
June 2022

German Ageing Survey (DEAS):
Brief description of the SUF DEAS2020/21, version 1.1

Stefan Stuth

Recommended citation: Stuth, Stefan (2022): German Ageing Survey (DEAS): Brief Description of the SUF DEAS2020/21, Version 1.1. Berlin: German Center of Gerontology.

German Center of Gerontology (DZA)
Research Data Center (FDZ-DZA)
Manfred-von-Richthofen-Straße
212101 Berlin
Phone+49 (0)30 - 26 07 40-0Fax +49 (0)30 -
26 07 40-33E-mail fdz@dza.de

TABLE OF CONTENTS

1.	The German Ageing Survey (DEAS) - a study on the second half of life	4
1.1	Sample	4
1.2	Topics of the main survey 2020/21	6
1.3	Changes and innovations in the main survey 2020/21	6
1.3.1	Mode change	6
1.3.2	Survey content	7
1.3.3	New occupational classification	7
1.3.4	Flag variables	7
1.3.5	Data provision	8
1.3.6	Anonymization	8
1.3.7	Regional data	8
1.3.8	Data documentation	9
1.3.9	Weights	9
1.3.10	The Complex Surveydesign of the DEAS	9
2	Response Rate and identity checks in the main survey 2020/21	9
3	Data weighting of the main survey 2020/21	10
4	Data formats	11
5	Systematics of variable names and missing codes in DEAS 2020/21	11
6	Meta-data on participation history of the DEAS respondents.	13
6.1	Participation profile	16
6.2	Whereabouts data	17
7	Modification of data in the SUF to preserve de facto anonymity.	17
8	The complex survey design of the DEAS	18
9	Other	19
9.1	Documentation for DEAS202 1	19
9.2	Regional data	19
	References	19

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

The German Ageing Survey (DEAS) is a long-term study by the German Center for Gerontology (DZA) on changes in the life situations and aging trajectories of people in the second half of life, funded by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ). The study is based on nationally representative cross-sectional and longitudinal surveys of several thousand participants aged 40 and older. The people involved are questioned comprehensively about their objective and subjective living conditions. In addition, test procedures are used to survey cognitive performance and physical health. The DEAS represents a unique data source in Germany for interdisciplinary research on the life situations and life courses as well as the subjective views and life plans of aging and old people because of its cohort-sequential longitudinal design, its breadth of topics, its national representativeness with a comparatively high number of cases, and the long observation period.

The microdata of the German Ageing Survey (DEAS) is only available for scientific research with non-commercial purposes. The Research Data Center gives researchers access to the information collected in the Ageing Survey and advises them on its use. The anonymized DEAS data of all completed survey waves and the documentation materials are available free of charge from the FDZ-DZA (<https://www.dza.de/forschung/fdz>).

The first wave of the DEAS survey was conducted in 1996, followed by survey waves in 2002, 2008, 2011, 2014, 2017, 2020, and 2020/21. The face-to-face oral interviews of panel participants and the interviews of a new-drawn baseline sample, originally prepared for 2020, could not take place as planned due to the Corona pandemic. To kick off the seventh wave of DEAS, a written-postal survey of panel participants was conducted at short notice in June and July 2020. In November 2020, the main survey started with an extensive telephone survey of panel participants, supplemented by a written drop-off for self-completion. The main survey ended in February 2021, with the last interview conducted on March 1, 2021. It was not possible to interview the new baseline sample, because only address data, but no telephone numbers, were available for the baseline sample and under Corona conditions the survey could not be conducted as a face-to-face interview.

In terms of content, the 2020/21 main survey followed the regular DEAS waves. Methodologically, however, adjustments had to be made in the survey instruments to account for changes in the survey mode (from face-to-face interviews to personal 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.

1.1 Sample

The DEAS combines cross-sectional and longitudinal sampling. Since 1996, new samples of the 40- to 85-year-old resident population in private households have been drawn at six-year intervals as stratified and clustered random samples stratified by age group, sex and part of the country (east/west)

from the population registers and interviewed personally by interviewers. Starting with wave 2 (2002), panel-ready respondents from the preceding surveys were also reinterviewed. This excludes respondents from the separately drawn sample of foreigners in 2002, who were not included in the longitudinal sample. After the 2008 survey, the interval between panel waves was shortened from six to three years.

