GLEN Data Manual

Release 1.0

2026

Philipp Schütze, Tessa Schick, Lisa Leßner & Anna Joraschek

GLEN Documentation

Cite as: Schütze, P., Schick, T., Leßner, L. & Joraschek, A. (2026) GLEN Data Manual. Release 1.0. GLEN Documentation.
https://doi.org/10.11588/heidok.00038807

Table of contents

List of abbreviations

BMI	Body Mass Index
CASMIN	Comparative Analysis of Social Mobility in Industrial Nations Scale
CAWI	Computer Assisted Web Interview
FSE	Factorial Survey Experiment
GLEN	German Longitudinal Environmental Study
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
KldB 2010	Klassifikation der Berufe 2010
PAPI	Paper And Pencil Interview
SUF	Scientific Use File(s)

1 Introduction

This Data Manual provides a comprehensive overview of the data structure and preparation of the German Longitudinal Environmental Study (GLEN). Key decisions and procedures for data editing are included, as well as explanations of the conventions used for variable naming and other technical details necessary for analyses with GLEN data.

1.1 Data access

The Scientific Use Files (SUF) are available through the GESIS Data Archive in Stata and SPSS format.[1] Access is granted upon completion of a user agreement with GESIS. After receiving access, data can be retrieved via the GESIS Data Archive. The documentation materials included in GLEN Release 1.0 are described in Table 1, together with information on the language(s) in which they are available.

Table 1: Overview of documentation materials in GLEN Release 1.0

Document	Content
Design and Cumulative Field Report (en) (Schmiedeberg et al., 2026)	Detailed information on sample design, recruitment procedures, survey modes, fieldwork implementation, and response rates
Quick Guide (en, de)	Brief overview of the data and its structure; Introduction on combining survey datasets
Variable Overview (en, de)	Excel file containing all survey-variables and labels in English and German
Screenshots (CAWI) (de)	Documentation of the presentation of all questions across different devices (desktop or smartphone)
Questionnaires (PAPI) (de)	PAPI questionnaires (for mixed-mode surveys)
Methodenberichte (de)	Documentation of the fieldwork process provided by the survey institute Verian
Technical Reports (en)	Reports on specific survey design components (e.g., survey experiments, weighting procedures)

In addition, codebooks for each survey are included in GLEN Release 1.0 and can be downloaded from the GLEN website. For future analyses, a secure data center in Kaiserslautern will provide researchers with the opportunity to link their own georeferenced data with the GLEN survey data (see Section 2.5 for more information).

[1] All data preparation was conducted using Stata 19.0.

1.2 Data collection and sample

The GLEN study was launched in 2024 with a large-scale, probability-based sample of 80,000 individuals. The target population comprises all adults (born before September 1, 2006) with a permanent primary residence in Germany, regardless of nationality or legal status. The panel began with a two-step process consisting of the Recruitment Survey, which established the panel sample, and the Starter Survey designed to increase online participation. GLEN Release 1.0 also includes the first two panel waves (Wave 1a and Wave 1b), as well as the Inter-Wave Survey 2025a. While the panel waves feature annually repeated core modules, the inter-wave surveys primarily focus on varying current topics and survey experiments.

GLEN is designed as a mixed-mode panel study, combining self-administered Computer-Assisted Web Interviews (CAWI) and self-administered Paper-and-Pencil Interviews (PAPI). The web-based questionnaires were programmed and implemented using SoSci Survey (Leiner, 2026). The PAPI questionnaires were distributed by the survey institute Verian, who also managed the return process, digitized the completed questionnaires, and handled the majority of respondent inquiries.

Given that PAPI questionnaires were returned via mail after completion, all PAPI cases are treated as complete by design; item nonresponse may occur, but unit nonresponse in the form of survey break-offs is not possible in this mode (see Section 2.2).

For more detailed information on sample design, recruitment procedures, survey modes, field work implementation, and response rates, please refer to the Design and Cumulative Field Report (Schmiedeberg et al., 2026).

1.3 Panel participation management

In general, respondents may temporarily withdraw their participation in the panel for any period of time (e.g., due to illness or temporary residence abroad). After such periods, respondents are invited to resume participation in subsequent surveys. Panelists may request the deletion of their survey data and/or contact data at any time. If only contact data are deleted upon request, the respondent will no longer receive survey invitations; however, their survey data will remain included in the SUF. Respondent data are fully excluded from the release data files if the respondent requests complete deletion of all data. As a consequence, survey data may be removed retroactively, including after GLEN Release 1.0.

