PASS User Guide

Madhu Chauhan et al.

November 06, 2026

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

FDZ-DATENREPORT

Documentation of labour market data

06|2026 EN Panel Study Labour Market and Social Security (PASS) User Guide // Version 3.0.

Madhu Chauhan, Sebastian Bähr, Jonas Beste, Matthias Collischon, Sandra Dummert, Anja Fehn, Corinna Frodermann, Sonja Malich, Marcel Müller, Valentina Prospero, Jens Stegmaier, Mark Trappmann, Claudia Wenzig.

Bundesagentur für Arbeit

Panel Study Labour Market and Social Security

(PASS) User Guide, Version 3.0.

Madhu Chauhan, Sebastian Bähr, Jonas Beste, Matthias Collischon, Sandra Dummert, Anja Fehn, Corinna Frodermann, Sonja Malich, Marcel Müller, Valentina Prospero, Jens Stegmaier, Mark Trappmann, Claudia Wenzig.

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

Getting started with the User Guide **6**

0.1 Data access . . . 7

0.2 Further tools and resources . . . 7

0.3 Terminology . . . 9

0.4 Feedback . . . 9

1 The Evolution of PASS Over Two Decades **10**

1.1 PASS background and objectives . . . 10

1.2 Development of Focal Issues Over the Waves . . . 13

1.3 Methodological Innovations to Improve Data Quality . . . 14

2 PASS Survey Design and Methodology **17**

2.1 Sampling Procedure . . . 17

2.1.1 Sampling Design in the First Wave . . . 18

2.1.2 Panel Retention and Sample Refreshment . . . 19

2.1.3 Replenishment Samples . . . 19

2.1.4 Refugee Sub-samples . . . 20

2.1.5 Sub-samples . . . 20

2.1.6 Sampling Frame and Auxiliary Data for Non-response Analyses and Post-Survey Adjustments 21

2.2 Mixed-Mode Design . . . 22

2.3 Respondent Incentives . . . 23

2.4 Call Records, Fieldwork Monitoring, and the Development towards an Adaptive Design . . . 26

3 PASS Survey Design and Methodology - Weighting **28**

3.1 Initial weighting process . . . 28
3.1.1 Stage 1: Design weights for the gross sample . . . 29
3.1.2 Stage 2: Modelling non-response . . . 29
3.1.3 Stage 3: Calibration . . . 30
3.2 Construction of the weights from wave 2 onwards . . . 32
3.2.1 Design weights for the wave n households in the $n+1$ th wave. . . . 33
3.2.2 Design weights for the wave $n+1$ refreshment sample . . . 34
3.2.3 Response propensity models for panel households . . . 34
3.2.4 Non-response weighting for households in the wave n refreshment sample . . . 35
3.2.5 Propensity models for temporary drop-outs . . . 35
3.2.6 Integration of weights by convex combination . . . 36
3.2.7 Response propensity models for panel persons . . . 36
3.2.8 Integration of the weights to yield the total weight before calibration . . . 36
3.2.9 Calibration to the household weight, wave $n+1$, cross-section . . . 37
3.2.10 Calibration to the person weight, wave $n+1$, cross-section . . . 37
3.2.11 Estimating BA cross-sectional weights for non-recipient Households and Individuals . . . 37
3.3 Integration of weights from replenishment samples . . . 38
3.4 Weighting datasets and weighting variables . . . 38
3.5 Web-survey weighting . . . 40

4 Survey Instruments **41**

4.1 Questionnaire . . . 41
4.2 Logic of questionnaire modules within PASS . . . 42
4.3 Key variables across the two questionnaires . . . 46
4.3.1 Household interview . . . 46
4.3.2 Individual interview . . . 47

5 PASS SUF Datasets and Variables: Collection, Editing and Usage **50**

5.1 Datasets in the SUF . . . 51
5.1.1 Classification of Datasets in the SUF . . . 51

5.1.2 Description of Datasets by levels . . . 54
5.1.3 Additional dataset for Websurvey mode . . . 56
5.1.4 Temporary datasets . . . 56
5.2 Variables in PASS . . . 57
5.2.1 System variables . . . 58
5.2.2 Surveyed variables . . . 58
5.2.3 Generated variables . . . 61
5.3 Data Editing in PASS . . . 61
5.3.1 Overview of Data editing . . . 62
5.3.2 Structure Checks . . . 63
5.3.3 Filter Checks and Assignment of Standardised Codes . . . 64
5.4 Record Linkage to Administrative Data . . . 66

6 PASS in Transition - Major developments and Innovations. 67

6.1 Additional Refugee Samples and Adaptations in Survey Languages . . . 67
6.2 Web survey . . . 69
6.3 Adaptive Design in PASS . . . 70
6.4 Impacts of COVID-19 Pandemic on Fieldwork, Questionnaire, Interviewer Training, and Other Areas 71
6.4.1 Fieldwork changes . . . 71
6.4.2 Questionnaire Adjustments . . . 72
6.4.3 Interviewer Training . . . 73
6.4.4 Potential Implications for Data Use and Analysis . . . 73
6.5 Interviewer Monitoring in the PASS Study: Integration into CAPI Fieldwork . . . 74

References 76

Imprint 83

List of Figures

1.1 Persons and households interviewed in PASS in each survey year (in thousand). Source: Data Report IAB. 16

2.1 (Sequential) Mixed mode switch survey administration of PASS 24

3.1 Generation of the weights for wave $n + 1$ given the weights of wave n 33

5.1 Variable naming scheme 58

List of Tables

0.1 Overview of PASS documentation and supporting resources 8

1.1 Timeline of important innovations in PASS 11

2.1 PASS variable samples in HHENDDAT across waves 20

3.1 Design weight variables in hweights dataset with meaning 29

3.2 Variables in the household weights dataset (hweights) 38

3.3 Variables in the person weights dataset (pweights) 39

4.1 PASS thematic blocks at the household level (Waves 1–18) 42

4.2 PASS thematic blocks at the individual level (Waves 1–18) 43

5.1 Classification of core datasets in PASS 51

5.2 Alphabetic list of subject-related indicators used in the variable names (cross sections) 59

5.3 Alphabetical list of subject-related indicators used in variable names (spell datasets) 61

5.4 Overview of the steps involved in editing the PASS data 63

5.5 Household types and interview requirements 64

5.6 Standardised missing codes and their explanations. 65

6.1 The use of foreign language instruments in PASS 68

6.2 Impact of the COVID-19 Pandemic on PASS (Waves 14–16) 71

Getting started with the User Guide

The *Panel Study Labour Market and Social Security (PASS)* User Guide provides a comprehensive reference for working with the PASS survey and its Scientific Use File (SUF) for research on labour market, poverty and means-tested income support in Germany. It is intended to support researchers in understanding the structure, content, analytical potential of PASS data and to facilitate its effective usage in empirical research.

This third edition of PASS User Guide is based on and extends the previous PASS User Guide published in 2013 which is available via the PASS/FDZ website, explicitly on the (see Bethmann et al. (2013)). The earlier edition was authored by Mark Trappmann, Jonas Beste, Arne Bethmann, Johannes Eggs, Benjamin Fuchs, Daniel Gebhardt, Stefanie Gundert, Gerrit Müller, Claudia Wenzig, Anja Wunder. The current version incorporates substantial revisions and additions, including updated methodological information, documentation of new survey content and datasets, and developments covering Waves 1 to 18 of PASS. The authors wish to explicitly acknowledge the contributions of the authors of the previous edition and the numerous contributors to PASS data reports, methods reports, questionnaires, and technical documentation whose work informed the preparation of this guide.

This third edition of PASS user guide places particular emphasis on the longitudinal character of PASS and the analytical opportunities arising from 18 waves of data (and continuing), capturing long-term developments in the living conditions and labour market trajectories of individuals and households in Germany.

The guide is designed for a broad range of users. It provides orientation for those engaging with PASS data for the first time, while also offering structured guidance for experienced researchers conducting advanced longitudinal analyses. While wave-specific publications such as data reports, methods reports, and field reports document survey implementation, data editing, and weighting procedures for individual waves in detail, this guide focuses on the concepts, structures, and methodological principles that remain comparable and relevant across waves. In doing so, it provides a stable foundation for working with PASS data longitudinally and for analysing developments over time.

Particular attention has been given to the organisation and accessibility of the guide. Its structure supports both sequential reading and targeted consultation, enabling users to navigate efficiently across chapters and locate relevant information with ease.

Chapter 1 outlines the study's objectives, thematic evolution, and key methodological innovations to situate PASS within its broader research context.

Chapter 2 presents the survey design, including sampling, data collection modes, incentives, and fieldwork processes.

Chapter 3 introduces the weighting framework and its development across waves, including extensions for newer components such as the web survey.

Chapter 4 turns to the survey instruments, describing questionnaire structures, modular design, and key variables at both the household and individual levels.

Chapter 5 provides a detailed overview of the SUF, including dataset structure, variables, data editing procedures, and possibilities for linking survey data with administrative records (PASS-ADIAB).

Chapter 6 highlights recent developments in PASS, including methodological adaptations, extensions in survey content, and innovations in data collection.

The User Guide is complemented by additional and evolving resources. A dedicated PASS Blog (currently under progress) serves as an extension of this guide and, in particular, is being developed to replicate and extend the Part II of the earlier User Guide versions. It provides a platform for methodological extensions, applied research examples, and updates on ongiong and recent developments, allowing the User Guide to remain focused on core documentation while ensuring that new insights can be disseminated in a flexible and timely manner.

0.1 Data access

Access to the PASS data is provided via a 'Scientific Use File' (SUF), which is made available by the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). Detailed information on application procedures, access conditions, and available datasets can be found on the FDZ website (Scientific Use Files - FDZ - Forschungsdatenzentrum). In addition, with PASS-ADIAB, a combined dataset comprising survey and administrative data is available. As PASS-ADIAB contains sensitive data, it is not available as SUF but via 'On-Site Use' and 'Remote Data Processing' (On-site Use and Remote Data Access - FDZ - Forschungsdatenzentrum)

0.2 Further tools and resources

The below tables summaries further resources and tools helpful for the users to work with PASS data, a short description about the resource, links to access these resources (where available) and language of publication for the resource.

Table 0.1: Overview of PASS documentation and supporting resources

Name	Content	Language
PASS User Guide	The User Guide offers general information on PASS that is not specific to certain waves. The following topics are covered: objectives and research questions of PASS; additions to existing data; survey and sampling design; instruments and interview programme; structure of the Scientific Use File and its datasets; general logic of data editing; weighting concept; and examples on how to use the datasets.	English
PASS Blog	The PASS Blog provides a flexible space for updates, examples, and learning resources that complement the User Guide. It replicates and extends the purpose of Part 2 of earlier versions of the User Guide by hosting content that may require more frequent updates.	English
Data Reports	For each wave, the respective data report provides wave-specific information on PASS, including the survey programme with wave-specific features and innovations, key statistics, generated variables, data editing, weighting, and tabulations of the surveyed variables until wave 16. In waves where a web survey is or was conducted, a separate data report on the web survey is also published.	German English
Methods and Field Reports	For each wave, the methods and field report describes the work of the field institute for the respective wave. It covers objectives and design of PASS, pretest, detailed fieldwork steps, data editing by the field institute, and weighting including modelling of non-response. In waves with a web survey, a separate methods and field report for the web survey is also published.	German
Questionnaires	For each wave, three questionnaires are available: household questionnaire, person questionnaire, and senior citizens' questionnaire.	German English
Interviewendennachbefragung	Data Report on Interviewer Surveys (PASS-INT) describes the interviewer survey (PASS-INT), which is completed by interviewers after a successfully conducted personal interview in the Panel Study Labour Market and Social Security (PASS). The interviewer survey has been conducted since wave 4 and provides interviewer-reported information relating to the interview and interview situation.	German
Scales and Instruments Manual	The manual provides information on questionnaire modules and scales implemented in waves 1 to 13 of the PASS study. It describes the theoretical background, development, and wording of individual and scale items. Items are displayed in German (original wording) and English (for international use). It also provides item and scale statistics including means, standard deviations, and internal consistencies for each survey year.	English
Correspondence List	Provides a user-friendly way to link questionnaire content, such as question numbers, to variable names and labels in the Scientific Use File. Structured by dataset, such as PENDDAT, alg2_spells, and KINDER, and by wave. Updated each wave to show in which waves specific items were included.	German

Table 0.1: Overview of PASS documentation and supporting resources (continued)

Name	Content	Language
Requesting observation numbers	Users can request the number of observations for research-relevant variables via email (iab.fdz@iab.de). For continuous variables: descriptive statistics and histograms. For categorical variables: frequency tables and bar charts. Outputs can be wave-specific or across waves.	German English

0.3 Terminology

PASS is particularly rooted in research on Social Code Book II (SGB II) and examines the effects of active labour market policies as well as income support.

Since the beginning of the reform, the name of the welfare benefit has changed several times: ‘‘Arbeitslosengeld II (ALG2)’’ / ‘‘Unemployment Benefit II (UB II)’’ from 01 January 2005 -- 31 December 2022, ‘‘Bürgergeld’’/ ‘‘Citizen’s Benefit’’ from 01 January 2023 to 30 June 2026, and now the ‘‘Grundsicherungsgeld’’ from 01 July 2026.

Throughout the userguide, these terminologies are used interchangeably or sometimes in their combined forms by various authors. However, they always refer to the welfare benefit in Germany and extend by definition to the recipients of these benefits.

Established variable names and dataset labels with now historical name references (e.g. alg2_spells.dta or alg2samp) have been deliberately retained. They likewise refer to the benefit under SGB II and therefore also cover information on ‘‘Citizen’s Benefit’’ now known as ‘‘Grundsicherungsgeld.’’

0.4 Feedback

The continued development of the PASS User Guide benefits greatly from feedback provided by its users. Suggestions for improvement, identification of gaps, or proposals for additional topics are highly appreciated and help ensure that the guide remains relevant and user-oriented. Users are welcome to share their comments, feedbacks and suggestions on the current version of the user guide at either IAB.HHP@iab.de via email.

Chapter 1

The Evolution of PASS Over Two Decades

Contact authors: ***Chauhan, Madhu; Trappmann, Mark; Wenzig, Claudia.***

The *Institute for Employment Research (IAB)* launched the *Panel Study Labour Market and Social Security (PASS)* in 2006 to address critical research gaps in the analysis of the labour market, the welfare state, and poverty in Germany. PASS is particularly rooted in research on *Social Code Book II (SGB II)* and examines the effects of active labour market policies as well as income support.

At the core of the study are five key research questions:

1. What pathways lead out of *Unemployment Benefit II*[1] receipt?
2. How do the social and financial situations of affected households change while receiving benefits?
3. What coping strategies do affected individuals develop, and how do their orientations and behaviours evolve?
4. How do interactions with agencies providing the benefits take place, and what do these institutional practices look like?
5. What employment trajectories and household dynamics contribute to long-term benefit dependency?

These five guiding questions provide a structured framework for the development of the survey (Achatz et al., 2007) and ensure that the collected data allows for both cross-sectional and longitudinal analyses.

1.1 PASS background and objectives

From its inception, the PASS design made it possible to link PASS data with administrative records, enabling in-depth longitudinal analyses of employment trajectories (see Section 5.4). In several respects, PASS extends

[1] Unemployment Benefits II is used interchangeably with various other terminologies across this user guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

and compliments existing administrative data from the *Federal Employment Agency (BA)* and representative surveys such as the *Socio-Economic Panel (SOEP)*, which provide valuable insights into the evaluation of SGB II.

Specifically, PASS extends the scope of administrative data by incorporating household dynamics before and after the receipt of *Unemployment Benefit II* (UB II) and encompassing all relevant population groups, including self-employed individuals and those marginally attached to the labour market, in addition to employed, unemployed, and welfare recipients. Moreover, the survey captures labour market-related attitudes and behaviours, offering a more comprehensive perspective on social security dynamics. Compared to existing panel surveys, the higher number of cases involving basic income recipients and low-income groups further strengthens the dataset, improving its capacity for detailed analyses such as assessing the impact of SGB II on specific target groups, including migrants, single parents, and supplemental benefit recipients (''*Aufstocker*''), thus enhancing the precision of statistical estimates and model coefficients.

However, the scope of PASS extends far beyond the evaluation of SGB II. The survey is representative of both benefit recipients and the general population, providing a comparative perspective that distinguishes between the experiences of social security benefit recipients and non-recipients. Furthermore, PASS covers broader socio-economic themes such as labour market participation, social inclusion, digitalisation, and other relevant changes of socio-economic conditions.

With 18 complete survey waves to date, PASS stands as one of Germany's most significant long-term studies on social security and labour market policy. Both its survey design and questionnaire content have evolved over time to reflect major societal changes, such as the refugee influx (2014--2015, 2022), the COVID-19 pandemic, and the introduction of the *Citizen's Benefit (Bürgergeld)* in 2023, to name a few.

PASS data provides valuable insights for both scientific research and policy advice. Researchers and policy-makers benefit from the panel's longitudinal structure, which enables tracking individual changes over time. Researchers can analyse transitions between employment and unemployment, explore poverty and household composition, and examine institutional experiences. The dataset offers a robust empirical foundation for evaluating the impact of welfare policies and supporting informed political decision-making.

The following table summarises major questionnaire-related and methodological innovations across PASS waves.

Table 1.1: Timeline of important innovations in PASS

		Important innovations	
Year	Wave	Questionnaire related	Methodology related
Waves 1–5: Initial data collection and first extensions focusing on labour market participation, social benefit receipt, household structure, and social situation.			

Table 1.1: Timeline of important innovations in PASS (continued)

Year	Wave	Questionnaire related	Methodology related
2006/07	1	Initial data collection focusing on labour market participation, social benefit receipt, household structure, and social situation.	
2007/08	2	Initial design of a more detailed survey on employment histories focusing on episodes (employment, unemployment, labour market-related activities).	Integration of new SGB II entrant samples. No longer an English-language survey version.
2008/09	3	Development and integration of focus modules on health and social networks (distinct from basic modules).	Introduction of dependent interviewing.
2010	4	Revision and finalisation of the employment episodes survey (biography module).	Data preparation: Design of cross-wave harmonisation for the biography module.
2011	5		Replenishment samples to improve panel stability and representativeness.
Waves 6–9: Expansion of thematic modules, particularly on job quality, childcare, health, and methodological refinements in data and fieldwork.			
2012	6	More detailed data collection on childcare and children's social participation.	Data preparation: Introduction of a child dataset. Fieldwork: Introduction of adaptive design: Three field phases with different control and incentive strategies.
2013	7	Refinements and add-ons in the employment biography module.	Increased use of preload information to reduce inconsistencies.
2014	8	Revision of the module on interactions with job centres.	
2015	9	Introduction of an improved module for surveying social networks and social support. Modification of the health module with more detailed questions on mental and physical health.	
Waves 10–17: Adaptation to current societal developments such as refugees, digitalisation, the COVID-19 pandemic, and the Citizen's Income reform.			
2016	10	Addition of migration-related questions, particularly focusing on refugees. Addition of child specific deprivation items.	Expansion of the sample to include Syrian/Iraqi refugees within SGB II. Arabic added as a survey language; Turkish discontinued.
2017	11	Introduction of specific modules on digitalisation and new forms of work.	
2018	12	Addition of questions on retirement planning and financial security.	
2019	13	Introduction of a module on political participation.	

Table 1.1: Timeline of important innovations in PASS (continued)

Year	Wave	Questionnaire related	Methodology related
2020	14	Addition of questions concerning the pandemic added into the ongoing fieldwork.	Modified fieldwork during the COVID-19 pandemic: No on-site CAPI interviews, increased share of telephone interviews (CATI), and extended field periods. Implementation of additional CAPI-interviewer monitoring system for quality control. Refinement of adaptive survey design for fieldwork management.
2021	15	Expanded questions on psychological stress during the COVID-19 pandemic. Introduction of a module on household practices to cope with financial strain.	Fieldwork: Continued absence of on-site CAPI interviews. Oversampling of the Syrian/Iraqi newcomer sample. Oversampling of self-employed recipients entering welfare benefits during the pandemic.
2022	16	Extended questions on internet usage.	Fieldwork: Resumption of CAPI fieldwork. Oversampling of the Syrian/Iraqi newcomer sample. Methodological experiment on gender-neutral phrasing. First PASS web survey.
2023	17	Additional questions on the introduction of Citizen's Income, including stigma awareness, interactions with job centres, and further training opportunities.	Expansion of the sample to include Ukrainian refugees within SGB II. New Ukrainian-language survey version. Replenishment samples.
2024	18		Second PASS web survey.

Table 1.1 outlines the key milestones in the development of PASS which are discussed in detail in the following sections.

1.2 Development of Focal Issues Over the Waves

In its initial years, the survey primarily focused on fundamental aspects of social security and labour market integration. The first five waves systematically captured employment trajectories, household composition, benefit receipt, and socio-economic conditions. Additionally, from wave 3 onwards, the survey incorporated initial modules on health and social networks.

