

DZA | Deutsches Zentrum für Altersfragen Forschungsdatenzentrum

March 2025

What's new? An Introduction to the 2023 Wave of the German Ageing Survey (DEAS) and Updates of Previous Waves (DEAS Version 1.0)

DOI 10.5156/DEAS.1996-2023.M.001

Stefan Stuth & Nicole Hameister

Citation suggestion:

Stefan Stuth & Nicole Hameister (2025). What's new? An Introduction to the 2023 Wave of the German Ageing Survey (DEAS) and Updates of the Previous Waves (DEAS Version 1.0) DOI 10.5156/DEAS.1996-2023.M.001. Berlin: German Centre of Gerontology (DZA).

Creative Commons CC-BY-Share Alike 4.0

German Centre of Gerontology (DZA) Research Data Center (FDZ-DZA) Manfred-von-Richthofen-Straße 2 12101 Berlin, Germany

www.fdz-dza.de fdz@dza.de

TABLE OF CONTENTS

| 1. | The German Ageing Survey (DEAS) – A Study on The Second Half | of Life | 3 |
|--------|--|---------|----|
| 1.1 | DEAS Wave 2023 | | 3 |
| 1.2 | DEAS Sampling | | 4 |
| 1.3 | Topics of DEAS 2023 | | 5 |
| 1.4 | Changes and Innovations in DEAS 2023 | | 6 |
| 1.4.1 | Random survey mode | 6 | |
| 1.4.2 | New, edited and deleted survey content | 6 | |
| 1.4.3 | New derived job characteristics | 7 | |
| 1.4.4 | Completed, comparable and differentiated job characteristics | 7 | |
| 1.4.5 | New edition of mortality data | 8 | |
| 1.4.6 | Data edition - deletion of cases | 8 | |
| 1.4.7 | Additional income variables | 8 | |
| 1.4.8 | New variables on education and qualification | 9 | |
| 1.4.9 | Regional home-ownership and rental data based on ImmoScout24 | 9 | |
| 1.4.10 |) Analysis of own and additional regional data | 9 | |
| 1.4.1 | More information about the interview | 10 | |
| 1.4.12 | 2 Data documentation | 10 | |
| 1.4.13 | 3 Weighting | 10 | |
| 1.4.14 | A baseline survey in the panel | 11 | |
| 1.4.15 | 5 Illnesses and accidents | 11 | |
| 2 | Response Rates and Panel Consistency Checks | | 12 |
| 3 | Data Weighting of DEAS 2023 | | 13 |
| 4 | Changes and Updates in Previous DEAS Waves | | 15 |
| 5 | Data Formats | | 16 |
| 6 | Classification of Variable Names and Missing Codes | | 16 |
| 7 | Modification of Data to Maintain de facto Anonymity | | 18 |
| 8 | The Complex DEAS Survey Design | | 19 |
| 9 | Documentation | | 21 |
| 10 | Outlook | | 21 |
| Litera | ture | | 22 |

1. THE GERMAN AGEING SURVEY (DEAS) – A STUDY ON THE SECOND HALF OF LIFE

The German Ageing Survey ("Deutscher Alterssurvey", DEAS) is a long-term study conducted by the German Centre of Gerontology (DZA) and funded by the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ). It focusses on the changes in the life situations and ageing processes of people in the second half of life. Due to its cohort-sequential longitudinal design, its range of topics, its national representativeness with a comparatively high number of cases and the long observation period, the DEAS is a unique data source in Germany for interdisciplinary research into the life situations and life courses as well as the subjective views and life plans of ageing and old people.

Microdata of the German Ageing Survey (DEAS) are available for scientific research for noncommercial purposes only. The Research Data Centre provides researchers with access to the information collected as part of the surveys and advises them on its use. The anonymised DEAS datasets of all completed survey waves and the documentation material are available free of charge from the FDZ website (www.fdz-dza.de).

1.1 DEAS Wave 2023

The base sample originally drawn for 2020 could not be surveyed as planned due to the Covid-19 pandemic. Based on the address data from the unused base sample, a self-administered paper survey ("Refreshment Sample 2021") was conducted in winter 2021/22. The aim of this survey was to supplement the aged DEAS sample with younger cohorts in order to be able to carry out comparative analyses across all age groups in the future. The 2023 survey combines the previous DEAS samples and the 2021 refreshment sample. In terms of methodology, the survey was conducted by a personal interview either in the respondent's home (CAPI) or by telephone ("CAPI-per-phone").

The 2023 survey is based on the regular DEAS waves in terms of content. Methodologically, the survey instruments were adapted so that they are as comparable as possible regardless of the survey mode (face-to-face vs. telephone interviews). As in the previous waves, the fieldwork, data collection and data weighting were carried out by the Institut für angewandte Sozialwissenschaft GmbH (Infas), Bonn (Schiel et al. 2023).