For detailed information on invitations and reminders, please refer to the Design and Cumulative Field Report (Schmiedeberg et al., 2026) as well as Section 2.3 of this manual covering gross datasets.

2 Data structure

The GLEN data are organized into three main dataset types, all of which are available in Stata and SPSS format: survey datasets, vignette datasets (where applicable), and gross datasets. All data included in the SUF are anonymized. Further details on anonymization are provided in Section 4.8. The survey datasets (see Section 2.1) contain participants' substantive responses and paradata variables (see Section 2.1.3) and form the core analytical files of the study. Vignette datasets (see Section 2.4) are provided for surveys that include factorial survey experiments and contain the corresponding vignette-level information. Gross datasets (see Section 2.3) document the fieldwork process and include information on invitation mode, reminders, and participation. Table 2 provides an overview of all datasets included in Release 1.0.

Table 2: Overview of GLEN data structure

Survey datasets	Vignette datasets	Gross datasets
01_recruitment_2024		01_recruitment_2024_gross
02_starter_2025	02_starter_2025_vig	02_starter_2025_gross
03_wave1a_2025		03_wave1a_2025_gross
04_interwave2025a	04_interwave2025a_vig	04_interwave2025a_gross
05_wave1b_2025		05_wave1b_2025_gross

2.1 Survey datasets

The survey datasets are provided separately for each survey. The central unit of analysis in the SUF is the individual respondent, identified with a unique identification number `id`. Using this variable, the datasets can be merged or appended across surveys. A brief introduction to merging datasets (e.g., in long or wide format) is provided in the GLEN Quick Guide. Table 3 summarizes the available survey datasets along with their respective field times and interview modes. Fieldwork followed a mixed-mode design, with CAWI serving as the primary mode. PAPI Questionnaires were administered as a follow-up option for respondents who were unwilling or unable to participate online. Consequently, field periods differ between modes in the Recruitment Survey following initial contact. Both interview modes were implemented only in the Recruitment Survey and in Wave 1a. In these cases, CAWI and PAPI data were integrated into a single dataset. All remaining surveys were conducted exclusively in CAWI mode. The survey mode can be identified using the variable `intmode`.

Table 3: List of survey datasets and field times

Dataset	Mode	Field time	
		Start	Finish
01_recruitment_2024	CAWI	November 2024	January 2025
	PAPI	November 2024	March 2025
02_starter_2025	CAWI	January 2025	April 2025
03_wave1a_2025	CAWI	May 2025	July 2025
	PAPI	May 2025	July 2025
04_interwave2025a	CAWI	September 2025	October 2025
05_wave1b_2025	CAWI	November 2025	January 2026

2.1.1 Central variables

The core variables listed in Table 4 are included in all survey datasets.

Table 4: List of central variables in the survey datasets

Variable name	Explanation
id	Respondent identification number
sample	Sample indicator 1 = Original sample (2024)
survey	Survey number in chronological order 1 = Recruitment Survey (2024) 2 = Starter Survey (2025) 3 = Wave 1a (2025) 4 = Inter-Wave Survey 2025a 5 = Wave 1b (2025)
surveytype	Survey type 1 = Wave 2 = Inter-Wave Survey 3 = Recruitment / Starter
intmode	Interview mode 1 = CAWI 2 = PAPI
inty, -m, -d	Date of completion: y = year, m = month, d = day

Each respondent is assigned a unique identification number (id) that remains constant across all waves to facilitate longitudinal data linkage. In addition, each dataset

includes the variable `survey` to identify the specific survey. As respondents can participate only once per survey, the combination of `id` and `survey` serves as a unique identifier (primary key) for each observation. Both variables are consistently available across the survey, vignette, and gross datasets, ensuring seamless combination (merging or appending) across all data components.

2.1.2 Variable name conventions

Variable names in the GLEN survey datasets follow a systematic naming convention, combining conceptual abbreviations (variable stem) with sequential numbers. For example, in the five variables `val1i1`–`val1i5`,

- `val` represents the concept or topic (e.g., "values").
- The `1` in `val1` indicates that all variables belong to the same item battery (here the first item battery on values).
- The suffix numbers `i1`–`i5` distinguish the individual variables within that battery.