As the survey evolved, its thematic scope gradually expanded. Between waves 6 and 10, the survey extended its coverage of employment biographies. Episode-based questions fostered precise tracking of job changes, unemployment durations, and other labour market-related activities in a structured manner. During this period, family-related topics drew attention. For instance, wave 6 introduced a child module to examine social inclusion and educational participation and childcare arrangements within households.

In later waves, the survey focused on contemporary societal developments. Starting from wave 10, the survey

introduced an expanded version of migration related questions to better capture the situation of refugees in Germany. In wave 11, the survey incorporated questions on the digital working environment, capturing topics such as remote work, and evolving forms of employment.

The outbreak of the COVID-19 pandemic in 2020 marked a significant thematic shift for the survey. The pandemic not only affected social life and labour market conditions, which led to content-related adjustments in the questionnaire, but also disrupted data collection methods. In response to the crisis, wave 14 introduced a dedicated module examining the effects of pandemic on the labour market. Key topics included short-time work, job loss, and the increased use of remote work, alongside changes in income due to pandemic. Additional questions explored childcare challenges due to school and day-care closures during pandemic. Mental health also became a key area of inquiry, prompting an extension of the health module to include items on anxiety, stress, and emotional exhaustion (see Section 4.3.2).

The introduction of the Citizen's Benefit (*Bürgergeld*) in 2023 (wave 17) led to further adjustments in the survey content, as the switch from *"Unemployment Benefit II" to "Citizen's Benefit"* involved substantial legal framework and policy changes beyond the re-labelling of a policy. Accordingly, the survey incorporated questions about changes in job centre services and support measures. A new module focused on questions regarding promotion of training and qualification measures was introduced, since the updated legal framework aims to enhance the integration of long-term unemployed individuals into the labour market. The survey also provides the possibility to examine the subjective perception of stigmatisation associated with receiving Citizen's Benefit.

These developments illustrate how PASS continuously realigns its focus in response to changing social and policy landscapes. While some modules, such as employment biographies, job quality, health, and agency contacts, have remained central, the survey has also incorporated several temporary modules, such as the evaluation of the Educational Package or the COVID-19 module (further details on survey instruments see Chapter 4). At the same time, topics that proved less relevant over time, such as crowd-working or pension provision, have been removed from the questionnaire. This ongoing revision ensures the survey remains aligned with current scientific, academic and policy priorities.

Alongside content-related expansions, PASS has implemented numerous methodological advancements over the years to enhance data quality and ensure representativeness. The most significant methodological innovations are outlined in the following section.

1.3 Methodological Innovations to Improve Data Quality

To maintain representativeness over time and counteract panel attrition, PASS has incorporated an annual refreshment sample of new SGB II entrants from the beginning. This approach allows the survey to include new households while ensuring continued and accurate representation of the SGB II recipient population. In addition to these annual refreshment samples, periodic replenishment samples are included in both the SGB

II and general population samples (further details in Chapter 2). The survey also introduced targeted refugee samples i.e., Syrian and Iraqi refugees in wave 10, and Ukrainian refugees in wave 17 (see Section 6.1). These inclusions were in response to changing societal dynamics and the evolving composition of the SGB II population. To effectively include these groups, PASS expanded its language offerings. It introduced an Arabic-language version in wave 10 and a Ukrainian-language version in wave 17.

Another key methodological improvement is the shift to an episode-based recording of employment biographies. While the first wave only documented respondents' current employment status, the revised survey methodology enables a comprehensive tracking of employment, unemployment, and other labour market-relevant episodes. This substantially improved the data foundation for analysing employment instability and transitions between different forms of work. The introduction of ''*Dependent interviewing*'' from wave 3 has further enhanced data quality by allowing respondents' previous answers to be incorporated into the current interview, reducing inconsistencies and minimising cognitive burden.

A major methodological shift occurred during the COVID-19 pandemic (see Section 6.4 and Sakshaug et al. (2020)). Since in-person interviews were not feasible during lockdown, the survey largely replaced *computer-assisted personal interviewing (CAPI)* with *computer-assisted telephone interviewing (CATI)* from wave 14 onwards. The surveys returned to pre-pandemic survey methods in wave 17. To minimise distortions caused by the change in methodology, the team conducted rigorous statistical consistency analyses, thus ensuring that the data remained comparable across wave 14 to wave 17.

PASS also employs innovative methods in fieldwork to continuously improve data collection quality. Since wave 6, the field time has been divided into three phases to maximise response rates. Initially introduced as an experiment, an adaptive survey design approach was systematically implemented in later waves (see Section 6.3 and Beste et al. (2023)). This method uses information from previous waves to identify hard-to-survey respondents and tailor incentive strategies accordingly.

From wave 14 onwards, PASS introduced new statistical procedures to enhance data quality (see Section 6.5 and Beste et al. (2021)). Longitudinal consistency checks ensure the coherence of responses across waves, while time stamp analyses help detect irregularities in interview conduct, by identifying unusually short or long interview durations. These quality control mechanisms further optimised data reliability and ensured comparability across survey waves.

Managing a high participation rate and minimising panel attrition remain key challenges for the survey. Therefore, in wave 16, PASS conducted its first web-based survey as a supplement to the main survey, including additional question modules. Another web survey followed in wave 18. These methodological experiments aim to explore whether a web survey mode, alongside CATI and CAPI, could help improve panel stability and counter rising costs, particularly in CAPI fieldwork (see Section 6.2).

Figure 1.1: Persons and households interviewed in PASS in each survey year (in thousand). Source: Data Report IAB.

Finally, Figure 1.1 presents wave specific numbers of individuals and households interviewed in PASS over the years.

Over the past two decades, PASS has continuously evolved both in content and methodology to serve as a reliable data source for research and policy advice. The ambition to remain indispensable in the future drives the ongoing integration of new topics and innovative methods, enabling PASS to respond flexibly and appropriately to societal changes.

Chapter 2

PASS Survey Design and Methodology

Contact authors: Beste, Jonas; Collischon, Matthias; Müller, Marcel; Prospero, Valentina; Trappmann, Mark; Fehn, Anja.

This chapter outlines the methodological framework of PASS, focusing on its design principles, data collection strategies, and adaptive developments that ensure representativeness and data quality across survey waves.

Section 2.1 describes the sampling design, including the dual-frame structure combining Unemployment Benefit II (UB II)[2] recipients and the general population, as well as the procedures for sample refreshment, replenishment, and inclusion of refugee sub-samples. **Section** 2.2 explains the mixed-mode data collection approach, which integrates computer-assisted telephone and personal interviews (CATI and CAPI) and details interviewer training, multilingual implementation, and mode-switching protocols. **Section** 2.3 outlines the respondent incentive strategy, tracing the transition from lottery-based to prepaid monetary incentives and recent adaptive adjustments based on response propensities. **Section** 2.4 discusses the use of call records, fieldwork monitoring, and the development of responsive and adaptive survey designs, including applications of machine learning to improve participation and reduce bias. This chapter further continues into the next **Chapter** 3 to discuss yet another important concept of PASS survey design and methodology i.e., Weighting.

These sections present a comprehensive view of how PASS combines robust sampling, flexible fieldwork methods, and adaptive design principles to maintain high data quality and longitudinal consistency.

2.1 Sampling Procedure

The sampling design of PASS is characterized by two key features: a **dual-frame approach** comprising the Unemployment Benefit II (UB II) recipients and the general population; and an **annual refreshment sample** of new UB II recipients entering the population.

[2] Unemployment Benefits II is used interchangeably with various other terminologies across this user guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

A comparison between benefit recipients and the general population is essential for analysing inflows into UB II, identifying differences between benefit-recipient and non-recipient households, investigating hidden poverty, and forming control groups. To facilitate such analyses, PASS combines a sample of UB II recipient households with a general population sample.

To capture inflows into UB II at an early stage and to ensure the representativeness of the benefit-recipient sample in cross-sectional analyses, a refreshment sample is drawn for this group in every wave (for details on the refreshment sample concept, see Trappmann et al. (2009), p.11).

2.1.1 Sampling Design in the First Wave

In wave 1, the sample consisted of two distinct sub-samples. They were connected at the first sampling stage through the selection of identical primary sampling units (*PSUs*) (for further details, see Rudolph & Trappmann (2007), p. 65). Within each PSU, the two subsamples were drawn independently.

- Welfare recipient/ UB II sub-sample (UB II recipients):

  A random sample of benefit units (''Bedarfsgemeinschaften'') comprising at least one individual receiving UB II in July 2006. The sample was drawn from administrative records provided by the Federal Employment Agency (BA). Since PASS is a household survey, the entire household in which a benefit recipient resided was included in the survey.

- General population sub-sample:

  This sub-sample consisted of private households in Germany. In wave 1, a random sample of addresses was drawn from the MOSAIC database held by the commercial provider *Microm*. The sample was disproportionately stratified by status to ensure that households with a lower socio-economic status, and therefore a higher risk of UB II entry, had an increased probability of being included (for results on stratification, see Rudolph & Trappmann (2007)).

At the first sampling stage, 300 postcode sectors were randomly drawn from the postcode register. These postcodes served as PSUs for PASS (for details on PSU selection, see Rudolph & Trappmann (2007), p. 77). The selection probability of each postcode was proportional to the number of households in that sector, based on MOSAIC data. Within each sampling point, benefit units (UB II recipient sample) or addresses (general population sample) were randomly selected.

For the UB II recipient sample, the number of benefit units selected depended on the proportion of UB II recipients within the sampling group (i.e., the number of benefit units according to BA administrative data divided by the number of households according to MOSAIC). On average, 100 benefit units per sampling group were included in the gross sample, ensuring a uniform selection probability across the BA sample (Rudolph & Trappmann, 2007,

p. 78). All members of households containing benefit units were surveyed, resulting in 6,804 interviews in the first wave.

For the general population sample, 100 addresses per sampling point were randomly selected. To over-represent lower-status households, addresses from lower-status groups were given a higher probability of inclusion. Field institute staff visited the selected addresses and manually recorded names from doorbell panels. At the institute, a random selection of names was then conducted, ensuring that each selected individual's entire household was surveyed. This process resulted in 5,990 completed household interviews in the first wave.

2.1.2 Panel Retention and Sample Refreshment

All households participating in a certain wave were targeted for an interview in the next wave if they provided panel consent. Additionally, split-off households (households that separated from originally sampled households) were followed and re-interviewed in subsequent waves. Each split-off household was assigned to the same sub-sample category as its origin household (either from wave 1 or from a refreshment sample in a later wave).

Starting with wave 2, a refreshment sample of newly entered UB II benefit units was added to the study in each wave. These refreshment samples consisted of households that received UB II at the reference date for sampling (i.e. July of the year before the survey wave, e.g. July 2007 for wave 2, up to July 2022 for wave 17) but were not recipients at any earlier sampling dates. The refreshment samples were drawn within the PSUs selected in wave 1, using the same procedure as the original sampling. On average, 1,000 newly interviewed households were added per wave. The cumulative sample of all BA households still receiving benefits at the most recent reference date can be projected to represent all households with at least one UB II recipient in Germany at that time.

2.1.3 Replenishment Samples

As is common in panel studies, attrition occurs over time, reducing the number of households and individuals available for analysis. To mitigate this effect, PASS introduced additional replenishment samples for both sub-samples before wave 5. These replenishment samples differ from the annual refreshment samples, which capture new UB II entrants. Instead, replenishment samples counteract panel attrition by adding new households to maintain representativeness.

For this purpose, 100 new PSUs were selected using probability proportional to size. Unlike previous samples, the wave 5 general population replenishment sample was drawn from municipal registers instead of the MOSAIC database and without stratification. Further details on this procedure are described in Berg et al. (2013), Chapter 6.

Additional general population replenishment samples were introduced in wave 11 and wave 17/18. Due to incomplete processing of cases in wave 17, the non-contacted cases were carried forward as new cases in the

gross sample of wave 18. While the wave 11 replenishment was drawn from the existing 400 PSUs, wave 17 introduced an additional 100 PSUs (Berg et al., 2022). Since these new PSUs were independently sampled, some were drawn twice, leading to a total count slightly below 500 PSUs.

Replenishment samples were also drawn for the UB II recipient sample, introduced alongside general population replenishment samples in wave 5 and wave 11. Additionally, a UB II recipient replenishment sample is planned for wave 19, covering new PSUs introduced in wave 17/18.

2.1.4 Refugee Sub-samples

To account for the increasing influx of refugees into SGB II, separate sub-samples for new SGB II entrants from Syria and Iraq were introduced in wave 10 (survey year 2016). Starting with wave 17 (survey year 2023), a separate group was added for Ukrainian refugees entering SGB II.

To ensure sufficient longitudinal case numbers for detailed analyses, replenishment samples were drawn for these refugee groups. The Syrian and Iraqi refugee replenishments were conducted in waves 14 and 15, while wave 19 replenishment samples will cover both refugee groups within the new 100 PSUs from wave 17. Additionally, these refugee-specific samples were disproportionately selected in selected waves to ensure adequate case numbers (see specific data reports for details).

2.1.5 Sub-samples

In the PASS dataset all sub-samples can be identified by the variable sample in the household dataset (*HHENDDAT*). These wave-specific values are tabulated below.

Table 2.1: PASS variable samples in HHENDDAT across waves

No.	Wave	Variable sample in HHENDDAT
1	Wave 1	UB II recipient sample
2	Wave 1	General population sample (Microm addresses)
3	Wave 2	Refreshment of UB II recipient sample (new entries)
4	Wave 3	Refreshment of UB II recipient sample (new entries)
5	Wave 4	Refreshment of UB II recipient sample (new entries)
6	Wave 5	General population replenishment (municipal registers)
7	Wave 5	Replenishment of UB II recipient sample (municipal registers)
8	Wave 5	Refreshment of UB II recipient sample (new entries)
9	Wave 6	Refreshment of UB II recipient sample (new entries)
10	Wave 7	Refreshment of UB II recipient sample (new entries)
11	Wave 8	Refreshment of UB II recipient sample (new entries)
12	Wave 9	Refreshment of UB II recipient sample (new entries)
13	Wave 10	Refreshment of UB II recipient sample (new entries)
14	Wave 10	Refreshment of UB II recipient sample (Syrian/Iraqi households)

Table 2.1: PASS variable samples in HHENDDAT across waves (continued)

No.	Wave	Variable sample in HHENDDAT
15	Wave 11	General population replenishment (municipal registers)
16	Wave 11	Refreshment of UB II recipient sample (new entries)
17	Wave 11	Refreshment of UB II recipient sample (Syrian/Iraqi households)
18	Wave 12	Refreshment of UB II recipient sample (new entries)
19	Wave 12	Refreshment of UB II recipient sample (Syrian/Iraqi households)
20	Wave 13	Refreshment of UB II recipient sample (new entries)
21	Wave 13	Refreshment of UB II recipient sample (Syrian/Iraqi households)
22	Wave 14	Refreshment of UB II recipient sample (new entries)
23	Wave 14	Refreshment of UB II recipient sample (Syrian/Iraqi households)
24	Wave 14	Replenishment of UB II recipient sample (Syrian/Iraqi households)
25	Wave 15	Refreshment of UB II recipient sample (new entries)
26	Wave 15	Refreshment of UB II recipient sample (Syrian/Iraqi households)
27	Wave 15	Replenishment of UB II recipient sample (Syrian/Iraqi households)
28	Wave 16	Refreshment of UB II recipient sample (new entries)
29	Wave 16	Refreshment of UB II recipient sample (Syrian/Iraqi households)
30	Wave 17	General population replenishment (municipal registers)
31	Wave 17	Refreshment of UB II recipient sample (new entries)
32	Wave 17	Refreshment of UB II recipient sample (Syrian/Iraqi households)
33	Wave 17	Refreshment of UB II recipient sample (Ukrainian households)
34	Wave 18	Refreshment of UB II recipient sample (new entries)
35	Wave 18	Refreshment of UB II recipient sample (Syrian/Iraqi households)
36	Wave 18	Refreshment of UB II recipient sample (Ukrainian households)
37	Wave 17/2	General population replenishment (municipal registers)

2.1.6 Sampling Frame and Auxiliary Data for Non-response Analyses and Post-Survey Adjustments

In addition to the survey design measures implemented to minimise non-response and panel attrition, PASS benefits from a comprehensive database that allows for detailed non-response analyses and post-survey adjustments.

The population samples were either drawn directly from the MOSAIC database by *Microm Consumer Marketing* in wave 1 or subsequently linked to it using address information in wave 5. The MOSAIC database provides auxiliary variables at the address level that can be used, for example, to predict survey non-cooperation based on indicators of social status or privacy concerns, or to assess the likelihood of successfully locating and contacting a sampled unit using information such as the proportion of households moving away from a given area within a year (Sinibaldi et al., 2014). These auxiliary variables were incorporated into the post-survey adjustments for the

wave 1 population sample. A more detailed description of the MOSAIC database is provided by Kueppers (2005).

In addition to the auxiliary data from MOSAIC, the administrative record data on benefit receipt, used for drawing the register samples, provides an even more detailed dataset. These administrative records contain individual-level information, including educational attainment, age, and current employment status, which can be used to analyse and adjust for initial non-response in the UB II recipient sample.

This rich data source ensures that non-response analyses and post-survey weighting adjustments can be conducted with a high level of precision, improving the overall representativeness and reliability of the PASS dataset.

2.2 Mixed-Mode Design

PASS employs a mixed-mode design that combines computer-assisted telephone interviewing (CATI) and computer-assisted personal interviewing (CAPI). In waves 1-3, CATI served as the default mode. This sequential mixed-mode design was adopted as a cost-effective approach to address challenges typically encountered when surveying low-income and welfare populations (Rudolph & Trappmann, 2007, pp. 91--92). These challenges include respondents' higher mobility compared with the general population, contacting difficulties due to limited landline coverage or frequent changes in mobile phone numbers and extremely low expected response rates in self-administered survey modes.

The sequential structure ensures that respondents who cannot be contacted or interviewed by telephone are subsequently visited in person for a CAPI interview. Following a tender process, the fieldwork agency responsible for data collection and preparation changed after wave 3 (A. Müller et al., 2011). This transition also marked a shift towards using CAPI as the default mode for all refreshment samples from wave 4 onwards.

In waves 1-3, interviewers first attempted to contact households by telephone whenever a valid number was available for the address, either through the sampling frame or through tracing prior to fieldwork. Households without a valid number were initially assigned to CAPI mode. From wave 4 onwards, CAPI became the default for refreshment samples, while panel households were first approached in the mode used during their previous interview. Across all waves, a mode switch from CATI to CAPI occurred if telephone contact attempts failed or if a household requested a face-to-face interview. Conversely, cases were switched from CAPI to CATI if personal contact attempts were unsuccessful or upon the household's request for a telephone interview.

Contact attempts in both modes were distributed across different weekdays and times of day to minimise non-response due to non contact. Further details on the organisation of fieldwork for each wave can be found in the respective field reports (e.g. Jesske et al. (2024), for wave 17).

In waves 1-3, interview mode was assigned at the household level, meaning that all members were interviewed using the same mode. From wave 4 onwards, mode assignment became more flexible and could be determined

at the individual level.

Refusal conversion attempts were also conducted by telephone towards the end of each fieldwork period. These targeted households or individuals who initially refused to participate due to reasons such as lack of time, lack of interest, or abrupt call termination. Specially trained CATI interviewers with above-average performance during regular fieldwork handled these cases (see Hartmann et al. (2008), p. 54-56 for waves 1-3; Jesske et al. (2024), p. 61 for wave 17). Indicator variables for interview mode are available in the PASS Scientific Use File to control for potential mode effects in empirical analyses.

Mode effects have been investigated in several articles, most comprehensively by Sakshaug et al. (2023) .

Given that a substantial proportion of the target population has a migration background, many respondents have limited proficiency in German. To ensure inclusivity, the questionnaire was translated into Turkish, Russian, Ukrainian, and Arabic which are the most common first languages among immigrants in Germany (see FDZ-Datenreport 7/2024 : Berg, Cramer, Dickmann, Gilberg, et al. (2024)).

Since PASS relies on telephone and in-person interviews, it requires a large pool of trained interviewers. Before conducting interviews, all interviewers undergo training tailored to their level of experience with or new interventions within the study. The content and structure of these training sessions are detailed in the data reports associated with the respective waves.

The figure below summarises the overall administration switch of PASS survey.

