willingness to participate in the panel (new)*: (new).	Information supplied by the survey institute regarding the households' willingness to participate in the panel.
pintdat	*Date of personal interview*: This generated variable indicates the date on which the personal interview was conducted in the format YYMMDD (new).	*pintjahr, pintmon, pinttag (PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

pintnum	*interviewer in personal interview*: The artificial identifier indicates the interviewer who conducted the interview. This information is consistent between *PENDDAT* and *HHENDDAT* as well as across waves. A definite characteristic of the label always identifies the same interviewer (new).	Information that is generated and supplied by the survey institute.
schul1	*Highest school qualification, excl. foreign qualifications and information from open-ended survey questions*: This variable records the highest academic qualification. Equivalent Eastern and Western German qualifications were combined (e.g., EOS and Abitur), but information from open-ended questions was excluded (fs).	For first survey: *PB0200; PB0220; PB0230; PB0300; PB0400 (PENDDAT)* After repeated survey: *schul1* from previous wave ; *PB0200; PB0220; PB0230; PB0300; PB0400 (PENDDAT)*
schul2	*Highest school qualification, incl. foreign qualifications and information from open-ended survey questions*: Defined as in *schul1* with the following differences: 1. inclusion of responses to open-ended questions; and 2. inclusion of information about foreign qualifications (fs).	For first survey: *PB0200; PB0220; PB0231; PB0300; PB0401 (PENDDAT)* After repeated survey: *schul2* from previous wave ; *PB0200; PB0220; PB0231; PB0300; PB0401 (PENDDAT)*
schulabj	*Year in which highest school qual. was attained*: Year in which the respondent attained his/her highest academic qualification (fs). Note: *Re-interviewed respondents for whom information regarding the highest school qualification was already available from a previous wave were not asked in the current wave about the year when this qualification was attained if they had attained a new qualification since the previous wave. In this case, the year in which the qualification was attained was estimated depending on the month and year of the interview.* Note: *If the interview in wave 17 was conducted before May 2023, it was assumed that the qualification was gained in 2022, if the interview was conducted later than May, the qualification was assumed to have been gained in 2023.*	For first survey: *PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)* After repeated survey: *schulabj* from previous wave ; *PB0220; PB0230; PB0410; pintjahr; pintmon (PENDDAT)*
statakt	*Current main status, generated (as of wave 2)*: Indicates which main status the TP had at the date of the personal interview of the respective wave (new).	*zensiert; spintegr; BIO0101; azges2 (bio_spells)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

stib	*Occupational status, code number, generated*: A generated of the detailed code number for occupational status from the individual variables. A generated variable using information from the module "employment" (*ET060*-ET120**). If there was more than one ongoing employment spell, the one with the most hours of work was selected. If there was more than one ongoing spell with exactly the same amounts of hours, the one that started first was selected (new).	*ET0615; ET0715; ET0815; ET0915; ET1015; ET1115; ET1215 (bio_spells)*
stibeewt	*Occupational status, first employment, code number, generated* : Detailed code number of the occupational status in the respondent's first regular employment. To generate the variable, information regarding the first regular employment was combined with information from the employment spells if the respondent had already reported his/her first regular employment during the questions on employment spells since January 2021 (uv).	For first survey: *PET3300; PET3400; PET3500; PET3600; PET3700; PET3800; PET3900 (PENDDAT)* *ET0615; ET0715; ET0815; ET0915; ET1015; ET1115; ET1215 (bio_spells)* After first survey: *stibeewt* from previous wave *(PENDDAT)*
stiblewt	*Occupational status, last employment, code number, generated*: Detailed code number of the occupational status in the respondent's last employment. Information from the employment spells were combined with information on the last employment for the generation if the respondent has been unemployed since January 2021 (fs).	For first survey: *PET1210; PET1220; PET1230; PET1240; PET1250; PET1260; PET1270 (PENDDAT)* *ET0615; ET0715; ET0815; ET0915; ET1015; ET1115; ET1215 (bio_spells)* After repeated survey: *stiblewt* from previous wave *(PENDDAT)* *ET0615; ET0715; ET0815; ET0915; ET1015; ET1115; ET1215 (bio_spells)*
vberuf1	*Highest vocational qualification attained by the father, incl. father in the HH, excl. open info., gen.*: A generated variable for father's highest vocational qualification analogous to *mberuf1* (uv).	For first survey: *PSH0600a-i (PENDDAT)* After first survey: *mberuf1* from previous wave *(PENDDAT)*

Table 13: Simple generated variables for wave 17 in the individual dataset (*PENDDAT*) (in alphabetical order) (continued)

vberuf2	*Highest vocational qualification attained by the father, incl. father in the HH, incl. open info., gen.*: A generated variable for father's highest vocational qualification (incl. information from open-ended survey questions) analogous to *mberuf1* (uv).	For first survey: *PSH0601a-i (PENDDAT)* After first survey: *mberuf2* from previous wave *(PENDDAT)*
vhh	*Control variable: father living in HH*: Variable indicating that the respondent's natural father, stepfather, adoptive father or father of non-specified status is living in the household (new).	Information on relationships between household members *(household grid)*
vschul1	*Highest general school qualification attained by the father, incl. father in HH, excl. information from* : A generated variable for father's highest general academic qualification analogous to *mschul1* (uv).	For first survey: *PSH0500 (PENDDAT)* After first survey: *vschul1* from previous wave *(PENDDAT)*
vschul2	*Highest general school qualification attained by the father, incl. father in household, incl. open info., gen.*: This generated variable records the father's highest general academic qualification (including information from open-ended survey questions) and is analogous to *mschul2* (uv).	For first survey: *PSH0501 (PENDDAT)* After first survey: *vschul2* from previous wave *(PENDDAT)*
vstib	*Father's occupational status, code number, generated*: The detailed occupational status of father is generated from individual variables (uv).	For first survey: *PSH0620; PSH0630; PSH0640; PSH0660; PSH0670; PSH0680 (PENDDAT)* After first survey: *vstib* from previous wave *(PENDDAT)*

Table 14: Wave 17 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset)

Variable	Label and description	Source var. for gen. var wave 17
bmonat	*Spell of UB II: start month, generated*: The month in which the spell of receiving Unemployment Benefit II began. If information was only available on the season when a spell began, the season was converted into a month to generate the variable. *Note: The generated date variables were both checked for plausibility and corrected when necessary. The dates originally reported by the respondent have been included in the source variables as of wave 2. The season in which the spell began were recoded into months as follows:* *21: beginning of year/winter = January;* *24: spring/Easter = April;* *27: middle of year/summer = July;* *30: autumn = October;* *32: end of year = December*	*AL20100 (alg2_spells)*
bjahr	*Spell of UB II: start year, generated*: The year during which the spell of receiving Unemployment Benefit II ended. Note: see *bmonat*	*AL20200 (alg2_spells)*
emonat	*Spell of UB II: end month, generated*: The month during which the spell of UB II receipts ended. To generate this variable, information about the season was converted into a month. For right-censored spells (i.e., spells that were ongoing when the household was interviewed), the interview month was entered. *Note: see bmonat*	*AL20300 (alg2_spells)* *hintmon (HHENDDAT)*
ejahr	*Spell of UB II: end year, generated*: The year during which the spell of Unemployment Benefit II ended. In the case of right-censored spells (i.e., spells that were ongoing when the household was interviewed), the interview year was entered. *Note: see bmonat*	*AL20400 (alg2_spells)* *hintjahr (HHENDDAT)*

Table 14: Wave 17 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

alg2kbma - alg2kbmi	*UB II: 1st cut: start month, generated* to *UB II: 9th cut: start month, generated:* The month during which Unemployment Benefit II was reduced. To generate this variable, information about the season was converted into a month. *Note: These UB II reductions are embedded in spells of UB II receipts. Information on an individual benefit reduction can be distinguished via the indicator at the end of the respective variable (a - h). The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent have been included in the source variables since wave 2.*	1st Benefit cut: *AL21000a (alg2_spells)* to 9th Benefit cut: *AL21000i (alg2_spells)* (only surveyed up to wave 11)
alg2kbja - alg2kbji	*UB II: 1st cut: start year, generated* to *UB II: 9th cut: start year, generated:* The year during which the Unemployment Benefit II reduction began. *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *AL21100a (alg2_spells)* to 9th Benefit cut: *AL21100i (alg2_spells)* (only surveyed up to wave 11)
alg2kema - alg2kemi	*UB II: 1st cut: end month, generated* to *UB II: 9th cut: end month, generated*: The month during which the Unemployment Benefit II reduction ended. To generate this variable, information on the season was converted into a month. If the respondent reported the duration of the benefit reduction, this information was used to calculate the end date of the benefit cut based on the generated start date. *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 9th Benefit cut: *alg2kbmi; alg2kbji; AL21200i; AL21201i; AL21202i (alg2_spells)* (only surveyed up to wave 11)
alg2keja - alg2keji	*UB II: 1st cut: end year, generated* to *UB II: 9th cut: end year, generated*: Year in which the Unemployment Benefit II cut ended. If the respondent reported a duration for the benefit cut, this information was used to calculate the end date of the benefit cut *Note: see alg2kma - alg2kbmi*	1st Benefit cut: *alg2kbma; alg2kbja; AL21200a; AL21201a; AL21202a (alg2_spells)* to 9th Benefit cut: *alg2kbmi; alg2kbji; AL21200i; AL21201i; AL21202i (alg2_spells)* (only surveyed up to wave 11)

Table 14: Wave 17 simple generated variables included in the spell dataset for Unemployment Benefit II *(alg2_spells)* (provided in the same order as in the dataset) (continued)

AL22150a - AL22150i	*ALG2: 1st Benefit cut: which HH member's benefit was cut, gen.* to *ALG2: 9th Benefit cut: which HH member's benefit was cut, gen.*: This variable records which household members experienced reductions in Unemployment Benefit II. This is a string variable with 15 positions. Starting from the left, each position in this variable represents the position of one individual on the household grid. The first position of the variable, for example, indicates whether Unemployment Benefit II was cut for the first individual in the household during the particular benefit reduction spell, the second position indicates whether the second individual's benefit was reduced (etc.). Because source information for the generated variable was collected from wave 2 to wave 4, all 15 positions are coded "I" (i.e., item not asked in wave) for all benefit cuts reported during the first and since wave 5 (see below). Each of the 15 positions of this variable, which represent one of a maximum of 15 individuals in the household, is assigned one of the following codes indicating each individual' benefit status. Codes: 1 = the household member's UB II was cut 2 = the household member's UB II was not cut W = don't know K = not specified T = not applicable (filter) F = question mistakenly not asked U = implausible value I = item not recorded in wave	Information which household member's benefit was cut in the respective benefit cut spell (only surveyed until wave 3)
zensiert	*Spell of UB II: spell ongoing at time of last HH interview (right-censored.), generated*: The censoring indicator shows whether a spell was still ongoing at the time of the last household interview. *Note: : A spell is regarded as censored if one of the following conditions is met:* *(a) It is a censored spell of a household from one of the previous waves that had not been re-interviewed in the subsequent waves up to the current wave.* *(b) A household surveyed in previous waves reports that a spell of UB II is still ongoing on the interview date in wave 17, or an end date is reported that is identical to the interview date in wave 17 and it is confirmed in the follow-up question that the benefit receipt is still currently ongoing.*	*AL20300; AL20400, AL20500 (alg2_spells)*

Table 15: Simple generated variables for wave 17 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset)

Variable	Label and description	Source var. for gen. var wave 17
bmonat	*Employment: start month, generated* The month during which the employment spell began. To generate the variable information on the season was converted into a month. *Note: The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables. Details regarding the season in which the spell began were recoded into months as follows:* *beginning of year/winter: January;* *spring/Easter: April;* *middle of year/summer: July;* *autumn: October;* *end of year: December*	*BIO0200 (bio_spells)*
bjahr	*Employment: start year, generated* The year during which the employment spell began. *Note: see bmonat*	*BIO0300 (bio_spells)*
emonat	*Employment: end month, generated* The month during which the employment spell ended, whereby the seasonal data was converted into specific monthly data for its generation and for right-censored spells (i.e., spells that were ongoing when the individual was interviewed) the interview month was used. *Note: see bmonat*	*BIO0400, BIO0600 (bio_spells); pintmon*
ejahr	*Employment: end year, generated* The year during which the employment spell ended. For right-censored spells (i.e., spells that were ongoing when the individual was interviewed), the interview month was entered. *Note: see bmonat*	*BIO0500, BIO0600 (bio_spells); pintjahr*

Table 15: Simple generated variables for wave 17 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset) (continued)

zensiert	*Employment: spell still currently ongoing (right censoring)* The censoring indicator shows whether a spell was ongoing at the time of the personal interview in the previous wave, i.e., whether it is a right-censored spell. *Note: A spell is considered censored if one of the following conditions is met:* *(a) the individual reports an end date of the BIO spell that the employment is ongoing on the interview date.* *(b) Alternatively, when a reported end date is identical to the interview date, the follow-up question confirms that the activity is ongoing. BIO0400; BIO0500;* *BIO0600 (bio_spells)*	
stib	*Occupational status, code number, generated* A detailed code for individual occupational status is generated from the individual variables.	Collection of spell information in wave 17: *ET0615; ET0715; ET0815; ET0915; ET1015; ET1115; ET1215 (bio_spells)* Otherwise, the value from the previous wave remains
az1	*Weekly contractual working hours*	Collection of spell information in wave 17: ET2015 *(bio_spells)* Otherwise, the value from the previous wave remains. Exception: az1 is coded -3 if the occupation was a dependent employment so far and the occupational status changed in self-employment/family worker, details refused or dont't know.

Table 15: Simple generated variables for wave 17 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset) (continued)

az2	*Weekly working hours incl. details in the case of ir-regular working hours, gen.* An integrated variable on weekly hours worked in the position held by the respondent, combining responses to open-ended questions on working hours and a categorical follow-up question. For the closed categories, the follow-up question utilised the mean values for the categories. For the open-ended category, the median of the weekly working hours reported (40 hours or more) was used.	Collection of spell information in wave 17: *ET2115; ET2215 (bio_spells)* Otherwise, the value from the previous wave remains.
alg1bm	*Receipt of UB I: start month, generated* The month during which the spell of Unemployment Benefit I began. To generate this variable, information on the season was converted into a month. *Note: Periods during which Unemployment Benefit I is received are embedded in the spells of registered unemployment. An individual can receive a maximum of one period of UB I per period of registered unemployment. The generated date variables were checked for plausibility and corrected if necessary. The dates originally reported by the respondent are included in the source variables.* For conversion to months, see *bmonat*.	*AL0800 (bio_spells)*
alg1bj	*Receipt of UB I: start year, generated* The year during which the spell of Unemployment Benefit I began. *Note: see alg1bm*	*AL0900 (bio_spells)*
alg1em	*Receipt of UB I: end month, generated* The month during which the spell of Unemployment Benefit I ended. To generate the variable information, the season was converted into a month. For right-censored spells (i.e., spells that were ongoing at the time of the interview), the interview date was entered. *Note: see alg2kma - alg2kbmi*	*AL1000; AL1200 (bio_spells) pintmon (PENDDAT)*
alg1ej	*Receipt of UB I: end year, generated* The year during which the spell of receiving Unemployment Benefit I ended. In right-censored spells (i.e., spells that were ongoing at the time of the interview), the interview date was entered. *Note: see alg2kma - alg2kbmi*	*AL1100; AL1200 (bio_spells) pintjahr (PENDDAT)*

Table 15: Simple generated variables for wave 17 in the BIO spell dataset *(bio_spells)* (in the same order presented in the dataset) (continued)

alg1akt	*Receipt of UB I: spell still currently ongoing (right censoring)* This variable indicates whether the spell of receiving Unemployment Benefit I was ongoing at the time of the personal interview during the previous wave, i.e., whether it is right-censored. *Note: A spell is considered censored if one of the following conditions is met:* *(a) the individual reports an end date for receiving Unemployment Benefit I that indicates that the benefits are ongoing.* *(b) Alternatively, an end date identical to the interview date is reported. The follow-up question confirms that benefits are ongoing. This variable is generated based on generated date variables, which have been checked for plausibility.*	*emonat; ejahr; AL1000; AL1100; AL1200 (bio_spells)*
br	*Gross income (incl. categorised info.), gen.* This variable is generated for spells that are ongoing during wave 17 using data from the current wave. For spells that ended or have not been updated in wave 17, information from wave 16 is used to calculate the variable.	*ET28*; ET29*; ET30*; ET31*; ET32*; ET33* (bio_spells)*
net	*Net income (incl. categorised info.), gen.* For ongoing spells during wave 17, this variable is generated using data from the current wave. For spells that ended or have not been updated in wave 17, the information from wave 16 is used to calculate the variable.	*ET34*; ET35*; ET36*; ET37*; ET38*; ET39* (bio_spells)*

Table 16: Wave 17 simple generated variables included in the one-euro spell dataset *(ee_spells)* (in the same order presented in the dataset)

Variable	Label and description	Source var. for gen. var in wave 17
bmonat	*Measure: start month, generated* Start month of the measure episode, whereby the seasonal data was converted into specific monthly data for its generation. *Note:The generated date variables were checked for plausibility and corrected if necessary. The dates reported by the respondent excluding identified implausible values) are included in the source variables. Seasons during which the spell began were recoded into months as follows:* *21: beginning of year/winter: January* *24: spring/Easter: April* *27: middle of year/summer: July* *30: autumn: October* *32: end of year: December*	*EE0600a (ee_spells)*
bjahr	*Measure: start year, generated* The year during which the active labour market policy spell began. *Note: see bmonat*	*EE0600b (ee_spells)*
emonat	*Measure: end month, generated* End month of the measure episode, whereby the seasonal data was converted into specific monthly data for its generation. For right-censored spells (i.e., spells that were ongoing at the time of the interview), the interview date was entered. *Note: see bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells); pintmon, pintjahr (PENDDAT)*
ejahr	*Measure: end year, generated* The year during which the active labour market policy spell ended. For right-censored spells (i.e., spells that were ongoing when the individual was interviewed), the interview date was entered. *Note: see bmonat*	*EE0600a; EE0600b; EE0700; EE0800a; EE0800b (ee_spells)*
zensiert	*Measure: spell still currently ongoing (right censored)* The censoring indicator records whether a spell was ongoing at the time of the personal interview during the previous wave, i.e., whether this is a right-censored spell.	*EE0700 (ee_spells)*

Table 17: Wave 17 simple generated variables included in the person register dataset *(p_register)* (in alphabetical order)

Variable	Label and description	Source var. for gen. var. wave 17
alter17	*Individual's age in wave 17 (2023)* The variable contains the best available age information of a person. This is either (a) the age calculated using the date of birth reported in wave 17, or (b) the age information from the household interview if no date of birth is available from wave 17. The information from alter17 was also transferred to the household dataset and corresponds to the information in *HD0200a* to *HD0200o*. This procedure is consistent with the procedure in the field. The age variable in the database was already filled with the best information during the field period. There, a variable in the database is initially filled with the age information according to the household interview. If a personal interview is conducted, this variable is overwritten in the database with the age calculated on the basis of the information from the personal interview (date of birth, personal interview date). Both the age information provided in the household dataset and the personal dataset are based on this variable in the database. The "best" age information contained in the household dataset for wave 17 was taken into account in the plausibility checks and in the generation of the benefit unit and household types..	*PD0100; pintjahr; pintmon; pinttag (PENDDAT); HD0200a to HD0200o (HHENDDAT)*
erwprox17	*Employment status according to HH interview in wave 17 (2023)* This variable is transferred unchanged as *HD1101** from the current wave from the *HHENDDAT* dataset.	*HD1101**
kinddat17	*Person included in the KINDER dataset in wave 17 (2023)* This variable indicates whether an individual is included in the *KINDER* dataset. Included in the *KINDER* dataset: All children aged under 15 years. In the waves 6 to 10 also all household members aged between 16 and under 25 years, for proxy variables surveyed in the modules social inclusion and education and participation packages.	*pnr (KINDER)*
korrsex	*Info. on sex was corrected between survey waves* For individuals who belonged to a sample HH in more than one wave, this variable indicates whether their sex was adjusted in the household interview.	*HD0100a to HD0100o of all waves (HENDDAT)*

Table 17: Wave 17 simple generated variables included in the person register dataset *(p_register)* (in alphabetical order) (continued)

lastint	*Survey wave of last interview at individual level* This variable indicates the wave in which the last individual interview was conducted (personal or senior citizen interview).	Personal interviews from all waves *(PENDDAT)*
neuj17	*Year in which individual joined current HH, reported in wave 17 (2023)* This variable indicates the year during which an individual joined the current household of which he/she is a member reported during wave 17. *Note: The wave 17 interview with the re-interviewed household provides that date when the individual moved or was born into the household since the previous wave.*	Information on the date since which an individual has belonged to a household. Surveyed in the household grid
neum17	*Month in which individual joined current HH, reported in wave 17 (2023)* This variable indicates the month that the individual joined the household of which he/she is a current member reported during wave 17. *Note: see neuj17*	Date an individual joined a household. Surveyed in the household grid.
wegj17	*Year since which individual has no longer been living in previous HH, reported in wave 17 (2023)* This variable indicates the year that the individual ceased to be a member of the household of the previous wave. *Note: Information on the date comes from the wave 17 interview with the household in which the individual was living in the previous wave.*	Date an individual ceased to belong to a household. Surveyed in the household grid.
wegm17	*Month since which individual has no longer been living in previous HH, reported in wave 17 (2023)* This variable indicates the month that the individual ceased to be a member of the household of the previous wave. *Note: see wegj17*	Date an individual ceased to belong to a household. Surveyed in the household grid.
zdub17	*Pointer: Personal identification no. of the individual doubled by the TP in wave 17 (2023)* Indicates that an individual from an original HH currently lives in a split-off HH without the original HH having reported the move of this individual. *Note: For matchings with the p_register via the personal identification number, one must first generate a match variable equalling zdub*, if it exceeds 0, or otherwise equalling pnr. Chapter 5.4.1.2 of the data report for wave 5 of PASS provides a detailed explanation on the reasons for the introduction of this variable.*	Information on all original household members of an original household and all of its split-off households are included in the household grid of the current and the previous waves.

Table 17: Wave 17 simple generated variables included in the person register dataset *(p_register)* (in alphabetical order) (continued)

zmhh17	*Pointer: Personal ID number of target person's mother in HH in wave 17 (2023)* Contains the personal identification number of the mother if she is living in the household. Biological mothers, stepmothers, adoptive or foster mothers and mothers whose status is not specified are considered mothers.	Relationships between household members (household grid).
zparthh17	*Pointer: personal ID number of target person's partner in HH in wave 17 (2023)* Contains the personal identification number of a partner living in the household. The following are considered partners: Spouse, registered partner, unmarried partner and partner with unspecified status.	Relationships between household members (household grid).
zupanel	*Survey wave in which individual joined panel* This variable indicates the wave in which the individual was a member of a sample household for the first time.	The individuals living in a household across waves (household grid).
zvhh17	*Pointer: Personal ID number of target person's father in HH in wave 17 (2023)* Contains the personal identification number of the father if he lives in the household. Biological fathers, stepfathers, adoptive or foster fathers and fathers whose status is not specified are considered fathers.	Relationships between household members (household grid).

The individual-level datasets contain a multitude of generated and constructed variables, including variables (e.g., occupational status) that are recorded in more than one dataset. Figure 3 provides an overview of both the simple and complex generated variables at the individual level.

Figure 3: Overview of generated variables for wave 17 at the individual level

	PENDDAT						BIO-Spells	EE_Spells
	Current status	Employment history		Social origin		€450 job	Employment and unem-ployment biography	One-euro job participation
		last employment	first employ-ment	mother	father			
Education	berabj							
	beruf1			mberuf1	vberuf1			
	beruf2			mberuf2	vberuf2			
	schulabj							
	schul1			mschul1	vschul1			
	schul2			mschul2	vschul2			
Education classification	casmin			mcasmin	vcasmin			
	isced97			misced97	visced97			
	bilzeit			mbilzeit	vbilzeit			
Information on current status	akt1euro							
	alakt							
	etakt							
	statakt						spelltyp	
Socio-economic position	egp	egplewt	egpeewt	megp	vegp		egp	
	esec	eseclewt	eseceewt	mesec	vesec		esec	
	isei1	iseilewt1	iseieewt1	misei1	visei1		isei1	
	isei2	iseilewt2	iseieewt2	misei2	visei2		isei2	
	mps	mpslewt	mpseewt	mmps	vmps		mps	
	siops1	siopslewt1	siopseewt1	msiops1	vsiops1		siops1	
	siops2	siopslewt2	siopseewt2	msiops2	vsiops2		siops2	
Occupational status	stib	stiblewt	stibeewt	mstib	vstib		stib	
	stibkz							
Date of employment			begmeewt			begmminj	bmonat	bmonat
			begjeewt			begjminj	bjahr	bjahr
		emonlewt					emonat	emonat
		ejhrlewt					ejahr	ejahr
Date of unemployment							alg1bm	
							alg1bj	
							alg1em	
							alg1ej	
Information on employment	befrist							
	azhpt1						az1	
	azhpt2						az2	
	azges1							
	azges2							
Occupation	isco88	isco88lewt	isco88eewt	misco88	visco88	isco88-minj	isco88	
	isco08	isco08lewt	isco08eewt	misco08	visco08	isco08-minj	isco08	
	kldb1992	kldb1992lewt	kldb1992-eewt	mkldb1992	vkldb1992	kldb1992-minj	kldb1992	
	kldb2010	kldb2010lewt	kldb2010-eewt	mkldb2010	vkldb2010	kldb2010-minj	kldb2010	
Employed in which industry	branche1					branche-minj1	branche1	
	branche2					branche-minj2	branche2	

	PENDDAT						BIO-Spells	EE_Spells
	Current status	Employment history		Social origin		€450 job	Employment and unem-ployment biography	One-euro job participation
		last employme nt	first employ-ment	mother	father			
Income	netges							
	brges							
	netto							
	nettokat							
	brutto							
	bruttokat							
Benefit receipt	alg1abez						alg1akt	
	aktgefbesch							
Household context and civil status	hhgr							
	famstand							
	vhh							
	mhh							
	apartner							
	epartner							
	ekind							
	ekin614							
	ekinu15							
	ekinu18							
	ekin1517							
	kindzges							
	kindzihh							
Migration background	ogebland							
	ostaatan							
	ozulanda							
	ozulandb							
	ozulandc							
	ozulandd							
	ozulande							
	ozulandf							
	migration							
Information on individual	gebhalbj							
	palter							
	zpalthh							
	zpsex							
General	altbefr							
	fb_vers							
	panel							
	pintdat							
	RegP0100							
	sample							

4.5 Constructed variables

Constructed variables are generated variables that require more extensive coding or recoding. In most cases, these variables have been empirically tested elsewhere and are based on theoretical concepts. At least some of these are standardized instruments used in social sciences or economics, such as the European Socio-economic Classification (ESeC), the International Standard Classification of Education (ISCED) or equivalised household income. This chapter provides detailed descriptions of the constructed variables made available in the PASS data, along with a short overview of the theoretical background and the most important references.