An overview of the survey variable stems in GLEN Release 1.0 and their explanations is presented in Table 5. Note that variable names remain constant/stable over all surveys and do not include dataset-specific identifiers. Although this consistency simplifies the appending of data (long format), merging datasets (wide format) requires renaming variables to ensure uniqueness, i.e., unique variable names in all merging datafiles to avoid overwriting of variables/data. For a brief introduction to combining datasets in Stata or R, please refer to the GLEN Quick Guide.

Table 5: List of variable stems

Stem	Topic
`area-`	Spatial or geographical information (e.g., residential area, region, or urban/rural classification)
`att-`	Attitudes (respondents' general attitudes, or opinions)
`consent-`	Informed consent / explicit agreement regarding data privacy, panel participation, and contact
`csm-`	Consumption patterns or consumer behavior (e.g., purchasing habits, goods, or services)
`dob-`	Date of birth
`econ-`	Economic situation, financial well-being, or perception of the economy

Continued on next page...

Continued from previous page

Stem	Topic
eng-	Civic or political engagement (e.g., protests, activism, memberships, volunteering)
env-	Environmental behavior and attitudes (pro-environmental actions, habits, or ecological footprint)
ewe-	Extreme weather experiences (exposure to or impacts of severe weather events and related attitudes)
exp-	Expectations (respondents' future outlook, predictions, or anticipated developments)
hlt-	Health status (physical or mental health, well-being, or health-related behaviors)
home-	Housing and living conditions (e.g., type of residence, household composition, or amenities)
incent-	Incentive choice
job-	Employment status, profession, or labor/employment related characteristics
kid-	Information regarding children in the household or in the family (e.g., number of children, presence of minors, or childcare details)
med-	Media usage (consumption of news, social media, TV, or information sources)
mob-	Mobility behavior (transportation choices, commuting habits, or vehicle usage)
newad-	New address
nut-	Nutrition and dietary habits (e.g., eating patterns, food choices)
pol-	Political orientation, political interest, party affiliation, or voting behavior
psy-	Psychological dispositions and climate-related mental states (e.g., perceived self-efficacy, trust)
rel-	Relationship status and partnership characteristics (e.g., marital status, cohabitation, or duration of relationship)
sat-	Satisfaction (e.g., life satisfaction, satisfaction with democracy)
sde-	Sociodemographic characteristics of the respondent (e.g., education level, professional choices)
sex-	Respondent sex
val-	Personal values, core beliefs, or socio-cultural value orientations
vig-	Vignette-based survey experiments (i.e., using descriptions of longer situational text or image scenarios)

Continued on next page...

Continued from previous page

Stem	Topic
web-	Internet usage, online behavior, or web-related specific questions

The survey question variables represent the actual data collected from respondents. The Variable Overview (Excel file) serves as a consolidated summary across all surveys, listing the variable labels in both German and English and indicating which questions were included in which specific survey. The original question wording (and value labels) as well as filters and information on experimental splits are documented in the survey-specific codebooks, which are available in separate German and English versions for each survey.

2.1.3 Paradata variables

Paradata refer to auxiliary data generated as a by-product of the online survey data collection process (i.e., only for CAWI mode) and provide information about how respondents interact with the online questionnaire rather than about their substantive responses. The paradata include timestamps for the start and end of the interview, indicators of questionnaire progress such as the last page accessed and the maximum page reached, as well as page-specific and total interview durations. See Table 6 for an overview.

Page-specific processing times are stored in the variables `sys_timeXXX` where `XXX` indicates the screen number (page number) of the online survey. Most screens contained only one question or item battery. Data users can identify the exact items (variable names) shown on page `XXX` by checking the variable label for `sys_timeXXX`. For more details on paradata in SoSci, see the SoSci Manual online.

In addition to these technical variables, paradata include weighting variables that can be used for analytical adjustments. Please refer to Section 4.9 and Leßner et al. (2026) for detailed information on weighting.

Table 6: List of paradata variables (only CAWI)

Variable	Explanation
sys_started	Start of interview in the format date and time
sys_lastdata	End of interview: time when participant most recently updated the data case

Continued on next page...