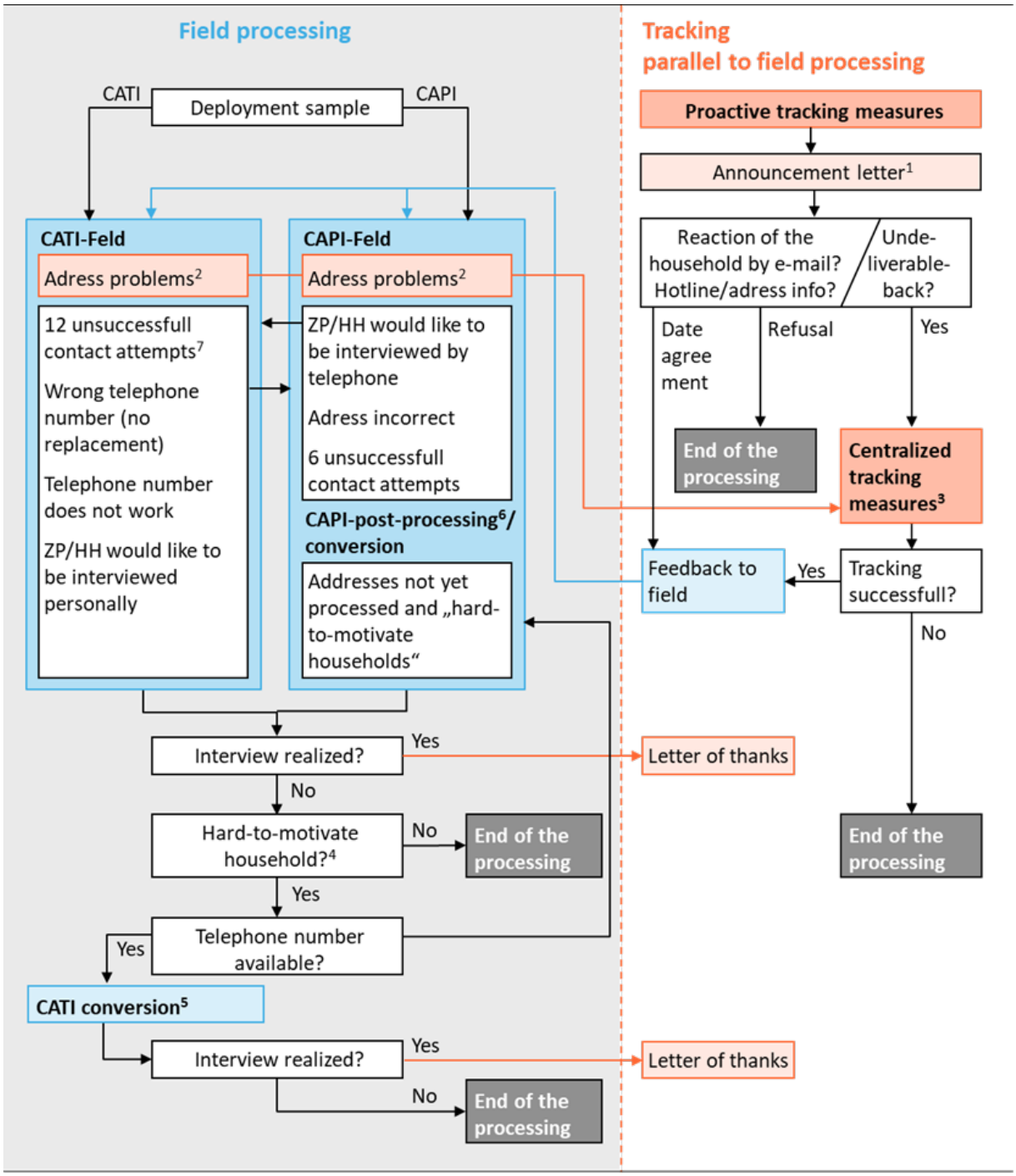
2.3 Respondent Incentives

In line with practices in other large-scale household surveys, PASS provides respondent incentives to improve participation rates and minimise non-response and attrition bias.

In wave 1, all sampled households received a special postage stamp as a small token of appreciation, included with the advance letter. The letter informed recipients that they would receive a lottery ticket for ‘‘Aktion Mensch’’ after completing the interview. Each respondent was subsequently sent a thank-you letter along with the ticket, valued at approximately 1.50 EUR.

In wave 2, the incentive strategy remained largely unchanged, except that the lottery ticket was now for the ‘‘*ARD-Fernsehlotterie*’’ and its value increased to around 5.00 EUR. Alongside extended fieldwork periods and intensified tracking efforts, wave 3 introduced a major change: a shift from conditional to unconditional monetary incentives as an effort to improve survey participation. Each household received a prepaid 10.00 EUR banknote with the advance letter, regardless of eventual participation. This prepaid incentive was provided to all panel households that had participated in the previous wave.

To evaluate the effects of this change on response rates, sample composition, and bias, a split-sample experiment was implemented (Felderer et al., 2018). Households (new participants) from the wave 3 refreshment sample



1 Different cover letter variants for refreshers and panel members
2 Processing status for initiating tracking measures: no connection under this number, incorrect telephone number, fax/modem, ZP moved, unknown under connection
3 Tracking measures: Residents' registration office research; address factory; BA inventory research, use of stable contact information for panel households
4 „Hard-to-motivate household" includes households with refusals for the following reasons: no interest, no time, hangs up immediately, other reasons
5 conversion studies were used at household level
6 In the CAPI, „hard to-reach households" and „hard-to motivate households" were contacted again as part of the follow-up.
7 Processing status: not picked up, not reached, answering machine

Figure 2.1: (Sequential) Mixed mode switch survey administration of PASS

were excluded from this experiment and continued to receive the conditional incentive i.e., a lottery ticket sent after participation for each respondent (Büngeler et al., 2010, pp. 18--20).

Between waves 4 and 6, the monetary incentive approach was maintained and extended in two ways. First, panel households received the unconditional 10.00 EUR incentive, but now at the individual level, meaning each household member eligible for interview received the amount. The prepaid incentive served a dual purpose: to foster respondents' commitment to the study and to increase their motivation to participate.

From wave 17 onwards, the prepaid incentive amount began to vary according to predicted participation probability (Beste et al., 2023). The threshold for participation was defined as 50 per cent of the panel households. The model used data on prior participation behaviour to classify households into two groups:

1. Individuals from high-propensity households (likely to participate) received a prepaid 10.00 EUR incentive with an advance letter.
2. Individuals from low-propensity households (less likely to participate) received an augmented payment of 20.00 EUR with an advance letter.

Household members who did not participate in the previous wave received neither an advance letter nor a prepaid incentive. Consequently, within some households, where not all individuals aged 15 and over could be surveyed in the preceding wave, it was possible that certain members received 10.00 EUR or 20.00 EUR in advance, while others received none. Since wave 17, prepaid incentives have also been utilised to address temporary non-response due to lack of time or soft refusals (Jesske et al., 2024, pp. 57--59).

As a secondary extension, certain cases from refreshment-sample respondents began receiving the same monetary incentive and a thank-you letter; after completing their interview i.e., conditional on response (Jesske & Quandt, 2011, pp. 31--32). The following categories of first-time interviewees are included in PASS: Respondents returning after a temporary drop-out due to unavailability, Members of split households not interviewed in the previous wave, and First-time participants in the refreshment samples.

Since wave 17, the post-interview incentive for these first-time respondents from the refreshment sample(s) was raised to 20.00 EUR per individual, to encourage participation among new entrants (Jesske et al., 2024, pp. 59--60).

Across all waves, face-to-face interviewers were additionally equipped with ''doorstep incentives'' alongside the incentives distributed centrally by mail, to encourage cooperation during household visits. Interviewers could offer small tokens of appreciation such as flowers, chocolates, or similar gifts, at their discretion. The standard guideline allowed 2.00 EUR per panel household and 5.00 EUR per refreshment household. Interviewers could either provide individual gifts or pool funds to purchase a larger token of appreciation for the household. Although this practice was used sparingly, it was recognised as a helpful strategy for fostering positive engagement during fieldwork (Jesske et al., 2024, p. 60).

2.4 Call Records, Fieldwork Monitoring, and the Development towards an Adaptive Design

Survey agencies routinely collect call history data including the date, time, and outcome of each contact attempt to monitor and improve data collection quality.

In interviewer-led surveys such as PASS, which combine face-to-face (CAPI) and telephone (CATI) modes, a ‘‘call’’ refers to either a household visit or a phone contact attempt. Call records play a central role in implementing responsive and adaptive survey designs (see Groves & Heeringa (2006); Laflamme et al. (2008)) by helping identify hard-to-reach or reluctant participants and guiding follow-up strategies. They also inform interviewer training, field supervision, and performance evaluation.

A major advantage of call data is that it is collected for both respondents and non-respondents. When combined with sampling frame information and auxiliary data, (often referred to as survey process data or ‘‘para data’’) (Couper, 1998; Kreuter, 2013), these records enhance the modelling and correction of non-response bias. Furthermore, call record data and other paradata have extensively been used to investigate PASS survey data quality, in particular measurement error bias and non-response bias(e.g., Kreuter et al., 2010, 2014; West et al., 2014)

This section initially outlines the development of field monitoring in PASS, followed by a discussion of its transition towards adaptive survey design.

Since wave 1, detailed call records have been available in PASS. These include the sample member identifier, date and time of each contact, call outcome, survey mode, and indicators showing whether the call was part of a refusal conversion effort or the foreign-language fieldwork. From wave 4 onwards, individual-level call data became available for both telephone and face-to-face interviews.

Initially, call data were used to monitor key fieldwork metrics such as contact and cooperation rates, numbers of fixed and vague appointments, and cases requiring tracking. From wave 4, these data were linked to sampling frame variables to assess and reduce non-response bias during ongoing fieldwork. By wave 5, PASS had introduced response propensity models and R-indicators (Schouten et al., 2009) for real-time bias monitoring within fieldwork.

Since wave 6, the core field monitoring system Trappmann & Müller (2014) has remained stable. In addition to weekly tabular reports from the survey agency, internal graphical analyses have been produced using call data. These visualise field progress through four key indicators at the household level i.e., contact rate, proportion of appointment address issues/ changes, cooperation rate, and response rate of successfully conducted interviews at the household level (panel households and the new sample households of the current wave). To evaluate fieldwork progress, wave-to-wave comparisons are also conducted. Complementary graphs display the share of

completed person interviews within households.

Around ten weeks into fieldwork (and for the foreign language field, around week 8/9), once all sample groups are in the field, R-indicators and bias analyses are produced. R-indicators, measuring contact and household interview success, are calculated for panel households and displayed as part of the field progress dashboard[3]. The accompanying bias analysis examines the distribution of key demographic variables such as age, education, gender, nationality, family status, and household type (e.g. single parent, couple with or without children), across panel households.

In wave 6, PASS first used call data to prioritise likely non-respondents and introduced enhanced interviewer incentives to mitigate bias. This wave also marked the initial effort to optimise call scheduling based on call history, aiming to reduce effort while improving contact and cooperation rates (Trappmann & Müller, 2014). These initiatives represented the first steps towards an adaptive survey design in PASS.

By wave 14, the design had evolved further. Beste et al. (2023) tested a machine learning-based adaptive approach, training models on past panel data to predict response propensities. The model identified households at higher risk of attrition, which were then prioritised during fieldwork. Increasing prepaid incentives from €10 to €20 for these high-risk cases led to a measurable reduction in drop-out rates and non-response bias.

For an overview of field monitoring indicators across major German social surveys, see Fieldwork Monitoring in Practice: Insights from 17 Large-scale Social Science Surveys in Germany (Meitinger et al., 2020). For theoretical perspectives on para data and its role in survey improvement, see Kreuter (2013).

For a deeper understanding of adaptive design, in addition to the publications mentioned above, refer to Chun et al. (2018) where the authors discuss four pillars of responsive and adaptive design (survey process data and auxiliary information, design features and interventions, explicit quality and cost metrics, and a quality-cost optimisation tailored to survey strata). They also highlight challenges in implementation.

Experimental evidence from Zhang & Wagner (2022) further evaluates the impact of adaptive design relative to post-survey adjustments, using data from the *Detroit Metro Area Communities Study*. Watson & Cernat (2023) compares short- and medium-term effects of fieldwork modifications in two household panels, namely, the Household, Income and Labour Dynamics in Australia (*HILDA*) Survey and the UK Household Longitudinal Study. Thus, showing that targeting decisions are more effective when made at the household rather than individual level.

[3] This was not done for wave 18 and the implementation is yet to be decided for wave 19.

Chapter 3

PASS Survey Design and Methodology - Weighting

Contact authors: Beste, Jonas; Trappmann, Mark.

Weighting constitutes an important part of the overall survey design and Methodology in PASS. Weighting methodology used in PASS ensures that survey estimates provide an unbiased representation of the target populations over time. **Section** 3.1 describes the construction of initial cross-sectional weights, including sample-design weights, non-response adjustments, and calibration procedures undertaken in the first wave. **Section** 3.2 explains how weights are updated for subsequent waves, detailing longitudinal weighting, the treatment of refreshment and replenishment samples, and adjustments for non-response and temporary drop-outs. **Section** 3.3 briefly discusses the integration of replenishment-sample weights into the ongoing panels. **Section** 3.4 summarises weighting datasets *hweights* (household weights) and *pweights* (person weights) and key variables they contain and their usage in analysis. Finally, **Section** 3.5 briefly describes the separate weighting scheme developed for the PASS web survey, which combines web-specific participation propensities with the existing PASS person weights to correct for selective web participation and to ensure comparability with the main survey.

3.1 Initial weighting process

PASS contains several sub-samples introduced across waves (see Section 2.1.5), including the initial welfare benefit (UB II)[4] recipient and general population samples, annual refreshment samples for new UB II entrants, and replenishment samples in Waves 5, 11, and 17. Each sub-sample receives its own weights in the wave in which it first appears. Its construction consists of three stages:

[4]Welfare benefit recipient sample and UB II recipient sample now replace the previously used terminology ‘‘BA sample’’ throughout this user guide. The 2 terminologies are used interchangeably to describe the Federal employment agency (German abbreviation BA) i.e., admin samples.

3.1.1 Stage 1: Design weights for the gross sample

Design weights equal the reciprocal of the selection probability for the gross sample. Detailed description of the generation of these variables is provided in Rudolph & Trappmann (2007) and in wave-specific data reports for the replenishment samples (e.g., Berg et al. (2013), Berg et al. (2019), Anker et al. (2025), chapter 6.1 for wave 5,11 and 17 respectively).

Design weights are stored in the dataset *hweights*. The individual design weights supplied, and their meanings are provided in the table below:

Table 3.1: Design weight variables in hweights dataset with meaning

Variable	Meaning
dw_ba	Design weight for welfare benefit recipient sample households (target population: households in which there was at least one benefit unit in joint receipt of benefits in accordance with Social Code Book II in any July since 2006)
dw_mi	Design weight for general population sample households (target population: households in the Federal Republic of Germany)
dw	Design weight for the combined sample (target population: households in the Federal Republic of Germany)

3.1.2 Stage 2: Modelling non-response

Two logit models estimate the participation probability for all households in the gross sample:

Model 1: Predicts the probability of **contact**

Model 2: Predicts the probability of **participation**, conditional on contact (at least one household interview + ≥1 complete personal interview)

These models are estimated separately for each sub-sample.

The set of variables used in the above-mentioned propensity models have evolved during the first waves of PASS. In Wave 1, the general population sample model included micro-geographical indicators provided by *Microm* alongside regional characteristics such as federal state and municipality size. Full model specifications can be found in the wave-specific method reports (Hartmann et al. (2008); Büngeler et al. (2009); and data reports from 2011 to 2025, eg. Anker et al. (2025)).

For the UB II recipient samples[5], additional characteristics on the level of the benefit unit from the sampling frames (*A2LL* and later *XSozial*) were included.

From Wave 3 onwards, *Microm* variables were dropped only variables on the level of benefit-unit, regional indicators, and the number of contact attempts were used for these samples. For the population replenishment

[5] Unemployment Benefits II is used interchangeably with various other terminologies across this user guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

samples age, gender and nationality were used in addition to regional predictors.

The product of the above-mentioned predicted probabilities from the two models is stored as *prop_t0* variable in *hweights* dataset.

Design weights divided by estimated participation probabilities yield modified design weights, which are used in the calibration stage (see below).

3.1.3 Stage 3: Calibration

A detailed documentation of the calibration process for Waves 1 and 2 can be found in Kiesl (2010). The calibration procedures and results reported by *TNS Infratest* in the method and field reports (Hartmann et al., 2008); (Büngeler et al., 2009) do not correspond to the weights provided in the Scientific Use File. From Wave 3 onward, calibration procedures are described in the wave-specific data reports produced by *infas* (e.g. Anker et al. (2025)). We therefore outline only the core principles here. The focus below is on calibration for the initial Wave 1 samples, as refreshment samples in later waves were calibrated jointly with the full sample (see Sections 3.2.7 and 3.2.8).

Household-level calibration (Wave 1)

In the first step, both sub-samples and the combined total sample were calibrated to official household statistics.

Total and UB II sample weights for benefit recipients were aligned with benchmark statistics from the Federal Employment Agency (as per the monthly reporting of July 2006).

Total and *Microm* sample weights were additionally calibrated to private-household benchmark statistics for Germany (as per the annual reporting of 2007) from the Federal Statistical Office. These benchmark figures are documented in Kiesl (2010).

All PASS weights are household weights, whereas the BA benchmarks refer to benefit units. To align these levels, *synthetic benefit units* were created as described in the Wave 1 data report (see Christoph et al. (2008), p. 49; variable *bgnr1* in *p_register* dataset). Households were first divided into synthetic benefit units, and calibration characteristics were generated at the benefit-unit level, including whether the unit received UB II at the sampling date. After calibration, multiplying these characteristics of benefit units in receipt of benefits, as of the sampling date, by household projection factors produced benchmark totals. All benefit units receiving benefits within the same household received identical projection factors.

Determining benefit-receipt of a household or even a benefit unit at the index date is not always straightforward. To support users, indicator variables are provided. For instance, the variable *alg2samp* at the household level is supplied in the *hh_register* dataset. This variable describes the benefit receipt as of the sampling date for all households in the categories:

0 - no receipt, 1 - receipt, 2 - no receipt according to survey (but included in UB II recipient sample and thus receipt according to register data), 3 - receipt unclear from survey (but included in UB II recipient sample and thus receipt

according to register data), 4 - receipt unclear from survey (general population sample).

In addition, every user can generate this variable him/herself using the unemployment benefit II spell data (*alg2_spells* dataset). Other useful variables are *AL20600* and *AL20700a--o* (describing the member/s in the household which receive benefits).

To generate the weights, it was imperative to determine which benefit units should be classified as receiving Unemployment Benefit II (UB II) at the sampling date. The following criteria were applied for this classification are as follows:

1. *UB II recipient sample households* (sample = 1)

   All households drawn from the BA register were classified as receiving UB II at the sampling date, even if the respondent denied benefit receipt, provided at least one household member was aged between 15 and 64 years.

2. General population sample households (sample = 2)

   In cases where survey data did not clearly establish benefit-receipt status, households were treated as receiving UB II for weighting purposes if both conditions held:

   - the household reported having ever received UB II (*HA0300 = 1*), and
   - the start or end date of at least one UB II spell fell in 2006 (including cases with unspecified start or end date).

Determining benefit receipt at the benefit-unit level presents greater uncertainty since it is not possible to obtain reliable retrospective data on benefit recipients in the household in July 2006. For most households, this is straightforward, as they consist of a single benefit unit that receives benefits if the household does. However, in households with multiple benefit units, determining benefit receipt becomes particularly difficult. So, the following procedure was applied:

- Information pertaining to individuals within the household currently receiving benefits (*AL20600* and *AL20700a--o*) was used.
- A benefit unit was classified as receiving UB II if at least one of its members was reported as a benefit recipient.
- In a household with more than one benefit unit and with no information as to which individuals the household is receiving benefits for (e. g. because the questionnaire responses state that no benefits are being claimed), all the synthetic benefit units were regarded as being in receipt of benefits.

This procedure results in the variable *bgbezs1* in the *p_register* dataset.

The household-level calibrated weights derived from this process are stored in the *hweights* dataset:

- **wqbahh** --- calibrated household weight for the welfare benefit recipient sample
- **wqmihh** --- calibrated household weight for the general population sample
- **wqhh** --- calibrated household weight for the total sample

Individual-level calibration (Wave 1)

Following household-level calibration, individuals who completed a personal or senior citizen's interview were calibrated to benchmark statistics at the person level. The calibrated household weights served as the starting point for this step.

The total and BA weights for benefit recipients in both sub-samples were calibrated to benchmark statistics from the Federal Employment Agency (as per the monthly reporting of July 2006). The total and general population sample weights were additionally calibrated to benchmark statistics from the Federal Statistical Office on private households in Germany for 2007. The benchmark figures used are detailed in Kiesl (2010).

Senior citizen's interviews were calibrated to population statistics in the same way as standard personal interviews. However, the BA statistics do not contain figures on the number of senior citizens in households receiving benefits, nor do they identify individuals living in benefit-receiving households that are not part of a benefit unit. Consequently, it was not possible to obtain the BA person weights for these individuals by means of calibration.

Instead, we estimated the participation probability of these individuals, conditional on their household participating in the survey, using a logit model with the following covariates: number of individuals aged 15 and over in the household; interview mode; age; and gender. The modified design weight was subsequently divided by this value.