In wave 1 (1996), the selection was limited to persons with German nationality; in wave 2, two separate samples of Germans and non-Germans were drawn for the baseline survey in 2002. Since 2008, new baseline samples have been drawn from the population aged 40 to 85 in private households (place of main residence) without distinguishing nationality.

Originally, the combined survey of a new-drawn sample with the survey of panel participants was planned for 2020. However, the Corona pandemic stood in the way of implementing this plan, as the population registration offices can only provide address data, but not telephone numbers. Under pandemic conditions, however, face-to-face interviews could not have been implemented without health risks. The interviewing of the new sample could not be realized accordingly in 2020. For this reason, the 2020/21 main survey targeted all panelists who had previously participated in the DEAS at least once. Hence, the basis for the longitudinal survey was all participants from the 1996 to 2014 baseline samples who could still be reached. In total, valid interviews are available for 5,402 persons aged 46 to 100. Of these, 4,103 people last participated in the brief written self-completion survey in summer 2020.

The SUF DEAS2021 dataset with anonymized data from the 2020/21 main survey contains information on survey participants from the following baseline samples:

Sample 2020/21	n
Panel respondents from B1996	554
Panel respondents from B2002	572
Panel respondents from B2008	1.717
Panel respondents from B2014	2.559
Total respondents	5.402

With the total available Scientific Use Files (SUF) of all previous survey waves (20,714 respondents, 50,256 valid interviews), individual and societal developments can be observed over the period of more than 24 years (see Klaus et al. 2017). The sample design of the German Ageing Survey opens up various analysis options in this context. There are the options of a cross-sectional investigation of life situations and life contexts in the respective survey year, for the analysis of social change between 1996 and 2020/21 (see Mahne et al. 2017), and of individual development over three to 24 years of life with the panel data (Vogel et al. 2019). Another perspective arises 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. Thus, the comparison of the development in a constant age period for different birth cohort groups is possible (cohort sequence analysis), for example, the prospective comparison of transition processes into retirement or the health development of individuals.

1.2 Topics of the main survey 2020/21

As in the previous waves, the study participants were also comprehensively interviewed about their living situation in wave 8. Information was collected on the following topics:

- Work and retirement
- Partnership, family and intergenerational relations
- Social networks and social support
- Recreational activities and civic engagement
- Housing and mobility
- Economic situation and economic behavior
- Subjective well-being
- Health and health behavior
- Need for help and care
- Attitudes, norms and values
- Basic sociodemographic data.

As in the previous waves, the survey was divided into two main parts. The bulk of the information was collected through a personal telephone interview with a comprehensive questionnaire in the target person's household. The second part of the survey consisted of a written questionnaire to be completed by the respondents themselves ('drop-off'). Here, more subjective topics - such as attitudes, values, images of age, well-being, more sensitive areas (e.g. details on material situation and health) and topics that did not require differentiated filtering were queried. Most of the items eliciting psychological scales were also collected in drop-off. 619 of the 4,419 respondents made use of the option to complete the drop-off as an online questionnaire.

1.3 Changes and innovations in the main survey 2020/21

1.3.1 Mode change

The standard survey mode of the DEAS is a face-to-face computer-assisted interview (CAPI). Due to the Corona pandemic, the survey mode had to be changed to a Computer Assisted Personal Interview on the phone (CAPI-per-Phone).¹ In the oral interview with an average duration of about 75 minutes, the

¹ Infas uses the term Capi-per-Phone to emphasize essential differences to the classic CATI interview. In contrast to regular CATI interviews, the survey was not conducted by anonymous telephone studios, but

questions were essentially used analogous to the 2017 wave. In the 2017 main survey, the proxy interview option was newly introduced: Panel subjects who felt unable to conduct the interview independently due to health reasons could specify that a person they trusted answer the questions in their place. Because of the switch to Capi-per-Phone this approach was not applicable and no proxy interviews were conducted in the 2020/21 survey. Due to the pandemic-related switch to CAPI-per-Phone, it was also not possible to conduct physical function tests to determine respiratory volume (pulmonary function test) or cognitive ability tests (number-and-digit test). Since the interviewers were unable to visit the respondents, it was not possible for them to provide information on the respondents' residential environment.