Continued from previous page

Variable	Explanation
sys_lastpage	Page that was edited most recently by the respondent; Each pagenumber corresponds to the specific pagetime available in sys_timeXXX
sys_maxpage	Last page of the questionnaire that was accessed by the respondent
sys_timeXXX	Time in seconds spent on questionnaire page XXX; The variable labels for sys_timeXXX list all variables displayed on page XXX
sys_timesum	Total time spent on the questionnaire (excluding outliers)
sys_timersi	Relative speed index, a metric quantifying an individual participant's completion velocity against the sample median; Values exceeding 1.0 denote faster-than-average respondents, while values below 1.0 indicate slower-than-average respondents
sys_missrel	Percentage of missing answers weighted by general participant behavior; Skipping frequently answered items increases the penalty, whereas rarely answered questions have minimal impact
sys_device	Device used to complete the survey: computer, tablet, or smartphone
cweight1	Calibrated weight, CAWI and PAPI, normalized, trimmed
cweight2	Calibrated weight, CAWI only, normalized, trimmed

2.1.4 Generated variables

For the convenience of GLEN data users, several variables are generated that may be useful for a range of research questions. Table 7 lists all generated variables included in Release 1.0. These variables are added to subsequent surveys unless the underlying variables were updated in that survey. In such cases, the variable label includes a note indicating the corresponding survey of data collection. The syntax for generating these variables is included in the SUF, and users are invited to adapt it to their needs. Users are strongly advised to confirm that the generation procedures meet their specific requirements before using any of the generated variables. The general missing code for all generated variables is -7 *Incomplete data* and is used if one or more of the variables that comprise the generated variable are missing.

Table 7: Generated variables Release 1.0

Variable	Content	Surveyed in
sex_gen	Sex, generated (register data imputed with recruitment data from respondents)	01_recruitment_2024, 02_starter_2025
doby_gen	Year of birth, generated (register data imputed with recruitment data from respondents)	01_recruitment_2024, 02_starter_2025, 03_wave1a_2025
migback	Migration background	01_recruitment_2024
isced	ISCED-classification in 8 categories	01_recruitment_2024, 05_wave1b_2025 *
isced_cat	ISCED-classification in 3 categories	01_recruitment_2024, 05_wave1b_2025 *
casmin	CASMIN-classification in 9 categories	01_recruitment_2024, 05_wave1b_2025 *
casmin_cat	CASMIN-classification in 3 categories	01_recruitment_2024, 05_wave1b_2025 *
bmi_gen	Body Mass Index	03_wave1a_2025
kldb2010	KldB occupation classification	03_wave1a_2025
kldb2010rel	KldB occupation classification respondent's partner	05_wave1b_2025
isco	ISCO-classification	03_wave1a_2025
iscorel	ISCO-classification respondent's partner	05_wave1b_2025

* Updated for respondents under 30 with a newly obtained educational degree in 05_wave1b_2025

Recategorization of open answers

For variables measuring education, the category *other* with an option for open entry is available for variables sde1o and sde2i7o. These entries are manually reviewed and, where possible, recoded into the appropriate existing categories. This also applies to the variables sde1, sde2iX, and sde3iX, which capture respondents' educational qualifications. To maintain data provenance and preserve the original responses, new (reclassification) variables were created using the _gen suffix (e.g., sde1_gen, sde2iX_gen), leaving the original variables unaltered.

Reclassification is only carried out when the information provided allows for a clear assignment, without requiring additional assumptions. For example, a respondent is only assigned to a specific vocational degree if the qualification is explicitly stated or if the reported occupation clearly implies a specific qualification.

In the Recruitment Survey, entries from the variable sde1o are coded into the school degree variable sde1_gen, as well as into the vocational degree variables sde2iX_gen.

Similarly, entries from sde2i7o are coded into the additional vocational degree variables sde2iX_gen and the university degree variables sde3iX_gen.
In Wave 1b, respondents aged 30 or younger were asked whether they had obtained any additional vocational qualifications. Open-ended responses to this question are also classified into the existing categories. As a result, information on respondents' educational and vocational qualifications is updated in the dataset of Wave 1b, and the CASMIN and ISCED classifications are revised accordingly.
As with all other variables that allow for open entries, the variables are anonymized as explained in Section 4.8.