The calibrated person weights are contained in the *pweights* dataset:

- **wqbap** --- calibrated person weight of the UB II recipient sample
- **wqmip** --- calibrated person weight of the Microm sample
- **wqp** --- calibrated person weight of the total sample

3.2 Construction of the weights from wave 2 onwards

The starting point for the weighting procedure in wave 2, and for the longitudinal section from wave 1 to wave 2, is the set of cross-sectional household and individual weights from wave 1. More generally, for wave $n+1$ and for the longitudinal section from wave n to wave $n+1$, the procedure begins with the cross-sectional weights from wave n for both households and individuals.

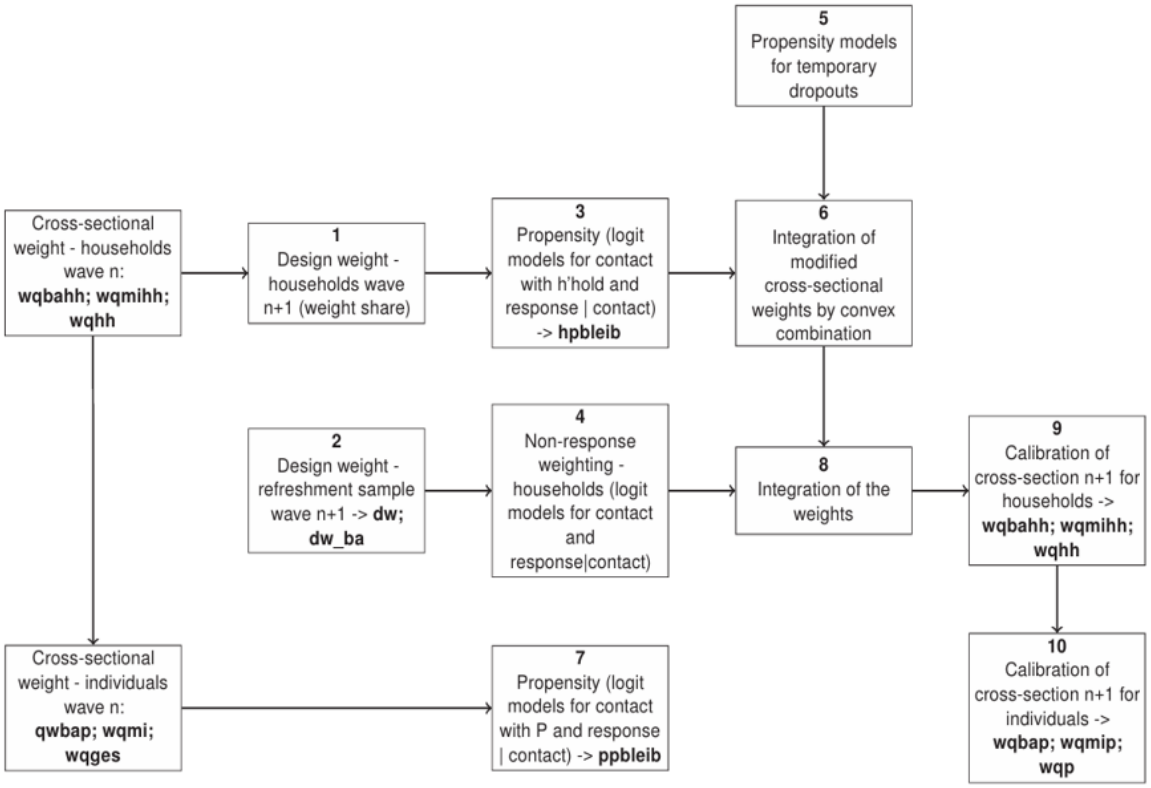


Figure 3.1: Generation of the weights for wave $n+1$ given the weights of wave n

In wave n ($n \geq 1$), each household has two weights: *wqhh* (the calibrated total weight) and depending on the sub-sample either *wqbahh* (calibrated UB II recipient sample weight) or *wqmihh* (calibrated general population weight). Similarly, each individual has two weights: *wqp* and depending on the subsample either *wqbap* (calibrated UB II recipient sample personal weight) or *wqmip* (calibrated general population personal weight). All four weights are updated for the subsequent wave ($n+1$)

Figure 3.1 presents the steps in the weighting procedure, which are described in detail below. This section provides a general overview. From wave 3 onwards, chapter 6 in each wave-specific data report (e.g. Berg, Cramer, Dickmann, Gerber, et al. (2024)) documents the exact models, variables, and coefficients applied.

For wave 2, detailed documentation is available in Büngeler et al. (2009). This section does not discuss the integration of replenishment-sample weights into the ongoing panel; Section 3.3 addresses this topic in detail.

3.2.1 Design weights for the wave n households in the $n+1$ th wave.

For wave $n+1$, new household design weights are generated from the wave n cross-sectional household weights, accounting for individuals moving into existing sample households from within Germany. This adjustment is carried out using a *weight-share* procedure.

Household events such as births, deaths, or moves out of the household do not affect the weight. By contrast, moves into a household from within Germany increases the inclusion probability, because these individuals also

had a chance of being sampled in previous waves. Therefore, for the weighting calculations, when individuals move into a sample household from within Germany, the household's prior inclusion probability is increased by the *mean inclusion probability* in the respective sub-sample (as the inclusion probabilities of the households in which new household members previously resided cannot be determined precisely for all earlier waves).

Consequently, the new design weight for sub-sample i $dw_i hh_{n+1}$ is therefore calculated from the old cross-sectional weight $wq_i hh_{n:}$

$$\frac{1}{dw_i hh_{n+1}} = \frac{1}{wq_i hh_n + \left(\frac{n_{sample\,i}}{n_{population\,i}}\right)}$$

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

3.2.2 Design weights for the wave $n+1$ refreshment sample

From wave 2 onwards, PASS draws an annual refreshment of the welfare benefit recipient (UB II) sample, selecting only benefit units (*Bedarfsgemeinschaften*) in which no member received benefits in July of the previous years. These refreshment samples are drawn from the same sampling points as the initial samples, with extensions to the 97 additional sampling points introduced in wave 5 and the 93 additional points added in wave 17.

Similar to the probability-proportional-to-size procedure used to draw the initial register data sample (Rudolph & Trappmann, 2007), the refreshment sample size in each sampling point is proportional to the share of new benefit recipients in the population at the time those sampling points were selected. The calculation of the design weights is further described in the same article. However, from wave 2 onwards, the number of benefit units within a household is no longer taken into account. The resulting design weight for refreshment sample is provided in the variable *dw_ba* for all cases.

3.2.3 Response propensity models for panel households

In this step, the probability of continued participation is estimated for each household that participated in the previous wave. This uses logit models capturing (1) panel consent, (2) risk of loss of contact, and (3) likelihood of refusal. These models include survey design features (e.g. interview mode, number of contact attempts), aspects of the previous interview (e.g. item non-response or partial non-response), household/respondent characteristics (e.g. gender, age, education, country of birth, labour force status, home ownership, household size), and area characteristics (e.g. municipality size), consistent with standard practice in longitudinal studies (Watson & Wooden, 2009).

The predicted propensities from the three models are multiplied. The reciprocal of this product is stored in the variable *hpbleib*, which serves two purposes:

- The longitudinal weight of a household for the period $wave_n$; $wave_{n+k}$ between waves can then be

calculated as the product of the cross-sectional weight for wave_n and the product of all *hpbleib* for wave n to wave $n + k - 1$.

- The product of the updated household design weight from step 1 (see Section 3.2.2) multiplied by *hpbleib* (which we call ‘‘modified cross-sectional weight’’) serves as a base for calculating a new cross-sectional weight for wave $n + 1$.

Note that this procedure works only for households with monotonous drop-out patterns and not for households that drop-out for one wave and return in the next wave. The treatment of those temporary drop-outs is specified in Section 3.2.5.

3.2.4 Non-response weighting for households in the wave n refreshment sample

For households drawn in the refreshment samples, non-response is modelled in a two-stage procedure, analogous to the approach used for the first wave. Full variable lists and model coefficients are documented in the respective wave-specific data reports. The estimated participation probability is stored in the variable *prop_t0.*

3.2.5 Propensity models for temporary drop-outs

From wave 3 onwards, some households in the PASS dataset returned to the panel after temporarily dropping out[6]. Longitudinal weights cannot be applied to these groups, since weighted longitudinal analyses can only be conducted using a balanced panel of households that participated in all waves within the relevant period. Allowing for non-monotonous participation patterns would result in an exponentially increasing number of weights as waves progress (Lynn & Kaminska, 2010).

For temporary drop-outs, the procedure is as follows:

First, the probability of dropping out in wave n given participation in wave $n - 1$ is derived from the propensity model for the transition from wave $n - 1$ to n [7]. Then, a simplified propensity model is estimated to predict the probability of returning in wave $n + 1$ given drop-out in wave n. This model includes only the final disposition code of the previous wave, interview mode, sample indicator, and whether the household is a split-off household.

The reciprocal of the product of these two predicted probabilities is multiplied by the calibrated household weight from wave $n - 1$. This produces a modified cross-sectional weight that serves as the basis for computing the new cross-sectional weight for wave $n + 1$.

[6] In PASS, households that fail to participate in two consecutive waves are no longer contacted.
[7] This can simply be calculated as 1–*hpbleib* for that wave.

3.2.6 Integration of weights by convex combination

Temporary drop-outs originate from the same population for which new base weights were calculated in Step 3 (Section 3.2.3). Therefore, integrated weights can be calculated as a convex combination of the modified cross-sectional weights for the two sub-samples (Spiess & Rendtel, 2000). The precise formulae are provided in Chapter 6 of the respective wave-specific reports (e.g. Anker et al. (2025)).

3.2.7 Response propensity models for panel persons

The key longitudinal weight in PASS is the individual-level weight, as individuals form the stable units over time. Participation propensities for individuals with monotonous drop-out patterns are modelled in the same way as the model for households shown in Section 3.2.3. Since household participation is a prerequisite for individual participation, the models include similar predictors, supplemented by individual-level characteristics (e.g. age, item non-response in the previous wave).

The predicted probabilities of the models are multiplied, and the reciprocal of this product is stored in *ppbleib*. The longitudinal individual weight for the period $wave_n$; $wave_{n+k}$ is calculated as the product of the cross-sectional weight at wave n and all *ppbleib* values from wave n to $n + k - 1$. Full model specifications and coefficients appear in the field and method report for wave 2 by *TNS Infratest* (Büngeler et al., 2009) and in the *infas* data reports (Berg et al., 2011) from wave 3 onwards. Temporary drop-outs again require separate treatment.

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

This step integrates the weights of the most recent refreshment sample with those of the panel households that were modified through non-response modelling (Sections 3.2.3, 3.2.4) and the temporary drop-out procedure (Section 3.2.6). The one-time integration of replenishment samples is discussed in Section 3.3.

In theory, a small group of new benefit recipients may have a double selection probability (if they were not in receipt previously but lived in the same household as earlier recipients without belonging to the benefit unit). PASS assumes this population is negligible and does not adjust for it.

This constellation of cases is expected to be extremely rare because it requires four conditions to be simultaneously met:

1. the individual must be in receipt of benefits at the reference date in the current wave;
2. they must not have received benefits at any previous reference date;
3. they must currently reside in a household with someone who was a benefit recipient at a previous reference date; and

4. they must not have lived with that benefit recipient at any earlier reference date.

Given the very low likelihood that all four criteria occur together, the two sampling frames can be treated as disjoint for practical purposes. Consequently, the register-sample weights remain unaffected by the integration of refreshment cases, and equally, the general population sample weights are unaffected.

In the cross-section, the new design weights for the benefit-recipient sample therefore project to all individuals who lived in a household containing at least one benefit unit at any of the reference dates (e.g. July 2006 for wave 1, July 2007 for wave 2 and so on). Adjustments are only required when deriving weights for the combined total sample, as all households with benefit receipt at any previous reference date must be assigned the correct inclusion probability.

For this adjustment, the inclusion probability in the other sampling frame is estimated. Specifically:

- for **refreshment-sample cases**, the mean selection probability and average participation probability of the general population sample within the same postcode sector are assumed;
- for **general-population cases** who (according to survey data) became benefit recipients between the wave 1 sampling date and the sampling date of any refreshment sample, the mean selection probability and average participation probability of the refreshment sample within the same postcode sector are assumed.

Finally, the weights derived in Sections 3.2.4 and 3.2.6 are combined to create a unified total weight.

3.2.9 Calibration to the household weight, wave $n + 1$, cross-section

After integration, a further calibration of weights from step 6 is performed. For households, raking (in wave 3) and a GREG estimator in all subsequent waves is used. This calibration aligns weights with official statistics from the Federal Statistical Office for the relevant survey year (for e.g. 2007 in wave 2) and, for benefit-recipient households, statistics from the Federal Employment Agency (reference month July). Detailed procedures appear in Kiesl (2010) for waves 1 and 2, and in the *infas* data reports from wave 3 onwards (e.g. latest report Anker et al. (2025)).

3.2.10 Calibration to the person weight, wave $n + 1$, cross-section

As in wave 1, person-level weights are calibrated under the constraint that they deviate as little as possible from the calibrated household weights. The calibration is therefore not based directly on the person weights of the previous wave. Details are provided in Kiesl (2010) and in wave-specific *infas* reports (e.g. Anker et al. (2025)).

3.2.11 Estimating BA cross-sectional weights for non-recipient Households and Individuals

In some cases, households and individuals cannot be assigned BA cross-sectional weights through calibration alone. These units did not receive Unemployment Benefit II at any reference date after Wave 1, but they still belong to the BA target population. That is, they lived in a household receiving UB II at one of the following reference sampling dates (for e.g., July 2006, July 2007, and so on). Three specific groups are affected:

1. Individuals in refreshment-sample households who are not members of a benefit unit

For these individuals, the BA person weight is derived from the calibrated BA household weight (*wqbahh*). It is calculated by dividing *wqbahh* by the proportion of such individuals who completed either a personal interview or a senior citizen interview, conditional on the household's participation.

2. Panel households no longer receiving UB II at the current reference date

The household retains its pre-calibration welfare benefit recipient sample weight. Individuals interviewed in both waves receive a new UB II recipient sample equivalent person weight equal to their previous UB II recipient sample equivalent person weight multiplied by the reciprocal of the re-participation probability (*ppbleib*). Individuals who did not complete an interview in the previous wave receive a UB II recipient sample equivalent person weight calculated by dividing the UB II recipient sample equivalent household weight for wave $n + 1$ by the participation proportion of such individuals, conditional on household participation.

3. Individuals not belonging to a benefit unit in panel households still receiving UB II

For these individuals, if they completed interviews in both waves, their new UB II recipient sample equivalent person weight equals their previous UB II recipient sample equivalent person weight multiplied by the reciprocal of the re-participation probability (*ppbleib*).

3.3 Integration of weights from replenishment samples

Replenishment samples were introduced in several PASS waves to counteract attrition and maintain representativeness. These include replenishment samples of the general population (Waves 5, 11, and 17) and of the benefit-recipient population (Wave 5). The procedures for integrating weights from these samples are documented in Chapter 6 of the wave-specific data reports for Waves 5, 11, and 17 (Berg et al. (2013); Berg et al. (2019); Anker et al. (2025)).

3.4 Weighting datasets and weighting variables

The weighting datasets *hweights* (household weights) and *pweights* (person weights) are organised as long files, consistent with the individual and household data files.

The *hweights* file contains the following variables:

Table 3.2: Variables in the household weights dataset (hweights)

Name	Label	Remarks
hnr	Household number (current)	Used together with welle for linking the datasets.
welle	Indicator for survey wave	Used together with hnr for linking the datasets.
sample	Subsample	Indicates whether UB II recipient sample weights or Microm weights are used.
dw_mi	Design weight – Microm sample	Selection probability (during sampling) in this subsample (gross).
dw_ba	Design weight – UB II recipient sample	Selection probability (during sampling) in this subsample (gross).
dw	Design weight – total sample	Selection probability (during sampling) in the total sample (gross).
prop_t0	Participation probability in the sampling year of the subsample	Probability that the household takes part in the year the subsample was drawn (logit model).
wqhh	Projection factor – household (total)	Cross-sectional projection factor for the respective wave (total).
wqmihh	Projection factor – household (Microm)	Cross-sectional projection factor for the respective wave (Microm).
wqbahh	Projection factor – household (UB II recipient sample)	Cross-sectional projection factor for the respective wave (UB II recipient sample).
hpbleib	Reciprocal re-participation probability – household w_n to w_n+1	Reciprocal value of the probability that the household participates again in the following wave, as predicted by a logit model.

The *pweights* file contains the following variables:

Table 3.3: Variables in the person weights dataset (pweights)

Name	Label	Remarks
pnr	Unchanging personal ID number	Used together with welle for linking the datasets.
welle	Indicator for survey wave	Used together with pnr for linking the datasets.
sample	Subsample	Indicates whether UB II recipient sample weights or Microm weights are used.
wqp	Projection factor – person (total)	Projection factor for the cross-section of the respective wave (total).
wqmip	Projection factor – person (Microm)	Projection factor for the cross-section of the respective wave (Microm).
wqbap	Projection factor – person (UB II recipient sample)	Projection factor for the cross-section of the respective wave (UB II recipient sample).
ppbleib	Reciprocal re-participation probability – person w_n to w_n+1	Reciprocal value of the probability that the individual participates again in the following wave, as predicted by a logit model.

3.5 Web-survey weighting

Alongside the standard weighting procedures for PASS, a separate weighting scheme was introduced for the PASS web survey, which was first administered after Wave 16 (2022). Since only a subset of Wave 16 respondents participated in the web survey, separate web survey weights were required to correct for selective participation.

To construct these weights, a statistical model was estimated to predict the probability that a Wave 16 participant would take part in the web survey. The inverse of the predicted probability served as the propensity weight. These weights were then multiplied by the existing PASS person weights from Wave 16 to produce calibrated weights for the web-survey dataset.

This procedure corrects for systematic differences between participants and non-participants in the web survey, aligning with established principles of propensity-score weighting and is conceptually in line with the established weighting procedures. The same approach will be applied in future waves of the PASS web survey.

Further details on the propensity model and its specification are provided in FDZ Data Report 13/2023 (Jesske & Gerber (2023)).

Chapter 4

Survey Instruments

Contact authors: Bähr, Sebastian; Collischon, Matthias; Müller, Marcel; Fehn, Anja.

This chapter provides an overview of the survey instruments used in the PASS, describing the structure, content, and evolution of its questionnaires. It explains how the survey captures multidimensional information on households and individuals, enabling analyses of welfare dynamics, employment, and living conditions over time. For a detailed information on important variables in PASS across waves, please refer to the scales manual (B. Müller et al., 2020).

Section 4.1 introduces the PASS questionnaire framework, outlining the separate household, individual, and senior citizen questionnaires and their core content. **Section** 4.2 explains the logic and development of questionnaire modules, including the rotation of thematic blocks, the integration of topical additions such as migration and COVID-19, and the adaptation of instruments to social and policy changes. **Section** 4.3 presents the key variables across the household and individual questionnaires, focusing on core modules related to welfare benefits, employment biographies, deprivation, health, and subjective well-being. These sections demonstrate how PASS combines consistent longitudinal measurement with flexible modular design to capture the evolving realities of labour market participation, welfare receipt, and social conditions in Germany.

4.1 Questionnaire

The PASS survey covers a broad range of information on individuals and their households. Similar to other household surveys, PASS employs different questionnaires for households and individuals.

First, the head of each household answers a household questionnaire, which covers a range of variables that are described in the following sections. Second, each household member aged 15 years or older participates in a personal interview that captures individual-level characteristics. For individuals above the legal retirement threshold in Germany, we use a shortened Senior Citizens Questionnaire to reduce respondent burden.

This chapter provides an overview of the questionnaire structure, the specific modules included, and key variables of potential interest to researchers.

4.2 Logic of questionnaire modules within PASS

PASS was conceived as a specialized survey of individuals and their household members receiving Unemployment Benefit II[8] (''Bürgergeld'' or UB II). Respondents are often unemployed or working-poor households suffering from (frequently multiple) impediments to finding self-sustaining employment. Consequently, central topics for the survey include their economic situation, deprivation, job search, and labour market biographies, details on the receipt of their benefits, and their interactions, and experiences with the job centres. As the welfare benefits system strives to foster the social integration of its recipients, PASS also includes survey instruments for multiple dimensions of social integration.

Since its inception, the PASS survey program has reflected the changing institutional, societal, and political environment surrounding the welfare state and German society in general. With an increase in refugee migration to Europe starting 2015, migrants have comprised a growing share of the new entrants into the UB II system (see 6.1). The ensuing need for data about this new recipient type led to extensions in the specific questionnaire programme. Similarly, the COVID-19 pandemic necessitated short-term question additions during the fieldwork in 2020 (see 6.4). PASS strives to inform social policy by collecting data on welfare reforms like the Educational package (''Bildungspaket'') introduced in 2012 or the Citizen's Benefit reform of 2023. In addition, PASS also acts as a vehicle for research projects on welfare benefits and labour market issues, which has led to the inclusion of experiments and instruments in a limited number of survey waves. Given this comprehensive list of objectives, balancing (longitudinal) research interests with a limited survey duration is necessary. So, in addition to essential core modules, more specialized modules are regularly rotated in and out of the survey or run for only a limited number of years. Table 4.1 and Table 4.2 provide a thematic overview of the household and individual modules and displays their availability by year and survey wave, while Section 4.3 highlights and describes essential survey modules.

