Innovations in the weighting of the DEAS

FDZ-DZA, March 2023

The following sections describe changes in the weighting that have occurred for three reasons:

A) The short survey from the summer 2020 must be integrated into the weighting scheme. However, the 2020 short survey does not contain all characteristics, which are normally included in the failure models for the weights of the regular surveys. This document presents the deviations.
B) The DEAS team, in collaboration with the FDZ, has considered whether it might be useful to slightly expand the default models to avoid systematic bias because of omitted-variable bias. These small adjustments are applied to the 2020/21 survey and to all subsequent surveys. They are documented here as well.
C) In the DEAS wave 2020 and in the subsequent wave 2020/21 it has been possible for the first time to include respondents that were 91-100 years old in the post-ratification. For this reason, the cross-sectional weighting factors from 2020 onwards are available in two versions: qsps_2020, qsps_21 as well as qsps-drop_20 and qspsdrop_21 cover the full age range of respondents up to 100 years of age. The weighting factors qsps_20_u91, qsps_21_u91, and qspsdrop_21_u91 include the same weights but do not include values for respondents 91 years of age or older.

The following remarks are based on the methodological report by Schiel et al. 2021.[1] Changes and deviations compared to the previous waves of the DEAS were, however, supplemented by additional tables and explanations by the FDZ of the DZA.

[1] 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/deutscher-alterssurvey/deas-dokumentation.

1 Cross-sectional weights –inclusion of respondents that are 91–100-year-old from the year 2020 onwards

Due to the aging population in Germany, the very old population group of 91–100-year-old is now also represented in official statistics in sufficient numbers that they can be used for the post-stratification of the DEAS data. While this was not possible for the previous waves, it has become possible as of the 2020 short survey. This has two consequences: 1) the cross-sectional weights qsps and qspsdrop cover the population under 91 years of age up to and including 2017. From 2020 onwards, qsps_20 and qsps_21 as well as qspsdrop_21 cover the population up to and including 100 years of age. For this reason, starting with the 2020 wave, the qsps cross-sectional weights exist in two versions.

Cross-sectional-weights for:

Wave	Respondents below 91 years:	Respondents below 101 years
2002	qsps_02, qspsdrop_02,	
2008	qsps_08, qspsdrop_08,	
2011	qsps_11, qspsdrop_11,	
2014	qsps_14, qspsdrop_14,	
2017	qsps_17, qspsdrop_17,	
2020	qsps_20_u91,	qsps_20
2021	qsps_21_u91, qspsdrop_21_u91	qsps_21, qspsdrop_21

Trend analyses not defining explicit age limits (e.g., on those up to 90 years old using if conditions) must pay attention to the correct choice of cross-sectional weights.

2 Longitudinal weights 2017-2021

2.1 Oral interviews

To calculate longitudinal weights, dropout models (logistic regression) are calculated to determine the sample members' probability of participation from wave 2017 to wave 2021. The population of the model is participants (panel cases) with valid interviews in the initial wave 2017 minus individuals known to have died between the two waves. The individual probability of participation in the current wave is calculated. Predictors of each model are the following characteristics included in the 2017 survey dataset edited by the DZA and provided by the DZA:

- Region: West (former federal territory and Berlin-West), East (new federal states and Berlin-East)
- County types, grouped: metropolitan counties, urban counties, rural counties
- Age, grouped by quartiles.
- Gender: male, female
- Educational attainment (ISCED), grouped: low, medium, high (with assignment of cases with missing information to the category "low")
- Network size, grouped: up to two persons, three to five persons, six and more persons
- Equivalent income grouped by tercile (with assignment of cases with missing data to the middle tercile).
- Subjective health status, grouped: very good/good, medium, poor/very poor (incl. missing)
- Drop-off status: no DO in initial wave/ DO completed in initial wave

The values of the respective variables in the 2017 wave are taken into account. Gender is constant, and information on education and region is derived from the first wave of the survey in each case.

For a multiplicative combination of the participation probabilities as a weighting factor, the reciprocal participation probability (1/participation probability) is required.

2.2 Drop-off

To calculate the longitudinal weights of the drop-off interviews, a failure model (logistic regression) is calculated to determine the sample members' probability of participation in the drop-off. The population of the model consists of panel cases with participation in the initial wave 2017 as well as participation in the interview in the observation wave 2021. Predictors of the models are:

- Region: West (former federal territory and Berlin-West), East (new federal states and Berlin-East)
- Gender: male, female
- Educational attainment (ISCED), grouped: low, medium, high (with assignment of cases with missing information to the category "low")
- Age, grouped by quartiles (i.e., split into roughly four equally populated age groups).
- Drop-off status: no DO in initial wave/ DO completed in initial wave

The region and education refer to the first survey wave of a case, gender is constant. Age is the age at the time of the observation wave. DO status refers to the initial wave (2017).