Sex

In addition to the variable sex collected in the surveys, two further variables are generated. Despite sex being measured in three categories in the GLEN surveys, including a *diverse* option, it was anonymized to include only two categories. More information regarding anonymization can be found in Section 4.8.
The variable sex_gen uses register data and represents the respondent's sex with missing values imputed from the variable sex of the Recruitment Survey. This variable is added to all subsequent survey datasets. If sex is collected again in a later survey (e.g., in the Starter Survey), an additional variable, sex_flag, is generated. This variable indicates discrepancies between the newly reported sex and sex_gen. More information on flag variables can be found in Section 2.1.5.

Year of birth

To complement the year of birth collected in the variable doby in the Recruitment Survey, two additional variables are generated. The variable doby_gen contains the respondent's year of birth based on register data, supplemented with information from doby in the Recruitment Survey if register data is missing. This variable is included in all subsequent survey datasets. In later surveys, if the year of birth is collected again (variable doby), a corresponding variable doby_flag is generated to capture any deviations from doby_gen. More information on flag variables can be found in Section 2.1.5.

Migration background

The variable migback indicates whether the respondent has a migration background, based on the country of birth of the respondent and their parents. It is coded as follows: 0 *No migration background*, meaning the respondent and both parents were born in Germany; 1 *First-generation*, meaning the respondent was not born in Germany; and 2 *Second-generation*, meaning that at least one parent was not born in Germany.

It should be noted that the official definition of migration background is based on nationality as defined by the Federal Statistical Office of Germany (DESTATIS, 2026b). However, as GLEN does not collect data on the nationality of the respondents' parents, the generated variable `migback` is based on country of birth. This corresponds to the concept of "immigration history" as also defined by the Federal Statistical Office of Germany (DESTATIS, 2026a).

ISCED classification

The variable `isced` contains the International Standard Classification of Education (ISCED) 2011 in eight categories. The classification follows ISCED-A, based on the documentation provided by Eurostat, the statistical office of the European Union (EUROSTAT, 2026).

The ISCED category is derived from the respondent's school qualifications as well as any additional vocational qualifications. For details on the coding procedure, please refer to the Stata syntax file `isced2011.do`. The resulting values of this variable are shown in Table 8.

Table 8: `isced` values and labels

Value	Label
-7	Incomplete data
0	Currently enrolled
1	Primary education
2	Lower secondary education
3	Upper secondary education
4	Post-secondary non tertiary education
5	Short-cycle tertiary education
6	Bachelor's or equivalent
7	Master's or equivalent
8	Doctorate or equivalent

Value 5 *Short-cycle tertiary education* does not appear in the GLEN dataset, as information on this type of education was not collected.

In Wave 1b, respondents up to the age of 30 were asked whether they had obtained an additional educational qualification. If additional qualifications were reported, they were used to update the respondent's ISCED classification.

Additionally, a categorized variable `isced_cat` is generated, classifying the ISCED codes into *low*, *medium* and *high* level of education.

CASMIN classification

The Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) scale is represented in the variable `casmin`, which classifies educational attainment into nine categories. The classification is derived from the respondent's school qualifications as well as any additional vocational qualifications. For details on the coding procedure, please refer to the Stata syntax file `casmin.do`. The resulting values of this variable are shown in Table 9.

In addition, the CASMIN classification is provided in a reduced form in variable `casmin_cat`, grouping educational attainment into primary, secondary, and tertiary levels. In Wave 1b, respondents aged 30 and under were asked whether they had obtained an additional educational qualification. If additional qualifications were reported, they were used to update the CASMIN classification.

Table 9: `casmin` values and labels

Value	Label
-7	Incomplete data
0	Currently enrolled
1	Inadequately completed (1a)
2	General elementary education (1b)
3	Basic vocational qualification (1c)
4	Intermediate vocational qualification (2a)
5	Intermediate general qualification (2b)
6	General maturity certificate (2c_gen)
7	Vocational maturity certificate (2c_voc)
8	Lower tertiary education (3a)
9	Higher tertiary education (3b)

Body Mass Index

The variable `bmi_gen` contains the respondent's Body Mass Index (BMI), calculated from the height and weight reported in `hlt8` and `hlt9`. The BMI is defined as body weight in kilograms divided by the square of body height in meters.