Even though the interview continued to be conducted orally, most of the questions had to be adapted with the mode change. While the wording of the questions was hardly affected, aids such as lists, person cards, which helped the interviewees to focus on specific family members or important members of their personal networks, could no longer be used. Furthermore, the scope of the answer categories had to be greatly reduced for some questions, since aids such as lists or cards were no longer possible to be used.

1.3.2 Survey content

Compared to the 2017 wave, there were few changes to survey content. Examples: Questions about the care and support needs of parents and in-laws still living were expanded. If respondents indicate that they support others, they are now asked if the person they support has dementia. Furthermore, there are new questions on the compatibility of care and work that cover the use of the new family care leave or the interest-free loan.

As a result of a call for modules by the DZA three new modules on prescriptive age-related withdrawal and activation norms, big five personality traits, and the preferred and subjective life expectancy are available. However, an embargo period applies to this new content. It will be available from January 2023 (for more information [see](#) the document "limited data availability for invited modules").

1.3.3 New occupational classification

In order to be able to analyze occupations and occupational careers using a domestic classification (instead of the ISCO08 classification designed for international comparisons), the classification of occupations (KldB 2010) is now available. Information on the current occupation, the very first occupation held or the last occupation held (KldB2010) is available for respondents (and in most cases for their partners) from the 2014 survey year to the survey year 2020/21.

1.3.4 Flag variables

In the course of processing and checking the survey data by the DZA, consistency and plausibility checks are carried out, the results of which will also be available to data users starting with the DEAS 2020/21.

by interviewers who formerly had conducted the survey in person in previous years at the interviewee's homes.

Flag variables highlight cases that have become conspicuous as a result of the checks. These checks take place for those variables for which either additional information from other questions is available to allow matching or thresholds can be defined for which it is implausible to fall below or to exceed the threshold. The variables for the flags are available in the instrument documentation and the underlying test rules are detailed. The flag variables start with the flag prefix followed by the variable name to which the flag variable refers. Example: The variable `kc105c`, records the amount of the deduction from the pension that pensioners have to accept when they retire early from the labor market. The amount of the deduction cannot exceed a legally fixed limit. However, some respondents indicated a deduction amount that exceeds this limit. The variable `flag_kc105c` marks these cases. All respondents who were inconspicuous get the value 0, while the value 1 is entered for cases that gave values above the limit.

The results of these tests are additionally cumulated block by block to highlight persons who were conspicuous more than once in a respective block of questions. The variables look like this: `flag_BLOCK-NRcum`. Example: The variable `flag_kc500cum` adds up the (repeated) inconsistencies in the health block (`kc501-kc556a`).

The variable `flag_kctotal` sums the number of all inconsistent or implausible statements in the entire oral interview.

More detailed information on the design, content and procedure of the 2020/21 survey can be found in the infas method report (Schiel et al. 2021). The instruments and method reports for all survey waves of the DEAS are available for download on the pages of the FDZ-DZA:

<https://www.dza.de/en/forschung/fdz>.

1.3.5 Data provision

With the release of the DEAS 2020/21, the mode of data provision changes. Users will automatically receive the complete data package with all waves from 1996 to the most recent survey. This way, the FDZ addresses a problem faced especially by data users who do not yet have experience with the DEAS data: DEAS collects certain data sparingly so as not to unnecessarily increase the question burden on respondents. For example, information is collected for occupation held by the respondents only if it has changed. If nothing has changed, sparsely collected variables receive the code for filtered. With the delivery of all waves, data users now have the opportunity to fill these gaps without having to decide on specific survey years at the time of application or, if necessary, to reorder additional data sets later.

1.3.6 Anonymization

In order to guarantee the anonymity of the respondents, the FDZ regularly checks whether existing protective measures are sufficient. In this course, only the following regional characteristics are still available in the SUF: State, BBSR county type and the "primary sampling units" (PSU). The PSUs are communities, which have been coarsened in a few cases to the county-level and are masked.