4.5.1 Individual Level

Table 18: Education in years

Variable name	*bilzeit*
Variable label	Duration of school education and vocational training in years, generated
Source variables	*schul2; beruf2*
Type / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	For many statistical models, a linear variable for education and training is more appropriate than a categorical variable. For school qualifications, it is easy to convert categorical data to linear data. The linear value simply corresponds to the time spent in school until attainment of the final qualification. Care must be taken to ensure that equivalent qualifications are assigned identical durations. An upper secondary school certificate, for example, should always be labeled with the same duration regardless of whether it was obtained after twelve or thirteen years of education. Final qualifications were assigned the following durations: Lower secondary school certificate, lower secondary school certificate from the former GDR (POS) after completion of grade 8: 8 years Intermediate secondary school certificate from the former GDR (POS) after completion of grade 10: 10 years Entrance qualification for university for applied sciences: 12 years General qualification for university or subject-specific higher education entrance (including EOS—similar qualification in the former GDR): 13 years Vocational qualifications differ because of their numerous, different requirements and potentially large differences in income even for qualifications with similar training duration. The training duration may not be subjected to a simple one-to-one conversion process.

Table 18: Education in years (continued)

	This problem can be avoided by attempting to operationalise the growth in human capital related to a particular vocational qualification (see e.g., Helberger, 1988). This study adopts a similar approach. Only the respondent's highest vocational qualification was considered, and the years estimated to represent the human capital growth resulting from this qualification were added to the years of education. Training as a semi-skilled worker: +1 year Apprenticeship, vocational school, school for health care occupations: +1.5 years Master craftsman certificate:+3 years Vocational academy: +3 years Applied sciences/Bachelor's degree: +3 years University/Master's degree: +5 years Ph.D.: +8 years Other German qualification: +1.5 years Other foreign qualification: +1.5 years
Literature:	Helberger (1988)

Table 19: Education in years, mother

Variable name	*mbilzeit*
Variable label	Duration of school education and vocational training of mother in years, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see "Education in years"

Table 19: Education in years, mother (continued)

	When generating the parents' years of education and training variables, the values added for vocational qualifications differ from those used to construct the corresponding variable for the respondents because information on vocational education/training was collected in less detail for parents (especially for tertiary education). The following values are assigned to particular courses of education/training: Training as a semi-skilled worker: +1 year Apprenticeship, vocational school, Health care occupations: +1.5 years Master craftsman certificate: +3 years Vocational academy: +3 years University, applied sciences: +3 years University: +5 years Other German qualification: +1.5 years Other foreign qualification: +1.5 years
Literature:	Helberger (1988)

Table 20: Education in years, father

Variable name	*vbilzeit*
Variable label	Duration of school education and vocational training of father in years, generated
Source variables	*vschul2; vberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see "Education in years" When generating the parents' years of education and training variables, the values added for vocational qualifications differ from those used to construct the corresponding variable for the respondents because information on vocational education/training was collected in less detail for parents (especially for tertiary education). The following values are assigned to particular courses of education/training:

Table 20: Education in years, father (continued)

	Training as a semi-skilled worker: +1 year Apprenticeship, vocational school, Health care occupations: +1.5 years Master craftsman certificate: +3 years Vocational academy: +3 years University, applied sciences: +3 years University: +5 years Other German qualification: +1.5 years Other foreign qualification: +1.5 years
Literature:	Helberger (1988)

Table 21: CASMIN

Variable name	*casmin*
Variable label	Education classified acc. to CASMIN, updated version, generated
Source variables	*schul2; beruf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	The CASMIN educational classification was developed within the framework of the CASMIN project (Comparative Analysis of Social Mobility in Industrial Nations) in order to compare academic and vocational qualifications internationally (König, Lüttinger & Müller, 1987). An updated version is now available (Brauns & Steinmann, 1999).

Table 21: CASMIN (continued)

	The procedures applied in the panel to recode qualifications according to the CASMIN classification, especially for problematic cases, follow the procedures described in Lechert, Schroedter and Lüttinger (2006) and Granato (2000). The slightly differing category values of the education variable in this dataset are considered. Details are presented in the table below. Cells containing valid CASMIN combinations are labeled with positive values, those with defined missing values are labeled with negative values.

Schule / Beruf	nicht erhob.	Schüler	n. gest.	TNZ	KA	WN	ohne Abschl.	Sonder-schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. aus. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Schüler	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlernausbild.	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Berufsfachsch.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Schul. d. Ges-wes.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH/ Bachelor	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni/Master	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
Dissert.	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus. Abschl.	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

Table 22: MCASMIN

Variable name	*mcasmin*
Variable label	Education of mother classified acc. to CASMIN, updated version, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph

Table 22: MCASMIN (continued)

Explanation	General description: see CASMIN (above). Because the education variable has different category values for respondents and their parents, the coding pattern for *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table details the differences (see CASMIN).

Schule / Beruf	nicht erhob.	PInt fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlernausbild.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

Table 23: VCASMIN

Variable name	*vcasmin*
Variable label	Education of father classified acc. to CASMIN, updated version, generated
Source variables	*vschul2; vberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	General description: see CASMIN (above). Because the education variable has different category values for respondents and their parents, the coding pattern for *mcasmin* and *vcasmin* differs slightly from the pattern used in *casmin*. The following table details the differences.

Schule / Beruf	nicht erhob.	PInt fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Anlernausbild.	-	-	-	-	-3	-2	-1	1a	1a	1b	2b	2c_gen	2c_gen	1b	1b
Lehre	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
Meister	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
BA	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
FH	-	-	-	-	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a	3a
Uni	-	-	-	-	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b	3b
And. dt. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c
And. aus. Abschl.	-	-	-	-	-3	-2	-1	1c	1c	1c	2a	2c_voc	2c_voc	1c	1c

Table 23: VCASMIN (continued)

Literature:	Brauns et al. (1999); Granato (2000); König et al. (1987); Lechert et al. (2006)

Table 24: ISCED 97

Variable name	*isced97*
Variable label	Education classified acc. to isced97, updated version, generated
Source variables	*schul2; beruf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	The ISCED-97, (International Standard Classification of Education) developed by the OECD (OECD 1999; for an outline, see also BMBF, 2003), is an education classification alternative to CASMIN. Note that the coding for the ISCED-97 classification includes categories that cannot reasonably be assigned to these data. The ISCED values “0” (pre-primary education/kindergarten) and “1” (primary education) do not apply because the respondents are at least 15 years old. Instead, a separate group was created for individuals with an education below ISCED level 2 (ISCED 2 = lower or intermediate secondary school certificate). Therefore, only ISCED levels 2 to 6 are coded in this dataset. Coding details are shown in the table below. Cells containing valid combinations according to ISCED are labeled with positive values, those with defined missing values are labeled with negative values.
Literature:	BMBF (2003); OECD (1999)

Schule / Beruf	nicht erhob.	Schüler	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. aus. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Schüler	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlernausbild.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Berufsfachsch.	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Schul.d.Ges-wes.	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
Meister	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH/Bachelor	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni/Master	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Dissert.	-	-	-	6	6	6	6	6	6	6	6	6	6	6
And. dt. Abschl.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus. Abschl.	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Table 25: MISCED 97

Variable name	*misced97*
Variable label	Education of mother classified acc. to *isced97*, updated version, generated
Source variables	*mschul2; mberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	For the theoretical background and variable generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to respondent education, it is not possible to generate 6 ISCED levels for parents because data on the corresponding qualifications (i.e., Ph.D. or equivalent) were not collected for parents. Therefore, only ISCED levels 2 to 5 are coded in this dataset. The following table provides the coding details.
Literature:	BMBF (2003); OECD (1999)

Schule Beruf	nicht erhob.	PInt fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlernausbild.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Meister	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
And. dt. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Table 26: VISCED 97

Variable name	*visced97*
Variable label	Education of father classified acc. to *isced97*, updated version, generated
Source variables	*vschul2; vberuf2*
Category / dataset	Education / individual-level data
Prepared by	Bernhard Christoph
Explanation	Zum theoretischen Hintergrund und zur Generierung vgl. ISCED-97.

Table 26: VISCED 97 (continued)

	For the theoretical background and variable generation details, see ISCED-97. In contrast to the ISCED-97 coding applied to respondent education, it is not possible to generate 6 ISCED levels for parents because data on the corresponding qualifications (i.e., Ph.D. or equivalent) were not collected for parents. Therefore, only ISCED levels 2 to 5 are coded in this dataset. The following table provides the coding details.

Schule Beruf	nicht erhob.	PInt fehlt	Elternt. unbek.	nicht gest.	TNZ	KA	WN	ohne Abschl.	Sonder-Schule	HS	RS	FHR	Abi	And. dt. Abschl.	And. au. Abschl.
nicht erhob.	-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
unplaus. Wert	-	-	-	-	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
PInt fehlt	-	-6	-	-	-	-	-	-	-	-	-	-	-	-	-
Elternt. unbek.	-	-	-5	-	-	-	-	-	-	-	-	-	-	-	-
nicht gest.	-	-	-	-4	-	-	-	-	-	-	-	-	-	-	-
TNZ	-	-	-	-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
KA	-	-	-	-	-3	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
WN	-	-	-	-	-3	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1
ohne Abschl.	-	-	-	-	-3	-2	-1	1	1	2	2	3a	3a	2	2
Anlernausbild.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
Lehre	-	-	-	-	-3	-2	-1	3b	3b	3b	3b	4a	4a	3b	3b
Meister	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
BA	-	-	-	-	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b	5b
FH	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
Uni	-	-	-	-	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a	5a
And. dt. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2
And. aus. Abschl.	-	-	-	-	-3	-2	-1	2	2	2	2	3a	3a	2	2

Literature:	BMBF (2003); OECD (1999)

Table 27: International Standard Classification of Occupations 1988 (ISCO88)

Generated:	Employment - Variable name - Source variables
	Current *(PENDDAT) - isco88 - ET2500, PET2500* Spell data *(bio_spells) - isco88 - ET2500* first *(PENDDAT) - isco88eewt - ET2500, PET1280, PET3950* last *(PENDDAT) - isco88lewt - ET2500, PET1280* of father *(PENDDAT) - visco88 - PSH0800* of mother *(PENDDAT) - misco88 - PSH0700* Minijob - *isco88minj - PMJ0900*
Variable label:	Current Empl.: Intern. Standard Classification of Occupations 88, current employment, gen. Spell data: *(bio_spells)*: Intern. Standard Classification of Occupations 88, gen. first Empl.: ISCO 88, first employment, gen.

Table 27: Internat. Standard Class. of Occupations 1988 (ISCO88) (continued)

	last Empl.: ISCO 88, last employment, gen. Father: ISCO 88 of the father, gen. Mother: ISCO 88 of the mother, gen. Minijob: ISCO 88, current Minijob, gen.
Category / dataset	Occupation / individual-level data
Prepared by	Bernhard Christoph
Explanation	The International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO) to allow international comparison. An advantage of the ISCO-88 is that in addition to the employment, the qualification level generally necessary to perform the job is also considered when assigning an occupation to a particular occupational code. This constitutes a major difference from the Classification of Occupations provided by the German Federal Statistical Office (KldB), which is also provided in this dataset.
Literature:	ILO (1990)

Table 28: International Standard Classification of Occupations 2008 (ISCO08)

Generated:	Employment - Variable name - Source variables
	Current *(PENDDAT) - isco08 - ET2500, PET2500* Spell data *(bio_spells) - isco08 - ET2500* first *(PENDDAT) - isco08ewt - ET2500, PET1280, PET3950* last *(PENDDAT) - isco08ewt - ET2500, PET1280* of father *(PENDDAT) - visco08 - PSH0800* of mother *(PENDDAT) - misco08 - PSH0700* Minijob - *isco08mini - PMJ0900* Apprenticeship aspiration, desired occupation - *isco08berufswunsch - PAA0100* Apprenticeship aspiration, intended occupation - *isco08angberuf - PAA1000*
Variable label:	Current Empl.: Intern. Standard Classification of Occupations 08, current employment, gen. Spell data: *(bio_spells)*: International Standard Classification of Occupations, gen.

Table 28: Internat. Standard Class. of Occupations 2008 (ISCO08) (continued)

	first Empl.: ISCO08, first employment, gen. last Empl.: ISCO08, last employment, gen. Father: ISCO08 of the father, gen. Mother: ISCO08 of the mother, gen. Minijob: ISCO08, current Minijob, gen. (from wave 17 incl. seniors) Apprenticeship aspiration, desired occupation: ISCO 08, job descriptions: Desired occupation, gen. Apprenticeship aspiration, intended occupation: ISCO 08, job descriptions: Intended occupation, gen.
Category / dataset	Occupation / individual-level data
Prepared by	Christian Dickmann
Explanation	The International Standard Classification of Occupations (ISCO) is an internationally comparable classification developed by the ILO. The ISCO-08 classification is an update of ISCO-88. The frame-work and the concepts on which ISCO-08 is based are essentially unchanged from those in ISCO-88. The definitions of these concepts have been updated and the guidelines for their application to the design of the classification have been revised in order to address deficiencies in ISCO-88. Reported occupations are coded in ISCO-08 if they concern employment spells that have been carried forward from the previous wave from the tenth survey wave onwards or if it is new information reported from wave 10 onwards. Employment spells reported before wave 10 and not carried forward into wave 10ff. are available only as ISCO-88 codes. When coding details on marginal part-time jobs (so-called minijobs), no information is available on occupational status. As the vast majority of these minijobs are low-skilled jobs, in all cases where the occupational status is usually used to decide between various possible occupational codes it was assumed that the job is not a managerial position. The occupation with the lower prestige was then always coded. From wave 17 onwards, information on mini-jobs from the interviews with senior citizens is also coded.

Table 28: Internat. Standard Class. of Occupations 2008 (ISCO08) (continued)

	The coding of the data on apprenticeship aspiration, which is available for waves 11 to 14, also does not contain any information on occupational status. Moreover, it is not possible to fall back on the training qualification because the query is directed at persons who have not yet obtained such a qualification. Here, it is generally the case that trainees are assigned to the occupation for which they would be trained if the training request were to be fulfilled in the future. If a distinction is made in the ISCO-08 codes between different qualification levels, those codes are avoided that stand for an activity that is possible without training or for a semi-skilled activity. If a distinction is made in the qualification levels between attending a vocational school on the one hand and attending a Fachhochschule or university on the other, the code aimed at attending a vocational school is selected.
Literature:	ILO (2012)

Table 29: Classification of Occupations 1992 (KldB92)

Generated:	Employment - Variable name - Source variables
	Current *kldb1992 - ET2500*
	Spell data *(bio_spells) - kldb1992 - ET2500, PET2500*
	First *(PENDDAT) - kldb1992eewt - ET2500, PET1280, PET3950*
	Last *(PENDDAT) - kldb1992lewt - ET2500, PET1280*
	of father *(PENDDAT) - vkldb1992 - PSH0800*
	of mother *(PENDDAT) - mkldb1992 - PSH0700*
	Minijob - *kldb1992minj - PMJ0900*
Variable label:	actual empl.: Classification of Occupations 1992, current employment, gen.
	Spell data: *(bio_spells)*: Classification of Occupations 1992, gen.
	first empl.: Classification of Occupations 1992, first employment, gen.
	last empl.: Classification of Occupations 1992, last employment, gen.
	Father: Classification of Occupations 1992 of the father gen.
	Mother: Classification of Occupations 1992 of the mother gen.
	Minijob: Classification of Occupations 1992, current Minijob, gen.
Category / dataset	Occupation / individual-level data
Prepared by	Bernhard Christoph

Table 29: Classification of Occupations 1992 (KldB92) (continued)

Explanation	The KldB92 is the current version of the Classification of Occupations published by the German Federal Statistical Office (Statistisches Bundesamt) from the year 1992. This classification system was developed to match the German occupational structure, which is based solely on employment.
Literature:	StBA (1992)

Table 30: Classification of Occupations 2010 (KldB2010)

Generated:	Employment - Variable name - Source variables
	Current - *kldb2010 - ET2500, PET2500*
	Spell data *(bio_spells) - kldb2010 - ET2500*
	First - *kldb2010eewt - ET2500, PET1280, PET3950*
	Last - *kldb2010lewt - ET2500, PET1280*
	of father - *vkldb2010 - PSH0800*
	of mother - *mkldb2010 - PSH0700*
	Minijob - *kldb2010minj - PMJ0900*
	Apprenticeship aspiration, desired occupation - *kldb2010berufswunsch - PAA0100*
	Apprenticeship aspiration, intended occupation - *kldb2010angberuf - PAA1000*
Variable label:	actual empl.: Classification of Occupations *2010*, current employment
	Spell data: *(bio_spells)*: Classification of Occupations *2010*, gen.
	first empl.: Classification of Occupations *2010*, first employment, gen.
	last empl.: Classification of Occupations *2010*, last employment, gen.
	Father: Classification of Occupations *2010* of the father, gen.
	Mother: Classification of Occupations *2010* of the mother, gen.
	Minijob: Classification of Occupations *2010*, current Minijob, gen. (from wave 17 incl. seniors)
	Apprenticeship aspiration, desired occupation: Classification of Occupations *2010* , job descriptions: Desired occupation, gen.
	Apprenticeship aspiration, intended occupation: Classification of Occupations *2010*, job descriptions: Intended occupation, gen.
Category / dataset	Occupation / individual-level data

Table 30: Classification of Occupations 2010 (KldB2010)(continued)

Prepared by	Christian Dickmann
Explanation	The KldB 2010 classification of occupations is a completely new product that depicts the current occupational landscape in Germany very realistically. With the KlbD 2010 it is now possible to portray the occupational structures that have changed substantially in the past decades far better than before in statistics and analyses. Another advantage of the KldB 2010 is its good compatibility with the international occupational classification, ISCO-08 (International Standard Classification of Occu-pations 2008), as this improves the international comparability of occupational information in official statistics and in research. Reported occupations are coded in KldB2010 if they concern employment spells that have been carried forward from the previous wave from the tenth survey wave onwards or if it is new information reported from wave 10 onwards. Employment spells reported before wave 10 and not carried forward into wave 10ff. are available only as KldB-1992 codes. From wave 17 onwards, information on mini-jobs from the interviews with senior citizens is also coded. The coding of the data on apprenticeship aspiration, which is available for waves 11 to 14, also does not contain any information on occupational status. Moreover, it is not possible to fall back on the training qualification because the query is directed at persons who have not yet obtained such a qualification. Here, it is generally the case that trainees are assigned to the occupation for which they would be trained if the training request were to be fulfilled in the future. If a distinction is made in the KldB 2010 codes between different qualification levels, those codes are avoided that stand for an activity that is possible without training or for a semi-skilled activity. If a distinction is made in the qualification levels between attending a vocational school on the one hand and attending a Fachhochschule or university on the other, the code aimed at attending a vocational school is selected.
Literature:	Federal Employment Agency (2011)

Table 31: Erikson, Goldthorpe and Portocarrero (EGP) Class Scheme

Generated:	Employment - Variable name - Source variables
	Current - *egp - isco88, stib* Spell data *(bio_spells) - egp - isco88, stib* First - *egpeewt - isco88eewt, stibeewt* Last - *egplewt - isco88lewt, stiblewt*

Table 31: Erikson, Goldthorpe and Portocarrero (EGP) Class Scheme (continued)

	of father - *vegp - visco88, vstib* of mother - *megp - misco88, mstib*
Variable label:	Current empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), current occupation, generated Spell data (*bio_spells*): Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), gen. First empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), first employment, gen. Last empl.: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), last employment, gen. Father: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of father, gen. Mother: Class scheme acc. to Erikson, Goldthorpe & Portocarrero (EGP), occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation	The class scheme developed by Erikson, Goldthorpe and Portocarrero (Erikson et al., 1979, 1982; Erikson & Goldthorpe, 1992) is among the most common instruments for operationalising class. For this variable, data are coded by ISCO-88 occupational classification and occupational status. The coding procedure is based on an earlier approach elaborated by Christoph et al. (2005), who provide a detailed description of the procedure. Here, in contrast, unpaid family workers were not coded as self-employed but as individuals in dependent employment consistent with the coding applied in the European Socio-Economic Classification (ESeC), which is described in the next section. One difference between the EGP coding applied here and the ESeC coding is that in the EGP coding procedure, cases are “missing” (-7) in which the occupational activity seemed incompatible with occupational status (e.g., “directors and chief executives” [ISCO=1210] who reported that they were “employees performing simple duties” [StiB=51]). To ensure compatibility with the standardised coding procedure we adopted, we did not apply a comparable revision procedure using the EseC codes. EGP was not created for occupation information of the mini job because the normally collected information about the occupational status was not gathered in the mini job module.
Literature:	Christoph et al. (2005); Erikson & Goldthorpe (1992); Erikson et al. (1982); Erikson et al. (1979)

Table 32: European Socio-economic Classification (ESeC)

Generated:	Employment - Variable name - Source variables
	Current - *esec - isco88, stib, PET2000, PET2700* Spell data *(bio_spells) - esec - isco88, stib,* *ET110*, ET130** First - *eseceewt - isco88eewt, stibeewt, PET1261* Last - *eseclewt - isco88lewt, stiblewt, PET3801* of father - *vesec - visco88, vstib, PSH0670* of mother - *mesec - misco88, mstib, PSH0370*
Variable label:	current empl.: European Socio-economic Classification (ESeC), current occupation, gen. Spell data *(bio_spells)*: European Socio-economic Classification (ESeC), gen. first empl..: European Socio-economic Classification (ESeC), first employment, gen. last empl.: European Socio-economic Classification (ESeC), last employment, gen. father: European Socio-economic Classification (ESeC), occupation of father, gen. mother: European Socio-economic Classification (ESeC), occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation	The European Socio-economic Classification is largely based on the EGP class scheme. Unlike the latter, great importance was attached to international comparability of the operationalisation and validation of the classification (for a general description, see Rose & Harrison, 2007; for Germany, see Müller et al. 2006, 2007). The Stata do-file required to generate the ESeC was kindly provided by Heike Wirth from GESIS-ZUMA (Fischer & Wirth 2007). We simply adjusted the file to meet the requirements of this study. This do-file, originally written in standard SPSS syntax by Harrison and Rose (2006) as a standard program to generate the ESeC, was converted into Stata. ESeC was not created for occupation information of the mini job because the normally collected information about the occupational status was not gathered in the mini job module.
Literature:	Fischer & Wirth (2007); Harrison & Rose (2006); Müller et al. (2006, 2007); Rose & Harrison (2007)

Table 33: Magnitude-Prestige Scale (MPS)

Generated:	Employment - Variable name - Source variables
	Current - *mps - isco88*
	Spell data *(bio_spells) - mps - isco88*
	First - *mpseewt - isco88eewt*
	Last - *mpslewt - isco88lewt*
	of father - *vmps - visco88*
	of mother - *mmps - misco88*
Variable label:	current empl.: Magnitude-Prestige Scale , current empl. gen.
	Spell data *(bio_spells)*: Magnitude-Prestige Scale , gen.
	first empl..: Magnitude-Prestige Scale , first employment, gen.
	last empl.: Magnitude-Prestige Scale , last employment, gen.
	father: Magnitude-Prestige Scale , occupation of father, gen.
	mother: Magnitude-Prestige Scale , occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation :	The MPS (Wegener, 1985, 1988) is the only Germany-specific instrument available to operationalize social prestige based on detailed occupation information. The scale was originally developed for the 1968 version of the International Standard Classification of Occupations (ISCO-68). Because occupation codes in this study were based on the more recent ISCO-88 classification and the Classification of Occupations (KldB) developed by the Federal Statistical Office, a variant of the scale adapted to the ISCO-88 was used (Christoph 2005). Infas merged the data as part of the occupational coding procedure. MPS was not created for occupation information of the mini job because the normally collected information about the occupational status was not gathered in the mini job module.
Literature:	Christoph (2005); Wegener (1985, 1988)

Table 34: Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala)- Basis ISCO-88

Generated:	Employment - Variable name - Source variables
	Current - *siops1 - isco88*
	Spell data *(bio_spells) - siops1 - isco88*
	First - *siopseewt1 - isco88eewt*
	Last - *siopslewt1 - isco88eewt*
	of father - *vsiops1 - visco88*
	of mother - *msiops1 - misco88*
Variable label:	aktuelle Ewt.: Standard International Occupational Prestige Scale (Basis ISCO-88), current empl., gen.
	Spell data *(bio_spells)*: Standard International Occupational Prestige Scale (Basis ISCO-88), gen.
	first empl.: SIOPS (Basis ISCO-88), first empl., gen.
	last empl.: SIOPS (Basis ISCO-88), last empl., gen.
	father: SIOPS (Basis ISCO-88), occupation of father, gen.
	mother: SIOPS (Basis ISCO-88), occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation:	The Treiman Prestige Scale, which was originally constructed by Treiman (1977) for ISCO-68, is the first and only prestige scale available for international comparative research on occupations. Since its adaptation to the ISCO-88 (Ganzeboom & Treiman, 1996, 2003), the scale has commonly been called the “Standard International Occupational Prestige Scale”. Infas merged the data as part of the occupational coding procedure. SIOPS was not created for occupation information of the mini job because the normally collected information about the occupational status was not gathered in the mini job module.
Literature:	Ganzeboom & Treiman (1996, 2003); Treiman (1977)

Table 35: Standard International Occupational Prestige Scale (SIOPS/Treiman-Skala) – Basis ISCO-08