1.2 DEAS Sampling

Since 1996, new base samples have been drawn every six years from the 40 to 85-year-old resident population in private households as stratified and clustered random samples - by age group, gender and region (East/West) - from the population registers and interviewed in person by interviewers. From wave 2 (2002) onwards, participants from the previous baseline surveys who were willing to participate in the panel have been interviewed repeatedly Excluded from this are the respondents from the separately drawn sample of non-German respondents in 2002 who were not included in the longitudinal sample. After the 2008 survey, the interval between the panel waves was shortened from six to three years.

In wave 1 (1996), the selection was limited to persons with German citizenship. In wave 2 (2002), two separate samples of Germans and non-Germans were drawn for the baseline survey. Since 2008, the basic sample has been drawn from the population of 40- to 85-year-olds in private households (place of main residence) without regard of nationality.

The 2023 survey was a panel survey and included all participants from the 1996 to 2014 base samples who were still available and willing to participate. It also included the respondents from the 2021 refreshment sample who were available and willing to participate. In total, valid interviews are available for 4.992 respondents aged 42 to 102. Of these, 4.248 people last took part in the telephone-only survey in winter 2020/21.

The SUF DEAS 2023 dataset with the anonymised data from the 2023 survey contains information on the respondents from the following samples:

| Sample 2023 | n |
|--|-------|
| Panel respondents from the 1996 baseline survey | 476 |
| Panel respondents from the 2002 baseline survey | 497 |
| Panel respondents from the 2008 baseline survey | 1.488 |
| Panel respondents from the 2014 baseline survey | 2.194 |
| Panel respondents from the 2021 refreshment survey | 337 |
| Total respondents | 4.992 |

Using the available Scientific Use Files (SUF) of all previous survey waves (21,760 respondents, 56,294 valid interviews), individual and social developments can be observed over a period of more than 27 years (e.g. Klaus et al. 2017). The sampling design of the German Ageing Survey opens up various analysis strategies: There are the options of cross-sectional examination of life situations and life contexts in the respective survey year, for analysing social change between 1996 and 2023 (e.g. Mahne et al. 2017) as well as individual development over three to 27 years of life with the panel data (e.g. Vogel et al. 2019). A further perspective results from the comparison of individual developments over six years in the periods 1996-2002, 2002-2008, 2008-2014 and 2014-2020 or 2014-2020/21. This makes it possible to compare the development in a constant age period for different birth cohort groups (cohort sequence analysis).

1.3 Topics of DEAS 2023

As in the previous waves, the respondents were asked comprehensive questions about their life situation. Information was collected on the following topics:

- Work and retirement
- Partnership, family and intergenerational relationships
- Social networks and social support
- Leisure activities and civic engagement
- Housing and mobility
- Economic situation and economic behaviour
- Subjective well-being
- Health and health behaviour
- Need for assistance and care
- Attitudes, norms and values
- Basic socio-demographic data.

As in the previous waves, the survey was divided into two main parts. The majority of the information was collected through a personal interview in the respondent's household, or alternatively through a telephone interview. The second part of the survey consisted of a self-administered paper questionnaire ('Drop-off'). In particular, subjective ideas (such as attitudes, values, views of old age, well-being), more sensitive areas (e.g. details on material situation and health) and topics that do not require differentiated filtering were surveyed here. Most of the items for the psychological scales were also collected in the Drop-off. The 2023 Drop-off was completed by 4,211 DEAS respondents, 772 of whom chose the web version (CAWI).

1.4 Changes and Innovations in DEAS 2023

1.4.1 Random survey mode

The regular DEAS survey mode is the computer-assisted personal interview (CAPI). Due to the Covid-19 pandemic, the survey mode for the 2020/21 wave had to be switched completely to a computer-assisted personal interview by telephone (CAPI-per-phone). ¹ In order to estimate the potential effect of the mode change in 2020/21, the survey mode of wave 2023 was randomly distributed (CAPI or CAPI-per-phone).² Half of the sample was assigned to the CAPI mode, the other half to the CAPI-per-phone mode. In pre-defined exceptional cases, the interviewers were permitted to switch to the alternative mode. In total, 73% of the interviews were conducted in the assigned start mode. Just under a third of respondents in CAPI start mode (31%) switched to telephone mode. Just under a quarter of respondents in telephone start mode (23%) switched to CAPI mode.

Respondents who were interviewed by telephone were unable to perform physical functioning tests to determine respiratory volume (lung function test) or cognitive performance tests (digit symbol substitution test). Furthermore, it was not possible for the interviewers to assess information on the living environment of the telephone respondents. Lists and cards were used again in the 2023 survey, if the survey was conducted in CAPI mode. Otherwise, the survey instrument did not differ between the two survey modes.

On average, the duration of an interview was approximately 81 minutes. The telephone interviews were slightly shorter at an average of 77 minutes, while the CAPI interviews lasted longer at around 86 minutes.