1.3.7 Regional data

Due to the high requirements for maintaining the anonymity of the respondents, it will no longer be possible to match additional regional information to the SUF data. Researchers who want to do

analyses with a regional focus will have to do the analyses on-site (at the DZA) at the guest researcher workstation. However, the FDZ is examining the possibility of making a coarser regional key indicator available (spatial planning regions) that could be used instead to match additional regional information to the scientific-use-files of the DEAS.

1.3.8 Data documentation

From now on, the detailed documentation of the generated variables is no longer located in this document, but in the variable documentation.

1.3.9 Weights

There have been minor adjustments in the default models that serve as the basis for the weighting factors. These can be found online in the document "Notes on innovations in DEAS weighting" at <https://www.dza.de/en/research/fdz/german-ageing-survey/deas-documentation>.

1.3.10 The Complex Surveydesign of the DEAS

As of now, additional features are available in the meta-data to allow researcher to take the complex survey design of the DEAS into account in their analyses (see chapter 8).

2 RESPONSE RATE AND IDENTITY CHECKS IN THE MAIN SURVEY 2020/21

From a gross input sample of 8,379 addresses, 5,418 CAPI interviews were realized. This corresponds to a response rate of 66 percent. In addition, 4,425 drop-offs were completed, which represents a response rate of 82 percent (cf. infas method report: Schiel et al. 2021).

In two steps, the information on year of birth/age and gender were used to check whether (1) the person interviewed in 2020/21 was identical to the person interviewed in the previous waves and whether (2) the person interviewed orally was identical to the person who completed the written supplementary questionnaire. Of the 5418 cases, 11 cases were excluded by Infas as not evaluable. Five additional cases discontinued their interviews. These cases were also deleted. No cases had to be deleted as part of the identity check with regard to previous waves: the individuals interviewed in 2020/21 were the actual panelists. After completion of the checks, 5402 valid interviews are available.

Five oral respondents did not match the individuals who completed the written supplemental questionnaire. The drop-off information was completely deleted for these cases. All variables were set to -7 (deleted in data preparation). Cases were from the paper pencil. One drop-off questionnaire was terminated very early. The little DO information available was completely deleted for this case. All variables in these cases were also set to -7 (deleted in Data Preparation). The case is from the CAWI. At the end of the drop-off & CAPI-per-Phone identity check, 4,419 drop-offs remained (619 CAWI and 3,800 Paper Pencil).

3 DATA WEIGHTING OF THE MAIN SURVEY 2020/21

In close coordination with the DEAS project team of the DZA, ifas carried out the data weighting. The procedure chosen is basically the same as that used in other panel studies.² The longitudinal weighting is based on failure models (logistic regression) to determine the probability of participation from one survey wave to the following survey wave (excluding persons who died between the two waves) with a limited set of predictors (see Schiel et al. 2021: chapter 8.1). The longitudinal weight is obtained by multiplying the cross-sectional weight of the initial wave by the inverse of the participation probability in the follow-up wave. This procedure has been used for every survey year since 2002. The datasets with the weights for the current and the older Scientific Use Files of the DEAS, contains also the longitudinal weights `ls20_21` and `ls17_21`.

Each data user has the possibility to make comparisons across several waves in a balanced panel design by combining the wave-specific participation probabilities (for details see Methodological Report: chapter 7). The longitudinal weighting variables and the variables of the reciprocal participation probabilities of all individual waves of the DEAS are provided in a separate file and are available for download to registered data users.

In addition, an *integrated cross-sectional weighting* is carried out for each survey wave from 2002 onward. For this purpose, all subsamples of a wave are integrated into a common cross-sectional sample. The subsamples are (a) panel cases with participation in the previous wave (re-participants), (b) panel cases without participation in the previous wave (returnees after temporary dropout) and (c) new baseline cases (first-time respondents if a new sample is added in the survey year). The initial weight for integration is the longitudinal weight of the observation wave for continued participation of respondents who also participated in the previous wave. For returnees after temporary dropout (panel respondents who skipped the previous wave), an auxiliary weight is calculated using two dropout models (probability of non-participation, probability of return) (for details see Schiel et al. 2021: chapter 7.2). The initial weights of the subsamples are combined in a two-step procedure by convex combination (composite weighting) to form an integrated cross-sectional weight by multiplying the respective initial weight by the relative share of the subsample in the total sample. The cross-sectional weight in the DEAS 2020/21 uses the 2020 short survey as the preliminary wave to determine the temporary dropouts and participants who participated in the last wave.