Generated:	Employment - Variable name - Source variables
	Current - *siops2 - isco08*

Table 35: Standard Internat. Occ. Prestige Scale (SIOPS/Treiman-Skala) (continued)

	Spell data *(bio_spells) - siops2 - isco08* First - *siopseewt2 - isco08eewt* Last - *siopslewt2 - isco08eewt* of father - *vsiops2 - visco08* of mother - *msiops2 - misco08*
Variable label:	aktuelle Ewt.: Standard International Occupational Prestige Scale (Basis ISCO08), current empl., gen. Spell data *(bio_spells)*: Standard International Occupational Prestige Scale (Basis ISCO-08), gen. first empl.: SIOPS (Basis ISCO08), first empl., gen. last empl.: SIOPS (Basis ISCO08), last empl., gen. father: SIOPS (Basis ISCO08), occupation of father, gen. mother: SIOPS (Basis ISCO08), occupation of mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Christian Dickmann
Explanation:	Ganzeboom and Treiman have also developed an updated version of the SIOPS for ISCO-08 and made available a syntax to generate it. For reported occupations, the SIOPS was generated on the basis of ISCO-08 if the occupations concern employment spells that have been carried forward from the previous wave from the tenth survey wave onwards or if it is new information reported from wave 10 onwards. For employment spells reported before wave 10 and not carried forward into wave 10ff. the SIOPS is available only on the basis of ISCO-88. The SIOPS was not generated for the occupation information on marginal part-time jobs and apprenticeship aspiration (surveyed in the waves 11 to 14) as the questions usually asked about occupational status were not asked in these modules.
Literature:	Ganzeboom & Treiman (2010, 2011)

Table 36: International Socio-Economic Index (ISEI) – Basis ISCO-88

Generated:	Employment - Variable name - Source variables
	Current - *isei1 - isco88* Spell data *(bio_spells) - isei1 - isco88*

Table 36: International-Socio Economic Index (ISEI) - Basis ISCO-88

(continued)

	First - *iseieewt1 - isco88eewt* Last - *iseilewt1 - isco88eewt* of father - *visei1 - visco88* of mother - *misei1 - misco88*
Variable label:	aktuelle Ewt.: International Socio-Economic Index (Basis ISCO88), current empl., gen. Spell data *(bio_spells)*: International Socio-Economic Index (Basis ISCO88), gen. first empl.: ISEI (Basis ISCO88), first employment, gen. last empl.: ISEI (Basis ISCO88), last employment, gen. father: ISEI (Basis ISCO88), occupation of the father, gen. mother: ISEI (Basis ISCO88), occupation of the mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation:	The ISEI is the most common indices of this kind, in part, due to the fact that, unlike most other SEIs, the ISEI is based on an original theoretical concept that considers the occupation and its socio-economic status as an intervening variable in the relationship between education and income. The ISEI was developed for the ISCO-68 (Ganzeboom, De Graaf & Treiman, 1992); it was later adapted to the ISCO-88 (Ganzeboom & Treiman, 1996, 2003). Infas merged the data as part of the occupational coding procedure. ISEI was not created for occupation information of the mini job because the normally collected information about the occupational status was not gathered in the mini job module.
Literature:	Ganzeboom et al. (1992); Ganzeboom & Treiman (1996, 2003)

Table 37: International Socio-Economic Index (ISEI) – Basis ISCO-08

Generated:	Employment - Variable name - Source variables
	Current - *isei2 - isco08* Spell data *(bio_spells) - isei2 - isco08* First - *iseieewt2 - isco08eewt*

Table 37: International-Socio Economic Index (ISEI)- Basis ISCO-08

(continued)

	Last - *iseilewt2 - isco08eewt* of father - *visei2 - visco08* of mother - *misei2 - misco08*
Variable label:	aktuelle Ewt.: International Socio-Economic Index (Basis ISCO08), current empl., gen. Spell data *(bio_spells)*: International Socio-Economic Index (Basis ISCO08), gen. first empl.: ISEI (Basis ISCO08), first employment, gen. last empl.: ISEI (Basis ISCO08), last employment, gen. father: ISEI (Basis ISCO08), occupation of the father, gen. mother: ISEI (Basis ISCO08), occupation of the mother, gen.
Category / dataset	socio-economic position / individual-level data
Prepared by	Christian Dickmann
Explanation:	The data records of the International Social Survey Programme (ISSP) for the years 2002 to 2007 form the basis for the ISEI-08 index. The data were merged by infas as part of the occupation coding procedure. For reported occupations, the ISEI was generated on the basis of ISCO-08 if the occupations concern employment spells that have been carried forward from the previous wave from the tenth survey wave onwards or if it is new information reported from wave 10 onwards. For employment spells reported before wave 10 and not carried forward into wave 10ff. the ISEI is available only on the basis of ISCO-88. The ISEI was not generated for the occupation information on marginal part-time jobs and apprenticeship aspiration (surveyed in the waves 11 to 14) as the questions usually asked about occupational status were not asked in these modules.
Literature:	Ganzeboom (2010)

Table 38: Classification of Economic Activities 2003 (WZ2003)

Generated:	Employment - Variable name - Source variables
	Current - *branche1 - ET2600*

Table 38: Classification of Economic Activities 2003 (WZ2003) (continued)

	Spell data *(bio_spells) - branche1 - ET2600* Minijob - *brancheminj1 - PMJ1300*
Variable label:	Current empl.: Current activity: economic sector/industry (WZ2003) Spell data *(bio_spells)*: economic sector/industry (WZ2003), generated Minijob: economic sector/industry, current minijob (WZ2003)
Category / dataset	socio-economic position / individual-level data
Prepared by	Bernhard Christoph
Explanation :	The information obtained from the open-ended survey question about the sec-tor/industry in which the respondent is employed was coded using the 2-digit Classification of Economic Activities of the Federal Statistical Office (WZ2003) code. At the two-digit level, this classification largely corresponds to the European Nomen-clature générale des Activités économiques dans les Communautés Européennes (NACE) in revision 1.1.
Literature:	StBA (2002); EG (2002)

Table 39: Classification of Economic Activities 2008 (WZ2008)

Generated:	Employment - Variable name - Source variables
	Current - *branche2 - ET2600* Spell data *(bio_spells) - branche2 - ET2600* Minijob - *brancheminj2 - PMJ1300*
Variable label:	Current empl.: Current activity: economic sector/industry (WZ2008) Spell data *(bio_spells)*: economic sector/industry (WZ2008), generated Minijob: economic sector/industry, current minijob (WZ2008)
Category / dataset	socio-economic position / individual-level data
Prepared by	Christian Dickmann
Explanation :	The responses to the open-ended question on the sector/industry in which the respondent is employed were coded using the two-digit code of the German Classification of Economic Activities compiled by the Federal Statistical Office (WZ2008).

Table 39: Classification of Economic Activities 2008 (WZ2008) (continued)

	The two-digit level is also termed the divisions level of the classification. It is based on the International Standard Industrial Classification of all Economic Activities (ISIC Rev. 4) of the United Nations and the Statistical Classification of Economic Activities in the European Community (NACE Rev. 2). These two industry coding bases are identical at the two-digit level. Reported industries are coded in WZ2008 if they concern employment spells that have been carried forward from the previous wave from the tenth survey wave onwards or if it is new information reported from wave 10 onwards. Industry details concerning employment spells reported before wave 10 and not carried forward into wave 10ff. are available only as WZ-2003 codes.
Literature:	StBA (2008); EG (2006)

Table 40: Physiological scale of SF12v2 (SOEP-Version, NBS)

Variable name	*pcs*
Variable label	Physiological scale of SF12v2 (SOEP-Version, NBS), generated
Source variables	*PG1200; PG1205; PG1210; PG1215**
Category / dataset	Health / individual-level data
Prepared by	Christian Dickmann
Explanation	The SF12 Questionnaire is an abbreviated version of the SF36 Questionnaire for measuring health-related quality of life. Since 2002 internationally renowned and applied SF12 indicators (version 2 – SF12v2) are used at SOEP. The SOEP-version of the questionnaire, however, differs from the original SF12v2 within formulation, order and layout of the questions. The SF12-indicators of PASS were surveyed analogous to SOEP. The generated pcs variable of PASS is based on the reproduced SPSS-Syntax of Nübling et al. (2006). So far the SF12-indicators were surveyed in waves 3,6,9,12 and 16 of PASS.
Literature:	Nübling et al. (2006); Andersen et al. (2007)

Table 41: Psychological scale of SF12v2 (SOEP-Version, NBS)

Variable name	*mcs*

Table 41: Psychological scale of SF12v2 (SOEP-Version, NBS) (continued)

Variable label	Physiological scale of SF12v2 (SOEP-Version, NBS), generated
Source variables	*PG1200; PG1205; PG1210; PG1215**
Category / dataset	Health / individual-level data
Prepared by	Christian Dickmann
Explanation	The SF12 Questionnaire is an abbreviated version of the SF36 Questionnaire for measuring health-related quality of life. Since 2002 internationally renowned and applied SF12 indicators (version 2 – SF12v2) are used at SOEP. The SOEP-version of the questionnaire, however, differs from the original SF12v2 within formulation, order and layout of the questions. The SF12-indicators of PASS were surveyed analogous to SOEP. The generated mcs variable of PASS is based on the reproduced SPSS-Syntax of Nübling et al. (2006). So far the SF12-indicators were surveyed in waves 3,6,9,12 and 16 of PASS.
Literature:	Nübling et al. (2006); Andersen et al. (2007)

4.5.2 Household or benefit unit level

Table 42: Equivalised household income, previous OECD weighting

Variable name	*incoecda*
Variable label	equivalised household income, old OECD weighting (rounded)
Source variables	*HD0200a-HD0200o; HA0100; hhincome*
Category / dataset	socio-economic position / household-level data
Prepared by	Bernhard Christoph/Christian Dickmann
Explanation	Equivalised household income considers the savings achievable through joint housekeeping in multiindividual households compared to single households. The per-capita income of the household is not divided by the actual number of individuals but by a divisor, which is usually less than this figure, and is calculated based on the assumed needs of household members (equivalised household size). According to the previous OECD scale, only the first household member (15 or older) is assigned a weighting factor of 1.0. Household members at least 14 years of age are assigned a weighting factor of 0.7, and children up to age 13 are assigned a weighting factor of 0.5 to calculate equivalised household size.

Table 42: Equivalised household income, previous OECD weighting (continued)

	From wave 17 onwards, the variable incoecda replaces (also retrospectively) the previous variable oecdinca, in which other age limits were used for the equivalence weighting (Berg et al., 2023).
Literature:	OECD (1982)

Table 43: Equivalised household income, modified OECD weighting

Variable name	*incoecdn*
Variable label	equivalised household income, modified OECD weighting (rounded) .
Source variables	*HD0200a-HD0200o; HA0100; hhincome*
Category / dataset	socio-economic position / household-level data
Prepared by	Bernhard Christoph/Christian Dickmann
Explanation	General description: see Equivalised household income, previous OECD weighting (above). The modified OECD equivalence scale assumes a weighting factor of 1.0 only for the first household member (15 or older). Household members at least 14 years old are assigned a weighting factor of 0.5, and children up to age 13 are assigned a weighting factor of 0.3 to calculate household size. For more information on the modified OECD scale, see Hagenaars, de Vos, and Zaidi (1994). From wave 17 onwards, the variable incoecdn replaces (also retrospectively) the previous variable oecdincn, which used different age limits for the equivalence weighting (Berg et al., 2023).
Literature:	Hagenaars et al. (1994)

Table 44: Deprivation index, unweighted

Variable name	*depindug3*
Variable label	All waves: deprivation index, unweighted (item total: 22) .
Source variables	*HLS0100a-HLS0400a; HLS0100b-HLS0400b; HLS0600a-HLS0900a; HLS0600b-HLS0900b; HLS1100a-HLS1200a* *HLS1100b-HLS1200b; HLS1400a-HLS2500a; HLS1400b-HLS2500b*

Table 44: Deprivation index, unweighted (continued)

Category / dataset	material situation / household-level data
Prepared by	Bernhard Christoph
Explanation	Following Ringen (1988), poverty researchers usually distinguish between direct and indirect measures of poverty. Indirect measurement focuses on the resources available to attain a particular standard of living, especially (equivalised household) income. This method is also called the resource-based approach to measuring poverty. In contrast, direct measurement attempts to record the household's ownership of goods and to determine the extent to which the households cannot afford certain goods or activities that are considered relevant. This method is also called the deprivation approach (see, e.g., Halleröd 1995). Previous scientific research suggests that the population classified as poor by the resource-based approach is not always identical to that identified by the deprivation approach. To define with precision who is to be considered poor, combining measures of resource poverty and deprivation is often been suggested i.e., to classify as poor only those individuals identified by both approaches (see Halleröd 1995; Nolan & Whelan 1996; Andreß & Lipsmeier 2001). The deprivation index is based on a list of 22 goods or activities. The surveyed households are asked to indicate whether they possessed these goods or participated in the activities mentioned. The unweighted index simply adds the number of items that respondents indicated they did not possess or in which they did not participate. However, only items that are missing for financial reasons are counted to prevent consumer preferences (e.g., a household choosing not to own a car or television) from being misinterpreted as a reduced standard of living. Additionally, an item was only accepted as missing for financial reasons if explicitly confirmed in the answers to both questions. "Don't know" or "details refused" answers were considered available goods or missing for a non-financial reason. This assumption does not apply to all cases. Alternatively, an index value for households that failed to answer a question for (at least) one particular good could be excluded (through listwise deletion). Of the 22 goods and activities surveyed, however, this method would quickly lead to a large number of missing index values. Therefore, the first method described was selected. Nevertheless, compared to the listwise deletion procedure, there is a risk that the number of goods missing for financial reasons is underestimated by this method. For waves 1 through 4, the variable *depindug* provides a version of the un-weighted deprivation index based on 26 items, i.e., adding to the 22 items mentioned above *HLS0500*, HLS1300** and *HLS2600**. These three items have not been asked since wave 5. And furthermore on item HLS1000*, which has not been surveyed since wave 13. Thus, depindug2 was newly integrated into the dataset in wave 5 and has been generated retroactively since wave 1.

Table 44: Deprivation index, unweighted (continued)

	For waves 1 to 12, there is also another version of the unweighted deprivation index, *depindug2*, which is based on 23 items instead of 22, namely *HLS1000** in addition to the items mentioned above. This item has not been collected since wave 13. Therefore, *depindug3* was newly added to the dataset in wave 13 and generated retrospectively since wave 1.
Literature:	Andreß & Lipsmeier (2001); Halleröd (1995); Nolan & Whelan (1996); Ringen (1988)

Table 45: Deprivation index, weighted

Variable name	*depindg3*
Variable label	All waves: deprivation index, weighted (item total until W7: 11.03, since W8: 10.53)
Source variables	*HLS0100a-HLS0400a; HLS0100b-HLS0400b;* *HLS0600a-HLS0900a; HLS0600b-HLS0900b; HLS1100a-HLS1200a; HLS1100b-HLS1200b;* *HLS1400a-HLS2500a; HLS1400b-HLS2500b;* *PLS0100-PLS0400; PLS0600-PLS0900; PLS1100-PLS1200;* *PLS1400-PLS2500;*
Category / dataset	material situation / household data
Prepared by	Bernhard Christoph
Explanation:	For a general description: see deprivation index, unweighted (above). Unweighted indices, such as the one described above, are often criticised for assigning all items included identical weightings. For example, the difference in asking whether a dwelling has an indoor toilet or whether there is a TV in the household immediately reveals the vast difference in the reduction of household's standard of living caused by the lack of an item. It therefore seems reasonable to weight the items. However, empirical research indicates that in most cases, weighted and unweighted index variants do not yield significantly different results (see Lipsmeier, 1999).

Table 45: Deprivation index, weighted (continued)

	For this survey, we weighted items according to the proportion of respondents who considered a particular item as necessary. We selected this procedure not only because it is conceptually convincing and commonly used (applied by Halleröd 1995, for example) but also because it can be implemented without unreasonable costs. The deprivation weightings determined for the individual questionnaire items are assumed highly stable over time, and these items only need to be administered once or in long intervals. Moreover, the large PASS sample allowed us to split the sample into several randomly selected subsamples, each of which classified only some items. Alternative weighting methods, such as restricting the indices to items that are considered necessary by a minimum proportion of the respondents (e.g., Andreß & Lipsmeier 1995, Andreß et al. 1996) or theoretically restricting the indices to a few fundamental items (e.g., Nolan & Whelan 1996), were not utilised in this survey but can be generated, if necessary, from the data provided. A discussion of the different methods of index weighting can be found in Andreß and Lipsmeier (2001, esp. p. 28 ff.). For waves 1 through 4, the variable *depindg* provides a version of the weighted deprivation index based on 26 items, i.e., in addition to the 22 items mentioned above, it includes the following items: *HLS0500*; HLS1300** and *HLS2600**. These three items have not been asked since wave 5. And furthermore on item *HLS1000**, which has not been surveyed since wave 13. Thus, depindg2 is newly integrated into the dataset in wave 5 and has been generated retroactively since wave 1. For waves 1 to 12, there is also another version of the unweighted deprivation index, *depindug2*, which is based on 23 items instead of 22, namely *HLS1000** in addition to the items mentioned above. This item has not been collected since wave 13. Therefore, *depindug3* was newly added to the dataset in wave 13 and generated retrospectively since wave 1.
Literature:	Andreß & Lipsmeier (1995, 2001); Andreß et al. (1996); Halleröd (1995); Lipsmeier (1999); Nolan & Whelan (1996)

Table 46: Household typology

Variable name	*hhtyp*
Variable label	Household type, generated
Source variables	Household information on age and relationships between household members.
Category / dataset	Category / dataset Household structure / household data
Prepared by	Daniel Gebhardt
Explanation	Various household typologies exist (see, e.g., Lengerer, Bohr & Jansen, 2005 for the Microcensus household typology; Porst (1984) and Beckmann & Trometer 1991 for the ALLBUS typology; and Frick, Göbel & Krause (n.d.) for the SOEP). The household typology used in PASS follows the latter typology. The decisive differentiation criteria are existing partnerships, number and age of children and existing generational relation-ships. Whereas the SOEP typology is based on the relationship of the household members to the head of the household, PASS uses information on the relationships among all household members. The PASS typology includes the ages of household members as indicated in the household interview and household size. Definition of relationships for generating the household type: • Married couples, registered partnerships, nonmarried partnerships and partner-ships whose status is not specified (missing value for the follow-up question about the type of partnership). • Child of an individual: biological child, stepchild, adopted/foster child or child whose status is not specified (missing value for the follow-up question about type of relationship to the child). • Parent of an individual: biological parent, stepparent, adoptive/foster parent or parent whose status is not specified (missing value in follow-up question about type of parenthood).

Table 46: Household typology (continued)

	Definition of household type: • One-person household: A household consisting of only one individual. • Couple without children: A household consisting of two individuals living as a couple. • One-parent household: A household consisting solely of one parent and his/her children. No restrictions apply to children's ages. • Couple with children under the age of 16: A household consisting of two individuals living as a couple and their respective and/or mutual children. All of the children are younger than 16. • Couple with children aged 16 or over: A household consisting of two individuals living as a couple and their respective and/or mutual children. All of the children are aged 16 or over. • Couple with children both under and over 16: A household consisting of two individuals living as a couple and their respective and/or mutual children. Some children living in the household are younger than 16 and others are older than 16. • Multigeneration household: A household consisting of members of at least three generations in linear succession. The core of the household is multigenerational, i.e., at least one individual in the household is both a child and a parent of another member of the household. Other people living in the household include parents, children, siblings, the central member's partner or a partner's siblings. • Other household: A household that could not be assigned to another household type. • Generation not possible (missing values): All households with at least one miss-ing value (-1, -2, -4) or implausible value (-8) in the main category of a relationship or age variable (except for households with three or fewer members in unambiguous relationship constellations for which the household type was generated even if ages were missing).
Literature:	Beckmann & Trometer (1991); Frick et al. (o.J.); Lengerer et al. (2005); Porst (1984)

Table 47: Wave 17 benefit unit ID

Variable name	*bgnr17*
Variable label	Benefit unit ID in wave 17 (2023)
Source variables	Household information on age and relationships between household members
Category / dataset	Benefit unit / person register
Prepared by	Gerrit Müller

Table 47: Wave 17 benefit unit ID (continued)

Explanation	The *bgnr17* variable is created at the individual level. It assigns an identification number to each household member that indicates the individual's relationship to a particular benefit unit. Consequently, household members with the same identification number constitute a benefit unit. The *bgnr17* variable is composed of the known household number and a two-digit indicator to identify the benefit unit with-in the household. The identification of a household member's relationship to a benefit unit is based solely on information about the relationships between household members from the household grid along with the ages obtained from the household interview. Therefore, the benefit units identified in this way are considered synthetic benefit units. The identification process does not consider information about actual benefits received, individual members' ability to work or qualification status, but it does identify groups of individuals in the same household who are or would be considered benefit units in jointly receiving benefits according to the provisions of Book II of the German Social Code in the event that such benefits are needed. This artificial allocation procedure is necessary because information about the existence of a benefit unit and the identification of individuals affiliated with that unit cannot be collected directly in the context of an interview. The allocation of an individual to a benefit unit is based on the latest version of the German Social Code, Book II, Section 7, Subsection 3. Each individual who has completed the age of 25 and has not yet reached the age limit according to § 7a, constitutes a separate benefit unit unless he or she is living in a partnership and/or has (a) child/children younger than 25 who has/have no partner/children of their own. In the latter case, the benefit unit consists of the individual, his/her partner and child(ren). If two individuals live in the same household with a mutual child but do not indicate that they are living in a partnership, a partnership is nevertheless assumed to exist according to Section 7, Subsection 3a. The corresponding individuals and their child(ren) are assigned to the same benefit unit. Individuals who are between the ages of 15 and 25 are generally assigned to their parents unless they are already living with a partner (or a child of their own) in a joint household. Individuals between the ages of 15 and 25 who live without their parents, partner or children constitute a separate benefit unit.

Table 47: Wave 17 benefit unit ID (continued)

	Individuals who reached the age limit according to § 7a, are not covered by Book II of the German Social Code and are therefore not considered members of a benefit unit (coded 0) unless they live with a partner who has not yet reached the age limit according to § 7a (or a child under 25). Likewise, children who have not reached age 15 who live in a household without their parents are not considered members of a benefit unit (code 0) because they are covered by the provisions of German Social Code Book XII. Benefit units were not assigned to households with missing information on relationships or the age of certain house-hold members. Instead, all members of these households were assigned code 99. By approximation, such households are interpreted as households consisting of only one benefit unit.
Literature:	German Social Code Book II – basic security for job-seekers (Sozialgesetzbuch, Zweites Buch - Grundsicherung für Arbeitssuchende (SGB II))

Table 48: Wave 17 benefit unit typology

Variable name	*bgtyp17*
Variable label	Type of benefit unit in wave 17 (2023)
Source variables	Household information on age and relationships between household members.
Category / dataset	Benefit unit / person register
Prepared by	Gerrit Müller
Explanation	The benefit unit typology is based on the same concept as the synthetic benefit unit used for variable *bgnr17*. Until age 25, children are considered members of their parents' benefit unit unless they themselves have a partner or child. BA statistics typologies are often still established based on reaching legal age (the 18th birthday). For example, according to our typology, households in which the youngest child is between 18 and 24 years old and that are classified as one-parent benefit units are considered single households in BA statistics. This difference must be noted when comparing PASS data with figures from the official statistics. Code 0, no benefit unit, was assigned to households in which one or more member(s) were not covered by Social Code Book II (see also code 0 for *bgnr17)*. Code 5, generation impossible (missing values), was assigned to households with missing information on relationships or the ages of individual household members (see code 99 for *bgnr17)*.
Literature:	-

Table 49: Benefit unit receiving Unemployment Benefit II on the wave 17 sampling date

Variable name	*bgbezs17*
Variable label	Benefit unit in receipt of UB II on the sampling date in wave 17 (2023)
Source variables	*HA0250*, HA0300, AL20100, AL20200, AL20300, AL20400,* *AL20616, AL20716*, HA0400, sample, hnr, bgnr17, hhgr*
Category / dataset	Benefit unit / person register
Prepared by	Mark Trappmann
Explanation	For each benefit unit that was identified according to the procedure described for variable *bgnr17*, this variable indicates whether the benefit unit on the sampling date in July of the previous year was actually receiving Unemployment Benefit II for wave 17.
Literature:	-

Table 50: Benefit unit receiving Unemployment Benefit II on the wave 17 survey date

Variable name	*bgbezb17*
Variable label	Benefit unit in receipt of UB II on the survey date in wave 17 (2023)
Source variables	*AL20616, AL20716*, zensiert (alg2_spells), sample, hhgr, bgnr17*
Category / dataset	Benefit unit / person register
Prepared by	Daniel Gebhardt
Explanation	For each benefit unit that was identified according to the procedure described for variable *bgnr17*, this variable indicates whether the benefit unit was actually receiving Unemployment Benefit II on the wave 17 survey date *(pinttag, pintmonat, pintjahr)*.
Literature:	-

Table 51: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date

Variable name	*bgbezbkorr10*
Variable label	Correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date

Table 51: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date (continued)

Source variables	*hintmon, hintjahr, kennungfbversW11 (HHENDDAT), AL20100-AL20400, AL20610, AL20710*, zensiert (alg2_spells), sample, hhgr, bgnr11, bgnr12, bgnr13, bgbezb10*
Category / dataset	Benefit unit / person register
Prepared by	Christian Dickmann
Explanation	See section below
Literature:	-

Table 52: Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date

Variable name	*bgbezbkorrflag10*
Variable label	Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 10 survey date
Source variables	*kennungfbversW11 (HHENDDAT), sample, bgbezb10, bgbezbkorr10*
Category / dataset	Benefit unit / person register
Prepared by	Christian Dickmann
Explanation	See section below
Literature:	-

Table 53: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date

Variable name	*bgbezbkorr11*
Variable label	Correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date
Source variables	*hintmon, hintjahr, kennungfbversW11 (HHENDDAT), AL20100-AL20400, AL20610, AL20710*, zensiert (alg2_spells), sample, hhgr, bgnr12, bgnr13, bgbezb11*
Category / dataset	Benefit unit / person register
Prepared by	Christian Dickmann

Table 53: Correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date (continued)

Explanation	See section below
Literature:	-

Table 54: Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date

Variable name	*bgbezbkorrflag11*
Variable label	Flag for correction of the Benefit unit receiving Unemployment Benefit II on the wave 11 survey date
Source variables	*kennungfbversW11 (HHENDDAT), sample, bgbezb11, bgbezbkorr11*
Category / dataset	Benefit unit / person register
Prepared by	Christian Dickmann
Explanation	See section below
Literature:	-

Due to the panel structure, PASS data are especially suited for analysing transitions into the sphere of Social Code Book II. The person register contains two variables – the generated variables *bgbezs** and *bgbezb** - that report the status of Unemployment Benefit II receipt at individual level at different points in time. *bgbezs** contains the benefit-receipt status as of the time when the sample was drawn, and *bgbezb** contains that at the time when the interview was conducted. The variable *bgbezb** is generated from the information provided in the interview for all subsamples and all waves and is therefore surveyed in a comparable manner over the entire period. The variable *bgbezs** , too, is generated from the details reported in the interviews for all subsamples and all waves. For all refreshment samples drawn from the registers of basic security benefit recipients of the Federal Employment Agency (all subsamples apart from the two population samples, *sample*=2 and *sample*=6 and *sample*=15), however, the register information is used as a correction factor in the first survey wave in which a new household is interviewed. In other words, in the first interview of each household in those samples it is set to one (benefit unit in receipt of basic security benefits) for at least one benefit unit, even if the information provided in the interview differs from this. In the subsequent waves this variable is then also generated solely on the basis of information provided in the interview. Due to the different sources of the variables, it is recommended to examine dynamics in basic security benefits either directly using the

spell data regarding receipt of basic security benefits or by means of the variable *bgbezb**. If the variable *bgbezs** is to be included, the first survey wave of any household should not be used, as then there would be a risk of possible measurement differences between administrative data and survey data being confounded with the genuine change. In the meantime a great deal of literature has been pub-lished about these measurement discrepancies on the basis of PASS data (see Bruckmeier et al. (2014); Bruckmeier et al. (2015); Bruckmeier et al. (2018); Eggs (2016); Kreuter et al. (2010); Kreuter et al. (2014)).