1.4.2 New, edited and deleted survey content

Compared to the 2020/21 wave, there were only a few changes to the survey content. There were a number of minor adjustments, new additions and deletions, some of which are highlighted here as examples: At various points in the interview, masculine formulations were replaced with gender-sensitive language.

There is a new initial question that the interviewer answers at the beginning: Will the interview be conducted on site or by telephone? Furthermore, for various groups of people (main caregivers, spouse, mother/father-in-law), the question whether and by whom they receive regular help, care or support was introduced. The wording of the question to record internet use

¹ Infas uses the term Capi-per-Phone to emphasise key differences to the classic CATI survey. In contrast to regular CATI interviews, the survey was not conducted by anonymous telephone studios, but by the interviewers who had already conducted the survey in person in previous years.

² One exception is respondents from the refreshment sample, who were assigned entirely to the CAPI mode (see also section 1.4.11).

in the interview was edited to avoid technical terms as far as possible, and the devices for internet use were also recorded.

As a result of a call for modules by the DZA, there were three new modules in the DEAS 2020/21 on the topics of prescriptive age norms, Big Five personality traits as well as subjective life expectancy and preferred life expectancy. The Big Five personality traits as well as subjective and preferred life expectancy have been included again in the 2023 Drop-off questionnaire. The Drop-off also includes questions about the perceived threat of Covid-19, vaccination status and attitudes towards Covid-19 vaccination.

1.4.3 New derived job characteristics

Occupations are a structuring and potentially socialising factor in the lives of employed and retired people. However, as a nominal characteristic with several hundred characteristics, occupations are not easy to integrate into empirical analyses. For this reason, DEAS already offers a number of established constructs with which occupations have been transformed into a metric (ISEI, SIOPs) or a categorical scale (social class, ESeC, ESeG).

In addition to ISEI and SIOPs, the DEAS now also offers the new **Occupational Reputation Scale (Berufliches Ansehen - BA)**. The core assumption of the Occupational Reputation Scale is that the social reputation of different occupations has a direct impact on the social status of those practising the occupation and their opportunities in life (Ebner/Rohrbach-Schmidt 2019).

As a new categorical scale, the DEAS now also offers the **occupational environments** according to Holland (1981). Based on the main activities carried out in the occupations, the occupations were assigned to one of six categories: Realistic (especially craft occupations), Investigative (research occupations), Artistic (artistic occupations), Social (e.g. caring and teaching occupations), Enterprising (e.g. occupations that focus on sales, management and advertising), Conventional (e.g. supervision and performance of largely pre-structured tasks in hierarchies, such as administrative activities) (Stuth 2022).

1.4.4 Completed, comparable and differentiated job characteristics

DEAS tries to keep the survey burden and interview duration as low as possible for the respondents. For this reason, occupations are only surveyed again if something has changed since the last individual survey. As a result, gaps occur in the data, because occupational data for panel participants is only collected in relatively few cases. These gaps were closed with the current data edition in all waves by filling in occupational information and the variables derived from it from information from the previous waves. In order to enable comparative analyses over time, the current data version now contains all occupational information for all waves (including

the derived occupational characteristics such as ISEI or SIOPS) as well as ISCO88 and ISCO08 variants.

Previously, derived occupational characteristics were generated for couples. This reference to couples has been dissolved in the current data edition for all DEAS waves: all occupations and the derived occupational characteristics are now available for all waves for the respondents, for the respondents' partners and, in the case of partnerless respondents who are divorced or widowed, also for their last spouse.

The naming of the generated occupational variables was also standardised across all waves.

1.4.5 New edition of mortality data

The mortality data contained in the metadata has been gradually expanded over previous DEAS waves. With the current survey wave, all information was merged and processed uniformly for all waves including information from panel maintenance measures and researching of death dates of the deceased. As a result, the updated dataset DEAS Meta now contains completely revised and new variables, such as year of death.

1.4.6 Data edition - deletion of cases

New quality controls were also established with the new edition of the mortality data. These controls revealed six cases that seemingly had participated in a DEAS wave after their death. Further checks revealed that respondents different from the original target persons had participated in these six cases. This can occur because the respondent's death sometimes is only reported years later (e.g. actively by relatives, or as feedback for the next wave or as feedback after panel maintenance measures). These cases are deleted in the current data edition: one case in DEAS 1996, three cases in DEAS 2008, one case each in DEAS 2011 and DEAS 2017.

1.4.7 Additional income variables

The regular SOEP-based income variables on income position, income poverty and income wealth available for all waves will be supplemented starting with DEAS wave 2023 by alternative income data from the SILC (Federal Statistical Office 2024) as the basis for reference values for new SILC-based poverty or wealth measures.

These reference values are used to determine whether people are above or below certain income thresholds (poverty risk, wealth). The data users can thus decide which variables are more suitable for their analyses. The SOEP reference values are based on the current monthly

income, while the SILC data is based on the annual income of the previous year. For reasons of consistency, we advise to use the SOEP-based income for time series comparisons and panel analyses.