The integrated cross-sectional weights of wave 2020/21 serve as initial weights for an adjustment to the characteristic distribution in the microcensus of 2020.³ The poststratification is done by means of "iterative proportional fitting" according to the characteristics age group*gender*country part. Due to the relatively small number of respondents aged 91 and older, we recommend not applying the poststratified cross-sectional weight to this group. In the Scientific Use File on the DEAS 2020/21, the

² See Hammon et al. 2016; Bethmann 2013; European Central Bank 2016.

³ Our thanks go to Robert Herter-Eschweiler from Destatis for providing us with the necessary aggregated data from the German microcensus.

variable *qsps_21_u91* contains this integrated cross-sectional weight. A poststratified cross-sectional weight without age limit and the integrated cross-sectional weights without poststratification are available in a separate file for registered data users.

The use of weights is the responsibility of the data user. Data weighting is an important way to deal with different drawing probabilities for sub-populations and the problem of selective participation and panel mortality of specific sub-populations. The use of cross-sectional weighting is primarily intended for general descriptive presentations and analyses that refer to a single point in time.

There have been minor adjustments in the default models that serve as the basis for the weighting factors. These can be found online in the document "Notes on innovations in DEAS weighting" at <https://www.dza.de/en/research/fdz/german-ageing-survey/deas-documentation>. The changes in weighting have occurred for two reasons. First, the short survey from summer 2020 must be integrated into the weighting scheme. However, the 2020 short survey does not contain all the characteristics that are normally included in the models of the regular DEAS surveys. The document presents the deviations. Second, the DEAS team, in collaboration with the FDZ of the DZA, had considered whether it would be useful to extend the default models slightly to avoid systematic bias due to omitted-variable bias. The adjustments, even if only minor, will apply to the 2020/21 survey and to all subsequent surveys. The document presents the adjustments.

4 DATA FORMATS

The Scientific Use File (SUF) 2020/21 as well as the SUF Meta (see section 6) are available in SPSS format as well as in Stata data format. If there are version-related problems with reading in the data, then please contact the FDZ-DZA consulting service (fdz@dza.de). All data and related documentation are available in [German](#) and [English](#).

5 SYSTEMATICS OF VARIABLE NAMES AND MISSING CODES IN DEAS 2020/21

The seventh wave of the DEAS names the variables using a system introduced in 2008. In the first two waves of the DEAS, the variable names were still assigned according to the coding tradition of a card system.⁴ Starting with the third wave variables were assigned names based on their question number.

⁴ For example, question 101 on employment status from the second wave face-to-face interview in 2002 (labeled 'w') is in the *second position* on Map 14 and is therefore *w14_12*.

The question numbers in the DEAS waves are grouped thematically (for example, the 300 question block contains all the data on family situation).

The variables of the 2020/21 survey begin with a 'k' as a wave marker, followed by a "c" or "d" to indicate the survey type (c = oral interview; d = written or online 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., kc427_1, kc427_2, etc.). Open statements are indicated by the suffix 'o' (kc103b_5o).

A complete overview of all variable names ever used in the DEAS datasets can be found in the variable correspondence list on the [documentation page of](#) the FDZ-DZA. The documentation page also contains the codebook with the basic enumeration of all variables present in the SUF DEAS 2020/21, including the labels.

The assignment of missing codes follows the new system developed from wave 4. This facilitates the automatic definition of missing values and standardizes the characteristics and causes for the presence of non-valid data. ⁵

⁵ In the data of 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 expression of the valid values. A uniform missing definition across all variables was thus not possible.

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

ValueSPS	Value Stata	Label [Explanation]
-1	.a	No answer
-2	.b	don't know
-3	.c	filtered (question)
-4	.d	filtered (sample) <i>[if the question is directed only at panel or baseline respondents; not applicable in wave 2020/21].</i>
-5	.e	no drop-off available
-6	.f	no specification (<i>drop-off</i>)
-7	.g	deleted in data preparation

In general, the missing codes in the dataset were formatted as missing values (user-defined-system missings). The exception is waves 1 to 3 (1996, 2002, and 2008), which do not follow the general system of missings naming. For these waves, we recommend performing frequency counts of the analyzed variables before the analysis to identify missing codes in the valid value range (e.g., 8, 9, or 98) in advance and declare or code them as missings.