During the fieldwork period for wave 11, evaluations of the data from wave 10 that were already available and feedback from the interviewers in the field indicated that the question about receipt of Unemployment Benefit II (UB II) in the household questionnaire was misunderstood by some of the individuals in the subsample of Syrian and Iraki households. In comparison with the other BA refreshment samples (from previous waves or the same wave without the Syrian and Iraqi households), the share of households reporting that they have never received UB II is especially large.

In order to address this problem, in the current fieldwork period (13 weeks after start of fieldwork and 3 weeks after start of the foreign language fieldwork and the new BA refreshment samples) changes were made to the module on receipt of Unemployment Benefit II (UB II). The changes concerned only the refreshment subsample of Syrian and Iraqi households of the waves 10 and 11 (sample = 14 or 17). For this group an additional explanation was added to the introductory text at the beginning of the module on receipt of UB II *(HABLK01)* and additional information was provided for the interviewer in question *HA0300*. The specific changes can be seen in the household questionnaire for wave 11. In the corresponding position there are two versions. Version 1 contains the set of questions prior to the changes (during the current fieldwork period), version 2 contains the revised set of questions. Using the variable *kennungfbversW11* in the household dataset *(HHENDDAT)* it is possible to identify which version of the question was asked in the household interview.

This change in the questionnaire leads to particularities for the data preparation of the information regarding receipt of Unemployment Benefit II (UB-II). The existing data preparation rules for the details reported by the panel households in the Syrian and Iraqi subsample at the start of UB II receipt from wave 11 are maintained. In the generated variables *bmonat* and *bjahr* in the UB II spell dataset *(alg2_spells)* the start date of the receipt of UB II continues to be set to the date of the previous interview if the date reported in the interview is earlier than that. The actual details on the benefit receipt period remain visible to the user in the variables *AL20100* and *AL20200*. The variable *bgbezb10*, which was already made available in the scientific use file of wave 10 in the person register *(p_register)*, is not corrected. Instead, in the scientific use file of wave 11 a new variable *bgbezb10_korr* is generated. For this, in addition to the details from wave 10, the information reported in wave 11 is also used to determine receipt of UB II at the time of the interview in this

subsample. If it is reported in the household interview of wave 11 that the household was drawing UB II at the time of the household interview of wave 10, this is recorded in variable *bgbezb10_korr*. The additional variable *bgbezbkorrflag10* indicates whether such a correction was made. For households that do not continue their participation in wave 11 or were still asked version 1 of the question, the future information from wave 12 is additionally taken into account so that it can be included in the variables *bgbezb10_korr* and *bgbezb11_korr* in the scientific use file of wave 12.

Table 55: Number of benefit units within the household

Variable name	*anzbg*
Variable label	Number of synthetic benefit units in the HH, generated
Source variables	*bgnr17, hnr*
Category / dataset	Benefit unit / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit units existing in the household. The benefit units were identified according to the procedure to generate the variable *bgnr17*.
Literature:	-

Table 56: Number of benefit units in the household receiving benefits on the sampling date

Variable name	*nbgbezug*
Variable label	Number of benefit units in the HH receiving benefits on the sampling date
Source variables	*bgbezs17, bgnr17, hnr*
Category / dataset	Benefit unit / household dataset
Prepared by	Daniel Gebhardt
Explanation	This variable indicates the number of benefit units within a household that were receiving benefits according to Social Code Book II on the sampling date. The value was calculated via the household number by aggregating the benefit units within a household that were actually receiving benefits according to variable *bgbezs17* from the person register.
Literature:	-

5 Data preparation

Since wave 3, infas, not the IAB, has been responsible for preparing the data. To guarantee consistent data preparation in the longitudinal section, infas was provided with the relevant syntax files for data preparation from wave 2, necessary sources, intermediary datasets and documentation of individual operations. Important decisions, such as the correction of structural problems in participating households or the development of the *bio_spells* dataset, which was first developed in wave 4, were made with the IAB. The IAB was also available for questions during data preparation.

The information gathered in the wave 17 interviews is available from infas as ASCII data. First, infas prepared the following datasets from the raw data[35]:

- Household dataset for the cross-section, including the spell-reshaped questions for the module „childcare“
- Household dataset for the longitudinal section (module „Unemployment Benefit II“)
- Dataset updating household composition (matrix)
- Dataset updating family relationships in the household (relationship matrix)
- Individual/senior citizen dataset for the cross-section
- Individual dataset for longitudinal section I (module „employment biography [spells]“)
- Individual dataset for longitudinal section II (module „measures“)
- Dataset for open texts (across household, personal and senior citizen interviews)

Second, a more detailed, formal and content-oriented verification of the data was performed. These data were then prepared as the scientific use file.

The data checks conducted at infas can be divided into three steps, which are detailed in the following sections. First, the household structure of the re-interviewed households was reviewed and when necessary, corrected. If serious problems were identified in the structure, the corresponding interviews were removed (see Chapter 5.1 on this issue). This step was followed by a detailed review of the filter questions (applying corrections if necessary). Filter errors were marked and specific codes were set for missing values (see Chapter 5.2 on this issue). Next, selected items were verified for plausibility. Clearly implausible or contradictory responses were marked by a specific missing code. However, such data corrections were limited.

The following table reviews the steps of the data preparation:

[35] The software packages Stata version 16 and SPSS version 26 were used for data preparation.

Table 57: Overview of the steps involved in preparing the data of wave 17 of PASS

No.	Procedure
1	Import the raw data into working datasets
2	Check the household structure (see Chapter 5.1)
3	Remove problematic interviews (household and/or individual levels) (see Chapter 5.1)
4	Integration of personal and senior data records
5	Correct the household structure of re-interviewed households (see Chapter 5.1)
6	Filter checks at household level (see Chapter 5.2)
7	Construct a household grid dataset and perform plausibility checks (see Chapter 5.3)
8	Generate synthetic benefit units (see description of variables, Chapter 4.5)
9	Generate new control variables based on the household data after filter checks, household grid dataset and plausibility checks
10	Filter checks at individual level (see Chapter 5.2)
11	Code information from open-ended survey questions (see Chapter 4.1)
12	Plausibility checks of household and individual-level data (excluding spell data) (see Chapter 5.3)
13	Prepare, plausibility check and construct spell datasets (see Chapters 5.6 to 5.8 and Chapter 5.3)
14	Simple generated variables (see Chapter 4.4)
15	Complex generated variables (see Chapter 4.5)
16	Generation of the data structure for the scientific use file (household, individual and register datasets)
17	Anonymisation (see Chapter 5.5)

5.1 Structure checks and removing interviews

A structure check was conducted before the filter checks. Here, interviews that were not considered successful were to be identified and if necessary, removed from the datasets. In addition, the structure of re-interviewed households was compared with the structure reported during the previous wave to identify and if necessary, to correct implausible or problematic changes in household composition and errors in the allocation of the personal interviews to their respective positions in the household. To observe households in the longitudinal section, it is essential that the individuals be assigned consistently to their position in the household and the respondents can be identified clearly across waves. A personal identification number must not be assigned to different individuals in different waves. If the correct household composition was unclear, all of the interviews conducted

with this household in wave 17 were removed from the dataset. If a personal interview was conducted with the wrong individual without further problems in household composition, then only the personal interview was removed.

Different processes identified problematic cases. The relevant cases were discussed as part of a formal procedure between infas and the IAB. The final decision on how to proceed with these cases was made by the IAB. The following specifies the extent of the checks conducted. Not every check in every wave identifies problems. The result of a check is usually that an issue occurs in few cases. Furthermore, known error sources are absorbed during the interviews. For example, the intention of the survey instrument is that not all known target persons can move out of a panel household at the same time and that at least one remaining individual is at least 15 years old.

- Comparison of first names (age and sex): By comparing the first names reported in the current and previous waves, changes in household composition that had not been recorded correctly were identified. Instead of recording moves into and out of a household in the relevant places during the house-hold interview, interviewers sometimes renamed household members or changed their age or sex. All cases in which a first name had been changed that could not be attributed to correcting the spelling and for which the year of birth reported in the previous wave differed by more than one year from that reported in the current wave were reviewed individually. A decision was made as to whether the interviewer made a simple change requiring correction of the first name, age or sex or an inadmissible change to the household structure.
- Comparison of dates of birth: Furthermore, whether more than one individual with the same date of birth was living in the household was reviewed. Whether these cases were plausible was decided in the context of the household, using two waves. The remaining cases then underwent an-other review. Households in which a date of birth was reported in the current and previous waves by individuals in different positions in the household structure were identified. Here, it seemed reasonable to suspect that a different individual provided the personal interview in the current wave. In the context of the household and individual-level data of the current and previous wave, individual decisions were made for each household and personal interview.
- Verification of age in person interview *(PD0100)*: In general, the date of birth from the personal/senior citizen interview of the current wave displaces all other age information on that individual, e.g., from the household grid, and is the basis for all generated variables utilising age. The date of birth is corrected in *PD0100*. If an individual's year of birth changes significantly according to *PD0100* but the day and month stay the same, the previously known date of birth has never changed according to *PD0100*, and at least two pieces of information about the date of birth from *PD0100* are available from previous waves, then the year of birth is reset to the value from the previous waves considering the whole household. Consider a hypothetical individual

whose date of birth is recorded as February 1, 1972 in at least two previous waves and whose date of birth is now recorded as February 1, 1992. This date of birth would make this individual younger than the other children in the household. Without a correction, such an arrangement leads to an implausible relationship structure, which would consequently mean that synthetic benefit units could not be generated. Hence, in the example above, the date is corrected to February 1, 1972 in the current wave.

- Matching household and person interviews: To identify households that are considered not successfully surveyed, the datasets at the household and individual level are merged. Personal interviews without a full household interview and household interviews for which no individual interview was available were marked[36].
- Plausible entries and excerpts: Moves into and out of a household are another important factor. Panel households with reported move-outs were generally inspected and correlated with the split-off households. Evaluations were made as to whether the remaining household of the panel household is plausible. Interviews from panel households in which all household members leave except individual children under 15 years old were discarded for the panel and split-off households. If more than one individual moved, whether these individuals formed a joint split-off or several different households was considered and whether this is plausible was determined. For instance, cases in which one partner left the panel household with young children but the children formed several split-off households were considered implausible. In cases of a non-realised split-off household, move-outs were considered plausible, but all individuals who moved out were remerged into one joint split-off household. Individual cases occurred in which the panel household indicates that individuals formed a split-off household, but all members could be identified in the split-off household. Alternatively, not all members of the panel household live in the split-off household, and at least one member of the panel household was not reported as having moved out or moved to a split-off household other than the one observed. Decisions were made as to which reported move-outs were considered valid and which were discarded as implausible. If a reported move-out was retroactively discarded as implausible, the individual who had allegedly moved out was retroactively re-integrated into the household panel.
- Split-off households and panel participants: In split-off households, individuals who, according to the field, are newly added to the PASS through the split-off household are checked to see whether they originate from the panel household. Two situations promote these cases. The first situation arises when a panel household reports several individuals moving out and the split-off individuals formed more than one household. In that case, a dynamic preload is created for the current file for all split-off households identified through the panel household. If, however, individuals who, according to the

[36] New sample households for which a household interview but no valid personal interview was available were removed from the dataset following the procedure used in wave 1. In contrast, the household interviews of re-interviewed households and split-off households were retained.

panel household, live in various split-off households are actually sharing a split-off household, those individuals who were not assigned to this split-off household by the panel household but to another split-off household do not have a preload in this split-off household. This is because they are recorded as a new person.
It is possible that individuals from a panel household move out of or into a household that was formed as split-off household during a previous wave and that was success-fully surveyed at that time. Thus, there is another move from the original panel household into this split-off household after the separation of the split-off household. Regardless of whether the panel household from which the split-off household emerged was successfully surveyed during the wave of the move, such cases cannot be controlled in the field. To do so, the split-off household would have to be provided with the personal information of all individuals from the panel household (and possibly all individuals in other split-offs from this panel household) as a preload. The few cases in which such a situation might occur do not justify such efforts in the field. Instead, these cases must be found during the structure checks. Note that in this context, split-off households must be considered in the waves following their first successful survey even if they are considered panel households in field control.
In both cases, the personal identification numbers *pnr* of the individuals in the split-off household are corrected retrospectively. It must also be considered that these individuals are treated as new respondents in the personal/senior citizen interview although they might have already participated in an interview. This deviation is generally not corrected (see also Chapter 4.4).

- Returns of panel participants over time: In panel households that reported a move-out as of wave 2, a return to the household can also occur as of wave 3. Recognising these individuals as moving back in and assigning them their former household position instead of a new household position is a function of the household grid. Whether these requirements were met in the field in all cases was also evaluated. For individuals who were identified in the current wave as moving back in by comparing the first name, age and sex with the members who previously moved out of the household, the household structure must be changed. These changes led to retroactive changes of the personal identification number of the individual and the individual information in the household interview - e.g., information about childcare - to the correct position within the structural check. Whether an individual who is marked in the field as moving back in is the same individual who moved out during a previous wave was also verified. If not, this change represents an individual who is new to PASS. Changes to the household structure are also made in this case.
  In case of moves back into a household, whether the split-off household in which the individual lived was successfully surveyed during the current wave and whether the split-off household reported that the individual moved out were verified. In addition, the status of individuals who moved back into their panel household during a previous wave must continue to be verified with the split-off household provided the split-off

household is part of the current panel sample. If an individual who moves back in is still considered a current household member in his/her split-off household, a decision was made as to whether this was plausible or whether either household structure should be corrected.

- Person movements between panel and split households (networks): Returns are not the only cases of individuals being considered current household members of several households. This situation can also occur when a member of a split-off household is not recorded as having moved out of the panel household. Individual cases can be acknowledged as plausible after examination of both household structures. These cases are documented in the *zdub** variables in the person register. For further explanation, please refer to Chapters 4.4 and 5.4.1.2 of the data report for Wave 5 of PASS (Berg et. al., 2012).
  Other issues concerning the relationship of a panel household and its split-off households can also arise. Individuals who joined PASS via a split-off household might move to the panel household. Another possibility is that individuals move from one split-off household to another. Generally, all individuals in a panel household and all of its split-off households must be considered a network. The structure checks are designed so that individual moves among the households of such a network are detected regardless of the direction in which an individual moves.
- Longitudinal household structure tests: Household structure verification generally evaluates the changes between waves, not the plausibility of the structure. Therefore, the household structure first-time interviews can only be verified to a limited extent. For first-time households, information concerning first name, age and sex is reviewed to determine whether individual household members are listed multiple times. In this case, only the initially reported household position is maintained. This situation might lead to other changes in the household structure. If, for example, in a household interviewed for the first time, there are four individuals and the individuals in positions 2 and 3 are identical, individual 3 is removed and individual 4 is retroactively moved to position 3. As a rule, in a household interviewed for the first time with X household members, positions 1 to X are to be filled without gaps. Someone retroactively recognised as moving back through a subsequent change in his or her personal identification number also makes it necessary to move the individual information in the household interview.
- Duplication of households: In very rare cases, feedback from field interviewers reveals households that are included twice in the panel sample as identical households. In wave 4, this concerned two households whereby the household interviewed later (with a higher hnr number) was deleted from the sample for the subsequent waves. Two duplicates were discovered in wave 13 and four duplicates in wave 17. Here too, the households surveyed later were deleted from the panel sample from the subsequent wave onwards. There is no retroactive removal of duplicates in the previous waves, as this would affect the weighting, for example. In the *hh_register*, duplicate households in

the wave in which the duplication is detected are flagged in the variable *hnettod** with the code 26 ”HH not realized, duplicate”, which makes the reason for the non-realization transparent. In the *p_register*, the household members of the duplicate household in the wave in which the duplication is detected are correspondingly marked with code 56 ”HH not realized, duplicate” in the variable *pnettod**.

- Merging households: For the first time, two previously unrelated households moved together in wave 15. The two households were not related to each other until now. Now the only person of one household moved into the other household. She transfers her person number pnr and her original household number uhnr to her new household.

Individual decisions were also made to address cases that proved to be problematic during the structure checks. Here, the seriousness of the particular problem was significant. In cases in which the correct household composition in wave 17 was unclear, all of the interviews from wave 17 were removed. In wave 18, these households will be treated as households that did not participate in wave 17. If in retroactively removed household interviews moves-out were reported, the split-off households were discarded. This removal affected both the interviews conducted in the current wave in these split-off households and the sample of the subsequent wave. Split-off households that developed from a discarded interview of a panel household are retroactively classified as not having been conducted and do not contribute to the panel sample of the subsequent wave. If there was merely a problem in assigning individuals to their respective positions in the household, i.e., if it was suspected that a personal interview had been conducted with the wrong individual in wave 17, then only that personal or senior citizen interview was removed. Structural problems with no serious consequences that could be solved, for example, by removing a personal interview, first name, age and sex were made at the household level. The incorrect information concerned was replaced with the last valid value from the previous wave or the value from the previous wave added to the number of years since the last valid interview.

In addition, all interviews with individuals for households with no complete household inter-view were removed. In the opposite case, i.e., households for which no individual-level interview was available, a distinction was made between re-interviewed households and households from the refreshment sample. Households from the refreshment sample that were not successfully surveyed were removed following the procedure used in the previous waves. In the case of re-interviewed households without interviews at the individual level, however, the household interview was not deleted.

The netto variables *(hnettok17, hnettod17, pnettok17, pnettod17)* in the household and person register datasets indicate removed interviews. Through the corresponding variables in the household register, it is possible to trace the re-interviewed households whose household interviews were later removed. Net variables in the person register allow for tracing the cases in which only single individual-level interviews or all of the interviews in

the household were deleted. In the case of households from the refreshment sample of wave 17 without at least one valid household and personal interview, it is not possible to trace deleted interviews in the register datasets because these households were not included in the datasets.

5.2 Filter checks and missing codes used

During the filter checks, the correct operation of the filter questions in the instruments was verified using a statistical program. If certain questions were asked when the value of the relevant filter variable would have required something else (for example, if detailed information was requested about vocational training although the respondent had stated that he/she did not have any vocational qualification), these variables were set to missing code "-3" (not applicable), which they would also have received through correct use of the filters[37]. Moreover, some items were not asked in individual cases when those questions would have been necessary according to the filter (e.g., if no further information was recorded about vocational training although the respondent had stated that he/she had under-gone such training). In these cases, the missing code "-4" (question mistakenly not asked) was assigned. An assignment of code"-4" can also be based on the household structure evaluation described in Chapter 5.1. If an individual's move-out is retroactively discarded as implausible and the individual is retroactively classified as belonging to his or her former household, then individual information about these individuals in the household interview must be coded retroactively as mistakenly not surveyed. Thus, the code "-4" does not always refer to a problem in the survey instrument. If code "-4" is assigned to a question that is relevant for filtering subsequent questions, then the subsequent questions are also coded "-4" in case these subsequent questions are not asked. If these questions were asked because, for instance, several filter questions linked to this subsequent question and another filter question triggered the question correctly, the value recorded there remains.

In an additional step, the missing codes assigned by the field institute and system missing codes were replaced by standard values for all variables. The following table provides an overview of the assigned values. Codes "-1" and "-2" are the standard "don't know" and "details refused" answers recorded during the survey, respectively. Code "-3" is the general "not applicable" code for questions not asked due to filters. As described above, code "-4" was as-signed if a question was not asked because of a filter error. Codes "-5" through "-7" are question-specific codes. These can be either specific missing codes (e.g., "Not applicable, not available for the labour market") or special categories for valid values (e.g., a category for an income of greater than € 99,999 in the open question on income). These codes were only assigned as required.

[37] As is customary in such cases, the filter checks were conducted beginning with the items that were asked first.

Table 58: Overview of the missing codes used

Code	Explanation
-1	“don’t know”
-2	“details refused”
-3	“not applicable (filter)” (question not asked due to filter)
-4	“question mistakenly not asked” (question should have been asked)
-5	question-specific code number 1, only assigned as required
-6	question-specific code number 2, only assigned as required
-7	question-specific code number 3, only assigned as required
-8	“implausible value”
-9	“item not surveyed in wave”
-10	“item not surveyed in questionnaire version” [38]

The value “-8” is a specific missing code assigned during the plausibility checks (see Chapter 5.3 on plausibility checks). The missing code “-9” became necessary for the first time in wave 2. It is assigned if an item was not asked during a specific wave. Because the dataset is prepared in long format, as was described above, variables that were no longer asked in any version of the questionnaire from one wave onward are coded “-9” for the observations from this wave onward. The same applies to variables collected for the first time in a particular wave. They are subsequently coded with “-9” for observations of previous waves without a survey. Code “-10” can be used to consider differences between questionnaires, that is, between the personal questionnaire and senior citizen questionnaire or between two versions of the household questionnaire until wave 3.

5.3 Plausibility checks

For the plausibility checks, an extensive list of theoretically possible contradictions in the respondents’ statements was checked. The checks conducted during the previous waves were adapted and extended for the current wave. Furthermore, the household structure and spell data were checked for plausibility - especially for inadmissible overlaps within the

[38] As of wave 4, code ”-10” has only been used to differentiate between personal and senior citizen questionnaires. Up to and including wave 3, there was an additional differentiation at the household level between first-time and repeatedly interviewed households. The differentiation at the household level is not continued in wave 4 due to the merger of the questionnaire versions into one comprehensive household questionnaire.

individual spell types. Generally, only the data gathered in the cross-section of wave 17 were verified. No checks were conducted in the longitudinal section, that is, to compare the information provided in the current wave with that provided in the previous wave.

In detail, the following steps were conducted:

- Contradiction check: In general, contradictions were only corrected if the implausibility could be defined as particularly serious and/or if the alteration was considered minor. The latter applied, for example, if only a small number of cases were affected or if one missing code (e.g., “-3”) was replaced by another (e.g., “-8”). Two strategies were used to filter implausible statements. Either the implausible responses were corrected directly, or they were assigned a specific missing code.
- Correction of implausible data in case of incorrect input: Implausible responses were only corrected if it was highly probable that the interviewer had entered information incorrectly: for example, if the interviewer entered a monthly total rent of EUR 9,998.-. Here, it was assumed in the plausibility check that the five-digit missing code “99998” (don’t know) was entered incorrectly. This response and other similar responses were recoded to the corresponding missing categories. If the recoded missing categories triggered a filter in subsequent questions, as is the case for the categorical question of income, then the categorical questions were retroactively set to code “-4” (question mistakenly not asked).
- Assignment of the missing code ”-8”:: However, it was rarely the case that a value could be recognised as an incorrect entry with certainty. In most cases, it was only possible to establish a contradiction between two statements but not to identify specific incorrect entries that had led to the implausible statement. Therefore, in these cases, no corrections were made, and the specific missing value code “-8” was assigned instead. It was decided on an individual basis whether the code was assigned to one of the two variables involved in the contradiction or to both of them.
- Plausibility check of the household structure: This check was conducted based on the information collected in the household interview about family relationships between household members, age, sex and first name. Prior to this check, information about relationships in the household was supplemented by information about partnerships reported in the personal interview.
- Comparison of relationship information and demographic data: To identify implausible household structures, the information on relationships was first combined with the demographic information for individual household members. For the households that were identified as implausible during these checks, individual decisions were made considering overall household structure and other information gathered during the interviews (e.g., on marital status in the personal interview). Implausible relationships were marked as such (“-8”) or corrected based on additional information on the household context if it was highly probable that an error had occurred. For example, in

the case of two people of the same sex who were both biological parents of a third member of the household, the sex was corrected based on the first name. If the first names also indicated two people were of the same sex and if there was no other relevant information available, then the relationship was marked as implausible based on the household structure.

- Comparison of relationships within the household: In a second step, checks were conducted comparing sets of three family relationships for plausibility. The following provides an example of a relationship structure that would be classified as implausible: individual A is individual B's spouse. Individual A is the biological parent of individual C. Individual C is a sibling of individual B. If such a combination or similarly implausible combination of relationships was identified, an attempt was made to make the relationship plausible based on the household context. In the case described, the relationship data were corrected by coding individual C as a child of individual B, whose status was not specified. The aim was to correct as many of the implausible entries as possible because a plausible and complete set of relationships is necessary to generate the benefit unit.
- Plausibility check of spell datasets: In addition, the spell datasets were subjected to a number of plausibility checks, as detailed in Chapters 5.6 through 5.8.