The individual probability of participation calculated in this way is multiplicatively linked to the calculated longitudinal weight as a reciprocal probability of participation (1/probability of participation). The resulting longitudinal weights for the drop-offs adjust the distributions for the panel cases participating in the drop-off to the

distributions in the previous wave (initial wave). They thus relate to the observation period from the 2017 baseline wave to the 2021 observation wave drop-off.

3 Special feature 2021: Longitudinal weighting 2020-2021

A brief survey took place in 2020. This survey was designed to quickly capture the impact of the Corona pandemic and was based on the gross DEAS sample. The likelihood that respondents participated in the regular DEAS survey in 2020/21 is very likely related to their (non)participation in the brief survey in summer 2020. Depending on the longitudinal study design chosen, longitudinal weights may therefore also be required from the brief survey to the current survey. For this reason, in addition to the longitudinal weighting for the period from 2017 to 2021, a weighting from the short survey in 2020 to the regular panel survey in 2021 is carried out. This is done analogously to the procedure outlined above, based on data from the 2020 short survey, and refers to the observation period from the 2020 baseline wave to the 2020/21 observation wave. However, two predictors used to calculate the 2017-2021 longitudinal weight are not included in the 2020 short survey. The short survey did not capture network size and it is also not possible to determine equivalized income. For this reason, two alternative measures are used: Household size and unweighted household net income per capita. The DO status variable is not included for the longitudinal weighting starting from the 2020 short survey. Participation in the short survey was already a de facto drop-off. For the drop-off weighting model short survey 2020 to CATI 2021, there were no deviations from the procedure described in point 1.2. The exception is the variable drop-off status, which does not refer to the initial wave, but to the last wave the respondents participated before the initial wave 2020.

Matching predictors of longitudinal weighting 2017-2021 and 2020-2021.

Predictors	Characteristics/Coding	2017 to CATI 2021	Short survey 2020 to CATI 2021
Region	West (former West Germany and Berlin-West), East (new Länder and Berlin-East), based on first interview	√	√
County types	grouped: Large urban counties, Urban counties, Rural counties.	√	√
Age	Grouped by quartiles (i.e., divided into approximately four equally populated age groups).	√	√
Gender	male, female	√	√

Education degree	grouped (ISCED): low, medium, high (with assignment of cases with missing information to the category "low")	√	√
Network size	grouped: Up to two people, three to five people, six and more people	√	*Not possible*
Household size	grouped: 1 person, 2 persons, 3 or more persons, missing/no specification	-	√
Equivalent income	grouped by terciles (with assignment of cases with missing information to the middle tercile	√	*Not possible*
Unweighted Household income per capita	grouped by terciles (with assignment of cases with missing information to middle tercile	-	√
subjective health	grouped: very good/good, medium, poor/very poor (incl. missing)	√	√
DO status	in output shaft no DO/ in output shaft with DO	√	

4 Cross-section weighting/ integration of samples into a common cross-section.

4.1 Oral interviews

To create cross-sectional weights, all participants in the current wave must be integrated into a common cross-sectional sample. These subsamples are in the current survey:

- Panel cases with participation in the previous wave 2017 (re-participants)
- Panel cases without participation in the 2017 pre-wave (temporary failures)

For the re-participating panel cases, the initial weight for integration is the longitudinal weight in the observation wave. For temporary dropouts, an auxiliary weight was created. The basis for this auxiliary weight was the cross-sectional weight in the wave in which the case last participated. This weight was adjusted using the results of two failure models: (1) The individual probability of non-participation in the follow-up wave after the last participation (1-participation probability) was calculated in the failure models of the longitudinal weighting. (2) Separate logistic regressions were calculated to determine the individual probability of participating in the wave under consideration, assuming not having participated in the previous wave(s) (probability of return).

The population of each model is all panel cases without participation in the initial wave (pre-wave of the observation wave). Predictors of each model are:

- Region: West (former federal territory and Berlin-West), East (new federal states and Berlin-East)
- Gender: male, female
- Age, grouped by quartiles
- Educational attainment ISCED grouped: low, medium, high (with allocation of cases with missing information to the category "low").

The region and education data refer to the first survey wave of a case, gender is constant. The age is the age at the time of the initial wave of the model.