6 META-DATA ON PARTICIPATION HISTORY OF THE DEAS RESPONDENTS.

In order to improve the overview, to simplify the merging of analysis data sets, and for history-event analyses, the FDZ provides a meta-data set that contains all persons who have ever been participated in a DEAS survey. In detail, the following variables can be found in the meta-dataset, some of which are also contained in the wave-specific SUFs (as of January 2021):

Variable	Label	Explanation
case number	Case number (anonymized respondent ID)	to merge information from respondents in all SUFs (1996-2020/21)
stich	Sample membership	Indicates the baseline sample from which the respondent originated
part_96	Participation in 1996	yes; no

Variable	Label	Explanation
part_02	Participation in 2002	yes - base 2002; yes - foreign 2002; yes - panel; no
part_08	Participation in 2008	yes - base 2008; yes - panel; no
part_11	Participation in 2011	yes - panel; no
part_14	Participation in 2014	yes - base 2014; yes - panel; no
part_17	Participation in 2017	yes - panel; no
part_20	Participation in short survey 2020	yes - panel; no
part_21	Participation in 2020/21	yes - panel; no
part	Participation profile, 1996-2020/21	Indicates in which survey years the person was interviewed
lastpart	Last participation	Year of last survey
n_obs	Number of participations	Number of individual survey participations (1-8)
entry	DEAS entry year	analogous to stich: calendar year of first survey
yob	Year of birth	Year of birth (possibly updated in the course of data cleansing)
gender	Gender	Gender
natGerman	German citizenship	Characteristic at first interview: German; non-German
migrat	Migration background	without MH; immigrated; MH without own immigration experience
education4	Education level, 4-level	determined from highest level of education and vocational training: low, medium, high, highest
isced	Educational level according to ISCED, 3 levels	low (ISCED 0-2), medium (ISCED 3-4), high (ISCED 5-6)
bland_96 to bland_21	Federal State	according to place of residence at the time of the survey
westost_96 to westost_21	West-East affiliation of the place of residence	with division of Berlin (former western part of Berlin = West and former eastern part of Berlin = east)
westost_21new	West-East affiliation of the place of residence	The whole city of Berlin is coded as part of eastern Germany

Variable	Label	Explanation
bbsr_district_type96 to bbsr_district_type21	district type	Structural district type of place of residence according to BBSR/INKAR
rlc2002_kat to rlc2021_kat	Return code	Reasons for not participating in wave 2, 3, 4, 5, 6, 7 (short), or 7 (regular) (or code for participation).
vitalstatus_last	Last known vital status	Indicates whether the person is still alive or has since died
vitalstatus_source	Year of vital status information	Calendar year from which the information on vital status originates
surv_birth	Survival time since birth	Months since birth until death or end of observation
surv1996	Survival time since survey 1996	Months since interview in 1996 until death or end of observation
surv2002	Survival time since survey 2002	Months since interview in 2002 until death or end of observation
surv2008	Survival time since survey 2008	Months since interview in 2008 until death or end of observation
surv2011	Survival time since survey 2011	Months since interview in 2011 until death or end of observation
surv2014	Survival time since survey 2014	Months since interview in 2014 until death or end of observation
surv2017	Survival time since survey 2017	Months since interview in 2017 until death or end of observation
surv2021	Survival time since survey 2021	Months since interview 2021 until death or end of observation
psu	Primary Sampling Unit	The PSUs are the 290 municipalities drawn randomly and in proportion to size from which respondents were randomly (and stratified) selected. In cases where the communities are very weakly populated ($N < 10$), the variable was coarsened for the SUF. Respondents were assigned to the parent county, along with respondents from municipalities that also belonged to the corresponding county. Of the original 290 PUSs, 266 PSUs remain after coarsening. The coarsened overlap-free PSU variable refers to the year of the initial survey and is masked.