5.4 Retroactive changes in waves 1 to 16

During the data preparation process for the scientific use file for wave 17, some changes were also made to the waves that had already been delivered. These changes included corrections of errors that were detected after the completion of the scientific use file of wave 16. The corrected data can now be used in the SUF datasets of the current wave, wave 17. The following tables provide an overview of the retroactive changes to the delivered waves of PASS[39].

[39] Adjustments to value or variable labels are only considered here if this changes the interpretation of variables or values.

Table 59: Overview of retroactive changes to the household dataset *(HHENDDAT, KINDER)*

Altered Variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*HD0200**	*HHEND-DAT*	5-16	correction	For individuals for whom no age in years is available from the household interview, but only the information that they fall either into the age group „15 to 64 years“ or into the age group „65 years and over“, an age of 26 years (lower age group) or 65 years (upper age group) should be assumed as control information in their interview at the individual level if these individuals refuse to state their year of birth at the beginning of their interview. However, according to the rules, the age of 26 or 65 years should not be stored in the age variables in the data records of the scientific use file, but the missing code -1 or -2. In fact, the age of 26 or 65 years was stored in the age variables in waves 5 to 16, contrary to the rules. This has now been corrected. This affects 52 ages.
incoecda \oecdinca	*HHEND-DAT*	1-16	regeneration and deletion	The variable oecdinca is deleted in the dataset from wave 17 (also retrospectively) and replaced by the new variable incoecda. When regenerating the means-weighted net household income (old OECD weighting), different age limits are used for the new variable incoecda.
incoecdn \oecdincn	*HHEND-DAT*	1-16	regeneration and deletion	The variable oecdincn is deleted in the dataset from wave 17 (also retrospectively) and replaced by the new variable incoecdn. When regenerating the means-weighted household net income (old OECD weighting), different age limits are used for the new variable incoecdn.
HA0100	*HHEND-DAT*	3	correction	In wave 3, the household size reported in variable HA0100 was too low by one person in three households and has now been corrected. Other places in the SUF were not affected.

Table 60: Overview of retrospective alterations in the individual dataset *(PENDDAT)*

Altered Variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
sample	*PENDDAT*	16	correction	One person was assigned an incorrect value in the sample identification in wave 16.
zpalthh palter	*PENDDAT*	15-16	correction	For individuals for whom no age in years is available from the household interview, but only the information that they fall either into the age group „15 to 64 years“ or into the age group „65 years and over“, an age of 26 years (lower age group) or 65 years (upper age group) should be assumed as control information in their interview at the individual level if these individuals refuse to state their year of birth at the beginning of their interview. However, according to the rules, the age of 26 or 65 years should not be stored in the age variables in the data records of the scientific use file, but the missing code -1 or -2. In fact, the age of 26 or 65 years was stored in the age variables in waves 5 to 16, contrary to the rules. This has now been corrected. This affects 7 ages.

Table 61: Overview of retroactive corrections to spell datasets *(bio_spells, alg2_spells, ee_spells)*

Altered Variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
ET2420 ET2421	*bio_spells*	16	correction	The filling of the variables ET2420 and ET2421 by the demand variables for „Other“ was not implemented correctly in the previous waves and was corrected in 4 cases.

Table 62: Overview of retrospective alterations to the register datasets *(hh_register; p_register)*

Altered Variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
*alter**	*p_register*	5-16	correction	For individuals for whom no age in years is available from the household interview, but only the information that they fall either into the age group „15 to 64 years“ or into the age group „65 years and over“, an age of 26 years (lower age group) or 65 years (upper age group) should be assumed in their interview at the individual level as control information if these individuals refuse to state their year of birth at the beginning of their interview. However, according to the rules, the age of 26 or 65 years should not be stored in the age variables in the data records of the scientific use file, but the missing code -1 or -2. In fact, the age of 26 or 65 years was stored in the age variables in waves 5 to 16, contrary to the rules. This has now been corrected. This affects 52 age details.
lastint	*p_register*	4 and 7	correction	For one person, the variable lastint previously contained the information that they had last given an interview in wave 4, although they had actually last been interviewed in wave 2. For another person, the variable lastint previously contained the information that they had last given an interview in wave 7, although they had never actually been interviewed.

Table 63: Overview of retrospective alterations to the weighting datasets *(hweights; pweights)*

Altered Variable	Dataset concerned	Altered wave	Type of alteration	Description of the alteration
-	-	-	-	-

5.5 Anonymisation

All data obtained by the IAB, a special department of the Federal Employment Agency (BA), are social data, which places high demands on data protection. It was therefore necessary to include some of the variables in the scientific use file in simplified form. These variables

are generally labeled with the flag “anonymised” in the variable label. For the same reason, it was also necessary to exclude available regional information, excluding the German states and information about East/West Germany. To protect the data, neither family relationships in the household nor the first names of the household members are part of the scientific use file. References to the household structure are provided, however, by generated variables. For example, the household and benefit unit type (*hhtyp*[40], *bgtyp*[41]), indicator variables on partners in the household (*apartner; epartner*[42]), indicator variables pointing to parents, partners in the household (*zmhh; zvhh; zparthh*[43]) and various indicator variables for parents (*mhh; vhh*[44]) or children of the target person (e.g. *ekind*[45]) living in the household are provided. The following table provides an overview of the variables concerned and the process of anonymisation[46] in each dataset. The following tables provide the anonymised variables for the employment spell dataset and the KINDER-dataset.

[40] Contained in the household dataset (*HHENDDAT*), see Chapter 4.5.2
[41] Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
[42] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[43] Wave-specific variables contained in the person register (*p_register*), see Chapter 4.4.
[44] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[45] Contained in the individual dataset (*PENDDAT*), see Chapter 4.4.
[46] If non-anonymised versions of one or several variables are indispensable for your research, please contact the Forschungsdatenzentrum (Research Data Center) to determine the possibility of obtaining access to the data. The form of this access will depend on the research project and the variables necessary.

Table 64: Overview of the anonymised variables in the individual dataset *(PENDDAT)* in wave 17

Varname	Variable label	Procedure
PD0100	Year of birth (date of birth, anon.)	The precise date of birth was shortened to year of birth.
gebhalbj	Half-year of birth, gen.	The precise date of birth was shortened to an indicator for the first or second half of the year.
PET1210	Last occupational status, simple classification (anon.)	For technical reasons, professional and regular soldiers were recorded separately. Due to the few case numbers and because this group is not usually asked about occupational status, this group was merged with civil servants and judges.
PET1250	Last occup. status civil servant: detailed info., incl. soldiers (anon.)	This variable contains additional cases. The professional and regular soldiers from *PET1240* were added to the corresponding civil servants category. The variable for professional and regular soldiers *PET1240* is not supplied.
PET1211	Last occup. status, simple class. (incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET1251	Last occup. status civil servant: detailed info., incl. soldiers (incl. spell info.) (anon.), gen.	Procedure as for *PET1250*. The variable for professional and regular soldiers *PET1240* is not supplied.
stiblewt	Occupational status, last employment, code number, gen.	When generating the occupational status variable, professional and regular soldiers were assigned to the corresponding civil servant category.
PET1510	Current occup. status, simple classification, surv. as of wave 2 (anon.)	Procedure as for *PET1210*.
PET1900	Current occup. status civil servant: detailed info., incl. soldiers (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers *PET1800* surveyed in the senior citizens' interviews is not supplied. For the personal interviews, no generated variable for professional and regular soldiers is incorporated into the individual dataset from the employment spells *ET09**.
stibkz	Current occupational status, simple classification, harmonised (anon.)	When generating the occupational status variable, professional and regular soldiers are assigned to the corresponding civil servants category.
stib	Occupational status, code number, gen.	Procedure as for *stiblewt*.
PET3300	First occup. status, simple classification (anon.)	Procedure as for *PET1210*.

Table 64: Overview of the anonymised variables in the individual dataset *(PENDDAT)* in wave 17

(continued)

PET3700	First occup. status civil servant: detailed info., incl. soldiers	Procedure as for *PET1250*. The variable for professional and regular soldiers *PET3600* is not supplied.
PET3301	First occup. status, simple class. (merged, incl. spell info.) (anon.), gen.	Procedure as for *PET1210*.
PET3701	First occup. status civil servant: detailed info., incl. soldiers, (merged, incl. spell info) (anon.), gen.	Procedure as for *PET1250*. The variable for professional and regular soldiers *PET3600* is not supplied.
stibeewt	Occupational status, first employment, code number, gen.	Procedure as for *stiblewt*.
PSH0320	Mother's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210*.
PSH0360	Mother's occup. status at that time, civil servant, incl. soldiers: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers *PSH0350* is not supplied.
mstib	Mother's occupational status, code number, gen.	Procedure as for *stiblewt*.
PSH0620	Father's occup. status at that time, simple classification (anon.)	Procedure as for *PET1210*.
PSH0660	Father's occup. status at that time, civil servant, incl. soldiers: detailed info. (anon.)	Procedure as for *PET1250*. The variable for professional and regular soldiers *PSH0650* is not supplied
vstib	Father's occupational status, code number, gen.	Procedure as for *stiblewt*.
PMI0200	Not born in Germany: country of birth	Countries with very low case numbers were grouped into larger categories.
ogebland	Country of birth, incl. open info., categories (anon.)	Procedure as for *PMI0200*.
PMI0500	No German nationality: which nationality? (anon.)	Nationalities of countries with very low case numbers were grouped into larger categories.
ostaatan	Nationality, incl. open info., categories (anon.)	Procedure as for *PMI0500*.
ostaatan-syr	Nationality, syr./iraq. HH, incl. open info., categories (anon.)	For the sub-samples of Syrian and Iraqi households, the Syrian nationality is shown separately.
os-taatanukr	Nationality, ukrain. HH, incl. open info., categories (anon.)	For the sub-samples of Ukrainian households, the Ukrainian nationality is shown separately.

Table 64: Overview of the anonymised variables in the individual dataset *(PENDDAT)* in wave 17

(continued)

PMI1000a	Father: country of res. before migration (anon.)	Countries of residence before migration with very low case numbers were grouped into larger categories.
PMI1000b	Mother: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000c	Father's father: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000d	Father's mother: country of res. before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000e	Mother's father: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
PMI1000f	Mother's mother: country of residence before migration (anon.)	Procedure as for *PMI1000a*.
ozulanda	Father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandb	Mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandc	Father's father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandd	Father's mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulande	Mother's father: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.
ozulandf	Mother's mother: country of residence before migration, incl. open info., categories (anon.)	Procedure as for *PMI1000a*.

Table 65: Overview of the anonymised variables in the BIO-spell dataset *(bio_spells)* in wave 17

Varname	Variable label	Procedure
ET0615	Wave 17, Occup. status, simple classification (anon.)	Procedure as for *PET1210*.
ET1015	Wave 17, Occ. status: civil servant/ judge/soldier, detailed information (anon.)	Procedure as for *PET1250*. The variable for professional and and regular soldiers is not supplied.
stib	Occ. status, code number, gen.	Procedure as for *stiblewt*.

Table 66: Overview of anonymised variables in the children dataset in wave 17 *(KINDER)*

Varname	Variable label	Procedure
alter12u14m	children in the age of 12 to less than 14 months old	Since wave 10 the age of children under 7 is asked once on a monthly basis. The information about month and year of birth was reduced to one indicator, if the child was in the age of 12 to less than 14 months old at the point of the interview. Based on this information the indicator was also filled for previous interview waves.

5.6 Receipt of Unemployment Benefit II

UB II is recorded at the household level in spell form in waves 1 to 16. This concept was continued in wave 17 but with a slightly revised set of questions.

5.6.1 Concept for updating the spells of Unemployment Benefit II receipt that were ongoing in the previous wave

To update spells for which UB II was ongoing during the previous wave and therefore were right-censored in the spell dataset, dependent interviewing questions are included. Households with ongoing spells from the previous wave start here again with the interview.

The households from the refreshment sample that were interviewed for the first time in wave 17 were asked about their receipt of UB II during the period since the last change in the household composition. If this change was before January 2021 or if no information was provided about changes in the household, then the household's receipt of UB II from January 2021 on was recorded. The reference date is adjusted by + 1 year in each wave and is always in January two years before the current survey year.

5.6.2 Structure of the Unemployment Benefit II spell dataset

The structure and contents of the spell dataset on UB II change due to the integration of the spells of UB II reported in wave 17. Here, it is necessary to distinguish among (1) new variables that refer to a particular wave, (2) new variables that do not refer to a particular wave and (3) variables that are no longer asked in wave 17.

1. Additionally, in wave 17, new wave-specific, cross-sectional variables were included in the UB II spell dataset. These variables include *AL20616, AL20716a* to *AL20716o* and *AL20816*. These variables refer to the interview date in wave 17. Cross-sectional variables also exist for the interview dates of the previous waves that contain the analogous information referring to the respective wave. The following table provides an overview of the cross-sectional information contained in the UB II spell dataset.

Table 67: Cross-sectional variables in the UB II spell dataset *(alg2_spells)*

	Wave 1	Wave 2	...	Wave 11	Wave 12	...	Wave 17
Does the HH receive UB II for all HH members?	*AL20600*	*AL20601*	...	*AL20610*	*AL20611*	...	*AL20616*
Does the HH receive UB II for individuals 1 to 15?	*AL20700a-AL20700o*	*AL20701a-AL20701o*	... ...	*AL20710a-AL20710o*	*AL20711a-AL20711o*	... ...	*AL20716a-AL20716o*
Amount of monthly UB II receipt?	*AL20800*	*AL20801*	...	*AL20810*	*AL20811*	...	*AL20816*
Has a cut of UB II begun?	*AL20900*	*AL20901*	...	*AL20910*			

2. Not available in wave 17.
3. Not available in wave 17.

5.6.3 Plausibility checks and corrections to the Unemployment Benefit II spell dataset

As in waves 1 to 16, the information on UB II was also subjected to a number of plausibility checks in wave 17. Inadmissible overlaps and dates of spells of UB II were corrected when necessary. In principle, changes were only made to the generated date variables (*bmonat; bjahr; emonat; ejahr*) of the spell of UB II receipt and the censoring indicator of the spell of UB II receipt (*zensiert*). If it was not possible to remove implausible data by correcting the dates, then in a small number of cases, spells of UB II receipt were merged or deleted.

5.6.4 Updating the Unemployment Benefit II spell dataset

After the spells of Unemployment Benefit II reported in wave 17 had been converted into spell format, and after inadmissible overlaps and implausible dates were corrected following the plausibility checks and corrections, the spells of UB II that were ongoing at the time of the interview in the previous wave were updated using the information gathered in wave 17. Two variants are to be distinguished here. In the first (1), only the censoring indicator *zensiert* is changed. The second variant (2) is an update of the spell that was censored during the previous wave using information gathered in wave 17. Here, the censoring indicator is integrated into the spell of receiving UB II, which was ongoing during the previous wave, as are the generated and recorded end dates and wave-specific cross-sectional information (see above). In addition to updating spells that were censored during the previous wave, new spells that were reported in wave 17 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Cases in which the household in wave 17 contradicts an ongoing spell of receiving UB II at the interview date in the previous wave.*
   If the household contradicted an ongoing spell of receiving UB II at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then *zensiert* was set to “2” (no). The information provided in the interview of the previous wave is assumed correct. Because it is not possible to make reliable statements about the continued duration of the benefit receipt beyond the date of the interview in the previous wave, it is assumed that the benefit receipt ended during the month of the interview in the previous wave. The reported and generated variables for the end date of the spell (*AL20300, AL20400* and *emonat, ejahr*) along with the question of whether a spell continues (*AL20500*)remain unchanged[47]. The generated end date of the UB II spell

[47] The same applies here. Only the censoring indicator is changed. The reported end date, the question for continuing spells and the generated end date remain unchanged.

(*emonat; ejahr*)) had been set to the interview date of the previous wave in the previous wave.

2. *Cases in which the household reports the end date of a spell of benefit receipt that was ongoing in the previous wave.*
   If information about the end date of a spell of UB II receipt that was censored in the previous wave is available in wave 17, then the spell that was censored in the previous wave was updated using the current information. First, the recorded end date (*AL20300; AL20400*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the receipt of UB II is ongoing (*AL20500*) and the censoring indicator (*zensiert*) are overwritten with the information gathered in the previous wave. Furthermore, the cross-sectional data referring to wave 17 (*AL20615; AL20715a to AL20715o, AL20815*) were included.
3. *Spells of UB II receipt reported for the first time during wave 17 that do not update any spells that were censored in the previous wave.*
   Spells reported for the first time during wave 17 were added to the UB II spell dataset. Next, the spell counter was generated new to create a variable *spellnr* without gaps.

5.7 Employment biographies

Since wave 4, the concept of an integrated survey of the employment biography has been implemented (employment, unemployment and gap periods at the individual level were recorded in spell form still in waves 2 and 3). For individuals who were asked for their employment biography for the first time in wave 17, the reference date for the start of the retrospective interval was adjusted again. The reference date is adjusted by + 1 year in each wave and is always in January two years before the current survey year. Accordingly, all employment and unemployment episodes since January 2021 were to be reported in wave 17. Individuals who were interviewed about their employment biography during the previous wave, however, should report all new spells since the date of the last interview.

5.7.1 Variables on the employment/inactivity status in PENDDAT

The concept for surveying employment spells has been revised several times over the various waves:

- Wave 1: Panel concept, i.e. surveying only the most recent information
- Wellen 2 und 3: Waves 2 and 3: modular survey of spells of employment and unemployment + filling of gaps of > 3 months and the most recent information

- Ab Welle 4: From wave 4 onwards: integrated survey of employment/unemployment/gap spells

Owing to the changes in the survey concept, the information available for the individual waves vary with regard to:

- the form of the available information (panel vs. spells)
- the degree of detail of the available information (main status vs. parallel states)
- the consistency of the existing parallelities (filling of gaps vs. full survey of parallel states)

The concept of the generated variables on the employment/inactivity status applied in waves 2 and 3 follows the survey logic of the first wave very closely. This logic – in a simplified form – was as follows:

- Is there a case of employment of at least 1 hour per week?
- If employment: one job or more?
- If employment (information reported for main employment): step-by-step identification of whether the employment is a mini job, a one-euro job or such like, or part of an apprenticeship
- If no employment (or main employment = mini job): determination of inactivity status (unemployment or other status))

The concept of the generated variables (*erwerb, erwerb2, nichterw, nichtew2*) follows this survey logic from wave 1 in the broadest sense. Whereas in wave 1 the interview logic did not permit competing states (respondents with employment that was not marginal part-time were not asked about other activities), from wave 2 onwards it became necessary to make decisions if there was more than one ongoing spell. When generating the variables on the employ-ment/inactivity status in waves 1 to 3 the following logic was applied:

Table 68: Logic of generation of *erwerb, erwerb2, nichterw, nichtew2*

Variable	Logic of generation wave 1	Logic of generation wave 2 and wave 3
erwerb	*(1) Differentiation main employment status* - no main employment - main employment: not apprenticeship/job creation scheme/mini job - main employment: part of apprenticeship - main employment: job creation scheme etc. - main employment: mini job *(2) Differentiation main employment status is the basis for further generation* - main employment: not apprenticeship/job creation scheme/mini job → employment as occupational status (Exceptions: apprentices (from PB0100) with arbzeit < 21 → apprentices; pupils (from PB0100) with arbzeit > 0 & arbzeit < 24 → pupils; students (from PB0100) with arbzeit > 0 & arbzeit < 21 → students; employed persons with arbzeit > 0 & arbzeit < 16 → other) - no main employment or main employment: mini job → take occupational status from PET0801 (meaning insert the status of economic inactivity) - no main employment + according to PB0100 pupil/student → take occupational status from PB0100 - main employment: job creation scheme etc. → Take as occupational status (job creation scheme, one-Euro job, etc.) *(3) Deciding in contradictory cases* - erwerb: job creation scheme etc. + PB0100: pupil/student/apprentice → -8 - erwerb: pupil + PB0100: student → -8 - erwerb: pensioner + PB0100: apprentice → -8 - erwerb: pupil + PB0100: apprentice → take status from PB0100 - erwerb: other + PB0100: pupil/student/apprentice → occupational status from PB0100	Not generated (-9)

Table 68: Logic of generation of *erwerb, erwerb2, nichterw, nichtew2* (continued)

erwerb2	*(1) Recode of erwerb* Merging categories: - unemployed + job creation scheme/one-Euro job etc. → unemployed - Apprenticeship/vocational training/further training - Retraining + student → (Vocational) apprenticeship/ university/ college	*(1) Recode of nichtew2* *(2) Integrate employment spells* - replace values, if current employment (>400 Euro from employment spells) is available *(3) Make adjustments* - erwerb2: employment + PB0100: student + working hours <= 20h → student - erwerb2: unemployment + PB0100: student → student - erwerb2: pupil + PB0100: student → status not clear
nichterw	*(1) Recode of PET0800*	*(1) Recode of LU0100 ((gap status without open answer) + current unemployment from unemployment spells)* - Combination of categories: - Registered as unemployed + not registered → Unemployed - (Vocational) apprenticeship/ university/ college + other → other - Determination MV from PET0151/ PET0911 + indicator for mistakenly not in the gap module filtered cases
nichtew2	*(1) Recode of PET0801* - Combination of categories: - Unemployed + job creation scheme/ one-Euro job etc. → Unemployed - Apprenticeship/ vocational training/ further training - Something different/ main status unclear → Other/ main status unclear - Retraining + student → apprenticeship/ vocational training/ studies	*(1) Recode of LU0101 (gap status with open answer)* - Combination of categories - Registered as unemployed + not registered → Unemployed *(2) Take pupil/ student/ apprentice from PB0100 into account* - If currently no valid status available → take the information from PB0100

The generated variables therefore continue the logic of the survey concept of wave 1, which is also the basic logic in the generated variable: Employment takes priority over all other states in principle (apart from a few exceptions); unemployment takes priority over all states apart from employment (apart from a few exceptions)

In wave 1 it would not have been possible to implement a different logic (e.g. unemployment taking priority over employment) as the survey logic prioritised the respondent's employment, and other states were only surveyed as alternatives. The

procedure followed for generating variables is therefore the same as that followed for surveying the information.

However, this procedure is not really useful for determining the person's main status and also ignores basic concepts that are found, for example, in the definition of unemployment (§§16, 119 Social Code Book III (SGB III); also applies for SGB II in accordance with §53a SGB II).

Unemployment has certain preconditions (according to the definition in Social Code Book III):

- being without work (i.e. no paid employment, or employment only up to a maximum of 15 hrs/week; fluctuations are possible) (§119 SGB III)
- availability (i.e. available for placement efforts on the part of the Federal Employment Agency (BA); seeking and willing to take up work >= 15hrs/week; able to follow up integration suggestions promptly; willing to participate in occupational integration measures) (§119 SGB III)
- own effort (i.e. making an effort to end unemployment) (§119 SGB III)
- registration (i.e. personally registered as unemployed at the BA) (§16 SGB III)
- not currently participating in a measure (§16 SGB III)

The logic followed so far, in which employment takes priority over unemployment, irrespective of the number of hours, is therefore driven more by the survey logic of wave 1 than by a consideration of what is actually to be regarded as the main status in terms of content.

Further criticism of the employment/inactivity variables concerns the fundamental objective of these variables. What are they intended to show? The person's main status? The employment status (if so, what exactly is that)? On closer examination, the objective appears inconsistent, as two concepts are combined:

- The statement regarding the TP's main status (i.e. in the case of competing states a decision is made as to which status takes priority over another under which conditions)
- The statement as to whether the TP currently has a certain status (even if this status is perhaps not the main status because another status takes priority)

There are essentially two possibilities for generating the employment/inactivity variables from wave 4 onwards:

- Continuing the previous logic for generating the variables but with a new data basis
- Revising the logic for generating the variables with the aim of:

- Defining the concepts more precisely (what exactly do the variables depict?)
- Improving the decisions that were made in the past against the background of the available data but are suboptimal in terms of content (i.e. not simply continuing the previous logic with a new data basis, but using the more detailed data basis with regard to content)
- Streamlining (i.e. removing variables with extremely limited additional benefit)

It was decided to fundamentally revise the variable-generating logic. The following procedure is used for the previous variables:

Table 69: Revision *erwerb, erwerb2, nichterw, nichtew2*

Variable	Decision	Explanation
erwerb	maintain (Wave 1: generated with regard to content) (Wave 2ff: -9)	The variable represents the survey concept of wave 1 optimally. The focus lies on employment (in a simplified way they beat unemployment, and this in turn beats everything else). Some considerations with regard to content seem to present an obstacle of the continuation, but this can be solved by a new concept due to the detailed database. For wave 1 the variable is maintained, because it is well-suited for the survey concept. The special characteristics (no parallelisms; concentration on employment; no differentiation of registered and unregistered unemployment) remain limited to wave 1.
erwerb2	dropped from SUF	The logic of the survey concept of wave 1 is continued in a harmonized way with this variable. But with it several problems arise: (1) There is a change in which employment spells are collected (wave 1: 1h/week vs. wave 2ff.: > 400 Euro) (2) Focus changes (wave 1: If employment [not mini job] available → no collection of parallel unemployment/gap-statuses; wave 2ff.: employment/unemployment/(partly also gap) simultaneously possible) (3) Due to adhering to the logic of wave 1 the opportunities of the new database cannot be used appropriately (e.g. in order to take more appropriate decisions with regard to content) Conclusion: A harmonized variable with focus on employment (as before in erwerb2) is the only possibility for a harmonized variable over all waves. A generation of these variables would be possible, but only on the base of inappropriate conceptual decisions. As the concept of wave 1 is regarded as problematic, an inclusion of the harmonized variable is omitted.