Table 4.1: PASS thematic blocks at the household level (Waves 1–18)

Thematic blocks at the household level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Household Composition																		
Household Structure	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Language Spoken in Household		x	x															
Living Situation		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

[8] Unemployment Benefits II is used interchangeably with various other terminologies across this user guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

Table 4.1: PASS thematic blocks at the household level (Waves 1–18) (continued)

Thematic blocks at the household level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Affectedness of the Household during COVID 19														x				
Income and Deprivation																		
Household income	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Funds and Debts	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Deprivation	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Deprivation of Children										x	x	x	x	x	x	x	x	x
Household Economics and Economic Practices															x	x	x	
Household Economics and Worries about the Future																		x
Social Security																		
Knowledge of Social Security Benefits																		x
Unemployment Benefits 2 / Citizens Benefits	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Non-Usage of Unemployment Benefits 2 / Citizens Benefits																		x
Costs of Accomodation																		x
Usage of Food Banks (”Tafeln”)								x										
Non-usage of Basic income in old age (”Grundsicherung im Alter”)																		x
Child Care and Child Welfare																		
Child Care	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Child Care During the COVID19 Crisis														x	x	x		
Social Participation of Children and Youths						x	x	x	x	x								
Receeiving measures from the Educational Package (”Bildungspaket”)						x	x	x	x	x								

Table 4.2: PASS thematic blocks at the individual level (Waves 1–18)

Thematic blocks at individual level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Demographics																		
Date of Birth	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Additional Demographics					x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 4.2: PASS thematic blocks at the individual level (Waves 1–18) (continued)

Thematic blocks at individual level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Migration	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Migration: Recognition of Foreign Qualifications											x	x	x	x	x	x	x	x
Social Origin	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Religion	x	x		x	x	x	x	x	x	x	x	x	x	x	x			
Affectedness of the Inividual during COVID 19																		
Affectedness of the Inividual during COVID 19														x	x	x		
Education and Vocational Training																		
Education	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Educational Aspiration											x	x	x	x				
Income (Vocational Training)		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Employment																		
Employment Biography (Spells)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Mini Job	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Minimum Wage					x	x	x	x	x	x								
Quality of Employment							x	x	x	x	x	x	x	x	x	x	x	x
Crowdworking												x	x					
Change of Working Hours									x	x								
Employment Biography (general)		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Latent Functions of Work														x	x	x		
Unemployment Benefits 2 / Citizens Benefits																		
Receipt of Unemployment Benefit (Arbeitslosengeld 2) Since 2005											x							
Agency Contacts	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x
Agency Contacts (Special Focus on Sanctions)														x				
Cooperation Plan																		x
Obligation to Look for a Job							x											
Measures	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Job Search	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Further Training																	x	x
Stigma Awareness							x										x	x
Attitudes																		
Attitudes (General Situation)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Attitude (Life)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 4.2: PASS thematic blocks at the individual level (Waves 1–18) (continued)

Thematic blocks at individual level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Attitudes (Self Efficacy)	x	x	x								x			x			x	x
Attitudes (Employment)	x		x	x	x	x	x	x										
Attitudes (Work Hours)					x				x									
Attitudes (Childcare)										x								
Opinions (Gender Role Models)	x	x			x			x			x			x			x	
Attitudes (Handling Money and Relationships)		x			x													
Attitudes (Job and Family)					x					x						x		
Attitudes (Reciprocity)									x	x								
Attitides (Standard of Living)	x							x										
Attitudes (Finances)								x			x							
Attitudes (Children's Leisure Time)							x	x	x	x								
Attitudes (Distribution)																		x
Worries about the Future																		x
Social Justice I						x												
Psychological Scales																		
Impulsiveness										x								
Personality: Big Five (BFI-K)					x								x					
Factorial Survey Instruments and Experiments																		
Vignettes (Job Attractiveness)					x													
Vignettes (Job-Seeking)												x						
Vignette Online Module (Maternal Employment)															x			
Trust Experiment														x				
Social and Societal Participation																		
Social Participation	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Political Participation													x	x	x	x	x	x
Social Trust											x	x	x	x	x	x	x	x
Networks	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Networks (Special Focus)			x		x				x			x						
Leisure Time		x	x	x	x	x	x	x	x	x								
Internet and Social Media																		
Internet and Device Onwership																x		
Social Media Use						x	x	x										x
Smartphone Ownership											x							
Health and Sports																		

Table 4.2: PASS thematic blocks at the individual level (Waves 1–18) (continued)

Thematic blocks at individual level	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14	W15	W16	W17	W18
Health	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Health (Special Focus)			x			x			x			x				x		
Memory - Experiment							x											
Sports						x	x	x										
Social security																		
Nursing Care	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Income (Pensions)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Income (Pensions - Special Focus)												x	x	x	x	x	x	x
Social Benefit		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

4.3 Key variables across the two questionnaires

4.3.1 Household interview

One of the main functions of PASS is to analyse individuals within households receiving long-term unemployment benefits, and the household survey contains a detailed questionnaire on this topic. Specifically, we collect the duration and dynamics of receiving welfare benefits in the form of UB II. The data is gathered retrospectively through the household questionnaire. For newly interviewed households (e.g., of the refreshment sample), information on benefit receipt is collected for approximately two years prior to the respective wave. For example, in wave 18 (2023), information regarding all episodes since January 2021 was collected. For re-interviewed households, dependent interviewing is used, and benefits receipt is always recorded for the period since the last interview date. Each episode contains information on its beginning and end (month and year).

Several income measures can be used to assess the material situation of households (‘‘resource approach’’). At the household level, information is available on the total disposable household income, savings, loans and debts, social security benefits, and household transfers.

As income measures are prone to short-term fluctuations (e.g., due to changes in household composition or employment status of household member/s), the standard of living is measured more comprehensively by accounting for economic deprivation. Deprivation can be defined as the lack of basic goods considered essential for an adequate standard of living. PASS includes a weighted deprivation index comprising 26 items until wave 4 and of 23 items in the subsequent waves, covering essential goods and technical equipment (like having an apartment with a separate bathroom or balcony, television or DVD player), spendings satisfying basic needs (such as having a hot meal per day or buying necessary medication), and leisure activities (like going to the movies, or inviting friends for dinner at home occasionally). The index reflects the number of items respondents

lack for financial reasons. In its weighted version, each item is weighted by the proportion of respondents who consider it essential. Since wave 10, the household questionnaire also includes questions about deprivation among children in the household, including four items concerning possessions or activities pursued by children in the household.

4.3.2 Individual interview

Due to the nature of PASS, the individual questionnaire also contains many questions on welfare benefits and related topics. Household-level information on UB II receipt is complemented by two questionnaire modules at the individual level, which all individuals living in UB II-recipient households answer. The first module includes questions about individual's contact with the agencies responsible for providing UB II (e.g., contact frequency, perceived quality, and intensity of support). The second module consists of retrospective questions about individual participation in labour market policy programs.The concept of surveying participation in *Active Labour Market Programmes (ALMP)* was thoroughly reworked in the first waves (see Gebhardt et al. (2009)). In particular, difficulties in identifying the exact type of employment and training measures necessitated this revision. Since the last revision (wave 4), the module exclusively relates to one-euro-jobs (start and end of the programme, evaluation of institutional support during the programme and perceived success).

Information on unemployment insurance payments (so-called UB I) is also collected at the individual level. Questions about the start and end date of the benefits receipt and reasons for cessation of unemployment are integrated into the biography module, where the employment history of all participants is gathered. All episodes of employment, unemployment, and non-employment (e.g., education, maternity leave, retirement) beginning approximately two years before an individual's first interview, are recorded, and are updated in each panel wave. The start and end dates (month/year) of each episode are surveyed to ensure a chronological and complete employment history (furthermore, parallel activities are to be reported here, and a control mechanism identifies gaps of more than one month between the reported activities), for each job episode detailed information is obtained like occupation, skill level, contract type, working hours, income, industrial sector, firm size. Similar information is also collected for the first job ever held. For mini-jobs, only cross-sectional data is collected (working hours and income for the current mini-job).

By combining and linking the history of UB II receipts and (un)employment history at both individual and household levels, PASS provides excellent data to explore the pathways that lead into and out of UB II receipts and for assessing the sustainability of these exits.

PASS also includes a module on quality of employment from wave 7 until wave 18. The questions are asked to both dependent employees and people with mini-jobs. In this module, respondents are asked about various characteristics of, and attitudes towards, the quality of their employment. For example, the module contains information regarding working time regulations, work flexibility, the content of the employment, and general

satisfaction with the employment. In addition, this module provides data about career opportunities and possible stress at work. The exact content of the questionnaire varies across respective survey waves.

Besides the above-mentioned information, data availability through agency contacts is considered key information that can be obtained via PASS. Except for wave 7, the *Agency Contacts Module* is collected annually. This module was updated in wave 9. The questions in this module concern the contact with the agency responsible for the *Citizen's Benefit*. The agency is known in Germany by various names, such as *''Jobcenter'', ''ARGE'', or ''Arbeitsgemeinschaft für Beschäftigung''.* A new focus on sanctions was introduced in wave 14 but removed in wave 15.

Before the 1 January 2023, the Citizen's Benefit (''*Bürgergeld*'') was known as ''*Arbeitslosengeld 2*'' or ''*Hartz 4*'' and sometimes followed different rules. For this reason, the wording of some questions in waves 17 and 18 differs from previous waves, and some new questions have been added since then. Agency contacts module generally focuses on the quantity and quality of contacts with the employment agency. In other words, how often did personal contact occur, which channels were used for these contacts (face-to-face, telephone, videophone, online, or by post), and what was the content of these conversations. It is also important to determine how satisfied recipients were with their agency contacts, which topics were discussed, and whether sufficient time was available for consultation. Questions regarding these elaborate conversations with the agency, cover both professional and private issues, such as psychological stress or physical illness. From wave 17 onwards, respondents were also asked whether they would have preferred more elaborate conversations about their professional and private situations. In addition, recipients were asked whether they were required to search for work or faced barriers preventing placement. The focus here is on the nature of these barriers to placement. Respondents were also asked whether they were offered different employment models or other financial, family, or general support. Items related to language support have been repeated from wave 17 onwards.

Since the introduction of the *Bürgergeld* reform in 2023, the survey has collected information on the *Kooperationsplan*, a contract between Jobcentres and benefits recipients, outlining duties and agreements on the steps that the individuals must take inorder to re-integrate into the labour market. The survey records whether such an agreement exists, what it includes, and details about the process. The specific questions can be found in the questionnaire.

The second main research question for PASS concerns individual's living conditions. Therefore, PASS includes a broad range of questions on individual health, elderly care, social networks, social inclusion and political integration. The following two dimensions should be described in detail: financial situation and health.

Individual income components are gathered mainly within the personal questionnaire. They refer to income from work (including earnings from self-employment or special payments), unemployment insurance payments, and statutory or civil servant's pensions.

A basic health module is included in each wave, supplemented by an extended health module in selected waves. The basic module includes self-reported health indicators referring to mental and physical well-being: number of consultations, hospital stays in the last twelve months, chronic diseases and disabilities, mental health problems, subjective health status, and health satisfaction. In addition, several waves include extended health questionnaires covering health-related behaviour, such as drinking and smoking, obesity, exercise, and health-related employment limitations. The module also includes the standardized SF-12 item battery as a global mental and physical health measure.

Referring to the third core dimension of the study (coping strategies, attitudes, and behaviour), various subjective indicators are gathered at the individual level, such as global aspects of personality traits and situation-specific behaviour or attitudes (see PASS Scales and Instruments Manual 2020). The basic questionnaire includes perceived self-efficacy, general life satisfaction, satisfaction with health and living conditions, religiosity and confession, employment orientations, and the respondent's subjective perception of social integration. The module on job search behaviour is another essential component of the questionnaire. It includes information on search duration, job-search activities, reasons for not looking for a job, preferences regarding job attributes (e.g., working hours), and willingness to make concessions (e.g., wages, commuting distance, qualification mismatches, reservation wages). Some subjective indicators are not included in every wave (see Tables 4.1 and 4.2), e.g., gender-role attitudes, financial behaviour, partnerships, or aspirations for children's education and leisure activities. Furthermore, waves 5 and 13 include the *Big Five personality-traits module*, a standardised 21-item battery covering five broad domains.

By addressing long-term benefit receipt, employment biographies, health, deprivation, and agency interactions, both at individual and household level, PASS provides a multidimensional view of the socio-economic conditions of individuals and households in Germany. The combination of retrospective and panel data enables longitudinal analyses of welfare dynamics, labour market participation, and living conditions, while also providing the flexibility to incorporate ad-hoc, short term addition of modules to capture additional areas of research interest.

Chapter 5

PASS SUF Datasets and Variables: Collection, Editing and Usage

Contact authors: Beste, Jonas; Chauhan, Madhu; Frodermann, Corinna; Dummert, Sandra; Malich, Sonja; Prospero, Valentina; Wenzig, Claudia.

The data collected in each wave of PASS is anonymised and made available to the scientific community for secondary analysis in the form of a *Scientific Use File (SUF)*. The raw survey data collected in each wave are processed through a series of editing, coding, and documentation steps to ensure consistency, accuracy, and comparability over time.

The SUF contains a rich combination of household- and individual-level information, covering both cross-sectional and longitudinal components, along with episode-based (''*spell*'') data.

In addition to the survey responses, the SUF incorporates structural identifiers, weighting factors, and metadata required to link datasets across waves. Its design allows researchers to conduct both descriptive and analytical studies, ranging from short-term behavioural assessments to long-term panel analyses of households and individuals in Germany.

This chapter provides a detailed guide to understanding and working with the SUF. There are four major sections within this chapter. **Section** 5.1 describes the structure of the SUF, including core and temporary[9] datasets, their levels, types, and formats, as well as the role of key and pointer variables in linking records. **Section** 5.2 explains the organisation of variables, covering naming conventions, classification into system and surveyed variables, and the use of standardised codes. **Section** 5.3 outlines the data-editing process carried out prior to the SUF release, including household and individual structure checks, filter verification, plausibility testing, and the integration of episode-based information. Finally **Section** 5.4 introduces the linkage of PASS data with

[9] These datasets have been referred to as ''discontinued'' datasets in the older PASS data reports, method reports and other PASS related documents.

administrative records and additional datasets such as PASS-ADIAB.

Together, these sections aim to equip users with the necessary knowledge to navigate the SUF efficiently, merge datasets accurately, and apply them in robust cross-sectional and longitudinal research.

5.1 Datasets in the SUF

5.1.1 Classification of Datasets in the SUF

The *Scientific Use File (SUF)* of PASS comprises multiple datasets that differ in content, frequency of collection, level, type, and format. A subset of these datasets is produced regularly in each survey wave; these are referred to as ***core datasets*** in this section. Core datasets form the backbone of the SUF and contain information collected consistently across waves, enabling robust longitudinal analyses.

In contrast, ***temporary datasets*** refer to those collected only for specific periods or for research needs and are no longer part of the current SUF release. However, they were briefly included within the previous SUF releases and hence briefly discussed.

An updated graphical overview of the dataset structure can be found in wave-specific data reports; for example, see Figure 2: Dataset Structure of the PASS in the PASS Datenreport Wave 17 (Anker et al., 2025).

Table 5.1 provides an overview of the core datasets and a detailed classification of these datasets based on their level, type, and format, as well as their key and pointer variables.

Table 5.1: Classification of core datasets in PASS

Dataset (filenames)	Level	Type	Format	Key variables	Pointer variables
Household dataset (”HHENDDAT.dta”)	Household level	Cross section	Long	(1) hnr (Household number); (2) welle (Indicator for survey wave)	(1) uhnr (Original household number)
Person dataset (”PENDDAT.dta”)	Individual level	Cross section	Long	(1) pnr (Constant person ID number); (2) hnr (Household number); (3) welle (Indicator for survey wave)	(1) uhnr (Original household number)
Household register (”hh_register.dta”)	Household level	Register	Wide	(1) hnr (Household number); (2) hnr$ (Household number in wave $)	(1) uhnr (Original household number); (2) pnrzp$ (Constant personal ID number of person who gave the household interview in wave $)

Table 5.1: Classification of core datasets in PASS (continued)

Dataset (filenames)	Level	Type	Format	Key variables	Pointer variables
Person register (”p_register.dta”)	Individual level	Register	Wide	(1) pnr (Constant personal ID number); (2) hnr$ (Household number in wave $); (3) zplfd$ (Serial number of the target person in the household in wave $)	(1) uhnr (Original household number); (2) zmhh$ (Constant personal ID number of target person’s mother living in the same household in wave $); (3) zvhh$ (Constant personal ID number of target person’s father living in the same household in wave $); (4) zparthh$ (Constant personal ID number of target person’s partner living in the same household in wave $)
Household weights (”hweights.dta”)	Household level	Weights	Long	(1) hnr (Household number); (2) welle (Indicator for survey wave)	No pointer variable
Person weights (”pweights.dta”)	Individual level	Weights	Long	(1) pnr (Constant person ID number); (2) welle (Indicator for survey wave)	No pointer variable
Citizen’s benefit spells (”alg2_spells.dta”)	Household level	Spells	Spell	(1) hnr (Household number); (2) spellnr (Spell number)	No pointer variable
Biography spells (”bio_spells.dta”)	Individual level	Spells	Spell	(1) pnr (Constant person ID number); (2) spellnr (Spell number)	No pointer variable
Children dataset (”KINDER.dta”)	Individual level	Cross-sect	Long	(1) pnr (Constant person ID number); (2) hnr (Household number); (3) welle (Indicator for survey wave)	(1) zmhh (Constant personal ID number of mother living in same household in respective wave); (2) zvhh (Constant personal ID number of father living in same household in respective wave)
One-Euro-Jobs spells (”ee_spells.dta”)	Individual level	Spells	Spell	pnr (Constant person ID number); (2) spellnr (Spell number)	No pointer variable

Before describing the datasets in detail, the following section explains the columns in the Table 5.1:

The first column lists all the **core dataset names** along with their corresponding file names as observed in the SUF (in parentheses).

Level classification, in the second column, distinguishes whether a dataset contains information at the household level or the individual level. Household-level datasets capture information about the household, collected through a structured household questionnaire administered to a knowledgeable household member at the beginning of the interview process. Individual-level datasets include data on persons aged 15 and above living in

these households, obtained through separate personal interviews. Since each respondent is linked to a specific household in every survey wave, the two levels are inherently connected.

Type classification, shown in the third column of the Table 5.1, refers to the nature and content of the datasets contained in the SUF. Across both household and individual levels, four main dataset types are distinguished: cross-sectional, register, weighting, and spell datasets.

- *Cross-sectional datasets* contain survey data collected during each household or individual interview, representing the situation at a specific point in time. They exclude the parts where the respondent was asked to report episodes (e.g., receipt of Citizen's Benefit[10]), which are instead stored in spell datasets.
- *Spell datasets* store episode-based information in which respondents report activities or events in the form of episodes. The respondent specifies a period starting in the past and reports all relevant activities or events up to the date of the interview. This way of collecting data particularly differs from the cross-sectional concept described above; therefore, it cannot be integrated directly in the cross-sectional datasets. Each reported episode forms a separate observation, includes a start date and end date, contains further information about its content, and allows multiple entries per respondent. Episodes may or may not overlap with the fieldwork period.
- *Register datasets* provide basic structural information on survey participation and wave-specific identifiers. The household register lists all households ever surveyed in PASS, while the person register contains all individuals within these households. These datasets include core identifiers and participation histories, serving as a reference framework for linking data across waves.
- Because of the complex sample design of PASS, weightings are necessary for all descriptive analyses. Hence, the SUF includes household- and individual-level *weighting datasets* that correspond structurally to the cross-sectional datasets. These contain weights for each wave in which a household or individual was surveyed, allowing researchers to project the sample data onto the target populations.

Format classification, shown in the fourth column of the Table 5.1, indicates how the data are structured within each dataset. In the SUF, datasets are prepared in one of three formats: *wide, long, or spell.*

- In the *wide format,* each unit (household or individual) is represented by exactly one observation (row) in the dataset. Wave-specific information is stored in separate variables (columns) ending with the corresponding wave number. For waves in which no information is available for a unit, these variables are filled with standardised missing- value codes. The register datasets of PASS follow this format, meaning that each observation uniquely represents a specific unit and can be identified using a single key variable.