Variable	Label	Explanation
		That is, the PSUs are differentiated from each other by a sequence of digits that no longer allow any inference about the original survey unit.
ssu	Secondary Sampling Unit	The variable ssu corresponds to the variable fallnum and represents a unique and unchangeable personal identification number.
strata_gemeinde	Stratification of the communities	The PSU belongs to the old or new federal states
strata	Stratification of the respondents	Assigns SSUs - respondents - to one of 12 stratification cells resulting from the combination of 3 age groups, gender, and old or new states. Assignment to one of the 12 cells refers to the time of their initial interview.

Above all, the creation of a longitudinal data set can be easily facilitated by using the meta data. The whereabouts data and drop-out information also open up the possibility of mortality or drop-out analyses. If there is interest in a differentiated analysis of the reasons for drop-out, even more detailed drop-out codes are also available on request from the FDZ.

The complex surveydesign of the DEAS can now (and should) be considered for analyses using the variables psu, ssu, strata_gemeinde and strata (see 8. The Complex Surveydesign of the DEAS).

6.1 Participation profile

The variable *part* summarizes the participation in all previous DEAS survey years. All respondents are assigned an eight-digit value: for example, the expression 11111111 means that the respondent has participated in all eight DEAS surveys, persons with the expression 01001111 participated for the first time in 2002, were not present in 2008 and 2011, but went through the questionnaire program again in 2014, 2017, 2020 and 2020/21. With the help of the participation profile, the analysis sample can be selected and compiled as preferred by the data user. If a selection is to be made only by the number of individual participation frequency, this can be done using the variable *n_obs*. Based on the pre-selected observations, the related information from the wave-specific SUFs can be added in further steps via the identification variable *fallnum*. If you have any questions about the use of the meta-dataset, please consult the FDZ: fdz@dza.de.

6.2 Whereabouts data

With each survey the return code as well as information from panel maintenance actions provide information on the respondents' vital status. In the case of known deaths, the date of death was also determined, if possible. The variable *vitalstatus_last* indicates the last known vital status of all DEAS participants ever interviewed (0=living, 1=dead). If respondents were interviewed only once and never provided a panel willingness, it refers to the date of the first and only interview, otherwise it refers to the last information status. The variable *vitalstatus_source* contains the calendar year in which information on the last known vital status is available. The variable *surv_birth* contains the months elapsed since birth to the date of death or to the last known date of survival. The variables *surv_birth*, *surv1996*, *surv2002*, *surv2008*, *surv2011*, and *surv...* on survival duration since birth and since the respective survey, form the basis for event analyses on mortality risks (together with the variable *vitalstatus_last* (death or censoring indicator)).

The variables *rlc2002_kat* to *rlc2021_kat* contain the categorized response codes on the occasion of the contacts for the DEAS surveys 2002 to 2021. They provide information on whether the target person could be interviewed or which reason was decisive for the respondents' failure to provide of a valid interview. Interviews that were subsequently canceled contain the code 6 for failure for other reasons.

7 MODIFICATION OF DATA IN THE SUF TO PRESERVE DE FACTO ANONYMITY.

In order to preserve the de facto anonymity of the study participants, some variables were not included in the DEAS2021 Scientific Use File (SUF) and coarsenings of the expressions were made for a number of variables.

- Among other things, all variables on the spatial context were deleted, which under certain circumstances would make it possible to determine the respondent's municipality of residence. The data on the respondent's birthday and- month were also removed.
- Individual variables with open answers were deleted or open answers containing very specific characteristics were generalized. Such changes in open data were indicated by the prefix '#'.
– Rare family demographic characteristics were coarsened.
- Name details that were recorded for auxiliary interview purposes e.g. Names of the persons of the social network were also deleted
- Rare year data were grouped into categories. This was done by recoding the value in the existing numeric variable to values of the middle of the summarized range or to the modal value (with appropriate labeling by value labels).
- Occasionally, maximum values were combined into a maximum category (e.g., household size).

At the DZA, specially secured PC workstations have been set up for guest users, which allow evaluations to be carried out with the complete DEAS datasets, e.g. for small-scale analysis at the district level or

including contextual characteristics of the immediate neighborhood. If you are interested, please contact the FDZ consulting service: fdz@dza.de.