Table 69: Revision *erwerb, erwerb2, nichterw, nichtew2* (continued)

nichterw	dropped from SUF	The previous division in labour status and economic inactivity status is given up and replaced by main status + indicator for current employment (subject to social insurance) + indicator for current registration as unemployed. Wave 1: Variable offers no additional information in comparison with the new main-status variable Wave 2ff.: Additional information in comparison with the new main-status variable is very limited Conclusion: In general rather additional complexity with very limited utility (e.g. students > 20h/working time per week). For the analysis a separate determination of sub-statuses probably more appropriate than previously included variables.
nichtew2	dropped from SUF	(see nichterw)

From wave 2 onwards the following variables are generated:

- *etakt: currently employed (above mini-job earnings limit), generated (from wave 2 onwards)*
- *alakt: currently registered as unemployed, generated (from wave 2 onwards)*
- *statakt: current main status, generated (from wave 2 onwards)*

The objectives of the revision were as follows:

- Separating the information on the main status (*statakt*) from the information on currently ongoing spell types (*etakt, alakt*)
- Documenting the rules more clearly when identifying the main status
- Differentiating between registered and not registered unemployment (where possible)

etakt (currently employed (above mini-job earnings limit), generated (from wave 2 onwards))
The variable indicates that the TP had an ongoing spell of employment at the time of the personal interview of the respective wave (i.e. an emp. above mini-job earnings limit). For wave 1 the variable cannot be generated as the survey concept differs between wave 1 and the subsequent waves (wave 1: at least 1 hr/week; wave 2ff. above mini-job earnings limit). A person is regarded as being currently employed if there is a censored employment spell in the spell record of the respective wave.

Values of the generated variable:

- -10 Item not surveyed in questionnaire version
- -5 Cannot be generated (missing values)
- -3 Not applicable (filter)
- 1 Currently in occupation (above mini-job earnings limit)
- 2 Currently not in occupation (above mini-job earnings limit)

alakt (currently registered as unemployed, generated (from wave 2 onwards))
The variable indicates that the TP was registered as unemployed at the time of the personal inter-view of the respective wave. For wave 1 the variable cannot be generated as the survey concept differs between wave 1 and the subsequent waves (wave 1: unemployment only surveyed if no employment reported; wave 1: unemployed; wave 2ff.: registered as unemployed). A person is regarded as being currently registered as unemployed if there is a censored (registered) unemployment spell in the spell record of the respective wave.

Values of the generated variable:

- -10 Item not surveyed in questionnaire version
- -5 Cannot be generated (missing values)
- -3 Not applicable (filter)
- 1 Currently unemployed
- 2 Currently not unemployed

statakt (current main status, generated (from wave 2 onwards))
The variable indicates which main status the TP had at the time of the personal interview in the respective wave.

This variable is generated on the basis of the spell records (waves 2 and 3: *employment/unemployment/gap spells*; wave 4ff.: *BIO-Spells*) and the status as pupil/student/apprentice in PB0100.

If a certain spell type is currently ongoing in the respective wave, then the corresponding state exists for that person. In waves 2 and 3 the spell type is determined via the respective spell record (employment/unemployment spells) or the gap state (*LU0101* in *gap-spells*) From wave 4 onwards the variable spelltyp can be used. In all waves only spells that were ongoing on the date of the interview (i.e. censored=1 in the SUF of the respective wave) are taken into account. The current status as a school pupil or as a student/apprentice from PB0100 is taken into account as if there were a currently ongoing spell in the respective spell.

Values of the generated variable:

- -10: Item not surveyed in questionnaire version
- -5: Cannot be generated (missing values)
- -3: Not applicable (filter)
- 1: In occupation with earnings above mini-job earnings limit
- 2: Unemployed, registered
- 3: Pupil/student (school)
- 4: Apprenticeship/Studying
- 5: Military or civilian service
- 6: Carrying out domestic duties
- 7: Maternity protection/parental leave
- 8: Pensioner/early retirement
- 9: Other/ main status unclear
- 10: Unemployed, not registered (since W4 from open item)
- 11: Ill/unfit to work/unemployable (open item)
- 12: Self-employed/family worker (open item)

The assignment of the codes should be conducted step-by-step:

Table 70: Basic assignment - Spell with higher priority beats spell with lower priority

Priority of a current spell (e.g. analogous status from PB0100)	Code in statakt (analogous to variable spelltyp)	Meaning
1	2	Registered as unemployed/ Participation in measure
2	1	In occupation with earnings above mini-job earnings limit
3	8	Pensioner/ early retirement
4	7	Maternity protection/ parental leave
5	5	Military or civilian service
6	4	Apprenticeship/ Studying
7	3	Pupil/ student (school)
8	12	Self-employed/ family worker
9	11	Ill/ unfit to work/ unemployable
10	10	Unemployed, not registered
11	6	Carrying out domestic duties
12	9	Other/ main status unclear

If no valid values are available for the additional information, the rough allocation remains unchanged.

Table 71: Detailed assignment for special cases

Basic assignment	Additional information	Decision
Registered as unemployed	In occupation with earnings above mini-job earnings limit + working hours (az2ges; actual working hours, sum over censored employment spells) >= 15h	In occupation with earnings above mini-job earnings limit
In occupation with earnings above mini-job earnings limit	Apprenticeship/ Studying + working hours (az2ges; actual working hours, sum over censored employment spells) <= 20h	Apprenticeship/ Studying

A current spell of registered unemployment exists if there is a censored spell of (registered) unemployment in the spell record of the respective wave (waves 2 and 3: *unemployment spells*; wave 4ff.: *BIO-spells*)

5.7.2 Income variables and working hours in the PENDDAT and in the BIO-spell dataset

In waves 1 to 4 the variables on current employment refer to the main employment [48]. An exception to this is the information on the gross/net income in waves 2 to 4 – this refers to all currently ongoing jobs above mini-job earnings limit (uncertainty with regard to wages in marginal part-time jobs). Spell-specific information is not available and is only surveyed from wave 5 onwards. The information is only surveyed as a total value for all jobs. This results in two problems:

1. From wave 2 onwards, the generated variables on working hours and gross/net wage refer to different jobs (main job and all jobs). If hourly wages are calculated on this basis, errors occur in TPs with more than one job.
2. The different earnings are not evident from the variable labels.

The generated variables on income and working hours are therefore revised accordingly in wave 4.

Income variables

[48] Waves 2 and 3; it concerns the censored employment in the employment spell record. If there was more than one censored spell, then the spell with the most hours was selected. If there was more than one censored spell with the same number of hours, the spell with the longest duration was selected. In the case of senior citizens, information was only gathered about one job.

The concept for surveying the income variables changed considerably between waves 1 and 2 without this leading to the creation of new variables: in wave 1 *gross income* (*bruttokat*) and *net income* (*nettokat*) report the income from the main employment, from wave 2 onwards it reports the income from all jobs that are not marginal part-time. This is inconsistent and potentially leads to errors in evaluations. This problem is to be corrected with the revision:

Table 72: Revision income variables

Variable - Content - Dataset	Generated for	Basis
	W1 - W2 - W3 - W4 - W5ff.	openA - CatA
bruttokat - Main employment, gross - PENDDAT	1 - 0 - 0 - 0 - 1	0 - 1
brutto - Main employment, gross - PENDDAT	1 - 0 - 0 - 0 - 1	1 - 1
nettokat - Main employment, net - PENDDAT	1 - 0 - 0 - 0 - 1	0 - 1
netto - Main employment, net - PENDDAT	1 - 0 - 0 - 0 - 1	1 - 1
brges - Total employment, gross - PENDDAT	0 - 1 - 1 - 1 - 1	1 - 1
netges - Total employment, net - PENDDAT	1 - 1 - 1 - 1 - 1	1 - 1*
br - Employment spell, gross - BIO-Spells	0 - 0 - 0 - 0 - 1	1 - 1
net - Employment spell, net - BIO-Spells	0 - 0 - 0 - 0 - 1	1 - 1

* In wave 1, only a categorical question for the net income of the main employment exists but not for the additional jobs. This is accepted in the generation of *netges* If the details (MV) of the net income of the additional jobs are missing, the variable *netges* cannot be generated.

Revised variables (already in the dataset in waves 1 to 3):

- **bruttokat** *(Current gross income main employment (without mini jobs, categorical), gen.)*
- **brutto** *(Current gross income main employment (without mini jobs, incl. cat. details), gen.)*
- **nettokat** *(Current net income main employment (without mini jobs, categorical), gen.)*
- **netto** *(Current net income main employment (without mini jobs, incl. cat. details), gen.)*

In wave 1 these variables refer to the respective main employment. From wave 2 onwards, however, it contained the cumulated responses for all jobs (above mini-job earnings limit), as only these were surveyed. The variable labels were adapted accordingly from wave 4 onwards. For waves 2 to 4 the variables are filled with the value -9 as it is not possible to generate the variable in the same way as in wave 1.

New variables in wave 4:

brges *(current total gross income (excl. marginal emp., incl. cat. info.), gen.)*

This variable contains the cumulated information on the gross income from all jobs (above mini-job earnings limit). For wave 1 the variable cannot be generated in this form as the gross income was only surveyed for the main employment. For waves 2 and 3 the variable is identical in terms of content to the variable brutto that was supplied in the SUF of wave 3 (i.e. prior to the revision described above). In waves 2 to 4 only the cumulated gross income was surveyed – the source variables used in waves 2 and 3 therefore already contain the corresponding information on the total income from all jobs (above mini-job earnings limit). For wave 4 the variable is to be created in the same way as in waves 2 and 3. From wave 5 onwards the variable is generated on the basis of spell-specific income details.

netges *(current total net income (excl. marginal emp., incl. cat. info.), gen.)*
This variable contains the cumulated information on the net income for all jobs (above mini-job earnings limit). For wave 1 the variable can be generated by combining the responses to the open-ended and categorical questions on the net income from the main employment with the responses for the other jobs (the categorical follow-up question is missing here, however). For waves 2 and 3 the variable is identical to the variable netto that was supplied in the SUF of wave 3. In waves 2 to 4 only the cumulated net income was surveyed – the source variables used in waves 2 and 3 therefore already contain the corresponding information on the total income from all jobs (above mini-job earnings limit). For wave 4 the variable was created in the same way as in waves 2 and 3. From wave 5 onwards the variable is generated on the basis of spell-specific income details.

Working hours
Owing to the correction of the variables on the (gross/net) income (see above in this section) it is no longer possible to generate hourly wages in the individual dataset, as the only information avail-able on working hours is the actual working hours of the main employment (*arbzeit* variable in the *PENDDAT* of the SUF of wave 3). Analogous to the revision of the income variables it is therefore necessary to revise the working hours variables in both the *PENDDAT* and the *BIO-spell dataset*.

Table 73: Revision working hours variables

Variable - Content - Dataset	Generated for	Basis	Remark
	W1 - W2 - W3	openA - CatA	
az1 - Employment spell, contractual - BIO-Spells	0 - 1 - 1	1 - 0	Cat. wave 2ff.
azhpt1 - Main employment, contractual - PENDDAT	0 - 1 - 1	1 - 0	Cat. wave 2ff.
azges1 - Total, contractual - PENDDAT	0 - 1 - 1	1 - 0	Cat. wave 2ff.
az2 - Employment spell, contractual - BIO-Spells	0 - 1 - 1	1 - 1	Corresponds to previous variable arbzeit (BIO-Spells); cat. wave 2ff.; Employment with max(az2) = main employment (if two identical: Employment with earliest start
azhpt2 - Main employment, contractual - PENDDAT	1 - 1 - 1	1 - 1	Corresponds until now to variable arbzeit (PENDDAT); cat. wave 1 != cat. wave 2ff.
azges2 - Total, contractual - PENDDAT	1 - 1 - 1	1 - 1*	Cat. wave 1!= Cat. wave 2ff.; in wave 1 no cat. for secondary employment

Revised variables (already in the dataset in waves 1 to 3):

arbzeit *(weekly working hrs. incl. details of irregular working hrs., gen.)*
Variable is dropped from *PENDDAT* and *BIO-spell dataset*. It is replaced in terms of content by *azhpt2* (*PENDDAT*) and *az2* (*BIO-spell dataset*).

New variables since wave 4:

az1 *contractual working hrs., gen.)*
The variable is generated for all spells in the *BIO-spell dataset*. It contains the most recent information on the contractual working hours for the respective spell (empl. above mini-job earnings limit). The cross-sectional variables for which details were asked most recently in the re-spective spell form the basis for generating the variable in each case.

E.g.:

- Spell created in wave 2, ended in wave 2: cross-sectional variables wave 2
- Spell created in wave 2, carried forward in waves 3 and 4: cross-sectional variable wave 4
- Spell created in wave 9, carried forward in waves 10, 11 and 12: cross-sectional variable wave 12

azhpt1 *(contractual current working hrs. of main emp. (excl. marginal emp.), gen.)*
The variable is generated for the *PENDDAT* . It contains the contractual working hours of the currently ongoing main employment in the respective wave from the spell data (empl. above mini-job earnings limit). For wave 1 the variable cannot be generated (-9), as the corresponding information was only surveyed from wave 2 onwards. From wave 2 the generated variable on the contractual working hours of the main employment (*az1*) from the respective spell data is transferred to the *PENDDAT*. Which currently ongoing spell is the main employment is determined on the basis of the actual working hours (generated variable az2 in the spell data; analogous to the procedure in waves 2 and 3, in which the variable *arbzeit* was used to determine the main employment).

azges1 *(total current contractual working hrs. (excl. marginal emp.), gen.)*
The variable is generated for the *PENDDAT*. It contains the cumulated contractual working hours of all currently ongoing jobs in the respective wave from the spell data (empl. above mini-job earnings limit). For wave 1 the variable cannot be generated (-9), as the corresponding information was only surveyed from wave 2 onwards. From wave 2 the variable is generated from the spell data on the basis of the generated variable on the contractual working hours (*az1*). To generate the variable the information in the generated variable on contractual working hours (*az1*) is cumulated across all spells that were currently ongoing at the time of the survey. This information is transferred to the *PENDDAT*.

az2 *(actual working hrs. incl. details of irregular working hrs., gen.)*
The variable is generated for all spells in the *BIO-spell dataset*. It contains the most recent information on the actual working hours for each spell and also integrates the responses to the categorical questions on irregular working hours. The variable is generated on the basis of the cross-sectional variables for which information was gathered most recently in the respective spell.

E.g.:

- Spell created in wave 2, ended in wave 2: cross-sectional variables wave 2
- Spell created in wave 2, carried forward in waves 3 and 4: cross-sectional variable wave 4
- Spell created in wave 9, carried forward in waves 10, 11 and 12: cross-sectional variable wave 12

The variable replaces the variable *arbzeit* that was previously generated in the employment spells (which is accordingly dropped). It is generated in the same way that *arbzeit* was generated in the data preparation process for waves 2 and 3.

Definition of main employment:
The variable *az2* serves to determine the main employment in a wave, for which various details are transferred to the *PENDDAT*. The main employment is the currently ongoing job with the most hours in the respective spell. If there is more than one job with the same number of hours, the one that began first is selected. If there is more than one job with the same number of hours and the identical starting date, the job that the respondent mentioned first is selected. Of the possible jobs, this one has the lowest spell number.

azhpt2 *(current actual working hrs. main emp. (excl. marginal emp., incl. cat. info.), gen.)*
The variable is generated for the *PENDDAT*. It contains the actual working hours of the currently ongoing main employment and also integrates the responses to the categorical questions on irregular working hours. In terms of content the vari-able replaces the variable *arbzeit* that was dropped from the *PENDDAT*. It is generated in the same way that the discontinued variables were generated for waves 1 and 2.

In wave 1 the variable is generated on the basis of the cross-sectional data. It therefore combines the responses to both the open-ended questions on the actual working hours and the categorical follow-up questions. One-Euro jobs, job-creation measures, minijobs and activities that are part of an apprenticeship are not taken into account here – for these cases the variable cannot be generated (-3), as analogous information was not gathered in waves 2 to 4.

From wave 2 onwards the generated variable on the actual working hours of the main employment (*az2*) from the respective spell data is transferred to the *PENDDAT*. Which currently ongoing spell is the main employment is determined here, too, on the basis of the actual working hours (generated variable *az2* in the spell data; analogous to the procedure in waves 2 and 3, in which the variable *arbzeit* was used to determine the main employment). The categorical follow-up question in the case of irregular working hours differs between wave 1 and the subsequent waves. Nonetheless the information is integrated across the waves.

azges2 *(current total actual working hrs. (excl. marginal emp., incl. cat. info.), gen.)*
The variable is generated for the *PENDDAT*. It contains the cumulated actual working hours of all currently ongoing jobs in the respective wave.

In wave 1 this is done by combining the hours of the main employment (after integrating the responses to the categorical questions on irregular working hours) with the responses on the actual working hours of the other jobs. One-Euro jobs, job-creation measures, mini jobs and activities that are part of an apprenticeship are not taken into account here – for these cases the variable cannot be generated (-3), as analogous information was not gathered in waves 2 to 4.

From wave 2 the variable is generated from the spell data on the basis of the generated variable on the actual working hours (*az2*). To generate the variable the information in the generated variable on actual working hours (*az1*) is cumulated across all spells that were currently ongoing at the time of the survey. This information is transferred to the *PENDDAT*.

5.7.3 Concept for updating the spells that were ongoing in the previous wave

Continuing employment, unemployment and gap spells were updated in wave 17. To update the spells that were ongoing during the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires.

5.7.4 Structure of the BIO-spell dataset

With respect to its structure, the BIO-spell dataset has oriented itself on the modular employment, unemployment and gap spell datasets of waves 2 to 3 since wave 4. employment-specific variables kept their names in the BIO-spell dataset compared to the employment SUF of wave 3, analogous to the unemployment- and gap-specific variables. Variables which are the same in employment, unemployment and gaps have been standardised (*BIO0100, BIO0101, BIO0200, BIO0300, BIO0400, BIO0500, BIO0600*) as of wave 4 or were already standardised in the original datasets of the SUF wave 3 (*bmonat, bjahr, emonat, ejahr, zensiert*). Furthermore, variables for type of activity (*spelltyp*), spell integration (*spintegr*) and comprehensive spell number (*spellnr*) are available.

Due to the integration of the employment and unemployment spells reported in wave 17 into the BIO-spell dataset, new employment- and unemployment-specific variables are added. Here, it is necessary to distinguish between (1) new variables that refer to a particular wave, (2) new variables that do not refer to a particular wave and (3) variables no longer surveyed in wave 17.

1. New variables that are related to a specific wave: The employment-specific variables in the BIO-spell dataset *ET0600* to *ET2200* are considered wave-specific, cross-section information that refer to wave 2; variables *ET0601* to *ET2201* refer to wave 3, *ET0552* to *ET2202* refer to wave 4, *ET0553* to *ET2203* refer to wave 5, *ET0554* to *ET2204* refer to wave 6, *ET0555* to *ET2205* refer to wave 7, *ET0556* to *ET2206* refer to wave 8, *ET0557* to *ET2207* refer to wave 9, *ET0558* to *ET2208* refer to wave 10, *ET0559* to *ET2209* refer to wave 11,

ET0560 to *ET2210* refer to wave 12, *ET0561* to *ET2211* refer to wave 13, *ET0562* to *ET2212* refer to wave 14, *ET0563* to *ET2213* refer to wave 15, *ET0564* to *ET2214* refer to wave 16 and *ET0565* to *ET2215* are cross-section information that refer to wave 17. Since wave 5, variables on income for currently persistent spells are also surveyed (*ET28** to *ET39**. Since wave 9, *ET41** and *ET42** are used to ask questions about overtime. The following table provides an overview of the ET-specific cross-section information in the BIO-spell dataset.

Table 74: ET-specific cross-section variables in the BIO-spell dataset (*bio_spells*)

	Wave 2	Wave 3	Wave 4	Wave 5	...	Wave 9	...	Wave 17
Occupational status			*ET0552*	*ET0553*	...	*ET0557*	...	*ET0565*
(simple and detailed	*ET0600*	*ET0601*	*ET0602*	*ET0603*	...	*ET0607*	...	*ET0615*
classification)	*ET0700*	*ET0701*	*ET0702*	*ET0703*	...	*ET0707*	...	*ET0715*
	ET0800	*ET0801*	*ET0802*	*ET0803*	...	*ET0807*	...	*ET0815*
	ET1000	*ET1001*	*ET1002*	*ET1003*	...	*ET1007*	...	*ET1015*
	ET1100	*ET1101*	*ET1102*	*ET1103*	...	*ET1107*	...	*ET1115*
	ET1200	*ET1201*	*ET1202*	*ET1203*	...	*ET1207*	...	*ET1215*
Supervisory function;	*ET1300*	*ET1301*	*ET1302*	*ET1303*	...	*ET1307*	...	*ET1315*
number of employees	*ET1400*	*ET1401*	*ET1402*	*ET1403*	...	*ET1407*	...	*ET1415*
supervised								
Cancellation of limi-	*ET1700*	*ET1701*	*ET1702*	*ET1703*	...	*ET1707*	...	*ET1715*
tation of an initially				*ET1753a*	...	*ET1757a*	...	*ET1765a*
limited employment				*ET1753b*	...	*ET1757b*	...	*ET1765b*
Working hours			*ET1952*	*ET1953*	...	*ET1957*	...	*ET1965*
(contracted; actual;	*ET2000*	*ET2001*	*ET2002*	*ET2003*	...	*ET2007*	...	*ET2015*
average for irregular	*ET2100*	*ET2101*	*ET2102*	*ET2103*	...	*ET2107*	...	*ET2115*
working hours)	*ET2200*	*ET2201*	*ET2202*	*ET2203*	...	*ET2207*	...	*ET2215*
Income for current				*ET2800-*	...	*ET2804-*	...	*ET2812-*
ongoing spells				*ET3900*	...	*ET3904*	...	*ET3912*
Overtime						*ET4100*	...	*ET4108*
						ET4200	...	*ET4208*

The BIO-spell dataset also includes an AL-specific variable which is understood as wave-specific cross-sectional information (*AL1300* for wave 2; *AL1301* for wave 3, *AL1302* for wave 4, *AL1303* for wave 5, *AL1304* for wave 6, *AL1305* for wave 7, *AL1306* for wave 8, *AL1307* for wave 9, *AL1308* for wave 10, *AL1309* for wave 11, *AL1310* for wave 12, *AL1311* for wave 13, *AL1312* for wave 14, *AL1313* for wave 15, *AL1314* for wave 16 and *AL1315* for

wave 17). The following table gives an overview of the cross-sectional information contained in the spell dataset.

Table 75: AL-specific cross-section variables in the BIO-spell dataset (*bio_spells*)

	Wave 2	Wave 3	...	Wave 17
Amount of monthly UB I receipt?	*AL1300*	*AL1301*	...	*AL1315*

2. New variables that are not related to a specific wave: Not available in wave 17 compared to wave 16.
3. Variables no longer collected: Not available in wave 17 compared to wave 16.

5.7.5 Plausibility checks and corrections of the spell datasets

At the individual level, the plausibility checks and corrections orient themselves by the waves since wave 2. As in the previous waves, checks were made only within one spell type. Cross-spell type checks were not conducted. As with the spell data on receiving UB II, correction and recoding were only conducted for the generated date variables. Here, details on seasons were recoded into months, "-8" values were set for implausible responses and date information was replaced or rendered plausible. Because only the generated date variables were edited, the original information gathered in the survey is available to the user in the date variables *BIO0200-BIO0500* and *AL0800-AL1100* thus permitting the user to conduct his/her own checks and corrections.

In some cases entire spells were deleted in the course of data cleaning. For example, spells that were obviously recorded twice were removed. Spells that are completely outside the survey period but for which data were nonetheless collected were also deleted.

5.7.6 Update of spell datasets

After the spells reported in wave 17 had been converted into spell format, plausibility checks and corrections for inadmissible overlaps and spells with implausible dates were corrected. The spells that were ongoing at the time of the previous interview wave were updated using the information recorded in wave 17.

Three variants are to be distinguished here. In the first (1), only the censoring indicator *zensiert* is changed. The second variant (2) is an update of the spell that was censored in the

previous wave using information gathered in wave 17 in the narrow sense. Here, the censoring indicator is integrated into the spell that was ongoing during the previous wave, as are the generated and recorded end dates and wave-specific cross-sectional information (see above).

In addition to updating spells that were censored during the previous wave, new spells reported in wave 17 are merged with the spell dataset (3). These three variants are outlined briefly below:

1. *Change of censoring indicator: Cases in which the individual in wave 17 contradicts an ongoing spell on the interview date in the previous wave.*
   If the individual contradicted the information that there was an ongoing spell at the time of the previous wave, either explicitly or implicitly (by reporting an end date that preceded the interview date in the previous wave) in the update question, then the censoring indicator *zensiert* was set to "2" (no). The information provided in the interview of the previous wave is assumed correct. Because it is not possible to make any reliable statements about the continued duration of the spell beyond the date of the interview in the previous wave, it is assumed that the spell ended during the month of the interview in the previous wave. The reported and generated variables on the end date of the spell (*BIO0400, BIO0500* and *emonat, ejahr*), along with the question of whether a spell continues (*BIO0600*) remain unchanged[49]. The generated end date of the spell (*emonat; ejahr*) was already set to the interview date of the previous wave in the previous wave.
2. *Continuation of the spell censored in the previous wave with wave 17 information in the narrow sense: Cases in which the individual reports the end date of a spell that was ongoing in the previous wave.*
   If information about the end date of a spell that was censored during the previous wave is available in wave 17, then the spell that was censored was updated using the current information. For employment spells, the recorded end date (*BIO0400; BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was ongoing (*BIO0600*), the reason for the cancellation of a work contract (*ET2300*), the generated variables on occupational status and weekly working hours (*stib, az1, az2*) and the censoring indicator (*zensiert*) were overwritten with the information gathered in wave 17. Furthermore, the cross-sectional data referring to wave 17 (*ET0565* to *ET4208*) were included.
   For AL spells, the recorded end date (*BIO0400; BIO0500*), the generated end date (*emonat; ejahr*), the follow-up question as to whether the spell was ongoing (*BIO0600*),

[49] Thus, the reported end date remains completed with the interview date of the wave in which the spell was censored or the special code "0" for continuing spells. In addition, the question about whether the spell continued (for the case that the end date corresponds with the interview date) is not changed. The generated date variables continue to contain the last valid in-formation, which here is the interview date for the wave in which the spell was censored.

the reason for the end of unemployment (*AL0600, AL0601*) and the censoring indicator (*zensiert*) were overwritten with the information gathered in wave 17. Furthermore, the cross-sectional data referring to wave 17 (*AL1315*) were included. AL spell data, moreover, feature the exception that the spell of UB I (receipt of UB I) is recorded within an AL spell. Which information is updated depends on whether UB I was already received during this spell of unemployment and whether this benefit was ongoing during the previous wave.