[10] Unemployment Benefits II is used interchangeably with various other terminologies across this user guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

- In the *long format,* each observation represents a specific wave in which the unit was surveyed, resulting in as many rows per unit as the number of waves it participated in. Wave-specific information appears as additional rows for the same unit. Cross-sectional and weighting datasets use this format. Variables remain consistent across waves, with each variable appearing only once as a column. Changes in question wording or survey design may result in the introduction of new variables, while variables that are surveyed only in specific waves are assigned the missing code ‘‘-9’’ in waves where they were not collected. Observations in long-format datasets are identified using a combination of key variables for the unit and the wave.
- The *spell format* is used exclusively for episode-based data. In this format, each reported episode corresponds to a separate observation. Episodes may span multiple waves, and updated information can be added in subsequent waves if the episode is ongoing. Units with no reported episodes are not represented in spell datasets, while those with multiple episodes have one row per episode. Observations in spell datasets represent specific episodes of specific units and are uniquely identified using a combination of key variables for the unit and the spell number.

All datasets include *key variables*, which are identifiers used to distinguish observations within and across datasets of the SUF. Therefore, the fifth column lists the key variables of the datasets. They allow users to reliably link the same unit - household or individual; across different datasets of the SUF and, where applicable, across survey waves. For example, the household number (*hnr*) identifies a specific household, while the personal identification number (*pnr*) uniquely identifies an individual. In long-format datasets, key variables are typically combined with the wave indicator (*welle*) to ensure that each observation is uniquely identified. In spell datasets, the spell number (*spellnr*) is included alongside the unit’s key variable to distinguish between multiple episodes for the same unit.

A second group close to the key variables is the *pointer variables*, which are displayed in the last column. While key variables identify the same unit and link it between datasets, pointer variables establish links between different, related units. For example, the original household number (*uhnr*) can be used to link a split-off household to its household of origin. Similarly, in the person register, pointer variables such as the personal ID numbers of a respondent’s mother (*zmhh$*), father (*zvhh$*), or partner (*zparthh$*) can be used to map family relationships within and across households over time. Compared with the other datasets, the spell datasets do not contain any pointer variables. These variables enable complex data linkages that are essential for analysing household composition changes, kinship structures, and social relationships from a longitudinal perspective.

5.1.2 Description of Datasets by levels

- **Household-level Datasets**

  The *household dataset* (*HHENDDAT.dta)* is a cross-sectional dataset in long format. It contains data collected at the household level, with each row representing a household surveyed in a specific wave. Key variables

include the household number (*hnr*) and the survey-wave indicator (*welle*). The original household number (*uhnr*) serves as a pointer variable for tracking household changes over time.

The *household register (hh_register.dta)* is a register dataset in wide format. It provides an overview of all surveyed households across waves, with wave-specific household numbers (*hnr$*) stored in separate columns. Key variables include the household number (*hnr)* and its wave-specific variations. Pointer variables include the original household number *(uhnr)* and the personal ID number of the respondent who provided the household interview in each wave *(pnrzp$)*.

The *household weights dataset (hweights.dta)* contains weighting information required for analysis due to the complex sample design of PASS. This dataset, in long format, provides household weights for each surveyed wave. Key variables include the household number (*hnr*) and the survey-wave indicator (*welle*). No pointer variables are included.

The citizen's benefit spells dataset (*alg2_spells.dta*) is a spell dataset in spell format. It records episodes of household-level Citizen's Benefit receipt. Each observation represents a separate spell, identified by the household number (*hnr*) and the spell number (*spellnr*). No pointer variables are included.

- **Individual-Level Datasets**

The *person dataset (PENDDAT.dta)* is a cross-sectional dataset in long format. It contains individual-level survey data, with key variables including the constant personal ID number (*pnr*), the household number (*hnr*), and the survey wave indicator (*welle*). The original household number (*uhnr*) is included as a pointer variable.

The *person register (p_register.dta)* is a register dataset in wide format, tracking individuals across survey waves. Each row represents a person, with key variables such as the personal ID number (*pnr*), wave-specific household numbers (*hnr$*), and the serial number of the target person in the household *(zplfd$*). Pointer variables include the original household number (uhnr) and the personal ID numbers of household members related to the target person, such as the mother (z*mhh$*), father (*zvhh$*), and partner (*zparthh$*).

The *person weights dataset (pweights.dta)* provides weighting information for individuals in long format. It includes the personal ID number (*pnr*) and the survey-wave indicator (*welle*). No pointer variables are included.

The *biography spells dataset (bio_spells.dta)* is a spell dataset in spell format, capturing individual life-course episodes collected since wave 2. Each observation represents a reported spell, with key variables including the personal ID number (*pnr*) and the spell number (*spellnr*). No pointer variables are included.

The *children dataset (KINDER.dta)* is a cross-sectional dataset in long format, containing information on

children within surveyed households since wave 6. Key variables include the personal ID number (*pnr*), the household number (*hnr*), and the survey-wave indicator (*welle*). Pointer variables include the personal ID numbers of the child's mother (*zmhh*) and father (*zvhh*), if they reside in the same household during the respective wave.

The *One-euro-job spells dataset (ee_spells.dta)* is a spell dataset at the individual level. It records episodes in which a person participated in or received an offer for a One-euro-job. Key variables include the personal ID number (*pnr*) and the spell number (*spellnr*). No pointer variables are included.

5.1.3 Additional dataset for Websurvey mode

In addition to CAPI and CATI, the web survey in PASS was conducted for the first time in wave 16 as an additional interview between two regular panels waves and is available as a single dataset named *"Web.dta".* One rationale for introducing the web survey was to test optimal procedures for a future use of the web mode as a regular data collection mode for PASS. In addition, it was used for additional questionnaire modules that did not get priority in the core questionnaire.

Regardless of the wave in which the web survey was implemented, this dataset is cross-sectional at the individual level. The key variables are *pnr* (Constant person ID number), *hnr* (Household number) and *welle* (Indicator for survey wave). The websurvey dataset has no pointer variables.

More information about the datasets of the websurvey can be found in the corresponding data reports which can be downloaded from the FDZ homepage[11].

5.1.4 Temporary datasets

In addition to the datasets collected annually, PASS includes several temporary datasets that were collected only once or, at most, twice. These include vignette datasets and spell datasets at the individual level.

Vignette datasets are identified by "VIGDAT" in the dataset name. There are currently three such datasets:

- Job acceptance readiness *(VIGDAT_SUB.dta)* -- collected only in wave 5.
- Assessment of job offers *(VIGDAT_KON.dta)* -- collected only in wave 12.
- Mothers in employment *(VIGDAT_MUK.dta)* -- collected only in wave 15.

The key variables for vignette datasets are typically *pnr* (constant personal ID number) and *vignr* (vignette number). They are generally presented in long format, and there are no pointer variables.

[11] Data reports are a bundle including the data report for the latest wave as well as all the other waves which can be downloaded from http://doku.iab.de/fdz/pass/FDZ-Datenreporte_PASS_EN.zip

At the individual level, there are three temporary spell datasets: *Unemployment Benefit I spell dataset and two active labour market programme datasets*. The *alg1_spells.dta* dataset, collected only in wave 1, contains episodes of Unemployment Benefit I receipt, including start and end dates and the total amount of benefits per month. The two *active labour market programme* datasets are *massnahmespells.dta* (wave 1) and *mm_spells.dta* (waves 2 and 3), both of which record episodes of participation in specific employment or training measures. All three temporary spell datasets share the same key variables: *pnr* (constant personal ID number) and *spellnr* (spell number). None contain pointer variables.

In addition to these spell and vignette datasets, there is an individual-level dataset on *retirement provision (PAVDAT.dta)*. This dataset, in long format, provides detailed individual-level information on retirement provisions. It was collected only in wave 3 and only for persons aged 40 to 64, or those with a partner in this age range. The key variables are *pnr* (constant personal ID number) and *welle* (indicator for survey wave). No pointer variables are included.

At the household level, there is a corresponding r*etirement-provision dataset (HAVDAT.dta)*, also collected only in wave 3. It contains detailed household-level information on retirement provisions, collected under the same age-related criteria as *PAVDAT.dta.* This dataset is in long format, with *hnr* (household number) and *welle* as key variables.

Finally, two datasets were collected only at the individual level in wave 1: *Proxydata* and *Refusing Individuals*. Further details on these temporary datasets can be found in data report for wave 1.

More information about the datasets of the temporary datasets is provided in the corresponding data reports or in the older versions of PASS User Guide (Bethmann et al., 2013, pp. 29--47), (see also Section 5.2).

5.2 Variables in PASS

Two principal approaches exist for naming variables in survey datasets. The first approach assigns variable names according to their respective order in the questionnaire. While this facilitates the identification of variables with their corresponding items within the questionnaire, it complicates longitudinal analyses and tracking across waves, as the sequence of questions often changes between waves.

The second approach, which has been adopted in PASS, assigns consistent variable names across waves, irrespective of their position in the questionnaire, supplemented by a wave indicator when necessary. This ensures that identical items retain the same variable name over time, thereby greatly facilitating longitudinal tracking and analysis. Although this reduces the questionnaire's usefulness as a direct documentation tool[12], the advantages, such as improved data management, comparability, and analytic efficiency in a long-term panel study, decisively outweigh this limitation. Moreover, the organisation of PASS data in long format makes the use

[12] To link questionnaire numbers with their corresponding fixed variable names, we provide a correspondence list. https://fdz.iab.de/en/pd_hd/panel-study-labour-market-and-social-security-pass-version-0623-v2/

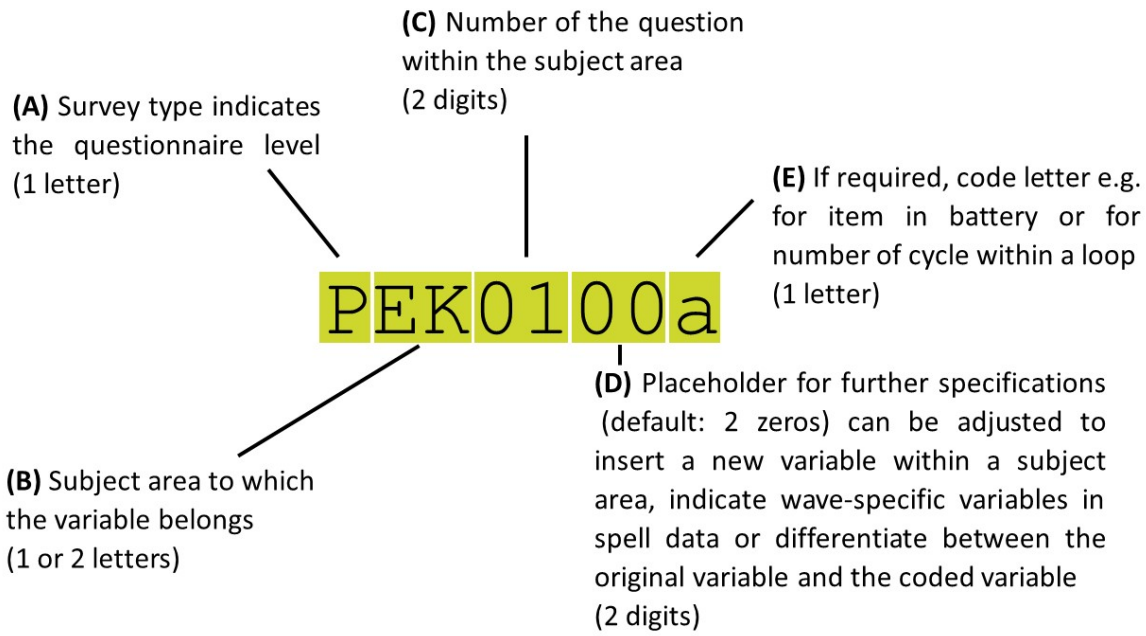


Figure 5.1: Variable naming scheme

of uniform variable names indispensable.

The codebook distinguishes between three different types of variables:

5.2.1 System variables

System variables are variables created during the survey process. They can be used, firstly, to comprehend the filters documented in the questionnaire. At least some of the system variables can also be of interest from a content-related or methodological point of view, for example the interview mode or the number of children of a certain age group living in the household. System variables are allocated individual names, for which lower-case letters and numbers are combined in some cases. The system variables also include the weights.

5.2.2 Surveyed variables

Surveyed variables are variables that were collected in this form directly in the questionnaire. These variables are given entirely new, abstract variable names. The concept behind this naming process is illustrated in the figure below using an example.

A. The first letter of the variable name indicates the questionnaire level, i. e. household or individual dataset, and is represented accordingly by the uppercase letter H or P.

B. This is followed by one or two upper-case letters which indicate the subject area to which the variable belongs (see Table 5.2 for a complete list).

The datasets which are processed in spell form, contain no introductory P or H variables. Instead, the variables in these datasets are given a uniform subject-based name consisting of two or three letters or two letters and

one digit (e.g. *AL20100, EE0100a*; see Table 5.3).

C. The introductory letter combination is then followed by two consecutively allocated numbers, which indicate the number of the question within the subject area.

D. The last two digits serve as placeholders for further specifications and are set to two zeros by default. However, they can be adjusted in specific cases. New variables introduced in later waves within a subject area can be assigned intermediate values (e.g. *PAS0850 between PAS0800 and PAS0900*). Additionally, this option has been used in cases where a second variant including coded information from an open-ended survey question or response category has been made available in addition to the original version of the variable (see Section 5.2.3). Furthermore, the placeholder is used to indicate wave-specific variables in spell data such as e. g. *AL20804 (wave 4), AL20805 (wave 5), AL20806 (wave 6)* and so on.

E. In the case of variables for items from multi-item batteries or in a looped sequence of questions, a further lower-case letter may be added to identify the item or the current cycle within the loop.

To clearly distinguish variables from the web survey, all variable names originating from the web survey are additionally prefixed with a lowercase ‘‘w.’’ Some of the items collected in the web survey are also included in the *PENDDAT* dataset from the main panel survey. However, the variables may differ in their response categories between the web survey and the main panel survey. For instance, the list of response categories for the question about practised sports (*wPSB0100/wPSB0101 vs. PSB0100/PSB0101*) has been both shortened and modified in the web survey (as compared to the main panel survey).

Table 5.2: Alphabetic list of subject-related indicators used in the variable names (cross sections)

Individual code	Individual subject area	Household code	Household subject area
PA	General	HA	General
PAA	Educational aspiration	HBT	Education and Inclusion Subsidies
PAC	Labor market opportunities	HCV	Corona
PAS	Job-search	HD	Demography
PB	Education	HEK	Income
PCV	Corona	HKI	Child-care
PD	Demography	HLS	Standard of living
PEE	1-Euro-Jobs	HR	Household economics and everyday economic practices
PEF	Dealing with finances	HT	Social participation
PEK	income	HW	Housing
PEO	Attitudes and orientations		
PER	Employment & retirement		
PET	Employment		

Table 5.2: Alphabetic list of subject-related indicators used in the variable names (cross sections) (continued)

Individual code	Individual subject area	Household code	Household subject area
PG	Health		
PGR / PPG	Perception about justice		
PKO	Cooperation plan		
PLA	Latent functions of work		
PLS	Standard of living		
PME	Memory		
PMI	Migration		
PMJ	Mini Job		
PML	Minimum wage		
PMV	Vignette: Childcare and employment		
PP	Care		
PPT	Political participation		
PQB	Quality of employment		
PSB	Sports		
PSH	Social origin		
PSK	Social relations		
PSM	Social media		
wPSB	Web: Sports		
PSU	Obligation for job search		
PSV	Stigma awareness		
PTK	Contact to social security institutions		
PTS	Internet use		
PV	Trust game		
PWB	Further training		
wPEA	Web: Employment		
wPEO	Web: Attitudes and orientations		
wPMK	Web: Media consumption		
wPQB	Web: Quality of employment		
wPSA	Web: Sanctions (Unemployment Benefit II)		

Table 5.3: Alphabetical list of subject-related indicators used in variable names (spell datasets)

Individual code	Individual subject area	Household code	Household subject area
AL (wave 1 only)	Receipt of Unemployment Benefit I	AL2	Receipt of Unemployment Benefit II
AL (from wave 2)	Spells of registered unemployment and receipt of Unemployment Benefit I since January 2005		
ALM (wave 1 only)	Employment and training measures		
EE (from wave 4)	One-Euro-Job		
ET (from wave 2)	Employment with earnings of more than € 400 per month since January 2005		
MN (waves 2–3)	Employment and training measures		
LU (waves 2–3)	Other activities since January 2005		

5.2.3 Generated variables

The generated variables in the strict sense are aggregated from various other variables, e. g. from open-ended and categorical income measures, or they are even more complex constructs such as equivalised household income or classifications for education (such as *ISCED or Casmin*) or status (e. g. *EGP, ESEC*). Generated variables in this strict sense are allocated individual names that are as clear and memorable as possible, in lower-case letters (e.g. *hhtyp or migration*).

Another group of generated variables includes those in which information from open-ended survey questions or response categories were added to another (closed) variable. Although these variables are, strictly speaking, also generated variables and are classified as such in the frequency tables of the codebook, they are not given clear names. Instead, their names are based on those of the original variable, while the final ‘‘0’’ is replaced by ‘‘1’’ (e. g. the original variable *PA0100a* and its coded counterpart *PA0101a*).

5.3 Data Editing in PASS

5.3.1 Overview of Data editing

The SUF of PASS is an outcome of a comprehensive and systematic data editing process. In each wave, the field institute collects raw data, which is then verified, coded for open-ended responses, and transformed into variables before being integrated into the SUF datasets. While the process is refined and adapted from wave to wave, the overall logic and sequence of steps have remained consistent over time. Wave-specific procedures are documented in the corresponding data reports (see, for example, Berg, Cramer, Dickmann, Gilberg, et al. (2024), for wave 17), but this section provides a general overview of the key steps and their sequence.

Data editing for the first two waves was conducted at the *Institute for Employment Research (IAB)*. From wave 3 onwards, the *Institut für Angewandte Sozialwissenschaft (infas)*, the new field institute for PASS, assumed responsibility for this task[13]. To avoid any unexpected changes in the procedure or inconsistencies in the SUF, several safeguards accompanied this transition.

First, the new contract with *infas* explicitly required adherence to the sequence and methodology of the data editing process as that of previous waves. Accordingly, *infas* received the relevant syntax files and datasets from wave 2, along with documentation for each processing step.

Second, *infas* carried out the editing process in continuous coordination with the IAB. Major decisions, such as handling problematic household structures or integrating spell datasets (such as the *bio_spells datase*t introduced in wave 4), were made in consultation with the IAB. In addition, IAB remained available for discussions and inquiries throughout the editing period.

Third, once the SUF for wave 3 was completed, IAB reviewed the final dataset, focusing on both its structure and content.

Besides the data editing, *infas* also performs initial transformations of the raw ASCII data from interviews to produce a range of internal datasets, including:

- a household dataset for cross-sectional questions (e.g., *childcare*),
- a household dataset for longitudinal modules (e.g. *ALG II)*,
- a dataset tracking household composition over time (matrix),
- a dataset representing household relationships (relationship matrix),
- a person/senior dataset for cross-sectional questions,
- Two separate person-level longitudinal datasets (*e.g. employment history and labour market policy measures*),

[13] The contract with the former field institute, TNS Infratest, was initially limited to three waves. As a result, fieldwork from wave 4 onwards was re-tendered. In the new call for proposals, the IAB decided to include the task of data editing starting with wave 3. Consequently, infas, as the new field institute from wave 4 on, also carried out the data editing for wave 3.

- a dataset containing open-ended responses across all survey components.

These datasets are then subjected to the formal and content-based checks described below before being incorporated into the SUF.

Apart from these institutional changes, the logic and sequence of the data editing process have remained consistent across waves. The process can be divided into the following steps:

Table 5.4: Overview of the steps involved in editing the PASS data

No.	Step of the procedure
1	Check of the household structure of re-interviewed households
2	Removal of problematic/incomplete interviews (household and/or individual level)
3	Integration of individual dataset and senior citizen's dataset
4	Correction of the household structure of re-interviewed households
5	Filter checks at the household level
6	Construction of a household grid dataset and plausibility checks
7	Generation of the synthetic benefit units (see description of variables in wave-specific data reports)
8	Generation of new control variables based on household data after filter checks and household grid plausibility checks
9	Filter checks at the individual level
10	Coding of information from open-ended survey questions
11	Plausibility checks of the household and individual-level data (excluding spell data)
12	Preparation, plausibility checks and construction of the spell datasets
13	Simple variable generations
14	Complex variable generations
15	Generation of the data structure for the scientific use file (household dataset, individual dataset, register dataset)
16	Anonymisation
17	Final check of the SUF datasets

5.3.2 Structure Checks

The first step involves verifying the household structure of re-interviewed households by comparing it with the structure reported in the previous wave. This step ensures the identification and, where necessary, correction of implausible or problematic changes in household composition, as well as mis-allocations of individual interviews within households. For consistent longitudinal analysis, individuals must retain their positions within the household across waves and remain uniquely identifiable. The same personal identification number must never be assigned to different individuals in different waves.

If the correct household composition cannot be determined, all interviews from the respective household in that wave are excluded from the SUF. If only a single individual interview has a mismatch without further

inconsistencies in household structure, then only that interview is removed.