1.4.8 New variables on education and qualification

Up to now, education variables have been available for all waves, combining school and vocational qualifications. Retrospectively for all waves, there are now two new variables on education that provide the school-leaving qualification (schledu_3) and vocational training qualification (vocedu_5). Furthermore, open-ended responses on educational qualifications were systematically included in the generation of these variables.

1.4.9 Regional home-ownership and rental data based on ImmoScout24

The RWI Essen has prepared data from ImmoScout24 (see Thiel 2024): price indices were calculated on the basis of ImmoScout24 data, which capture the costs of home ownership and rental apartments. The district-based index values were edited and are now available as quintiles from the 2008 wave to the current 2023 wave. Category 1, i.e. the first quintile, is assigned to respondents who live in one of the districts/cities whose rental and home ownership prices are in the bottom 20% compared with the rest of Germany. Accordingly, quintile 5 is assigned to respondents whose districts feature the top 20% of rental and home ownership prices.

The real estate price index is not part of the regular SUFs but is published separately on direct request to the FDZ team. Interested users should contact <u>fdz@dza.de</u>.

1.4.10 Analysis of own and additional regional data

In principle, regional analyses are also possible with self-compiled regional data at the district or municipality level. However, these must be carried out on site at the DZA at a specially secured workstation for guest researchers. The FDZ is responsible for matching the data. For this purpose, researchers must prepare their regional data in advance and make it available to the FDZ team. If analyses are carried out at municipality level (Gemeinden), additional costs may be incurred as infas has to provide the municipality code for all respondents of the respective waves according to the current territorial status. These costs must be covered by the visiting researchers.

In addition, MICROM regional data is available for the DEAS waves 2002, 2008, 2011 and 2014. Additional regional data from infas360 is available for 2017. The original ImmoScout24derived price indices can also be analysed at the district or municipality level at the guest researcher workstation.

1.4.11 More information about the interview

In order to be able to research methodological questions, for example on the influence of interviewers on the interviews, the dataset DEAS Meta now offers all available information on the interviewers. However, these variables differ across waves. Furthermore, information on how long the interview lasted is now available for each respondent in each wave.

| Interviewer | 1996 | 2002 | 2008 | 2011 | 2014 | 2017 | 2020/21 | 2023 |
|--------------------------------------|------|------|------|------|------|------|---------|------|
| ID | | | | | | | | |
| Age | | | | | | | | |
| Age groups | | | | | | | | |
| Gender | | | | | | | | |
| Vocational training qualification | | | | | | | | |

Available information about the interviewees in DEAS Meta:

Green = information available

1.4.12 Data documentation

The detailed documentation of the variables in the metadata set (participation history and mortality data) can no longer be found in this document, but in the extensive <u>Documentation of</u> <u>Instruments and Variables</u>.

1.4.13 Weighting

Due to the lack of a baseline sample, which should have taken place regularly in 2020, DEAS data for panel respondents is becoming increasingly selective over time (panel mortality and real mortality). For this reason, we provide a new cross-sectional weighting factor in the DEAS 2023 which, in addition to the previous post-stratification characteristics of gender, birth cohort and region, also takes education into account for post-stratification. Researchers who want to analyse DEAS 2023 cross-sectionally should use this weighting. For trend analyses (i.e. comparing several cross-sections side by side), we recommend using the classic cross-

sectional weights from infas for reasons of consistency. The new extended cross-sectional weights have the suffix "_dza" and are (so far) only available for the 2023 wave.

1.4.14 A baseline survey in the panel

As the 2021 refreshment survey was a self-administered paper questionnaire, there is no information available on many relevant characteristics of the respondents. In order to fill the gaps, the refreshment respondents were interviewed as baseline respondents in the 2023 wave. For this reason, people in the refreshment sample were not randomly assigned to an interview mode, but were completely assigned to the CAPI mode.

1.4.15 Illnesses and accidents

In the personal interview, respondents have the opportunity to provide open information about illnesses or accidents they have had in the last three, six or ten years. (The reference period is ten years for first-time respondents, six years for panel respondents in 2008 and three years for panel respondents from 2011 onwards). We have coded and ranked the open responses from the waves from 2008 onwards. The three most frequent responses were each stored in three new variables, each of which shows the three most frequently mentioned illnesses and the three most frequently mentioned accidents dichotomously (no/yes). Respondents can be assigned to more than one of the dummy variables in case their open answers contained information on more than one illness or accident. A new illness appeared in the data that only existed since 2020: Covid-19. Thus a fourth illness dummy was included in DEAS 2020/21 and DEAS 2023. Further information can be found in the Documentation of Instruments and Variables.