If, in the previous wave, there was also an ongoing receipt of UB I in the AL spell to be updated, then the recorded end date of the receipt (*AL1000, AL1100*), the indicator as to whether the spell is ongoing (*AL1200*), the generated end date of the receipt (*alg1em, alg1ej*) and the censoring indicator of the receipt (*alg1akt*) were overwritten with the information obtained in wave 17.

If no UB I was received in previous waves in the AL spell to be updated, then the information on UB I receipt was overwritten with the information obtained in wave 17. In addition to the indicator as to whether UB I was received in the AL spell (*AL0700*), the reported start and end date (*AL0800, AL0900, AL1000, AL1100*), the indicator for ongoing receipt (*AL1200*) and the respective generated variables (*alg1bm, alg1bj, alg1em, alg1ej, alg1akt*) were replaced with the newly recorded information.

If there was UB I receipt in the AL spell to be updated in the past but that ended in the previous wave, no changes were made to these spells.

3. *Merging of newly reported spells: cases where people report new spells in wave 17, i.e. spells that do not update any spells that were censored in the previous wave.*
   Spells reported for the first time in wave 17 were added to the BIO-spell dataset. Next, the spell counter was generated anew to create a variable *spellnr* without gaps. Updating the spell datasets does not affect the spell numbers of the previous wave's SUF. Spells already included in the wave 16 SUF (*spellnret, spellnral, spellnrlu, spellnr*) maintain their spell number. The new spells from wave 17 are added to the respective dataset and the spell numbers are updated.

5.8 One-Euro job spell dataset (***ee_spells)***

In wave 4, the concept for surveying participation in employment and training measures was thoroughly revised. The MN spell dataset has been replaced by the One-Euro job spell dataset (*ee_spells*) as of wave 4. This was updated in wave 17. The reference date as of which to consider one-Euro jobs was January 2022 for wave 17. The reference date is adjusted by + 1 year in each wave and is always in January of the previous year to the current survey year. The questions on the one-euro job were asked for the last time in wave 17. The one-euro job spell dataset will therefore not be continued after wave 17.

5.8.1 Concept for updating the spells that were ongoing in the previous wave

Continuing *ee_spells* were updated in wave 17. To update the spells that were ongoing in the previous wave and were therefore right-censored in the spell dataset, dependent interviewing questions are included in the personal questionnaires.

5.8.2 Structure of the EE spell dataset

By integrating the One-Euro jobs (OEJ) reported in wave 17 in the OEJ spell dataset (*ee_spells*), new variables are added that refer to a specific wave. The following table gives an over-view of the cross-sectional information contained in the OEJ spell dataset.

Table 76: Cross-sectional variables in the EE spell dataset (*ee_spells*)

	Wave 4	Wave 5	...	Wave 12	...	Wave 17
Weekly working hours in the OEJ	*EE1100*	*EE1101*	...	*EE1108*	...	*EE1113*
Field of employment			... ...	*EE1600* *EE1600z*	... ...	*EE1605* *EE1605z*
OEJ is the same work permanent co-workers do	*EE1200*	*EE1201*	...	*EE1208*	...	*EE1213*
Which kind of training necessary for OEJ	*EE1300*	*EE1301*	...	*EE1308*	...	*EE1313*
Only work or also training/ classes?	*EE1400*	*EE1401*	...	*EE1408*	...	*EE1413*
Assessment OEJ	*EE1500a-* *EE1500h*	*EE1501a-* *EE1501h*	...	*EE1508a-* *EE1508h*	... ...	*EE1513a-* *EE1513h*

For the OEJ spell dataset, it must be considered that there are also spells if the OEJ was not performed, i.e., if there was no participation.

5.8.3 Plausibility checks and corrections in the EEJ spell dataset

The OEJ spell dataset on the participation in OEJ was both checked for plausibility and corrected. The plausibility checks contained checks for dates, for the reference date for the

newly integrated spells in wave 17 (January 2022) and for logical inconsistencies in cases of respondents with several OEJ spells.

Only the generated date variables (*bmonat, bjahr, emonat, ejahr*) were corrected and recoded. Details on seasons were recoded into months, furthermore "-8" values were assigned for implausible responses as well as date information was replaced or rendered plausible. Next, a spell counter *spellnr* was generated. The variable generation was performed analogously to the chronological counters in the BIO-spell datasets. Non-participating spells were not included in the sorting and thus kept their original position within the survey wave. Spells from wave 16 maintained their spell number for the wave 17 SUF.

6 Weighting Wave 17

The weighting concept for wave 17 generally follows the concepts developed in previous waves (see Berg et al., 2023). The starting point for the wave 17 weighting procedure and for the longitudinal section from wave 16 to wave 17 were the cross-sectional weights from wave 16 for households and individuals. The two weights for each household and two weights for each individual were updated. This chapter of the data report documents the technical details and exact models used to generate the weights for wave 17. An overview of the weighting concept used in PASS can be found in chapter 8 (Trappmann, 2013a) of the PASS User Guide (Bethmann, Fuchs, and Wurdack, 2013). Examples of how to use the weights can be found in Chapter 12 (Trappmann, 2013b).

6.1 Design weights for the panel replenishment (municipal register sample) in wave 17

The design weights for the panel replenishment (from the municipal registers) of the general population sample (sample=30) are defined as the reciprocal value of the selection probabilities at the different levels of the sampling design. The selection probabilities are determined via three selection stages. The selection probability of the point or zip code, the selection probability of the people in the point and the disproportionate selection of people by nationality and age for the deployment sample (see Sample section).

The multiplication of these three selection probabilities results in the overall selection probability of the selected person in the deployment sample. The transformation of the sample of individuals into a household sample represents an additional step in the replenishment of the population sample, which can only be carried out for the realized cases. This additional weighting step, which corrects the different selection probabilities due to the different (reduced) household size, was carried out after the calculation of the participation propensities, i.e. after the transition from the gross sample to the net sample, by multiplying the selection probability of the individuals by the estimated participation propensity and the number of target persons in the household.

A detailed description of the selection steps and the calculation of the selection probabilities for the structurally identical panel replenishment (from the municipal registers) in wave 5 can also be found in the data report of wave 5.

6.2 Integration of the design weights for the panel replenishment (from municipal registers (EWO)) using the existing weights of the population sample

The integration of the design weights for the panel replenishment (from the municipal register) using the existing weights of the general population sample (Microm, replenishment from the municipal registers, waves 5 and 11) was done as in previous waves after the propensity models but before calibration.

The weights of the combined population samples should project the Microm sample from wave 1, the replenishment based on the municipal registers from wave 5, the replenishment based on the municipal registers from wave 11 and the new municipal-register replenishment from wave 17 to all the households in Germany. Therefore, separate weights were first calculated for the general population sample and the replenishment from the municipal registers following the concept used in previous waves. Then the Microm sample plus the replenishment sample from the municipal registers from the waves 5 and 11 were integrated with the municipal-register replenishment from wave 17 (sample =30) via a convex combination to obtain the population weight before calibration

After that the population weights and the BA weights were integrated to create overall weights as was done in previous waves.

6.3 Design weights for the panel households in wave 17

New “household design weights” were generated for wave 17 from the cross-sectional weights for households of wave 16, taking into account people moving into households from within Germany. This step was performed by using the weight share procedure as described in wave 2 (see Gebhardt et al., 06/2009). Births, deaths or move-outs from households have no influence on weight; moves into households from within Germany, however, increase the inclusion probability of a household because the individuals who moved into the household also had the opportunity to be included in the sample in waves 1 to 16. The new design weight for subsample i $dwihh_{17}$ is therefore calculated from the old cross-sectional weight $wqihh_{16}$:

$$1/dw_ihh_{17} = 1/wq_ihh_{16} + (n_{sample_i}/n_{population_i})$$

The new design weight is only an intermediate step and therefore is not included in the data.

6.4 Design weights for the refreshment sample in wave 17

In wave 17 the panel was refreshed by sampling new households from new inflows to benefit receipt. All households that were receiving benefits in July 2022 but had no probability of being selected for the register data sample in the same month of the previous years 2006 to 2021 had a likelihood of being selected. This refreshment could be achieved by selecting only benefit units in which no member was receiving benefits in July of the previous years. The refreshment sample was drawn from the 300 points of the first wave and the 100 replenishment points of wave 5. Analogous to the special PPS procedure used to draw the first register data sample, which is described in Rudolph and Trappmann (2007), the sample size was proportional to the share of new benefit recipients in the population in the sampling point (at the time when the sampling points were selected). The calculation of the design weights is also described in the same article. For cases in *sample* = 31 (usual refreshment sample), *sample* = 32 (refreshment sample Syrians/Iraqis) and *sample* = 33 (refreshment sample - Ukrainians), the design weight of the refreshment sample for wave 17 is included in the variable *dw_ba*.

6.5 Propensity to participate again - households

In this step, again similar to the procedure in wave 16, the probability of re-participation in wave 17 was estimated for each household that participated in wave 16 based on logit models for willingness to participate in the panel, availability and participation. Additionally, households that participated in wave 15 but not in wave 16 (temporary nonresponses) were considered in the modeling for wave 17. In addition to variables from the household and personal interviews with the head of the household conducted during the previous wave, other fieldwork variables were included, e.g., number of contact attempts. The estimated propensities of all three models were multiplied. The reciprocal value of this product can be found in the variable *hpbleib* for each wave. The longitudinal weight for a household from one of the samples of wave 1 for the total period possible [t1 to t17] across all seventeen waves can be obtained as the product of the cross-sectional weight to t1, *hpbleib* (wave 1 to wave 2) and hpbleib (wave 2 to wave 3, etc.) (see also the PASS User Guide section 12 (Trappmann, 2013b)).

→ **Table A9** gives an overview of the variables, codes and reference categories for the logit models of re-participating households.

The logit models on re-participation for willingness to participate in a panel, availability and participation are shown in → **Table A10**.

6.6 Propensity to participate - first-time interviewed split-off households

This step calculated the propensities to participate for new split-off households, i.e., households that are included in the panel due to the relocation of one individual of the panel sample in a new household. Here, only split-off households that had not been interviewed in the previous waves were considered. This condition means that the participation propensities for first-time participating split-off households were modeled separately following the criterion of originating in wave 16 (split-off W16 households) and split-off household created in wave 17 (split-off W17 households). The probability of re-participation was estimated via logit models. For the split-off households no separate modelling of the probability of accessibility and participation was carried out, as accessibility was available for almost all households. Only one overall model was estimated for all households. Missing time-stable information on the household reference person (HRP) was added from the previous wave when necessary. The reciprocal value of the probability of participation for the split-off households can also be found in the variable *hpbleib*.

→ **Table A11** gives an overview of the variables, codes and reference categories for the logit models of the split-off households participating for the first time (waves 16 and 17).

The logit models on the first participation of split-off wave 16 households are shown in → **Table A12**.

The logit models for the first participation of split-off wave 17 households are shown in → **Table A13**.

6.7 Nonresponse weighting for households from the refreshment sample of BA new inflows wave 17

A nonresponse modelling for the households from the refreshment sample of BA new inflows into UB II receipt in wave 17 (*sample* = 31, normal sample, *sample* = 32, refreshment sample Syrians / Iraqis and *sample* = 33, refreshment sample Ukrainians) was performed (participation) similar to the wave 16 refreshment sample, each for accessibility and participation. An integrated model was estimated for the three subsamples, the variable (*samaufftyp* indicates the affiliation to the subsamples. The participation probability derived from the model can be found in the variable *prop_t0*.

→ **Table A14** gives an overview of the variables, codes and reference categories for the logit models of the BA refreshment sample of wave 17.

The logit models on the first participation for availability and participation of the BA refreshment sample and BA replenishment sample of wave 17 are shown in → **Table A15**.

6.8 Nonresponse weighting for households from the panel replenishment (municipal register sample) wave 17

A two-step nonresponse modelling (availability and participation) was performed for the panel replenishment (municipal register sample) of the general population (sample==30). The participation probability derived from the model can be found in variable *prop_t0*.

→ **Table A16** gives an overview of the variables, codes and reference categories for the logit models of the BA refreshment sample of wave 17.

The logit models on the first participation for availability and participation of the BA refreshment sample and BA replenishment sample of wave 17 are shown in → **Table A17**.

6.9 Propensity to participate again - individuals

The decisive longitudinal weight is not the household but the individual-level weight because these units are stable over time. The propensities to participate again for individuals in wave 17 were estimated using additional personal characteristics via logit models for willingness to participate in the panel, availability and participation. The dependence of the personal sample conveyed via the household context and correction of the estimation of standard errors made necessary by it were considered in these models by clustering the error terms at the household level. The predicted propensities of the models were multiplied. The reciprocal value of this product can be found in variable *ppbleib*. The longitudinal weight for an individual for the period [t1 to t17] across all seventeen waves can be obtained as the product of the cross-sectional weight to t1, *ppbleib* (wave 1 to wave 2) and *ppbleib* (wave 2 to wave 3, etc.).

→ **Table A18** gives an overview of the variables, codes and reference categories for the logit models of re-participating individuals.

The logit models on re-participation for willingness to participate in a panel, availability and participation are shown in → **Table A19**.

6.10 Integration of the weights to yield the total weight before calibration

This step again involved combining the household weights of the new replenishment of BA new inflows in wave 17 and panel household samples (including the refreshments from waves 2 to 16) that were modified by the nonresponse modeling. The multiple selection probability of a sampled benefit recipient living in the same household as a benefit recipient in previous years without being a member of the benefit unit himself/herself was ignored. The new design weights of the benefit recipient sample are projected in the cross-section to all individuals who were living in a household that included at least one benefit unit in July of each of the years 2006 to 2022. It is only when calculating new weights for the total sample that it becomes necessary to adjust the weights for all households receiving benefits in July 2022. For this adjustment, the inclusion probability in the other sample was estimated for cases from the Microm sample (wave 1), EWO replenishment samples (wave 5, wave 11 and wave 17) and new refreshment sample of BA new inflows (wave 17). For cases from the refreshment sample of BA new inflows, the mean wave 1 selection probability in the Microm sample, the wave-specific mean selection probability of the EWO refreshments in the respective postcode area and the average participation probability (for waves 1 to 17) in that sample were assumed. For cases from the Microm or the EWO refreshment samples, if they are (according to survey data) new recipients of UB II who first received the benefit, the mean selection probability of a household in the refreshment sample (BA new inflows wave 17) in the respective postcode area and the average participation probability in that sample were assumed. The two weights were then integrated to form a new total weight.

6.11 Integration of temporary non-responses (households)

Households that skipped one wave - i.e., did not participate (temporary nonresponses) - could participate again in wave 17, as was possible in previous waves. No longitudinal weights are calculated for these households, i.e., (weighted) longitudinal evaluations can only be made with participants across all waves in question. Non-participation of a household can only occur in one wave; if a household skips two consecutive waves, it will no longer be contacted. To calculate mutual cross-sectional weights including the temporary nonresponses, there was a convex combination of the modified household weights of the temporary non-responses and the modified household weights of the panel household

sample (not of the refreshment sample) before calibration. Thus, the convex combination of the household weights was made before calibration; the calibration was then made with the new combined household weights.

Although the household weights modified by nonresponse modeling already serve as projection factors for the panel and refreshment sample, it was necessary to calculate such modified household weights as an estimator for the respective population again for the temporary nonresponses. The starting point was the calibrated household weights of wave 15 (wave 16 is the temporary non-response).

For temporary nonresponses, the probability of non-participation in wave 16 in case of participation in wave 15 (non-participation propensities wave 16) and the probability of participation in wave 17 in case of a non-participation in wave 16 (participation propensities wave 17) was determined. The probability of non-participation in wave 16 is calculated from 1–participation probability in wave 16.

The described propensities for participation and non-participation were estimated via logit models. The estimated probabilities of the respective models were multiplied. The modified household weight of the temporary nonresponses was then calculated by multiplying the calibrated household weights of wave 15 by the reciprocal value of this product.

→ **Table A20** gives an overview of the variables, codes and reference categories for the logit models of the temporary nonresponses.

The logit models of temporary nonresponses are shown in → **Table A21**.

The convex combination of the weights of the participants across all waves (panel household sample) and the temporary nonresponses was made for the weights of all three sub-samples i (Microm, BA and total) by multiplying the respective modified household weights by the share of the panel household sample or the temporary nonresponses from the total sample, i.e., the sum of the panel household sample and temporary nonresponses:

$dw_{ihh_{temp.Ausfall}} * (n_{temp.Ausfall_i}/(n_{temp.Ausfall_i} + n_{Bestand_i}))$ for temporary nonresponses and

$dw_{ihh_{Bestand}} * (n_{Bestand_i}/(n_{temp.Ausfall_i} + n_{Bestand_i}))$ for the panel household sample.

6.12 Calibration to the household weight, wave 17, cross-section

Another calibration of the modified design weights, including the non-response weighting at the household level using the GREG procedure to the benchmark values of the Federal Statistical Office for 2022, followed. For households receiving benefits the weights were adjusted to the statistics of the Federal Employment Agency for July 2022. As in the previous year, the increase in UB II receipt since the previous year at the level of benefit units (407,765) was also included as an additional benchmark value in the total sample. Cases in the previous samples from waves 1 to 17 that, according to wave 17 of the survey, were receiving UB II in July 2022, will be projected to the benchmark statistics of the Federal Employment Agency on UB II.

The main objective of weighting is to balance distortions arising from the sample design (with different selection probabilities) and through selective participation or non-participation. By using the weights, population values from the sample can be estimated in an unbiased way. If the weights show a high variance, a large variance of the estimation functions can result. This is the trade-off between bias and variance so typical for statistics. The weighting reduces the bias; however, a too-severe increase in the variance caused by weighting is also to be avoided. Therefore, attempts are made to avoid very large weighting factors (and subsequently, very small factors) whenever possible and to make appropriate corrections to the weights if necessary. Within the framework of the calibration at hand, these corrections are made at two points:

- The input weights for the calibration (the modified design weights after considering non-response analyses) were trimmed before calibration, i.e., they were replaced by new input weights. The maximum and minimum of the trimmed design weights were determined by using particular percentiles of the distribution depending on the distribution of the design weights.
- In addition, the interval of weights was limited during calibration, i.e., a maximum and a minimum limit for weights was determined. Here, the total width of the weights was determined; the range of the pure calibration weights can be calculated from the relation of original weights to the trimmed input weight. Notably, narrower limits for the weights result in less variance of the weights and thus less variance of the estimations; too-narrow limits can, however, make the calibration of all benchmark values impossible.

To evaluate the weights, in addition to the average value and the standard deviation, the efficiency measure (E) is described as follows. The efficiency measure E is based on the variance of the weighting factor. The efficiency measure indicates the size of the effective

case number of a passive characteristic that does not correlate with active characteristics when using the weight. The effective case number is the number of respondents who would have produced the same sample error in an unlimited random sample given the variance of the characteristic in the sample. The efficiency measure expresses the relation of n to n‘ as percentage.

6.13 Calibration of the BA sample

The population of the cumulated BA sample of all seventeen waves consists of all of the households in Germany with at least one benefit unit receiving benefits according to SGB II at one of the (until now) seventeen drawing dates (in July of each of the years 2006 to 2022). In wave 17, only the benchmark values of the BA statistics from July 2022 are calibrated. The calibration thus only influences the weights of the households from the BA sample in which at least one benefit unit receiving benefits according to SGB II was living in July 2022. The starting points for the calibration were modified design weights, including the nonresponse weighting. The modified design weights were trimmed at the 10% percentile and 90% percentile of their distribution and then rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 194.5627 to 5124.039. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.0. Thus, the total projection factors after calibration lie between a minimum of 19.45627 and a maximum of 10419.92.

A calibration was made for the following characteristics:

Benefit unit basis BA statistics:

- Number of BCs receiving benefits according to SGB II by federal states
- Number of BCs receiving benefits according to SGB II by number of individuals under 65 years of age in the benefit unit and by west/east
- Number of BCs receiving benefits according to SGB II by number of children under 15 years of age in the benefit unit and by west/east
- Number of BCs receiving benefits according to SGB II consisting of a single parent with child(ren), by west/east

As in the previous year, an additional benchmark was included. This is the increase in UB II recipients since the previous year at the level of benefit units (407,765).

For the calibration, the benchmark variable for each household must have a valid value.

Therefore, the very low nonresponse item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the nonresponse item thus leads to slight deviations from the values as presented in the table appendix.

The nominal distributions and distributions after calibration (BA sample, households) are shown in → **Table A22**.

The characteristics of the distribution of weights (BA sample, households) are shown in → **Table A23**.

6.14 Calibration of the population sample

All private households in Germany form the population. The starting points for the calibration were modified design weights, including the nonresponse weighting. The modified de-sign weights were trimmed at the 10% percentile and 90% percentile of their distribution and after that rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 3743.43 to 30705.29. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.5 and upwards to 3.0. Thus, the total projection factors after calibration lie between minimal 1976.897 and maximal 56408.25.

A calibration was made for the following characteristics:

1. Households based on Mikrozensus 2022:
    - Number of households by federal state and BIK type
    - Number of households by household size and west/east
    - Number of households by “children under 15 years of age in the household yes/no” and west/east

For the calibration, each benchmark variable for each household must have a valid value. Therefore, the very low nonresponse item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated

weights without considering the nonresponse item thus leads to slight deviations from the values as presented in the table appendix.

The nominal distributions and distributions after calibration (population sample, households) are shown in → **Table A24**.

The characteristics of the distribution of weights (population sample, households) are shown in → **Table A25**.

6.15 Calibration of the total sample

All of the private households in Germany form the population. The starting points for the calibration were modified design weights, including the non-response weighting. The modified design weights were trimmed at the 10% percentile and 90% percentile of their distribution and after that rescaled so that they totaled the untrimmed design weights. The projection factors of the trimmed design weights range from 212.829 to 19175.63. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.0. Thus, the total projection factors after calibration lie between min. 21.2829 and max. 31924.15.

A calibration was made for the following characteristics:

1. Benefit unit basis BA statistics:
    - Number of BCs receiving benefits according to SGB II by federal states
    - Number of BCs receiving benefits according to SGB II by number of individuals under 65 years of age in the benefit unit and by west/east
    - Number of BCs receiving benefits according to SGB II by number of children under 15 years of age in the benefit unit and by west/east
    - Number of BCs receiving benefits according to SGB II consisting of a single parent with child(ren), by west/east
2. Household basis Mikrozensus 2022:
    - Number of households by federal state and BIK type
    - Number of households by household size and west/east
    - Number of households by “children under 15 years of age in the household yes/no” and west/east

In addition, the increase in UB II recipients since the previous year at the level of benefit units (407,765) was included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each household must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item thus leads to slight deviations from the values in the table appendix.

The nominal distributions and distributions after calibration (total sample, households) are shown in → **Table A26**.

The characteristics of the distribution of weights (total sample, households) are shown in → **Table A27**.

6.16 Calibration of the person weight, wave 17, cross-section

As in previous waves, the person weights were calibrated under the restriction that they differ as little as possible from the calibrated household weights. The calibrated household weights were quasi-inherited by the individual household members. These input weights were calibrated at the individual level.

As in the previous year, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years (550,617) was also included as an additional benchmark value in the total sample. Again, those cases in the previous samples from all waves of the survey who, according to wave 17 of the survey, were receiving UB II in July 2022 are projected to the benchmark statistics of the Federal Employment Agency on receipt of UB II.

Before calibration, the calibrated household weights that formed the input weight were also trimmed. For the calibration of person weights, the range of weights was determined to a certain interval.

6.17 BA sample

The population of the cumulated BA sample of all seventeen waves consists of all individuals aged 15 and over who are living in a household in which there was at least one

benefit unit receiving benefits according to SGB II at one of the (until now) seventeen drawing dates (in July of each of the years 2006 to 2022). Only those individuals aged 15 and over who were living in a benefit unit that received benefits according to SGB II in July 2022 were considered for calibration. Individuals living in a household that did not receive benefits and individuals living in a household with at least one benefit unit according to SGB II but who were not part of a benefit unit themselves were removed from the dataset for the calibration. The weighting of these individuals was calculated in a different way (see below).

The starting point for the calibration is the calibrated household weights of the BA sample. They were trimmed at the 10% percentile and 90% percentile of their distribution and then re-scaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors range from 119.7777 to 3864.634. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.2 and upwards to 4.0. Thus, the total projection factors after calibration lie between a minimum of 31.33221 and a maximum of 17052.99.

A calibration was made for the following characteristics:

Benefit recipients basis BA statistics:

- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by federal states
- Number of individuals in benefit units receiving benefits according to SGB II, by age (15-24 and 25-64)
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II by sex and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by “single parent yes/no” and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by nationality (German/non-German)

As in the previous year, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years (550,617) was included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation only serves the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the nonresponse item thus leads to slight deviations from the values in the table appendix.

The nominal distributions and distributions after calibration (BA sample, individuals) are shown in → **Table A28**.

The characteristics of the distribution of weights (BA sample, individuals) are shown in → **Table A29**.

6.18 Population sample

All individuals over 14 years of age in private households in Germany form the basic population. The starting points for the calibration were calibrated household weights of the population sample. These weights were trimmed at the 10% percentile and 90% percentile of their distribution and after that rescaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors lie between a minimum of 3697.123 to a maximum of 26708.72. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.5. Thus, the total projection factors after calibration lie between a minimum of 492.9775 and a maximum of 123316.2.
A calibration was made for the following characteristics:

1. Population based on Mikrozensus 2022:
    - Number of individuals aged 15 and over in private households by federal state
    - Number of individuals aged 15 and over in private households, by age, sex and west/east region
    - Number of individuals aged 15 and over in private households, by household size and west/east region
    - Number of individuals aged 15 and over in private households, by academic qualifications and west/east region
    - Number of individuals aged 15 and over in private households, by marital status and west/east region
    - Number of individuals aged 15 and over in private households, by nationality
2. Population based on BA statistics:
    - Number of unemployed individuals including participants in measures, by west/east region
    - Number of employees subject to social security, by west/east region

The source for the benchmark value of employment status was the BA statistics because the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low nonresponse item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation only serves the feasibility of the calibration, the imputed values were set to missing values after the calibration. A projection with the calibrated weights without considering the nonresponse item therefore leads to slight deviations from the values in the table appendix.