To detect such problematic cases, multiple internal consistency checks are conducted, assessing, for example, the plausibility of household changes and the accuracy of individual assignments. These cases are then reviewed through a formalised process between *infas* and IAB, with IAB making the final decisions. However, not all checks result in exclusions, as many potential inconsistencies are pre-emptively avoided through built-in validations in the survey software (e.g., preventing all target persons from leaving a household simultaneously, or requiring at least one remaining person aged 15 or older).

The wave-specific data reports (e.g., Berg, Cramer, Dickmann, Gilberg, et al. (2024)) provide further details on these structural checks. Additionally, the household register (variables *hnettok**, *hnettod**) and person register (*pnettok**, *pnettod**) indicate removed cases across waves. It should be noted that not all deleted interviews are traceable in the SUF due to the way in which the register files are constructed[14].

In addition, incomplete interviews at both the household and individual levels are excluded from the SUF[15]. Households that do not meet PASS's definition of a successfully surveyed household (see Table 5.5)[16] are also excluded. These cases are not recorded in the register datasets, as they were never deemed eligible to begin with, unlike the removed interviews described above.

Table 5.5: Household types and interview requirements

Type of household	Household level interview	Individual level interview(s)
New household (interviewed for the first time and drawn for the initial sample or a refreshment sample)	Yes (completed)	Yes (at least one completed)
Re-interviewed household (household already interviewed in a previous wave of PASS)	Yes (completed)	None required
New split-off household (interviewed for the first time and split-off from another household in PASS)	Yes (completed)	None required

5.3.3 Filter Checks and Assignment of Standardised Codes

Every variable included in the SUF undergoes a thorough filter check. During this process, the system flags filter violations and assigns standardised missing codes. Following Table 5.6 presents an overview of the standardised codes used in PASS:

[14] In PASS, the SUF's register files are net files. The household register includes all households ever successfully surveyed. The person register contains all persons living in those households at the time of interview. Interviews removed from households or persons not included in these registers e.g., first-time interviews in refreshment samples are not visible in the SUF.

[15] Thus, interviews that are terminated before completion do not appear in the SUF datasets.

[16] As the criteria for a "successfully surveyed" household vary by household type, some households in the SUF appear without individual interviews in certain waves.

Table 5.6: Standardised missing codes and their explanations.

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 No. 1, only allocated as required.
-6	Question-specific code No. 2, only allocated as required.
-7	Question-specific code No. 3, only allocated as required.
-8	Implausible value.
-9	Item not administered in wave.
-10	Item not administered in questionnaire version.

These codes fall into the following categories:

- Missing values due to respondent answers (‘‘-1’’ = don’t know, ‘‘-2’’ = refused),
- Filter-based missing values (‘‘-3’’ = not applicable, ‘‘-4’’ = mistakenly not asked),
- Question-specific codes (‘‘-5’’ to ‘‘-7’’),
- Implausible answers (‘‘-8’’),
- Items not part of the questionnaire/wave (‘‘-9’’, ‘‘-10’’).

Except for implausible responses, which were flagged in a later step, all other missing-value categories are addressed during this phase. Variables in the raw datasets are checked sequentially, reflecting the order in which they were collected. In this process, the codes ‘‘-3’’ and ‘‘-4’’ are assigned accordingly.

A variable receives the code ‘‘-3’’ if it should not have been asked based on filter logic. Likewise, questions that were mistakenly asked despite filter conditions are corrected to ‘‘-3’’[17].

While erroneously collected data can be recoded as ‘‘-3’’, it was not possible to retroactively supply missing answers. If a question was skipped despite being required by the filter, the missing code ‘‘-4’’ (‘‘*mistakenly not asked*’’) is applied.

It is important to note that ‘‘-4’’ codes may also result from retrospective corrections in household structure. For example, if a person was wrongly recorded as having moved out but was later reassigned to the household, missing household data for this person is marked with ‘‘-4’’.

If a ‘‘-4’’ appears on a filter-relevant variable, then all subsequent questions dependent on it are also coded ‘‘-4’’, provided they were not asked. If a follow-up was still triggered by an alternative filter path, its valid value

[17] For instance, if detailed vocational training data was collected although the respondent indicated no vocational qualification, this information is replaced with ‘‘-3’’.

remains.

Codes ‘‘-5’’ to ‘‘-7’’ are question-specific and include both special missing codes and valid categories (e.g., top-coded income values). Codes ‘‘-9’’ and ‘‘-10’’ are used for items not included in specific questionnaires or waves: ‘‘-9’’ denotes absence in a given wave, while ‘‘-10’’ covers differences between questionnaire versions (e.g., regular vs. senior or variant household questionnaires used in waves 1 to 3).

5.4 Record Linkage to Administrative Data

A distinctive feature of PASS is the possibility of linking survey data to administrative records, provided respondents give their consent for record linkage in each wave. The individual survey responses are linked to administrative data of the Institute for Employment Research (IAB) containing additional information on episodes of Unemployment Benefit I receipt, Citizen’s Benefit receipt, employment, job search, and participation in active labour-market programmes.

The *Research Data Centre (FDZ)* of the German Federal Agency at IAB produces the *PASS-ADIAB* dataset, which combines all PASS respondents with the relevant administrative data. This combined dataset enables both substantive research where administrative data serve as a supplementary source of information and methodological research, while administrative records are used as a validation source (Kreuter et al., 2010). Linking *PASS-ADIAB* to the PASS survey datasets is possible via the variable *pnr* (constant personal ID number).

For the waves 1 to 16, approximately 85.3% of people who responded to the person-level questionnaire and consented to linkage were successfully linked (Dummert & Sauer, 2024).

For further details on *PASS-ADIAB*, see Antoni & Bethmann (2019), which provides an overview of research opportunities, data sources, record linkage, and data structure, as well as the current methods and data reports available on the FDZ website (https://fdz.iab.de/unsere-datenprodukte/personen-und-haushaltsdaten/pass/). For technical details regarding the linkage methodology, refer to Bachteler (2008). Analyses of determinants of consent to record linkage and potential selection biases are available in Beste (2011).

As *PASS-ADIAB* contains sensitive data, specific user conditions apply. This integrated file is available to researchers for onsite use at the FDZ in Nuremberg or at one of the other FDZ locations in Germany, France, Spain, Italy, Poland, Luxembourg, the UK, the USA and Canada (see https://fdz.iab.de/en/data-access/on-site-use/ for details and locations) and via remote execution.

Chapter 6

PASS in Transition - Major developments and Innovations.

Contact authors: Beste, Jonas; Collischon, Matthias; Stegmaier, Jens; Wenzig, Claudia.

Over nearly two decades, PASS has evolved continuously to address changing social realities, methodological challenges, and innovations in survey research. This chapter highlights the major developments that have shaped the design and implementation of the study, reflecting PASS's commitment to maintaining methodological excellence, representativeness, and data quality.

Section 6.1 describes the expansion of survey languages and the integration of additional refugee samples to ensure adequate representation of new migrant populations within the SGB II framework. **Section** 6.2 presents the introduction of a web-based survey mode, designed to explore digital data collection approaches and broaden thematic coverage. **Section** 6.3 explains the adoption of an adaptive fieldwork design, which employs predictive modelling and targeted incentives to counteract non-response and attrition. **Section** 6.4 outlines the adjustments implemented during the COVID-19 pandemic, which required rapid adaptations in fieldwork operations, questionnaire design, and interviewer training. Finally, **Section** 6.5 introduces the interviewer-monitoring system developed for PASS, detailing its integration into CAPI fieldwork and its role in safeguarding data integrity through statistical and procedural quality-control measures.

6.1 Additional Refugee Samples and Adaptations in Survey Languages

The PASS study was established to serve as a central and representative data source for research on basic income security in Germany. To achieve this objective, the composition of the SGB II[18] population, including its demographic dynamics, must be continuously monitored and methodologically adjusted where necessary.

[18] Unemployment Benefits II is used interchangeably with various other terminologies across this User guide and more widely across German welfare benefits research. For a detailed description refer 0.3.

Table 6.1: The use of foreign language instruments in PASS

Foreign-language version	Introduced in	Last used in
English	Wave 1	Wave 1
Turkish	Wave 1	Wave 9
Russian	Wave 1	Ongoing (Wave 19)
Arabic	Wave 10	Ongoing (Wave 19)
Ukrainian	Wave 17	Ongoing (Wave 19)

Because migrant groups are overrepresented among SGB II recipients, the inclusion of individuals with a migration background has been a methodological priority from the outset. To enable participation of migrants with limited German language skills, the survey was initially administered in Turkish, Russian, and English alongside German language, as these languages reflected the largest linguistic groups among SGB II households at the time. The English version was discontinued after the first wave due to low uptake (nine households). Similarly, the Turkish version was discontinued after wave 9, as participation declined from 163 households in wave 1 to only 14 in wave 9 (see Berg et al. (2010), p. 23).

The large influx of refugees, particularly from Syria since 2015, and from Ukraine since 2022, significantly changed the composition of SGB II recipients. To maintain representativeness and avoid non-response bias, PASS expanded its language offerings. Arabic was introduced in wave 10 to facilitate participation among Syrian and Iraqi refugees. Following Russia's full-scale invasion of Ukraine in 2022, the share of Ukrainian refugees in the SGB II population rose sharply; as a result, Ukrainian was added in wave 17. Because both Ukrainian and Russian are widely spoken in Ukraine, respondents may choose between these languages during interviews, which has helped sustain high participation rates and methodological consistency. The Table 6.1 summarises the use of foreign language instruments that were introduced in PASS across waves.

Foreign-language interviews are primarily conducted via CATI (telephone interviews) by native-speaking interviewers. As most CAPI (face-to-face) interviewers are German-speaking, non-German-speaking respondents are encouraged to transition from CAPI to CATI to improve cost efficiency and ensure data quality (see *FDZ-Datenreport 07/2013, User Guide, Ch. 3.3.2*, for language use until wave 6). All translations undergo a multi-stage quality assurance process aligned with international standards such as those of the European Social Survey (Dorer, 2018). Since wave 10, survey instruments have been designed to allow interviewers to switch languages at any point during the interview.

To ensure sufficient case numbers for subgroup analyses, Syrian and Iraqi households receiving SGB II were oversampled in the early years following the 2015 refugee arrivals. As their representation in the main sample increased, this oversampling was gradually reduced and later discontinued. For Ukrainian refugees, oversampling was unnecessary, as they were adequately represented through the new entries sample. All Syrian and Ukrainian refugee households are identifiable in the dataset via dedicated sample variables.

The inclusion of refugees as a relevant sub population required additional methodological adjustments beyond language expansion. These included the development of targeted questionnaire modules (e.g. recognition of foreign qualifications), measures to ensure comparability across language versions (German and foreign-language interviews) and interview modes (CAPI and CATI). Cultural sensitivities in interview conduct and new approaches to panel maintenance also played a crucial role. As part of panel maintenance, efforts were intensified to collect stable contact information (email addresses, social media accounts, third-party phone numbers) to counteract the high mobility of refugees, who frequently changed residence after arrival due to administrative relocation or personal decisions.

In summary, the integration of refugees represents one of the most significant methodological adaptations in the 20-year history of PASS. These adjustments in sampling, language provision, and survey methods have significantly enhanced the representativeness of the survey, providing a robust basis for analysing both SGB II dependency and the social and economic integration of refugees in Germany.

6.2 Web survey

To gain experience with online data collection and to provide an opportunity for collecting information on specific topics that were too extensive for inclusion in the main PASS survey, a web-based survey was introduced following wave 16. All respondents from the CAPI and CATI main survey were invited to participate.

The web survey is conducted in autumn each year, no later than six months after the main interviews (but in some cases even sooner). To date, it has been implemented following PASS waves 16 and 18. Unlike the main interviews, which are administered via telephone or face-to-face, the web survey is self-administered. Each wave is designed to last approximately 10--20 minutes, and participants receive a post-paid incentive of 5 euros, sent by post upon completion of the questionnaire.

The web survey questionnaire contains no standard recurring modules; rather, it focuses on special topics. The survey following wave 16 included questions on sanctions for individuals receiving unemployment benefits, employment, leisure activities, and political participation. It also repeated selected questions from the main survey to enable analyses of survey-mode effects (Mühlbacher & Trappmann, 2024). The survey following wave 18 included an extended household deprivation module compared to the main questionnaire, as well as a survey experiment on unemployment-benefit uptake.

Further details on the design, content, and implementation of the web survey are provided in the corresponding data reports (Dickmann et al., 2023).

6.3 Adaptive Design in PASS

Ensuring high data quality and minimising attrition are central objectives of the PASS. Beginning in wave 7 and standardised from wave 14 onwards, PASS introduced an adaptive design approach to fieldwork aimed at improving response rates and reducing bias by prioritising cases with a high risk of non-response. This innovation marked a shift towards a more targeted, data-driven fieldwork strategy that has since been integrated into standard survey operations (for further methodological details and evaluation results, (see Beste et al. (2023)).

The foundation for adaptive design in PASS lies in the extensive longitudinal data accumulated over previous waves. These data include survey responses, para-data, and interviewer assessments, which enable the PASS team to predict the response propensity of each household. While protocols of the adaptive design changed often between waves 7 and 13 (cf. Trappmann & Müller (2014), Kreuter et al. (2014)), the standardized approach that is currently used, was first implemented in wave 14 as part of an experimental study, in which households with a low predicted likelihood of participation were randomly assigned to receive increased prepaid incentives (20 euros instead of the standard 10 euros). The aim was to assess whether targeted incentives could reduce drop-out rates and mitigate non-response bias.

Prediction models were developed using machine learning algorithms trained on data from waves 4 to 12 and validated against wave 13. Random forest models proved particularly effective in identifying households at high risk of attrition. In wave 14, these predictions informed case prioritisation and treatment assignments in both CATI and CAPI modes. The experimental results demonstrated that doubling the incentive significantly increased participation among low-propensity households and yielded modest improvements in overall data representativeness (Beste et al., 2023).

Following these successful results, adaptive design became a regular component of PASS fieldwork from wave 16 onwards. In place of randomised experimentation, the prioritisation process is now systematically applied using updated prediction models based on the most recent data. Households flagged as low propensity are routinely targeted with enhanced fieldwork measures, most notably through increased incentives.

This transition to a standard operational practice illustrates PASS's commitment to evidence-based survey methodology and continuous quality improvement. The adaptive design helps preserve the longitudinal integrity of the panel by counteracting attrition in key subgroups. It also ensures a more efficient allocation of fieldwork resources by targeting efforts where they are most needed.

For users of the PASS SUF, the integration of adaptive design strengthens both the reliability and validity of the data. The prioritisation process is carefully documented, and its effects on fieldwork outcomes are regularly evaluated. Although the current focus is on targeted incentive adjustments, the adaptive design framework allows for future expansion to other targeted measures, such as interviewer assignments or tailored contact

strategies.

6.4 Impacts of COVID-19 Pandemic on Fieldwork, Questionnaire, Interviewer Training, and Other Areas

The COVID-19 pandemic (2020--2022) had a profound impact on the PASS survey, particularly during wave 14, when numerous methodological and content-related adjustments were necessary to respond to the exceptional circumstances while safeguarding data quality and maintaining panel stability. Wave 15 was still heavily influenced by the pandemic, while wave 16 marked the gradual return to standard survey operations. By wave 17, PASS had fully resumed its regular routines.

This section outlines the major pandemic-related adaptations in fieldwork, questionnaire content, and interviewer training. A detailed discussion of the same follow after the Table 6.2.

Table 6.2: Impact of the COVID-19 Pandemic on PASS (Waves 14–16)

Category	Wave 14 (2020)	Wave 15 (2021)	Wave 16 (2022)
Pandemic events	First nationwide lockdown (16 March 2020); school closures until May; gradual easing from May; stable summer period	Second lockdown (winter 2020/21); gradual reopening from Spring 2021; schools operating with restrictions	Restrictions lifted from spring 2022; partial mask mandates; high vaccination rate
Fieldwork / Survey mode	Stoppage of CAPI fieldwork and transfer to CATI; all CAPI halted mid-March; introduction of “CAPI by phone”	“CAPI by phone” fully replaces face-to-face interviews; used as a survey mode from start	Partial return to on-site CAPI (Jun 2021); full resumption by May 2022; “CAPI by phone” optional
Questionnaire	COVID-related updates introduced 8 April 2020: home isolation & quarantine (HCV0100, HCV0200); childcare (HKI0900–HKI01100); employment (PCV0200–PCV0600); living situation (PCV0100); mental health (PG1215*)	Continuation of COVID-specific questions	Continuation/adaptation of COVID-specific modules; later removal
Interviewer training	Traditional in-person training; after lockdown: webinars; training for “CAPI per phone”	Full shift to online training; digital training concept developed	Consolidation of digital training; continued use of webinars and online modules
Other areas	Fieldwork extended until 29 Sept 2020; suspension of interviewer caps	Expanded refreshment sample; oversampling of self-employed SGB II recipients	CAPI vignette module implemented as separate online survey

6.4.1 Fieldwork changes

Preparations for wave 14, including questionnaire development, sampling, household contact strategies, and interviewer training were carried out under standard PASS procedures. Fieldwork began on 14 February 2020, but the nationwide lock down on 16 March 2020 brought abrupt and far-reaching changes affecting both private and professional life as well as operational fieldwork.

To avoid drop-outs caused by the sudden replacement of interviewers, especially among long-standing panel households previously interviewed in-person, the ''CAPI by phone'' method was introduced in mid-May 2020. This allowed former CAPI interviewers to conduct interviews by telephone, maintaining interviewer's continuity and preserving rapport with respondents, albeit in an auditory format. We are aware that the name may be misleading as CAPI by phone interviews are telephone interviews and not personal interview. However, they differ from regular CATI interviews in that they are not administered from a centralized call-centre with regular supervision and are typically administered by CAPI interviewers who previously interviewed the same households face-to-face. In this sense they are hybrids between CATI and CAPI.

To address reduced accessibility, the field period for wave 14 was extended until 29 September 2020. In addition, the usual interview caps per interviewer (CATI: 100 interviews; CAPI: 40 households) were suspended for wave 14 and handled more flexibly in subsequent waves (see Method reports wave 14 (Beste et al., 2021, p. 129)*;* 15 (Jesske et al., 2022, p. 137)*;* 16 (Jesske & Gerber, 2023, p. 128)).

Fieldwork in waves 15 and 16 remained affected by pandemic-related restrictions (see Method report wave 15 (Jesske et al., 2022, Ch. 5.1)*;* 16 (Jesske & Gerber, 2023, Ch 5.1)). In wave 15, in-person contact with households was prohibited at the start of the field period, and ''CAPI by phone'' served as the full substitute for face-to-face interviews. A vignette experiment on maternal employment and childcare, originally planned as a CAPI module, was implemented as a separate online survey (Frodermann et al., 2024); and (Jesske et al., 2022, Ch. 5.8).

From June/July 2021, in-person household interviews were gradually reintroduced under strict hygiene protocols, though ''CAPI by phone'' remained vital for established panel households. It was only in wave 16 that in-person interviewing resumed widely. From May 2022, CAPI interviews were once more conducted on-site, with ''CAPI by phone'' retained as a complementary method for hard-to-reach households. In wave 17, PASS fully returned to standard procedures, with CAPI reserved for in-person interviews and CATI for telephone interviews (with very few exceptions).

To compensate for drop-outs and reduced case numbers in wave 14, the sampling strategy was adjusted. In wave 15, the refreshment sample was expanded (see Berg et al. (2023), p. 18). Moreover, the sample was extended to targeted subgroups, notably self-employed persons receiving SGB II benefits, whose numbers had increased sharply during the pandemic. Oversampling enabled detailed subgroup analyses.

6.4.2 Questionnaire Adjustments

In addition to methodological adaptations, wave 14 introduced new modules to capture the realities of life during the pandemic (see Berg et al. (2022), Ch. 1.2). From 8 April 2020 onwards, pandemic-specific questions supplemented both household and person questionnaires.

Household-level topics included the extent of household affectedness (e.g., home isolation or quarantine) and the organisation of childcare during the pandemic. The personal questionnaire included items on employment (short-time work, job loss, remote work), life changes, and mental health.

Certain items were accompanied by interviewer instructions clarifying whether responses should refer to pre-pandemic or current conditions. These changes during ongoing fieldwork were implemented across all language versions (German, Arabic, and Russian).

The variable *kennungfbversW14* allows users to distinguish interviews conducted before and after the introduction of pandemic-related questions and interviewer instructions. In waves 15 and 16, most COVID-19 questions were continued or modified to reflect evolving conditions, though the home isolation/quarantine module was discontinued after wave 15. Since wave 17, all pandemic-related questions have been removed from the questionnaire.

6.4.3 Interviewer Training

Interviewer training also underwent major transformations. While wave 14 training initially took place in person, all subsequent sessions after the lockdown were moved online. This ensured that interviewing quality remained high under restrictive conditions.

In waves 15 and 16, a fully digital training concept was further developed (Jesske et al., 2022, Ch. 7.1). Training materials were provided as self-learning modules with integrated knowledge checks, and interactive webinars reinforced the content. Digital training has since become a permanent feature of PASS interviewer preparation.