2 RESPONSE RATES AND PANEL CONSISTENCY CHECKS

Out of the gross sample of 8.175 contacts, 5.022 interviews were completed, which corresponds to a response rate of around 62%. In addition, 4.212 Drop-offs were completed, which represents a response rate of 84% (cf. infas methodological report: Schiel et al. 2023).

In two test steps, the year of birth/age as well as gender were used to check whether (1) the people interviewed matched the people from the previous waves and whether (2) the person interviewed orally was identical to the person who completed the Drop-off questionnaire. Of the 5.022 interviews, 15 cases were excluded by infas because they were not conducted properly. In a further 15 cases, the panel consistency check revealed that it was not the targeted respondent who completed the interview. These 15 interviews were also excluded, so that DEAS 2023 contains 4.992 valid cases.

In six cases the respondents interviewed orally did not match the persons who completed the Drop-off questionnaire. The existing drop-off information was completely deleted for these cases. All Drop-off variables have been recoded to the missing values of -7 or .g (deleted in data edition). Four cases originate from the PAPI and two cases from the CAWI questionnaire. At the end of the Drop-off checks, 4.205 valid Drop-offs remain (770 CAWI and 3.435 PAPI).

3 DATA WEIGHTING OF DEAS 2023

The field institute infas carried out the data weighting in close consultation with the DEAS project team at the DZA. The chosen procedure corresponds in principle to the procedure used in other panel studies.³ The *longitudinal weighting* is based on default models (logistic regression) to determine the probability of participation from one survey wave to the following survey wave (excluding people who died between the two waves) with a limited set of predictors (see Schiel et al. 2023: Chapter 7). The longitudinal weight is obtained by multiplying the cross-sectional weight of the initial wave by the reciprocal of the probability of participation in the follow-up wave. This procedure has been used for every survey year since 2002. The weighting dataset for the 2023 wave already contains longitudinal weighting variables for the 2020/21 to 2023 panel: *ls21_23* and *lsdrop21_23*.

In principle, data users have the option of making comparisons across several waves in a balanced panel design by combining the wave-specific participation probabilities (for details, see method report: Chapter 7). The longitudinal weighting variables and the variables of the reciprocal participation probabilities of all individual waves of the DEAS and the cross-sectional weights are provided in the weights dataset and are available for download to registered data users.

In addition, an *integrated cross-sectional weighting* is available for each survey wave from 2002 onwards. For this purpose, all subsamples of a wave are integrated into a joint cross-sectional sample. The subsamples are (a) panel cases with participation in the previous wave (reparticipants), (b) panel cases without participation in the previous wave (returnees after temporary absence) and (c) new baseline cases (first-time respondents, if a new sample is added in the survey year). The starting weight for the integration is the longitudinal weight of the observation wave for the participants from the previous wave. An auxiliary weight is calculated using two drop-out models (probability of non-participation, probability of return) for returnees after temporary drop-out (panel respondents who skipped the previous wave). For details, see Schiel et al. 2023: Chapter 7.

The initial weights of the subsamples are combined in a two-stage process by convex combination (composite weighting) to form an integrated cross-sectional weight by multiplying the respective initial weight by the relative share of the sub-sample in the total sample. The cross-sectional weight in DEAS 2023 regards the 2020/21 survey as a preliminary wave in order to determine the temporary defaults.

The integrated cross-sectional weights of the 2023 wave serve as starting weights for an adjustment to the distribution of characteristics in the 2022 microcensus. Post-stratification is carried out using "iterative proportional fitting" according to the characteristics age

³ See Hammon et al. 2016; Bethmann 2013; European Central Bank 2016

group*gender*region. Due to the relatively small number of respondents aged 91 and above, we recommend not applying the post-stratified cross-sectional weighting to this group. In the Scientific Use File for DEAS 2023, the variables *qsps_23_u91* and *qspsdrop_23_u91* contain the integrated cross-sectional weight for respondents under 91 years of age. A post-stratified cross-sectional weight without age limit and the integrated cross-sectional weights without post-stratification are available in the weighting dataset.

Due to the lack of a base sample in 2020, the DEAS data in the panel will become increasingly selective over time (panel mortality and real mortality). For this reason, there is a new cross-sectional weighting factor in the DEAS 2023 which, in addition to the previous post-stratification characteristics of gender, birth cohort and region, which also takes education into account for post-stratification.

Researchers who want to analyse DEAS 2023 cross-sectionally should use this weighting. For trend analyses (i.e. comparing several cross-sections side by side), we recommend using the classic cross-sectional weights from infas for reasons of consistency. The new extended cross-sectional weights have the suffix "_dza" and are (so far) only available for the 2023 wave.

The use of weights is at the responsibility of the individual user. Data weighting is one of several ways of dealing with different sampling probabilities and the problem of selective participation and panel mortality. The use of cross-sectional weighting is primarily intended for general descriptive presentations and analyses that relate to a single point in time.

4 CHANGES AND UPDATES IN PREVIOUS DEAS WAVES

Some changes were made in the previous waves, so that all SUFs for the years 1996-2023 will be issued with a new version number.