To form the auxiliary weight for integration into the cross-section, the following elements are multiplicatively linked: cross-sectional weight of the wave in which the case last participated, reciprocal probability of non-participation in the follow-up wave after the last participation (1-participation probability from the failure models used to calculate longitudinal weights), reciprocal probability of return in the wave under consideration.

The starting weights of all subsamples are combined convexly. This procedure takes into account the fact that the individuals also had a selection probability in the respective other samples.

In addition, the distributions of year of birth, gender and region in the sample are adjusted to known distributions in the population (microcensus) by means of iterative proportional fitting (IPF). The basis for this is the reference table already provided by the DZA with the distribution figures from the Microcensus 2020[2] , differentiated by birth cohort groups, gender and region.

4.2 Drop-off

To determine the cross-sectional weights of the drop-off, drop-off models (logistic regression) are calculated based on all participants in the interview to determine the probability of participating in the drop-off interview in the same wave. Predictors of the model are:

- Region: West (former federal territory and Berlin-West), East (new federal states and Berlin-East)
- Gender: male, female
- Educational attainment ISCED grouped: low, medium, high (with allocation of cases with missing information to the category "low").
- Age, grouped by quartiles
- DO status: in the last wave no DO/ in the last wave with DO.

[2] The FDZ of the DZA would like to thank Peter Krause of the Federal Statistical Office for providing the reference table based on the German microcensus.

The region and education data refer to the time of the initial survey; gender is constant. Age is the age at the time of the observation wave. Drop-off status refers to the wave in which respondents last participated before the current survey. The 2020 short survey does not count because no additional drop-off was possible in this wave.

The cross-sectional weight for the drop-off was formed by multiplying the cross-sectional weight of the observation wave by the reciprocal probability of participation in the drop-off (1/probability of participation). In addition, an adjustment of the distributions year of birth*sex*region in the sample to known distributions in the population (microcensus) is carried out by means of Iterative Proportional Fitting (IPF). The basis for this is the reference table already provided by the DZA with the distribution figures from the Microcensus 2020, differentiated by birth cohort groups, gender and region.

5 Additions to the failure models of the weights

In principle, the failure models should not be changed on a large scale to ensure comparability between waves. However, a moderate extension to further reduce bias may be justifiable. The prerequisite for this was that additional predictor variables have a significant and, in particular, substantial effect on prediction of (re)participation probability. This was the case for the characteristic drop-off status, i.e., whether the person in the previous wave also completed the additional written questionnaire after the oral interview. Therefore, in almost all drop-off models, the DO status of the initial wave or the last participation wave was taken into account. Completing the drop-off indicates higher motivation or loyalty to the DEAS and increases the probability of participating again or completing a drop-off again.

ISCED education level had already been used as a predictor of longitudinal panel participation in the oral interview in previous waves. As an important predictor, it is now also used in the cross-sectional model to predict reparticipation in the oral interview (temporary dropouts).

Overview of the supplemented predictors in the different failure models.

	Weighting					
Predictors	Longitudinal 2017 to CATI 2020/21	Longitudinal 2020 short to CATI 2020/21	Cross-sectional CATI 2020/21	DO longitudinal 2017 to 2020/21	DO longitudinal 2020 short to 2020/21	Cross-sectional DO 2020/21
Region	√	√	√	√	√	√
County types	√	√				
Age	√	√	√	√	√	√
Gender	√	√	√	√	√	√
Education	√	√	√	√	√	√
Network size	√					
Household size		√				
Equivalent income	√					

Unweighted House-hold per capita income		√				
Subjective health	√	√				
DO status	√			√	√	√

Note: The additions to the default models were not applied retrospectively and currently affect only the survey year 2020/21. To examine whether the inclusion of DO status might lead to breaks or distortions, the FDZ of the DZA calculated trend analyses and examined whether results differ when the DO status is also controlled for beyond the weighting factors. However, the results of the analyses did not differ.

6 Poststratified cross-sectional weighting

The cross-sectional weights of the oral interviews and the drop-off were adjusted to the known distributions year of birth x sex x region according to the German microcensus (2020) by means of "iterative poportional fitting" (IPF). In addition, two further post-stratified cross-sectional weights were created that did not include participants with a birth year before 1930, as these birth cohorts have very low case numbers. This was done for both the CAPI-per-phone interview and the drop-off interviews. A minor adjustment was also made here: from the waves 1996 to 2020 DEAS short survey, Berlin was divided into East and West Berlin on the basis of postal codes, and each was separately assigned to the East German region and the West German region, respectively. For the poststratified cross-sectional weighting of the 2020/21 survey, Berlin was completely assigned to the region "East Germany".

For questions, problems or hints and ideas, please contact:

Stefan Stuth

stefan.stuth@dza.de