8 THE COMPLEX SURVEY DESIGN OF THE DEAS

The DEAS is not based on simple random sampling, but on multiple stratified and clustered random sampling. To prevent both random bias (e.g., men might be disproportionately sampled by chance in simple random samples) and systematic bias (because, for example, the very old are less likely to participate in the survey than younger respondents for health reasons), the population is divided into strata in advance. In the DEAS, these strata are created from a combination of gender, three age groups, and whether the respondent belongs to the old or the new federal states. Individuals belonging to the oldest age group and the new federal states were oversampled. The random drawing of respondents takes place within these strata. Since the DEAS is (usually) conducted as a face-to-face interview, it is logistically and cost-wise almost impossible to cover the entire Federal Republic. The travel and organizational costs for conducting the face-to-face interviews would be disproportionately high. For this reason, the DEAS is clumped. From the 14627 municipalities available for selection in 1996, a size-proportional sample was drawn from the old federal states and a size-proportional sample from the new federal states. Size-proportional means that larger municipalities had a higher drawing probability than smaller municipalities. As a result, the stratified individual sample was drawn from 90 municipalities in the new states and 200 municipalities from the old states.

If this complex sampling design is not taken into account, clustering, for example, leads to respondents being counted as independent units of observation. However, respondents who come from the same communities may be similar to each other (e.g., respondents from poor/rich communities are likely to be very similar with regard to their income). If clumping is not taken into account, the variance will be underestimated and significant effects/relationships will be found in analyses, which would not be found if the survey design was taken into account correctly.

For this reason, additional variables are provided in DEAS to allow the statistical analysis programs to take the complexity of the survey design into account. In Stata, this can be done with the svy suite. In the dataset containing the meta-data, the following characteristics are provided for this purpose: psu (Primary Sampling Unit), ssu (Secondary Sampling Unit), strata_gemeinde (stratification of the PSUs), strata (stratification of the SSUs).

The svyset command in Stata, which should be used to define the DEAS's surveydesign features, looks like this:

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

The appropriate weight must be chosen or created by the data users according to their design of the analysis. Following analyses that include the -svy- suffix take complex surveydesign of the DEAS into account that was defined using the svyset command.

9 OTHER

9.1 Documentation for DEAS202 1

On the [website](#) of the FDZ-DZA, various documentation materials for all available DEAS waves are available for download in the DEAS documentation section: the complete survey instrument, the infas method report, the codebook, various working aids (such as on grandchildren's characteristics), a variable correspondence list as tables, the detailed variable documentation, and an explanation of data weighting. Most of these documents are additionally available in English.

Also listed on the FDZ-DZA page are all publications based on the DEAS known to us - as a complete PDF document or for convenient browsing with filter options.

9.2 Regional data

Regional characteristics at the level of counties and independent cities, and neighborhoods can be added to the survey data and used for analyses – but only in the especially secured guest-researcher workstation at the DZA in Berlin. Interested parties should contact the FDZ-DZA consulting service directly.

REFERENCES

Bundesamt für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (2012). BBSR spatial delineations and spatial types. Bonn: BBSR. URL: (<https://www.inkar.de/>) [last viewed 01.05.22].

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: <https://www.ecb.europa.eu/pub/research/statistics-papers/html/index.en.html> [last viewed 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: <https://www.neps-data.de/Datenzentrum/Publikationen/NEPS-Survey-Papers> [last viewed 01.06.22].
- 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: <https://doi.org/10.1093/ije/dyw326>
- Mahne, K., Wolff, J. K., Simonson, J., & Tesch-Römer, C. (Eds.) (2017). *Aging 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., Schulz, S., Dickmann, C. & Fahrenholz, M. (2021). Methods report. German Ageing Survey (DEAS): Implementation of the 7th survey wave 2020/21. Bonn: infas. URL: <https://www.dza.de/forschung/fdz/fdz-publikationen> [last viewed 01.06.22].
- Trappmann, M. (2014). In: 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*. (FDZ-Datenreport, 07/2013 (en)), Nuremberg: IAB (pp. 81-99). URL: <https://fdz.iab.de/187/section.aspx/Publikation/k131218302> [last viewed 01.06.22].
- Vogel, C., Wettstein, M., & Tesch-Römer, C. (Eds.) (2019). *Women and men in the second half of life. Aging in social change*. Wiesbaden: Springer VS. Open Access. doi: <https://doi.org/10.1007/978-3-658-25079-9>