A number of variables were added, such as

- the occupational classifications and their derived variables (see sections 1.4.3 and 1.4.4, for waves 1996-2023)
- the harmonised education variables schledu3 and vocedu5 for waves 1996-2023
- the most frequently mentioned three or four diseases and the most frequently mentioned three injuries, forwaves 2008-2023
- the Big5 personality traits, for waves from 2020/21 onwards
- the four AgeCog scales (see <u>Documentation of Instruments and Variables</u>), for waves 1996-2023
- the Real Etate Price Index (only available on separate request from the FDZ), for waves 2008-2023

In addition, some variables were removed, such as westost_neu or couple-related occupational characteristics. These characteristics can be easily reproduced by dichotomising the federal states or by combining the occupational information of the respondents with the information of their (last)spouses. Long variable names were shortened uniformly, and there were some corrections and adjustments in the variable and value labels.

5 DATA FORMATS

The Scientific Use Files (SUF) for the years 1996 to 2023, the weights dataset and the SUF Meta (see Section 6) are available in both SPSS and Stata data format. If there are any version-related problems with importing the data, please contact the FDZ team directly (fdz@dza.de). Most data sets and the corresponding documentation are available in German and English, the remaining data sets and documentation will be provided soon in English as well.

6 CLASSIFICATION OF VARIABLE NAMES AND MISSING CODES

The 2023 wave also follows the variable naming system introduced in 2008. In the first two DEAS waves, variable names were still based on a card system coding tradition.

To make variables easier to identify, they have been assigned consecutively according to the question number since the third wave. The question numbers are grouped thematically, for example, the 300 block contains all information related to family situation.

The variables of the 2023 survey begin with an 'l' as a wave marker, followed by a "c" or "d" to indicate the type of survey (c = personal interview; d = Drop-off questionnaire). The following number corresponds to the respective question number. In the case of multiple answers, the answer options are distinguished by a consecutive numerical ending (e.g. lc427_1, lc427_2, etc.). Open-ended answers are indicated by the suffix 'o' (lc103b_5o).

In the case of multiple responses, the answer options are distinguished by a consecutive numerical suffix (e.g. kp34_1, kp34_2, etc.). Open-ended responses are identified by the suffix 'o' (e.g. kp9o).

A complete overview of all variable names ever used in the DEAS datasets is available in the variable correspondence list on the <u>FDZ-DZA documentation page</u>. This page also includes the codebook, which provides the basic frequencies for all variables available in the SUF DEAS Supplementers 2021, including the corresponding labels.

The assignment of missing codes follows the new system developed from wave 4 onwards. This facilitates the automatic definition of missing values and the understanding of the sources of non-valid data.⁴

⁴ In waves 1 to 3, missing values were coded either as 0, -1, -2 or as 7,8,9 or 97, 98, 99 etc. - depending on the characteristics of the valid values.

| SPSS | Stata | Label [Explanation] |
|------|-------|--|
| -1 | .a | Denied |
| -2 | .b | Don't know |
| -3 | .C | Filtered (question) |
| -4 | .d | Filtered (sample) [if the question is only aimed at panel or baseline respondents; not applicable in the 2020/21 wave] |
| -5 | .e | No Drop-off available |
| -6 | .f | Not specified (Drop-off) |
| -7 | .g | Deleted in data edition |

The following table describes the codes for missing values that occur in the data set.

In general, missing codes in the dataset are defined as user-defined system missings. Exceptions apply to waves 1 to 3 (1996, 2002, and 2008), in which this system for labelling missing values had not yet been introduced. For these waves, it is recommended to run frequency counts of the variables to be analysed prior to analysis, in order to identify missing codes within the valid value range (e.g. 8, 9, or 98) and to define them accordingly as missing values.

7 MODIFICATION OF DATA TO MAINTAIN DE FACTO ANONYMITY

To preserve the de facto anonymity of respondents, several variables were excluded from the SUF DEAS 2023, and the values of certain variables were edited:

- Almost all variables relating to spatial context were removed. Information on respondents' month and year of birth was also deleted.
- Individual variables containing open-ended responses were either deleted or edited to generalise highly specific details (such as years or locations). These changes are marked with a '#' prefix in the value labels.
- Rare family demographic characteristics were top-coded.
- Names recorded for survey administration purposes—for example, in mapping social networks—were also deleted.
- Rare annual data were grouped into broader categories ("top-coding") by recoding values within the existing numerical variable to either a central value in the new range or to the modal value. These were relabelled accordingly.
- In some cases, maximum values were combined into an open-ended top category (e.g. for household size).

For research requiring full access to the unmodified DEAS datasets—such as small-area analyses at the district level or studies involving neighbourhood context data (e.g. Microm data)—the DZA provides specially secured PC workstations for guest users. If you are interested, please contact the FDZ team at: <u>fdz@dza.de</u>.

