
AES Germany Quality Report

Annex 1: Methods
(Extracted parts of the German report)

29 November 2023



Contents

A1 Methodological Outline of Sampling, Realized Interviews, Editing, Classification and Weighting ..	3
A1.1 Sampling and Stratification	3
A1.2 Realized Interviews and Response Rate	4
A1.3 Validity of Completed Interviews	5
A1.4 Checking and Editing of AES data	5
A1.5 Classification and Coding	5
A1.6 Weighting and extrapolation.....	6
Literatur	8

A1 | Methodological Outline of Sampling, Realized Interviews, Editing, Classification and Weighting

A1.1 Sampling and Stratification

The universe of the German AES 2022 covers the German population aged 18-69 years, living in private households. In Germany, in addition to the main AES study, so-called supplementary state studies can be carried out in parallel. Federal states that take advantage of this option can use the services of the nationwide survey, i.e. the interviews in their own federal state, as part of the supplementary study and, if necessary, increase the sample size to obtain precise statements. The following three federal states made use of this option in 2022: North Rhine-Westphalia, Saxony and Baden-Wuerttemberg. Against this background, a subsample of the German AES data with the German population aged 25-64 years and without the addresses of the increase for the federal states was used in order to provide German AES data for Eurostat.

A sequential mixed-mode design was used: An online survey was carried out in the first step, the initial processing (CAWI: Computer Assisted Web Interview), and face-to-face interviews were carried out in the second step, the post-processing (CAPI: Computer Assisted Personal Interview). The CAPI post-processing improves data quality by balancing web survey selectivity and leads to higher response rates.

Sampling was realized using a multi-stage, multi-stratified sampling procedure. In a first step, individual communities were selected as a clustered sample. Within these communities, at the second stage, samples of persons were drawn for the population identified.

First Stage Selection

In a first step, the universe of all communities in Germany was stratified according to the following regional characteristics:

- Federal States (Bundesländer)
- Administrative Districts
- County
- Type of Community¹

¹ Type of community (BK-Regions)

0 = 500,000 and more (structure type 1)

1 = 500,000 and more (structure type 2, 3)

2 = 100,000 to less than 500,000 (1)

3 = 100,000 to less than 500,000 (2, 3)

4 = 50,000 to less than 100,000 (1)

5 = 50,000 to less than 100,000 (2, 3)

6 = 20,000 to less than 50,000

7 = 5,000 to less than 20,000

8 = 2,000 to less than 5,000

9 = less than 2,000

structure type 1 = core area

These characteristics were used in combination to identify equally large PSUs (Primary Sampling Units). In January 2022, a random sample of 400 PSUs in 368 communities was drawn. The selection total included all communities of the Federal Republic of Germany as of 30 June 2019. The sampling units of the survey are selected randomly proportionate to the number of private households per PSU (PPS design: probabilities proportional to size). So, the respective margins of the sample correspond to the distribution of private households in the universe.

The community selection was realized as a systematic sampling with a random start per shift. If a community was selected more than once, i.e. more than one PSU falls into this community, more addresses in this community were selected accordingly. Typically, 300 addresses were ordered per PSU to ensure that 89 people per PSU could be selected.

In addition to the Verian (then still Kantar Public)² cover letter, a BMBF cover letter, the so-called sampling rules and a study profile letter were sent to the sampled communities. In the sampling rules, taking into account the disproportionate approach, a structurally appropriate random sampling by age and sex was requested. In order to have the addresses available in time for the field start at the beginning of July 2022, the letters to the communities were sent in February 2022. All addresses received from the communities were formally examined and cleaned.

Second Stage Selection

The sample of persons for the AES 2022 was drawn from the cleaned address basis. According to the design, an identical number of addresses (n=89) per PSU had to be sampled. This was not done by simple random sampling. Instead, a stratification and adaptation of the sample to the structures of the population compensated for any deviations in the selected community sample. The distribution "Age group x sex per federal state" was stored as the structure matrix for the sampling. Two samples with different weights were taken from the total sample for the two sub-samples defined by design (18- to 24-year-olds and 25- to 69-year-olds), each PSU being as structurally similar as possible.

The address basis for the AES-2022 survey was as follows:

- Sample of 18- to 24-year-olds: 400 PSU with 22 addresses each = 8,800 addresses
- Sample of 25- to 69-year-olds: 400 PSU with 67 addresses each = 26,800 addresses

The entire address sample thus comprised 35,600 target person addresses, which were included in the fieldwork of the survey.

A1.2 Realized Interviews and Response Rate

The fieldwork of AES 2022 started on 12 July 2022. The field was terminated with the approval of the BMBF on March 20, 2023. Thus, fieldwork took around eight months, i.e. longer than the six months recommended by the EU-Manual. However, within the specified time frame of July 1, 2022 to March 31, 2023 (European Union 2021a, Article 8).

structure type 2 = agglomeration area

structure type 3 = transition area

² The Kantar Public Group is operating under the new company name Verian Group since November 2023.

In the fieldwork time, a total of 9,901 interviews were collected for AES 2022, including 6,883 interviews in CAWI and 3,004 interviews in CAPI mode. The agreed case numbers were thus approximately achieved in total and separately by mode.