The nominal distributions and distributions after calibration (population sample, individuals) are shown in → **Table A30**.

The characteristics of the distribution of weights (population sample, individuals) are shown in → **Table A31**.

6.19 Total sample

All individuals aged 15 and over in private households in Germany form the population. The starting point for the calibration was the calibrated household weight of the total sample. That weight was trimmed at the 10% percentile and 90% percentile of their distribution and then rescaled so that they totaled the untrimmed calibrated household weights. The trimmed projection factors range from 285.9404 to 23184.16. The relation between the total projection factors after calibration and the trimmed design weights was limited downwards to 0.1 and upwards to 3.5. Thus, the total projection factors after calibration lie between a minimum of 28.59404 and a maximum of 79577.31.
A calibration was made for the following characteristics:

1. Benefit recipients basis BA statistics:
    - Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by federal states
    - Number of individuals in benefit units receiving benefits according to SGB II, by age (15-24 and 25-64)
    - Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by sex and by west/east

- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by “single parent yes/no” and by west/east
- Number of individuals aged 15 and over in benefit units receiving benefits according to SGB II, by nationality (German/non-German)

2. Population based on Mikrozensus 2022:

   - Number of individuals aged 15 and over in private households, by federal state
   - Number of individuals aged 15 and over in private households, by age, sex and west/east
   - Number of individuals aged 15 and over in private households, by household size and west/east
   - Number of individuals aged 15 and over in private households, by academic qualifications and west/east
   - Number of individuals aged 15 and over in private households, by marital status and west/east
   - Number of individuals aged 15 and over in private households, by nationality

3. Population based on BA statistics:

   - Number of unemployed individuals including participants in measures, by west/east
   - Number of employees subject to social security, by west/east

The source for the benchmark value of employment status was the BA statistics because the definition of unemployment and employment subject to social insurance in PASS does not correspond to the ILO concept.

In addition, the increase in UB II recipients since the previous year at the level of individuals between 15 and 64 years of age (550,617) was included as an additional benchmark value in the total sample.

For the calibration, each benchmark variable for each individual must have a valid value. Therefore, the very low non-response item was imputed before calibration. The imputation was made by means of the average value and the modal value of the respective variable.

Because the imputation is only required for the feasibility of the calibration, the imputed values were set back to missing values after the calibration. A projection with the calibrated weights without considering the non-response item therefore leads to slight deviations from the values in the table appendix.

The nominal distributions and distributions after calibration (total sample, individuals) are shown in → **Table A32**.

The characteristics of the distribution of weights (total sample, individuals) are shown in → **Table A33**.

6.20 Estimating the BA cross-sectional weights for households and individuals not in receipt of Unemployment Benefit II

Finally, in wave 17, some households and individuals remained that could not be assigned a BA cross-sectional household weight or a BA cross-sectional person weight by means of calibration. The number of these households is larger again in wave 17 than in the previous waves because a larger part of the BA sample of waves 1 to 16 has withdrawn from benefits. These are the following three groups that were not receiving benefits in July 2022 but that belong to the population of the BA sample (households or individuals in households receiving UB II in July 2006, July 2007, July 2008, July 2009, July 2010, July 2011, July 2012, July 2013, July 2014, July 2015, July 2016, July 2017, July 2018, July 2019, July 2020, July 2021 or July 2022).

- From the refreshment sample: Individuals in the household who are not members of a benefit unit: Here, the person weight was obtained from the BA household weight in wave 17 after calibration (*wqbahh*) by dividing it by the proportion of these individuals who gave a personal or senior citizen interview - provided that their household was participating.
- Panel households in which nobody continued to receive UB II in July 2022: The house-hold retains the BA weight before calibration. Individuals in households with interviews in both waves were assigned a new BA person weight, which is obtained by multiplying their old BA person weight by the reciprocal re-participation probability *ppbleib*. Individuals in these households who did not provide a personal interview in wave 16 are as-signed a new BA person weight calculated by dividing the BA household weight of their household for wave 17 by the proportion of such individuals who participate if their household is taking part.
- Individuals who are not members of a benefit unit in panel households that continued to receive UB II in July 2022: Individuals in these households with interviews in both waves were assigned a new BA person weight, which is obtained by multiplying their BA person weight from the previous wave by the reciprocal re-participation probability *ppbleib*.
- The individuals and households were also adjusted to a benchmark figure for the individuals or benefit units that did not continue to receive UB II. The exact population of this group is unknown but can be approximated from the total of all cumulated BA subsamples minus the individuals or benefit units currently receiving benefits. This benchmark was reduced additionally by the estimated cumulative death rate of this

group of people for the time period 2005 to 2022 by 3.4%. The number of individuals who are no longer receiving UB II at wave 17 is 8,304,924. The number of benefit units that are no longer receiving UB II is 6,203,092.

7 Appendix: Brief description of the dataset

A brief description of the dataset can be found in the → **table appendix**.

Literature

AAPOR (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7. Auflage. Lanexa: AAPOR.

Achatz, J., Hirseland, A. & Promberger, M. (2007). Rahmenkonzept für das IAB-Panel „Arbeitsmarkt und Soziale Sicherung". In M. Promberger (Hrsg.), Neue Daten für die Sozialstaatsfor-schung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung", *IAB-Forschungsbericht 12/2007* (S. 11-32), Nürnberg.

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., & Wagner, G. G. (2007). Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Schmollers Jahrbuch*, 127(1), 171-182.

Andreß H.-J., Burkatzki, E., Lipsmeier, G., Salentin, K., Schulte, K. & Strengmann-Kuhn, W. (1996). *Leben in Armut. Analysen der Verhaltensweisen armer Haushalte mit Umfragedaten*. Endbericht des DFG-Projekts „Versorgungsstrategien privater Haushalte im unteren Einkommensbereich (VuE)". Bielefeld.

Andreß, H.-J. & Lipsmeier, G. (1995). Was gehört zum notwendigen Lebensstandard und wer kann ihn sich leisten? Ein neues Konzept zur Armutsmessung. *Aus Politik und Zeitgeschichte* 31-32, S. 35-49.

Andreß, H.-J. & Lipsmeier, G. (2001). *Armut und Lebensstandard. Gutachten im Rahmen des Armuts- und Reichtumsberichts der Bundesregierung*. BMAS. Bonn.

Beckmann, P. & Trometer, R. (1991). Neue Dienstleistungen des ALLBUS: Haushalts- und Familientypologien, Klassenschema nach Goldthorpe. *ZUMA-Nachrichten, 28*, S. 7-17.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Bethmann, A., Fuchs, B., Gebhardt, D.(2011). Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung" (PASS). Bd. 1: Datenreport Welle 4. *FDZ Datenreport 08/2011*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Bethmann, A., Fuchs, B., Trappmann, M., Wurdack, A. (2012). Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung" (PASS). Bd. 1: Datenreport Welle 5. *FDZ-Datenreport 06/2012*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, Beste, J., Dummert, S., Frodermann, C., Fuchs, B., Schwarz, S., Trappmann, M. & Trenkle, S. (2017). Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung" (PASS) Bd. 1: Datenreport Welle 10. *FDZ-Datenreport 07/2017*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, Beste, J., Dummert, S., Frodermann, C., Schwarz, S., Trappmann, M., Bähr, S., Coban, M., Friedrich, M., Gundert, S., Müller, B. Teichler, N., Unger, S. & Wenzig, C. (2019). Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) Bd. 1: Datenreport Welle 12. *FDZ-Datenreport 09/2019*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, Beste, J., Dummert, S., Frodermann, C., Schwarz, S., Trappmann, M., Bähr, S., Coban, M., Friedrich, M., Gundert, S., Müller, B., Teichler, N., Unger, S. & Wenzig, C. (2020). Codebuch und Dokumentation des „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) Bd. 1: Datenreport Welle 13. *FDZ-Datenreport 12/2020*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Wenzig, C., Trappmann, M., Bähr, S., Collischon, M., Coban, M., Friedrich, M., Gleiser, P., Gundert, S., Malich, S., Müller, B., Schwarz, S., Stegmaier, J., Teichler, N., & Unger, S. (2021): Codebuch und Dokumentation des Panel „Arbeitsmarkt und soziale Sicherung„ (PASS) Bd. 1: Datenreport Welle 14. *FDZ-Datenreport 14/2021*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Schwarz, S., Wenzig, C., Trappmann, M., Altschul, S., Bähr, S., Collischon, M., Coban, M., Gleiser, P., Gundert, S., Küfner, B., Mackeben, J., Malich, S., Müller, B., Stegmaier, J., Teichler, N., & Unger, S. (2022): Codebuch und Dokumentation des Panel „Arbeitsmarkt und soziale Sicherung„ (PASS) Bd. 1:. Datenreport Welle 15. *FDZ-Datenreport 10/2022*. Nürnberg.

Berg, M., Cramer, R., Dickmann, C., Gerber, V., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Malich, S., Schwarz, S., Wenzig, C., Trappmann, M., Bähr, S., Bömmel, N., Coban, M., Collischon, M., Gundert, S., Küfner, B., Mackeben, J., Müller, B., Stegmaier, J., Teichler, N. & Wunder, A. (2023): Codebuch und Dokumentation des Panel „Arbeitsmarkt und soziale Sicherung„ (PASS) Bd. 1 Datenreport Welle 16. *FDZ-Datenreport 10/2023*. Nürnberg.

Beste, J., Olbrich, L., Schwanhäuser, S. (2021). Interviewer:innenkontrolle im Panel Arbeitsmarkt und soziale Sicherung (PASS). FDZ-Methodenreport 04/2021. Nürnberg.

Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide „Panel Labour Market and Social Security“ (PASS). Wave 6. *FDZ Datenreport 07/2013*. Nürnberg.

Bonin, H. & Rinne, U. (2017). Omnibusbefragung zur Verbesserung der Datenlage neuer Beschäftigungsformen. Kurzexpertise im Auftrag des Bundesministeriums für Arbeit und Soziales. IZA Research Report. Bonn. 80.

Brauns, H. & Steinmann, S. (1999). Educational Reform in France, West-Germany and the United Kingdom. Updating the Casmin Classification. *ZUMA-Nachrichten, 44*. S. 7-45.

Bruckmeier, K., Müller, G. & Riphahn, R. (2014). Who misreports welfare receipt in surveys? *Applied Economics Letters, Vol. 21, No. 12*, S. 812-816.

Bruckmeier, K., Müller, G. & Riphahn, R. (2015). Survey misreporting of welfare receipt - respondent, interviewer, and interview characteristics. *Economics Letters, Vol. 129*, S. 103-107.

Bruckmeier, K., Hohmeyer, K., Schwarz, S. (2018): Welfare receipt misreporting in survey data and its consequences for state dependence estimates: new insights from linked administrative and survey data. *Journal for Labour Market Research, Vol. 52, No. 1*, 21 S.

Bundesagentur für Arbeit (2011). Klassifikation der Berufe 2010. Nürnberg.

Bundesministerium für Bildung und Forschung [BMBF] (2003). Berufsausbildung sichtbar gemacht. Schaubildsammlung. 4. Auflage. Bonn: BMBF.

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R. & Tschersich, N. (2009). IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 2 (2007/2008). Methoden- und Feldbericht. *FDZ- Methodenreport 08/2009*. Nürnberg.

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R. & Tschersich, N. (2010). IAB-Haushaltspanel im Niedrigeinkommensbereich Welle 3 (2008/2009). Methoden- und Feldbericht. *FDZ- Methodenreport 10/2010*. Nürnberg.

Christoph, B. (2005). Zur Messung des Berufsprestiges: Aktualisierung der Magnitude-Prestige-skala auf die Berufsklassifikation ISCO88. *ZUMA-Nachrichten, 57*. S. 79-127.

Current Population Survey staff (2018). Electronically mediated work: New questions in the Contingent Worker Supplement. Monthly Labor Review, 1-32. Retrieved from https://www.jstor.org/stable/26502639.

Dickmann, C., Gerber, V., Jesske, B., Beste, J., Dummert, S., Frodermann, C., Malich, S., Mühlbacher, V., Schwarz, S., Wenzig, C., Trappmann, T., Zins, S., Bähr, S., Collischon, M., Gundert, S., Küfner, B., Mackeben, J., Müller, B., Stegmaier, J., Teichler, N. & Wunder, A. (2023): Codebuch und Dokumentation des Websurveys im Panel „Panel Labour Market and Social Security“ (PASS). Datenreport Websurvey 2022. *FDZ-Datenreport 13/2023*. Nürnberg.

Eggs, J. (2016). Patterns and impact of longitudinal measurement error for welfare receipt. Bielefeld: Bertelsmann 136 S. Reihe / Serie: IAB-Bibliothek Nr. 362.

Europäische Gemeinschaften [EG] (2002). Verordnung (EG) Nr. 29/2002 der Kommission vom 19. Dezember 2001 zur Änderung der Verordnung (EWG) Nr. 3037/90 des Rates betreffend die statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft. Amtsblatt der Europäischen Gemeinschaften L6/3-L6-33. Brüssel.

Europäische Gemeinschaften [EG] (2006). Verordnung (EG) Nr. 1893/2006 des europäischen Parlaments und des Rates vom 20. Dezember 2006 zur Aufstellung der statistischen Systematik der Wirtschaftszweige NACE Revision 2 und zur Änderung der Verordnung (EWG) Nr. 3037/90 des Rates sowie einiger Verordnungen der EG über bestimmte Bereiche der Statistik. Amtsblatt der Europäischen Gemeinschaften L393/1-L393-39. Brüssel.

Erikson, R. & Goldthorpe, J. (1992). *The Constant Flux. A Study of Class Mobility in Industrial Society*. Oxford: Clarendon Press.

Erikson, R., Goldthorpe, J. & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western Societies: England, France and Sweden. *British Journal of Sociology, 30*, S. 415-441.

Erikson, R., Goldthorpe, J. & Portocarero, L. (1982). Social Fluidity in Industrial Nations: England, France and Sweden. *British Journal of Sociology, 33*, S. 1-34.

Fischer, A. & Wirth, H. (2007): *Constructing Version 4 of ESEC Classes from 3-digit ISCO (Stata-do file)*. Mannheim: Gesis-ZUMA.

Frick, J., Göbel, J. & Krause, P. (o.J.). $HGEN: Generated Household Level Variables. [http://www.diw.de/documents/dokumentenarchiv/17/60053/hgen.pdf (8.11.2007)].

Fuchs, B. (2013). Structure of the scientific use file and its datasets. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide ”Panel Labour Market and Social Security” (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 27-47). Nürnberg.

Ganzeboom, H. & Treiman, D. (1996). Internationally Comparable Measures for Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research, 25*, S. 201-239.

Ganzeboom, H. & Treiman, D. (2003). Three Internationally Standardised Measures for Comparative Research on Occupational Status. In H. Jürgen, P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in Cross-National Comparison. A European Working Book for Demographic*

and Socio-Economic Variables (S. 159-193), New York: Kluwer Academic / Plenum Publishers.

Ganzeboom, H. (2010). A new International Socio-Economic Index (ISEI) of Occupational Status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with Data from the ISSP 2002-2007; With an Analysis of Quality of Occupational Measurement in ISSP. Paper presented at Annual Conference of International Social Survey Programme, Lisbon, May 1 2010. [http://www.harryganzeboom.nl/pdf/2010-ganzeboom-isei08-issp-lisbon-(paper).pdf (01.05.2010)].

Ganzeboom, H. & Treiman, D. (2010). Occupational Status Measures for the new International Standard Classification of Occupations ISCO-08; with a Discussion of the new Classification [www.harryganzeboom.nl/isol/isol2010c2-ganzeboom.pdf (24.05.2010)].

Ganzeboom, H. & Treiman, D. (2011). International Stratification and Mobility File: Conversion Tools [http://www.harryganzeboom.nl/ismf/index.htm (ohne Datum)].

Ganzeboom, H., De Graaf, P. & Treiman, D. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research, 21*, S. 1-56.

Gebhardt, D., Müller, G., Bethmann, A., Trappmann, M., Christoph, B., Gayer, C., Müller, B., Tisch, A., Siflinger, B., Kiesl, H., Huyer-May, B., Achatz, J., Wenzig, C., Rudolph, H., Graf, T. & Biedermann, A. (2009). Codebuch und Dokumentation des „Panel Arbeitsmarkt und Soziale Sicherung“ (PASS). Datenreport Welle 2 (2007/2008). *FDZ Datenreport 06/2009*. Nürnberg.

Granato, N. (2000). Mikrodaten-Tools: CASMIN-Bildungsklassifikation. Eine Umsetzung mit dem Mikrozensus 1996. *ZUMA-Technischer Bericht 2000/12*. Mannheim.

Hagenaars, A., de Vos, K. & Zaidi, M. (1994). *Poverty Statistics in the Late 1980s: Research Based on Micro-data*. Luxembourg: Office for Official Publications of the European Communities.

Halleröd, B. (1995). The Truly Poor: Direct and Indirect Consensual Measurement of Poverty in Sweden. *Journal of European Social Policy, 5*, S. 111-129.

Harrison, E. & Rose, R. (2006). ESeC User Guide, Appendix 6 (SPSS-Syntax: Esec Full) [http://www.iser.essex.ac.uk/esec/guide/docs/Appendix6.sps (31.11.2007)].

Hartmann, J., Brink, K., Jäckle, R. & Tschersich, N. (2008). IAB-Haushaltspanel im Niedrigeinkommensbereich. Methoden- und Feldbericht. *FDZ Methodenreport 07/2008*. Nürnberg.

Helberger, C. (1988). Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie. In H.-J. Bodenhöfer (Hrsg.), *Bildung, Beruf, Arbeitsmarkt* (S. 151-170), Berlin: Duncker & Humblot.

International Labour Office [ILO] (1990). *International Standard Classification of Occupations. ISCO-88*. Geneva: International Labour Office.

International Labour Office [ILO] (2012). *International Standard Classification of Occupations. ISCO-08*. Geneva: International Labour Office.

Jäckle, A. (2008). The Causes of Seam Effects in Panel Surveys. *ISEP Working Paper Series 2008-14*. Essex.

Jesske, B. & Quandt, S. (2011). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 4. Erhebungswelle – 2010 (Haupterhebung). *FDZ-Methodenreport 08/2011*. Nürnberg.

Jesske, B. & Schulz, S. (2012). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 5. Erhebungswelle – 2011 (Haupterhebung), *FDZ Methodenreport 11/2012*. Nürnberg.

Jesske, B. & Schulz, S. (2013). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 6. Erhebungswelle – 2012 (Haupterhebung), *FDZ Methodenreport 10/2013*. Nürnberg.

Jesske, B. & Schulz, S. (2014). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 7. Erhebungswelle – 2013 (Haupterhebung), *FDZ Methodenreport 11/2014*. Nürnberg.

Jesske, B. & Schulz, S. (2015). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 8. Erhebungswelle – 2014 (Haupterhebung), *FDZ Methodenreport 11/2015*. Nürnberg.

Jesske, B., Knerr, P. & Schulz, S. (2016). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 9. Erhebungswelle – 2015 (Haupterhebung), *FDZ Methodenreport 04/2016*. Nürnberg.

Jesske, B., Knerr, P. & Kraft, L. (2017). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 10. Erhebungswelle – 2016 (Haupterhebung), *FDZ Methodenreport 09/2017*. Nürnberg.

Jesske, B. & Schulz, S. (2018). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 11. Erhebungswelle – 2017 (Haupterhebung), *FDZ Methodenreport 13/2018*. Nürnberg.

Jesske, B., Schulz, S., Quandt, S. (2019). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 12. Erhebungswelle – 2018 (Haupterhebung), *FDZ Methodenreport 05/2019*. Nürnberg.

Jesske, B., Schulz, S., Quandt, S. (2020). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 13. Erhebungswelle – 2019 (Haupterhebung), *FDZ Methodenreport 06/2020*. Nürnberg.

Jesske, B. & Schulz, S. (2021). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 14. Erhebungswelle – 2020 (Haupterhebung), *FDZ-Methodenreport, 06/2021*. Nürnberg.

Jesske, B., Schulz, S. & Gerber, V. (2022). Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 15. Erhebungswelle – 2021 (Haupterhebung), *FDZ-Methodenreport 06/2022*. Nürnberg.

Jesske, B., Schulz, S. & Gerber, V. (2023): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 16. Erhebungswelle – 2022 (Haupterhebung), *FDZ-Methodenreport 06/2023*. Nürnberg.

Jesske, B., Schulz, S. & Gerber, V. (2024): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 17. Erhebungswelle – 2023 (Haupterhebung), *FDZ-Methodenreport 05/2024*. Nürnberg.

König, W., Lüttinger, P. & Müller, W. (1987). Eine vergleichende Analyse der Entwicklung und Struktur von Bildungssystemen. Methodologische Grundlagen und Konstruktion einer vergleichbaren Bildungsskala. *CASMIN-Projekt. Arbeitspapier Nr. 12*. Mannheim.

Kreuter, F., Müller, G., Trappmann, M. (2010). Nonresponse and measurement error in employment research. Making use of administrative data. *Public Opinion Quarterly, Vol. 74, No. 5*, S. 880-906.

Kreuter, F., Müller, G., Trappmann, M. (2014). A note on mechanisms leading to lower data quality of late or reluctant respondents. *Sociological Methods and Research, Vol. 43, No. 3*, S. 452-464.

Lechert, Y., Schroedter, J. & Lüttinger, P. (2006). Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus- Zusatzerhebung 1971 und die Mikrozensen 1976-2004. *ZUMA-Methodenbericht 2006/12*. Mannheim.

Lengerer, A., Bohr, J. & Janßen, A. (2005). Haushalte, Familien und Lebensformen im Mikrozensus – Konzepte und Typisierungen. *ZUMA-Arbeitsbericht 2005/05*. Mannheim.

Lipsmeier, G. (1999). Die Bestimmung des notwendigen Lebensstandards – Einschätzungsunterschiede und Entscheidungsprobleme. *Zeitschrift für Soziologie*, 28, S. 281-300.

Müller, W., Wirth, H., Bauer, G., Pollak, R. & Weiss, F. (2006). ESeC – Kurzbericht zur Validierung und Operationalisierung einer europäischen sozioökonomischen Klassifikation. *ZUMA-Nachrichten, 59*, S. 111-119.

Müller, W., Wirth, H., Bauer, G., Pollak, R. & Weiss, F. (2007). Entwicklung einer europäischen so-zioökonomischen Klassifikation. *Wirtschaft und Statistik, 5*, S. 527-530.

Müller, B., Bähr, S., Gundert, S., Teichler, N., Unger, S. & Wenzig, C. (2020). PASS Scales and Instruments Manual, FDZ-Methodenreport, 07/2020 (en). Nürnberg.

Nolan, B. & Whelan, C. (1996). Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches. *Journal of European Social Policy, 6*, S. 225-240.

Nübling, M., Andersen, H. H. & Mühlbacher, A. (2006). *Entwicklung eines Verfahrens zur Berechnung der körperlichen und psychischen Summenskalen auf Basis der SOEP-Version des SF 12 (Algorithmus)* (No.16). DIW Data Documentation.

Organisation for Economic Co-Operation and Development [OECD] (Hrsg.) (1999). *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries. 1999 Edition*. Paris: OECD.

Organisation for Economic Co-Operation and Development [OECD] (Hrsg.) (1982). The OECD List of Social Indicators. Paris: OECD.

Porst, R. (1984). Haushalt und Familien 1982. Zur Erfassung und Beschreibung von Haushalts- und Familienstrukturen mit Hilfe repräsentativer Bevölkerungsumfragen. *Zeitschrift für Soziologie, 13*, S. 164-175.

Rendtel, U. & Harms, T. (2009). Weighting and calibration for household panels. In P. Lynn (Hrsg.), *Methodology of Longitudinal Surveys* (S. 265-286), Chichester: Wiley.

Ringen, S. (1988). Direct and Indirect Measurement of Poverty. *Journal of Social Policy, 17*, S. 351-365.

Rose, R. & Harrison, E. (2007). The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. *European Societies, 9*, S. 459-490.

Rudolph, H. & Trappmann, M. (2007). Design und Stichprobe des Panels „Arbeitsmarkt und Soziale Sicherung“ (PASS). In M. Promberger (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“, *IAB-Forschungs-bericht 12/2007* (S. 60-101), Nürnberg.

Sozialgesetzbuch Zweites Buch [SGB II]: Grundsicherung für Arbeitssuchende.

Statistisches Bundesamt [StBA] (1992). Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt [StBA] (2008). Klassifikation der Wirtschaftszweige 2008 (WZ 2008) mit Erläuterungen. Wiesbaden: Statistisches Bundesamt.

Trappmann, M., Christoph, B., Achatz, J., Wenzig, C., Müller, G. & Gebhardt, D. (2009). Design and stratification of PASS. A New Panel Study for Research on Long Term Unemployment. *IAB-Discussion Paper 5/2009*. Nürnberg.

Trappmann, M. (2013a). Weighting. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide ”Panel Labour Market and Social Security” (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 56-66). Nürnberg.

Trappmann, M. (2013b). Weights. In: Bethmann, A., Fuchs, B. & Wurdack, A. (Hrsg.)(2013). User Guide ”Panel Labour Market and Social Security” (PASS). Wave 6. *FDZ Datenreport 07/2013* (S. 81-99). Nürnberg.

Treiman, D. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Wegener, B. (1985). Gibt es Sozialprestige? *Zeitschrift für Soziologie, 14*, S. 209-235.

Wegener, B. (1988). *Kritik des Prestiges*. Opladen: Westdeutscher Verlag.

Imprint

FDZ-Datenreport 07|2024

Publication Date
17. March 2025

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

Rights of use
This publication is published under the following Creative Commons licence: Attribution – ShareAlike 4.0 International (CC BY-SA 4.0)
https://creativecommons.org/licenses/by-sa/4.0/deed.de

Download
https://doku.iab.de/fdz/reporte/2024/DR_07-24_EN.pdf

Documentation version
PASS-SUF0623_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2407.en.v1

Dataset version
PASS-SUF0623, v1, DOI: 10.5164/IAB.PASS-SUF0623.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/en/startseite-en/

Corresponding author
Jonas Beste
Telefon +49 (0) 911/179-2279
E-Mail Jonas.Beste@iab.de