8 THE COMPLEX DEAS SURVEY DESIGN

The DEAS uses a complex multi-stratified and clustered random sampling method to ensure that the sample is representative and reduces both random and systematic biases. Here's a breakdown of the methodology:

- 1. Stratification: Before sampling, the population is divided into different strata based on:
 - \circ Gender
 - Three age groups
 - Region (old or new federal states) Additionally, people from the oldest age group and the new federal states are oversampled to ensure adequate representation of these groups.
- 2. Clustering: To make the survey logistically feasible and cost-effective, it is clustered. Instead of randomly selecting individuals from the entire country, the sample is drawn from specific municipalities, which are divided into two groups: those from the old federal states and those from the new federal states. Within each group, municipalities are randomly selected based on size, meaning larger municipalities are more likely to be selected.
- 3. Sampling Process:
 - Proportional to size: Larger municipalities have a higher chance of being selected.
 - From 14,627 municipalities: A sample of 90 municipalities from the new federal states and 200 from the old federal states was drawn in 1996.
 - The refreshment sample 2021 is based on a new sample of municipalities. A sample of 59 municipalities from the new federal states and 142 from the old federal states was drawn in 2020 from 11,014 municipalities.
- 4. Why it matters:
 - Clustering and dependence: Respondents within the same community may share similar characteristics, which can lead to dependencies between observations. For instance, people in wealthy areas may have similar income levels.
 - Impact of clustering on analyses: If the clustered sampling method isn't taken into account, the analysis would treat respondents as independent, which could lead to an underestimation of variance and potentially produce misleading results..

The svyset command in Stata is used to define the survey design for complex survey data, taking into account stratification, clustering, and weights.

svyset psu, strata(strata_psu) fpc(fpc_1) vce(linearized) singleunit(scaled) || ssu [pw=weight], strata(strata_ssu)

- psu: This specifies the primary sampling units (PSUs), which are the clusters of the survey sample. In the case of DEAS, this corresponds to the municipalities or other units of clustering. Each PSU represents a group of respondents that might be similar to each other.
- strata(strata_psu): This option defines the strata for the primary sampling units (PSUs).
 These strata are defined by region (old or new federal states).
- fpc(fpc_1): This refers to the finite population correction (FPC) factor, which adjusts for the fact that the primary sampling units are drawn from a finite population. Whereas the impact of the correction is usually negligible, it comes with a great advantage: This option ensures that the variance is correctly adjusted when multiple waves of data are considered, if respondents repeatedly took part.
- vce(linearized): This option specifies the method for variance estimation. The linearised option is commonly used in survey data to compute robust standard errors that account for the complex design of the survey, including clustering and stratification.
- singleunit(scaled): This option adjusts for single units (clusters) that may have only one observation (for example, if a community has only one respondent sampled). The scaled option ensures that the variance for these units is correctly calculated by scaling it up based on the design.
- || ssu: This part indicates the secondary sampling units (SSUs) which are the individual respondents, sampled within each PSU. It tells Stata that there are multiple respondents within each primary sampling unit.
- [pw=weight]: This specifies the weight variable that should be used to adjust for unequal probabilities of selection, nonresponse, and other design features. The weight variable adjusts the analysis to ensure that the sample is representative of the population. You will need to create or select the appropriate weight variable based on the design of the DEAS.
- strata(strata_ssu): This option defines the strata at the level of the secondary sampling units (SSU), that are defined by gender and age, and (implicitly) region.

The appropriate weight must be selected or created independently according to the analysis design. All subsequent analyses that include the suffix -svy take into account the complex survey design of the DEAS, which was defined using the svyset command.

If fpc_1 is also taken into account, Stata also knows that the same SSU (the respondents) occur several times (e.g. if several cross-sections are analyzed). The variance calculations are corrected accordingly for analyses with svy commands.

In summary, the svyset command is crucial for correctly handling the DEAS survey design in Stata, ensuring that the data is analysed properly by accounting for complex design features such as stratification, clustering, and weighting.

9 DOCUMENTATION

In the <u>DEAS documentation section</u> of the FDZ-DZA website, you can view and download a range of valuable materials related to all available DEAS waves. These include:

- Complete Survey Instruments: The full questionnaires used in each wave of the DEAS, providing insight into the exact questions asked during data collection.
- Method Reports (infas): Detailed reports on the methodology used in the survey, explaining the sampling methods, data collection procedures, and other methodological considerations.
- Codebooks: A comprehensive listing of all the variables collected in each wave, along with their coding and definitions, essential for understanding how the data is structured.
- Working Aids: Tools to assist with specific aspects of the data, such as guides on interpreting characteristics like grandchild data.
- Variable Correspondence List (Excel Format): This document helps users identify how variables correspond across different waves, crucial for longitudinal analysis.
- Detailed Variable Documentation: In-depth descriptions of each variable, including its definition, coding, and any important considerations when analysing it.
- Explanation of Data Weighting: Information on how to apply the appropriate weights to ensure the data is representative and your analyses are methodologically sound.
- Publications based on DEAS: The website also offers a catalogue of publications based on DEAS data, available in a complete PDF format or as an interactive list with filter options, making it easier to find relevant research.