For the initial survey in CAWI mode, the target value of the response rate of 19.9 percent (n=7,100) was minimally undershot at 19.3 percent (n=6,883). For the follow-up in CAPI mode, 7.6 percent (n=2,700) of the gross addresses were expected to be processed. This value was slightly exceeded at 8.4 percent (n=3,004). This compensates for the slightly lower number of cases from CAWI mode. The overshoots and undershoots are in the expected range.

A1.3 Validity of Completed Interviews

The next step was to clarify which interviews were considered to be valid and thus can be included in the net data record. The following tests were carried out to verify the validity of interviews:

- interview break-off,
- speeding,
- straight lining,
- item nonresponse analyses and
- comparison of the collected information on age, sex and nationality with those of the communities.

The total sample realized was reduced by a total of 79 interviews. Two further interviews had to be excluded due to subsequent deletion requests by the target persons. The valid total net data set therefore comprises 9,820 cases.

A1.4 Checking and Editing of AES data

The advantage of computer assisted surveys is the possibility to obviate some sources of error with data entry while interviewing. Filtering of questions occurs automatically, and proper values are defined. Additionally, plausibility checks are programmed, leading to questions if answers given by the respondent are inconsistent across the interview. These checks cause alarm messages and give the interviewer the possibility to ask for correct answers if necessary.

Due to computer assistance effort in editing data after fieldwork is reduced, though not all possible sources of error can be eliminated. Therefore, data checks, especially plausibility checks, have to be done ex post. Additionally, open answers have to be inspected (e.g.: "other, please name") and then be classified to values of closed-ended questions if possible.

A1.5 Classification and Coding

The AES contains information that must be classified according to the international classification systems ISCED-Level 2011, ISCED-Fields 2013, NACE rev. 2, ISCO 2008.

The information on respondent's occupation asked in an open question has to be classified according to the "International Standard Classification of Occupations" (ISCO 2008), the sectors of economic activity according to the „Nomenclature des Statistiques des Activités économiques de la Communauté Européenne" (NACE rev. 2). For both classifications Verian built up computer assisted dictionaries that allow a fast and high-quality process of classifying. But, as not all open answers can be classified automatically, the remaining answers must be classified manually.

The educational activities covered by the questionnaire are partly covered by ISCED level 2011 (UNESCO 2014) and consistently coded according to ISCED-Fields 2013 (UNESCO 2012; FED, NFE and INF activities). For ISCED-Fields coding, the coding process developed within the framework of AES

2007 and revised by Kantar Public or Verian, respectively, within the framework of AES 2016 based on ISCED-FIELDS 2013 was used. In the TFAES, it was also decided in November 2020 to provide an adapted ISCED-Fields 2013 coding for AES 2022, in which the so-called interdisciplinary codes are omitted. Instead of the interdisciplinary codes, those of the main area or, if identification is not possible, of the first topic should now be used. This proposal has already been tested in the ISCED fields coding under AES 2020 and has now again been implemented in AES 2022.

The ISCED level 2011 coding is not, unlike the other standard encodings, based on open nominations, but on closed answers to the educational background. The German Federal Office of Statistics (Statistisches Bundesamt) and Verian created a differentiated assessment of the ISCED level for AES 2016. This instrument takes into account about 450 individual variables. The implementation of the amount of information of a variable on the ISCED level is very labor-intensive due to the many plausibility work at the individual case level.

A1.6 Weighting and extrapolation

The weighting of the AES-2022 data is necessary to compensate for design-related disproportionalities and to compensate for any structural effects due to unsuccessful interviews.

Weighting at the person level

Weighting takes place in three stages. At the first stage, the disproportionality inherent in the design was balanced. On the second stage, the losses from the gross to net sample were corrected using a logit model based on the gross information. This compensates for the cases in which target persons did not participate in an interview (Rohwer 2011, Särndal et al. 1992). This step resulted in a sample effectiveness of 83.28 percent. At the third stage, the calibration was carried out using rim-weighting (Cochran 1968; Deming & Stephan 1940; Deville, Särndal & Sautory 1993; Kish 1965).

The optimum set of weighting characteristics was determined taking weighting effectiveness into account. A total of 15 edges were defined:

- federal state x bik-type
- bik-type
- occupational status
- vocational qualification
- ISCED level
- east/west x school education
- age groups x school education
- school education
- east/west x employment
- age groups x employment
- scope of employment - full-time
- federal state x sex
- federal state x age group
- east/west x sex x age group
- east/west x German/non-German

Extrapolation on population level

For several parameters, it is of interest not only to show percentage distributions or averages in the sample, but to extrapolate the results to the absolute figures in the population ("projection"). According to official population statistics, the population of 18 to 69-year-olds is 55.905 million people. The extrapolation factor is determined by multiplying all cases of the overall AES-2022 weighted sample by a constant factor that reproduces population numbers.

Literatur

Cochran, W.G. (1968): The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 24, 295-313.

Deming, W. E. & Stephan, F. F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*. 11. 427-444.

Deville, J.-C., Sarndal, C.-E. & Sautory, O. (1993): Generalized Raking Procedures in Survey Sampling, *Journal of the American Statistical Association*, Vol. 88, No. 423, pp.1013-1020.

Kish, L. (1965): *Survey sampling*. New York: John Wiley & Sons.

Rohwer, G. (2011): *Uses of Probabilistic Models of Unit Nonresponse*. NEPS Working Paper No. 5.

Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer Verlag.