6.4.4 Potential Implications for Data Use and Analysis

Through early and coordinated measures by the IAB and the fieldwork institute, PASS successfully adapted to the challenges posed by the COVID-19 pandemic between 2020 and 2022, maintaining high-quality data and panel stability.

Key to this success were the flexible survey design enabling a rapid shift to telephone interviewing (CATI or CAPI per phone) and the timely inclusion of pandemic-relevant questionnaire content. Flexible sampling strategies targeting affected subgroups and digital interviewer training ensured robust and unbiased data collection.

For data users analysing waves 14 to 16, it is essential to consider these methodological and content-related adaptations systematically. Relevant features, particularly ‘‘CAPI per phone’’ as a supplementary survey mode,

and the wave 14 questionnaire version should be accounted for in analyses. Pandemic-related changes affected multiple aspects of data collection, including interview duration, contact quality, interviewer continuity, and response behaviour. These factors may influence data comparability across waves and should ideally be controlled for, in longitudinal analyses.

6.5 Interviewer Monitoring in the PASS Study: Integration into CAPI Fieldwork

Maintaining high data quality is a core principle of PASS. In recent years, interviewer monitoring has become an increasingly important component of this effort. As interviewer-administered surveys can be vulnerable to both intentional falsification and deviations from the prescribed standardised interview procedure, PASS has developed a systematic approach to identify and address such risks - particularly within its CAPI fieldwork.

In contrast to CATI (Computer-Assisted Telephone Interviewing), where interviewers operate from supervised call centres with real-time oversight, CAPI interviewers work independently in the field. This autonomy limits direct insight into interview conduct and creates the need for dedicated quality-assurance mechanisms tailored to CAPI's decentralised nature.

Audio recordings of all CAPI interviews have been collected (given respondent consent), since wave 10, providing valuable material for monitoring and for understanding interview dynamics in the field. An interviewer's recording rate additionally serves as an indicator of adherence to survey protocols.

A comprehensive monitoring strategy was piloted following the field period of wave 14 and throughout wave 15. From wave 16 onwards, interviewer checks have been embedded directly into the data collection process. These checks allow for the timely identification of suspicious behaviour and enable targeted actions, such as clarifying anomalies, retraining interviewers, excluding them from fieldwork, or initiating follow-up interviews when necessary. This proactive approach prevents further data collection by problematic interviewers and minimises the risk of error propagation across waves.

The integration of interviewer monitoring into PASS represents a significant step towards safeguarding data quality in decentralised CAPI settings. It enables structured, evidence-based responses to potential irregularities and complements existing quality controls in CATI. As a result, users of the PASS SUF can rely on data that benefit from both robust preventive measures and an active commitment to continuous quality assurance.

The quality-control strategy in PASS combines statistical diagnostics and procedural checks to ensure comprehensive oversight (Beste et al., 2021):

- Analysis of timestamps:

  Timestamps recorded during interviews are analysed to detect unusually short durations at the item or

module level, which may indicate skipped questions or interviewer-entered responses.

- Duration comparisons:

  Interviews with and without audio recordings are compared, as interviewers may alter their behaviour when aware that a session is being recorded. Suspicious deviations in timing can therefore reveal irregularities.

- Longitudinal consistency checks:

  For relatively stable variables (e.g., political attitudes), large deviations across waves can signal data irregularities or potential falsification.

- Statistical indicators:

  PASS employs a set of 23 statistical indicators designed to identify anomalies in response patterns (e.g., high rates of item non-response, violations of Benford's law, non-differentiation). A meta-indicator (Schwanhäuser, 2024) aggregates these measures at the interviewer level to flag unusual data characteristics.

- Procedural checks by *infas* (the survey institute):

  These include systematic evaluations of audio recordings, postal validations with respondents, and targeted re-contacts to verify interview authenticity.

Overall, the monitoring system focuses not only on detecting outright falsifications, but also on identifying partial falsification and non-standardised interview conduct, both of which can compromise data integrity.

Interviewer monitoring is carried out in close collaboration between the PASS study team at the IAB and the survey institute *infas*, ensuring continuous exchange and alignment throughout the fieldwork process.

The implementation of interviewer monitoring has already led to measurable improvements in CAPI quality assurance. As detailed by Beste et al. (2021), the first applications of the system in waves 14 and 15 resulted in the exclusion of two interviewers and their associated interviews from the dataset due to clear protocol violations.

References

Achatz, J., Hirseland, A., & Promberger, M. (2007). *Panel study labour market and social security (PASS): Concept and design*. Institute for Employment Research (IAB).

Allaire, J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2024). *Rmarkdown: Dynamic documents for r*. https://github.com/rstudio/rmarkdown

Anker, F., Berg, M., Cramer, R., Dickmann, C., Gerber, V., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Malich, S., Prospero, V., Wenzig, C., Trappmann, M., Bähr, S., Collischon, M., Gundert, S., Mackeben, J., Müller, M., Stegmaier, J., & Wunder, A. (2025). *Codebuch und dokumentation des panels ‘‘arbeitsmarkt und soziale sicherung’’ (PASS). Datenreport welle 18* (No. 10/2025; p. 194). https://doi.org/10.5164/IAB.FDZD.2510.de.v1

Antoni, M., & Bethmann, A. (2019). PASS-ADIAB -- linked survey and administrative data for research on unemployment and poverty. *Jahrbücher für Nationalökonomie Und Statistik*, *239*(4), 747--756. https://doi.org/10.1515/jbnst-2018-0002

Bachteler, T. (2008). *Dokumentation record linkage IEB--PASS* [Technische Dokumentation].

Berg, M., Cramer, R., Dickmann, C., Gerber, V., Gilberg, R., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Malich, S., Schwarz, S., Wenzig, C., Trappmann, M., Bähr, S., Collischon, M., Gundert, S., Küfner, B., Mackeben, J., Müller, M., Prospero, V., Stegmaier, J., Teichler, N., & Wunder, A. (2024). *Codebuch und dokumentation des panels ‘‘arbeitsmarkt und soziale sicherung’’ (PASS). Datenreport welle 17* (No. 07/2024; p. 194). https://doi.org/10.5164/IAB.FDZD.2407.de.v1

Berg, M., Cramer, R., Dickmann, C., 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. (2024). *Codebook and documentation of the panel study ‘‘labour market and social security’’ (PASS). Datenreport wave 16* (No. 12/2023; p. 176). https://doi.org/10.5164/IAB.FDZD.2312.en.v1

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., 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). *Codebook and documentation of the panel study ‘‘labour market and social security’’ (PASS). Datenreport wave 12* (No. 09/2019; p. 173). https://doi.org/10.5164/IAB.FDZD.1909.en.v1

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., 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). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 13* (No. 12/2020; p. 170). https://doi.org/10.5164/IAB.FDZD.1220.en.v1

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. (2023). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 15* (No. 10/2022; p. 175). https://doi.org/10.5164/IAB.FDZD.2210.en.v1

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. (2022). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 14* (No. 14/2021; p. 176). https://doi.org/10.5164/IAB.FDZD.2114.en.v1

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Bethmann, A., Fuchs, B., Trappmann, M., & Huber, M. (2013). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 6* (No. 06/2013; p. 195).

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Marwinski, K., Gebhardt, D., Wenzig, C., & Wetzel, M. (2010). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 3* (No. 06/2010; p. 200).

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Marwinski, K., Gebhardt, D., Wenzig, C., & Wetzel, M. (2011). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Datenreport wave 3. FDZ-datenreport 06/2010*. Institut für Arbeitsmarkt- und Berufsforschung.

Beste, J. (2011). *Selektivitätsprozesse bei der verknüpfung von befragungs- mit prozessdaten: Record linkage mit daten des panels ''arbeitsmarkt und soziale sicherung'' und administrativen daten der bundesagentur für arbeit* (No. 09/2011; p. 28).

Beste, J., Frodermann, C., Trappmann, M., & Unger, S. (2023). Case prioritization in a panel survey based on predicting hard-to-survey households by machine learning algorithms: An experimental study. *Survey Research Methods*, *17*(3), 243--268. https://doi.org/10.18148/srm/2023.v17i3.7988

Beste, J., Olbrich, L., & Schwanhäuser, S. (2021). Interviewer:innenkontrolle im panel arbeitsmarkt und soziale sicherung (PASS). In *FDZ-methodenreport 04/2021* (pp. 1--11). https://doi.org/10.5164/IAB.FDZM.2104.de.v1

Bethmann, A., Fuchs, B., Wurdack, A., Beste, J., Trappmann, M., Müller, G., Eggs, J., Gundert, S., Wenzig, C., & Gebhardt, D. (Eds.). (2013). *User guide ''panel study labour market and social security'' (PASS), wave 6* (No. 07/2013; p. 112). https://doku.iab.de/fdz/reporte/2013/DR_07-13_EN.pdf

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R., & Tschersich, N. (2009). *IAB-haushaltspanel im*

niedrigeinkommensbereich welle 2 (2007/08). Methoden- und feldbericht (p. 102).

Büngeler, K., Gensicke, M., Hartmann, J., Jäckle, R., & Tschersich, N. (2010). *IAB-household panel of those with low income: Wave 3 (2008/2009). Methods and field report* (p. 113). Institute for Employment Research. http://doku.iab.de/fdz/reporte/2010/MR_10-10.pdf

Christoph, B., Müller, G., Gebhardt, D., Wenzig, C., Trappmann, M., Achatz, J., Tisch, A., & Gayer, C. (2008). *Codebook and documentation of the panel study ''labour market and social security'' (PASS). Vol. 1: Introduction and overview, wave 1 (2006/2007)* (No. 05/2008; p. 75).

Chun, A. Y., Heeringa, S. G., & Schouten, B. (2018). Responsive and adaptive design for survey optimization. *Journal of Official Statistics*, *34*(3), 581--597. https://doi.org/10.2478/jos-2018-0028

Couper, M. P. (1998). Measuring survey quality in a CASIC environment. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 41--49.

Dickmann, C., Jesske, B., Gerber, V., Beste, J., Dummert, S., Frodermann, C., Malich, S., Mühlbacher, V., Schwarz, S., Wenzig, C., Trappmann, M., Zins, S., Bähr, S., Collischon, M., Gundert, S., Mackeben, J., Küfner, B., Müller, B., Stegmaier, J., Teichler, N., & Wunder, A. (2023). *Codebuch und dokumentation des websurveys im ''panel arbeitsmarkt und soziale sicherung'' (PASS). Datenreport websurvey 2022* (No. 13/2023; p. 22). https://doi.org/10.5164/IAB.FDZD.2313.de.v1

Dorer, B. (2018). *ESS round 9 translation guidelines*. https://www.europeansocialsurvey.org/docs/round9/methods/ESS9_translation_guidelines.pdf

Dummert, S., & Sauer, I. (2024). *PASS-befragungsdaten verknüpft mit administrativen daten des IAB (PASS-ADIAB) 1975--2022* (FDZ-Datenreport No. 05/2024; p. 82). FDZ im IAB. https://doi.org/10.5164/IAB.FDZD.2405.de.v1

Felderer, B., Müller, G., Kreuter, F., & Winter, J. (2018). The effect of differential incentives on attrition bias: Evidence from the PASS wave 3 incentive experiment. *Field Methods*, *30*(1), 56--69. https://doi.org/10.1177/1525822X17726206

Frodermann, C., Hipp, L., & Bünning, M. (2024). Money matters! Evidence from a survey experiment on attitudes toward maternal employment across contexts in germany. *Gender & Society*, *38*(3), 436--465. https://doi.org/10.1177/08912432241252601

Gebhardt, D., Gundert, S., & Wurdack, A. (2009). *Revising the active labour market programmes (ALMP) module in PASS: Methodological report*. Institute for Employment Research (IAB).

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 439--457.

Harron, K. L., Doidge, J. C., Knight, H. E., Gilbert, R. E., Goldstein, H., Cromwell, D. A., & Meulen, J. H. de. (2017). A guide to evaluating linkage quality for the analysis of linked data. *International Journal of Epidemiology*, *46*(5), 1699--1710. https://doi.org/10.1093/ije/dyx177

Hartmann, J., Brink, K., Jäckle, R., & Tschersich, N. (2008). *IAB-haushaltspanel im niedrigeinkommensbereich. Methoden- und feldbericht* (No. 07/2008; p. 156).

Jesske, B., & Gerber, V. (2023). *Panel arbeitsmarkt und soziale sicherung (PASS) -- websurvey haupterhebung 2022* (No. 07/2023; p. 105). https://doi.org/10.5164/IAB.FDZM.2307.de.v1

Jesske, B., & Quandt, S. (2011). *Methodenbericht panel arbeitsmarkt und soziale sicherung (PASS). 4. Erhebungswelle 2010 (haupterhebung)* (No. 08/2011; p. 137).

Jesske, B., & Schulz, S. (2021). *Panel arbeitsmarkt und soziale sicherung (PASS) -- erhebungswelle 14 (2020, haupterhebung)* (No. 06/2021; p. 153). https://doi.org/10.5164/IAB.FDZM.2106.de.v1

Jesske, B., Schulz, S., & Gerber, V. (2022). *Panel arbeitsmarkt und soziale sicherung (PASS) -- erhebungswelle 15/2021 (haupterhebung)* (No. 06/2022; p. 163). https://doi.org/10.5164/IAB.FDZM.2206.de.v1

Jesske, B., Schulz, S., & Gerber, V. (2024). *Panel arbeitsmarkt und soziale sicherung (PASS) -- erhebungswelle 17/2023 (haupterhebung)* (No. 05/2024; p. 201). https://doi.org/10.5164/IAB.FDZM.2405.de.v1

Kiesl, H. (2010). *Kalibrierte hochrechnung für das panel ''arbeitsmarkt und soziale sicherung''*.

Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic uses of process information*. John Wiley & Sons. https://doi.org/10.1002/9781118596869

Kreuter, F., Müller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: Making use of administrative data. *Public Opinion Quarterly*, *74*(5), 880--906. http://www.jstor.org/stable/40985409

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*, *43*(3), 452--464. https://doi.org/10.1177/0049124113508094

Kueppers, R. (2005). MOSAIC von microm. In G. Grözinger & W. Matiaske (Eds.), *Deutschland regional: Sozialwissenschaftliche daten im forschungsverbund* (pp. 95--104). Hampp.

Laflamme, F., Maydan, M., & Miller, A. (2008). Using paradata to actively manage data collection survey process. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 630--637.

Lynn, P., & Kaminska, O. (2010). *Weighting strategy for understanding society* (Understanding Society Working Paper Series Nos. 2010-05). University of Essex.

Meitinger, K., Stadtmüller, S., & Silber, H. (2020). Fieldwork monitoring in practice: Insights from 17 large-scale social science surveys in germany. *Survey Methods: Insights from the Field*. https://surveyinsights.org/?p=12237

Mühlbacher, V., & Trappmann, M. (2024). *Implementierung eines webbasierten datenerhebungsmodus in einem bestehenden panel der allgemeinbevölkerung: Eine analyse der selektions- und messeffekte* (No. 01/2024; p. 74). https://doi.org/10.5164/IAB.FDZM.2401.de.v1

Müller, A., Rebien, M., & Stops, M. (2011). *Einschaltungspotenzial für den arbeitgeber-service der bundesagentur für arbeit: Ergebnisse aus der IAB-erhebung des gesamtwirtschaftlichen stellenangebots* (No. 10/2011; p. 17).

Müller, B., Bähr, S., Gundert, S., Teichler, N., Unger, S., & Wenzig, C. (2020). *PASS scales and instruments manual* (FDZ-Methodenreport Nos. 07/2020 (en); p. 111). FDZ im IAB. https://doi.org/10.5164/IAB.FDZM.2007.de.v1

Müller, K., & Wickham, H. (2025). *Tibble: Simple data frames*. https://tibble.tidyverse.org/

Peycheva, D., Ploubidis, G., & Calderwood, L. (2021). *Determinants of consent to administrative records linkage in longitudinal surveys: Evidence from next steps*.

Qiu, Y., & See file AUTHORS for details., authors/contributors of the included fonts. (2024a). *Sysfonts: Loading fonts into r*. https://github.com/yixuan/sysfonts

Qiu, Y., & See file AUTHORS for details., authors/contributors of the included software. (2024b). *Showtext: Using fonts more easily in r graphs*. https://github.com/yixuan/showtext

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rudolph, H., & Trappmann, M. (2007). *Design und stichprobe des panels "arbeitsmarkt und soziale sicherung" (PASS)* (No. 12/2007; pp. 60--101).

Sakshaug, J., Beste, J., Coban, M., Fendel, T., Haas, G.-C., Hülle, S., Kosyakova, Y., König, C., Kreuter, F., Küfner, B., Müller, B., Osiander, C., Schwanhäuser, S., Stephan, G., Vallizadeh, E., Volkert, M., Wenzig, C., Westermeier, C., Zabel, C., & Zins, S. (2020). Impacts of the COVID-19 pandemic on labor market surveys at the german institute for employment research. *Survey Research Methods*, *14*(2), 229--233. https://doi.org/10.18148/srm/2020.v14i2.7743

Sakshaug, J., Beste, J., & Trappmann, M. (2023). Effects of mixing modes on nonresponse and measurement error in an economic panel survey. *Journal for Labour Market Research*, *57*. https://doi.org/10.1186/s12651-022-00328-1

Schnell, R., Bachteler, T., & Reiher, J. (2010). Improving the use of self-generated identification codes. *Evaluation Review*, *34*(5), 391--418.

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, *35*(1), 101--113.

Schwanhäuser, S. (2024). *Detecting interviewer falsification in surveys: A guide for improving quality controls* (p. 260).

Sinibaldi, J., Trappmann, M., & Kreuter, F. (2014). Which is the better investment for nonresponse adjustment: Purchasing commercial auxiliary data or collecting interviewer observations? *Public Opinion Quarterly*, *78*(2), 440--473. https://doi.org/10.1093/poq/nfu003

Spiess, M., & Rendtel, U. (2000). *Combining an ongoing panel with a new cross-sectional sample* (DIW Discussion Paper No. 198). DIW Berlin, German Institute for Economic Research. https://ideas.repec.org/p/diw/diwwpp/dp198.html

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* (p. 19) [(IAB-Discussion Paper 05/2009),].

Trappmann, M., & Müller, G. (2014). *Introducing adaptive design elements in the panel study "labour market and social security" (PASS)* (pp. 1--7).

Ushey, K., & Wickham, H. (2025). *Renv: Project environments*. https://rstudio.github.io/renv/

Watson, N., & Cernat, A. (2022). Simulating the consequences of adaptive survey design in two household panel studies. *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smab050

Watson, N., & Cernat, A. (2023). Simulating the consequences of adaptive survey design in two household panel studies. *Journal of Survey Statistics and Methodology*, *11*(4), 806--828. https://doi.org/10.1093/jssam/smab050

Watson, N., & Wooden, M. (2009). *Identifying factors affecting longitudinal survey response*. https://doi.org/10.1002/9780470743874.ch10.

West, B. T., Kreuter, F., & Trappmann, M. (2014). Is the collection of interviewer observations worthwhile in an economic panel survey? New evidence from the german labor market and social security (PASS) study. *Journal of Survey Statistics and Methodology*, *2*(2), 159--181. https://doi.org/10.1093/jssam/smu002

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & van den Brand, T. (2025). *ggplot2: Create elegant data visualisations using the grammar of graphics*. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. https://dplyr.tidyverse.org

Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. https://readr.tidyverse.org

Wickham, H., Pedersen, T. L., & Seidel, D. (2025). *Scales: Scale functions for visualization*. https://scales.r-lib.org

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. https://tidyr.tidyverse.org

Xie, Y. (2014). Knitr: A comprehensive tool for reproducible research in R. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman; Hall/CRC.

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. https://yihui.org/knitr/

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown

Xie, Y. (2025a). *Bookdown: Authoring books and technical documents with r markdown*. https://github.com/rstudio/bookdown

Xie, Y. (2025b). *Knitr: A general-purpose package for dynamic report generation in r*. https://yihui.org/knitr/

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook

Zhang, S., & Wagner, J. (2022). The additional effects of adaptive survey design beyond post-survey adjustment: An experimental evaluation. *Sociological Methods & Research*, *53*(3), 1350--1383. https://doi.org/10.1177/

00491241221099550

Zhu, H. (2024). *kableExtra: Construct complex table with kable and pipe syntax*. http://haozhu233.github.io/kableExtra/

Imprint

FDZ-Datenreport 06|2026 EN

Date of publication

30 September 2026

Publisher

Research Data Centre (FDZ)

of the Federal Employment Agency (BA)

in the Institute for Employment Research (IAB)

Regensburger Str. 104

D-90478 Nuremberg

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/2026/UserGuide/PASS-User-Guide.pdf

Documentation version

PASS-SUF0624_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2606.en.v1

Dataset version

PASS-SUF0624, v1, DOI: 10.5164/IAB.PASS-SUF0624.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/

Questions about the content

E-Mail: IAB.HHP@iab.de