Most of these documents are available in English, allowing broader access for international researchers.

These resources are invaluable for anyone working with DEAS data, offering clear guidance on how to use the data correctly and ensuring your analyses are both accurate and rigorous.

10 OUTLOOK

The next DEAS survey will take place in 2026 and will combine a complete baseline survey with the survey of panel participants. The 2026 wave will bring some fundamental changes:

- a) There will no longer be an upper age limit for the sampling of respondents.
- b) There will also be a small sample of respondents in care facilities.

c) The survey instruments are comprehensively revised, simplified and supplemented with new questions.

LITERATURE

- 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 Labor Market and Social Security" (PASS) wave 6. (FDZ-Datenreport, 07/2013 (en)), Nuremberg: IAB (pp. 81-99). URL: <u>https://doku.iab.de/fdz/reporte/2013/DR_07-13.pdf</u> [last viewed on 28.02.25].
- Federal Office for Research on Building, Urban Affairs and Spatial Development (BBSR) in the Federal Office for Building and Regional Planning (2012). Spatial delineations and spatial types of the BBSR. Bonn: BBSR. URL: (https://www.inkar.de/) [last viewed on 01.05.22].
- Ebner, Christian, & Rohrbach-Schmidt, Daniela (2019). Occupational prestige in Germany for the Classification of Occupations 2010: Description of the methodological approach, first descriptive results and quality of measurement. Version 1.0. Bonn: Federal Institute for Vocational Education and Training. URL: <u>https://lit.bibb.de/vufind/Record/DS-183865</u> [last viewed 03.03.25]
- European Central Bank (2016). The Household Finance and Consumption Survey (HFCS): Methodological report for the second wave, ECB Statistical Paper Series No. 17). Frankfurt a. M.: European Central Bank. URL: <u>https://www.ecb.europa.eu/pub/research/statistics-papers/html/index.en.html [last</u> viewed on 01.06.22].
- Hammon, A., Zinn, S., Aßmann, C., & Würbach, A. (2016). Samples, weights, and nonresponse: the adult cohort of the National Educational Panel Study (Wave 2 to 6) (NEPS Survey Paper No. 7). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study. URL: <u>https://www.nepsdata.de/Datenzentrum/Publikationen/NEPS-Survey-Papers [last viewed on 01.06.22].</u>
- Holland, J. L. (1997). Making vocational choices: A theory of work personalities and work environments (3rd ed.). Odessa, FL: Psychological Assessment Resources.
- Klaus, D., Engstler, H., Mahne, K., Wolff, J. K., Simonson, J., Wurm, S., & Tesch-Römer, C. (2017). Cohort Profile: The German Ageing Survey (DEAS). International Journal of Epidemiology, 46(4), 1105-1105g. doi: <u>https://doi.org/10.1093/ije/dyw326</u>
- Mahne, K., Wolff, J. K., Simonson, J., & Tesch-Römer, C. (Eds.) (2017). Ageing in transition. Two decades of the German Ageing Survey (DEAS). Wiesbaden: Springer VS. Open Access: doi: https://doi.org/10.1007/978-3-658-12502-8.
- Schiel, S., Ruiz Marcos, J., Dickmann, C., & Fahrenholz, M. (2023). Methodological report: German Ageing Survey (DEAS): Implementation of the 8th survey wave 2023. Bonn: infas. URL: <u>https://www.dza.de/forschung/fdz/deutscher-alterssurvey/deasdokumentation</u>
- Federal Statistical Office (2024). Statistical report. Community statistics on income and living conditions (microcensus sub-sample on income and living conditions). Federal

Statistical Office: Wiesbaden. Stuth, Stefan (2022). Occupational Environments - An Activity-Based Approach Based on Holland. WZB - Social Science Research Center Berlin. Data file version 1.0.0. doi: https://doi.org/10.7802/2502.

- Thiel, P. (2024). FDZ Data Description: Real-Estate Data for Germany (RWI-GEO-RED v11) -Advertisements on the Internet Platform ImmobilienScout24 2007-06/2024. Essen: RWI. URL: <u>https://www.rwi-</u> <u>essen.de/fileadmin/user_upload/RWI/FDZ/FDZ_Datensatzbeschreibung_RED_v11.pdf</u> [last viewed on 28.02.25]
- Vogel, C., Wettstein, M., & Tesch-Römer, C. (Eds.) (2019). Women and men in the second half of life. Ageing in social change. Wiesbaden: Springer VS. Open Access doi: <u>https://doi.org/10.1007/978-3-658-25079-9</u>