To ensure anonymity, values of *18* or lower are recoded to *18*, and values of *40* or higher are recoded to *40*.

Variables based on Klassifikation der Berufe 2010

As a measure of occupation, the variable `kldb2010` is provided, which contains the

Klassifikation der Berufe 2010 (KldB 2010) of the Federal Employment Agency.[2] The open-ended responses in `sde14` and `sde15` were coded into KldB categories by Verian. The variable `kldb2010rel` is constructed in the same way, based on the open-ended responses in `rel7` and `rel8`, and captures the occupation classification of the respondent's partner. Further information on the coding process can be found in the Methodenbericht Welle 1a (Brix & Steinacker, 2026). As occupation is considered sensitive data, the SUF includes only an anonymized version. Details on the anonymization procedure are given in Section 4.8. Access to the full, non-anonymized data, as well as additional variables, is provided through the Safe Data Center (see Section 2.5).

Variables based on ISCO08 classification

The variable `isco08` contains a classification of occupations created by Verian, based on the International Standard Classification of Occupations (ISCO). This classification was published in 2009 by the International Labor Office (ILO, 2023). To generate the variable, the open entries in variables `sde14` and `sde15` are manually categorized.
The variable `isco08rel` is generated based on the open entries in `rel7` and `rel8` in the same way to reflect the occupation classification of the respondent's partner.
To learn more about the classification process, please refer to the Methodenbericht Welle 1a (Brix & Steinacker, 2026). Similar to `kldb2010`, the SUF includes only an anonymized version due to data sensitivity. Details on the anonymization procedure are given in Section 4.8. Access to the full, non-anonymized data, as well as additional variables, is provided through the Safe Data Center (see Section 2.5).

2.1.5 Flag variables

Flag variables are provided as an additional service to help users identify potential data quality issues. Rather than deleting information, inconsistent entries are retained and documented through corresponding flag variables whenever possible. These variables begin with the name of the variable(s) to which they refer and have the suffix `_flag`. They are generated only when the original value remains unchanged. If a value is replaced, the corresponding variable is coded as missing with -5 *Invalid entry*. See Section 4.1 for more information on the usage of this missing value.
Flag variables are created in the following cases:

- implausible values (see Section 4.2),
- inconsistent responses within a survey (see Section 4.3 for more information on consistency checks),

[2] English: German Classification of Occupations 2010

- inconsistencies in response options between PAPI and CAWI modes,
- inconsistencies between generated variables and survey variables (see Section 2.1.4 for more information on generated variables), and
- imputations (see Section 4.7 for information on specific data editing decisions).

Because only selected inconsistencies are flagged, users are strongly recommended to carefully review all relevant variables before conducting analyses. Table 10 provides an overview of all flag variables.

Table 10: Flag variables

Flag Variable	Dataset(s)	Description
`doby_flag`	`All`	Year of birth in `doby_gen` differs from year of birth reported in the current survey (`doby`)
`sex_flag`	`All`	The value of `sex_gen` differs from the sex reported in the current survey (`sex`)
`sde2_flag`	`01_recruitment_2024`	Inconsistency PAPI/CAWI: Unlike in PAPI, selecting `sde2i1` (no vocational training yet) together with one or more additional vocational qualifications was not possible by design in CAWI mode; Therefore, combinations of `sde2i1` with `sde2i2-sde2i7` were flagged in PAPI
`sde3_flag`	`01_recruitment_2024`	If respondents reported a higher academic degree without reporting the prerequisite degree (combinations of `sde3iX`), e.g., a Master's degree without a Bachelor's degree
`hlt7_flag`, `hlt8_flag`, `hlt9_flag`, `home16o1_flag`, `home16o2_flag`, `mob11i1_flag`, `mob11i2_flag`, `mob11i4_flag`	`03_wave1a_2025`	Implausible value in the flagged variable (see Section 4.2 for detailed information on value checks)

Continued on next page...

Continued from previous page

Flag Variable	Dataset(s)	Description
bmi_gen_flag	03_wave1a_2025	Implausible combination of height and weight leading to implausible value in bmi_gen
home7_flag	03_wave1a_2025	Contradictory entry: More people in household under 14 (home7) than people in household (home2)
mob6_flag	03_wave1a_2025	Inconsistency PAPI/CAWI: CAWI allowed responses for 3 cars and 3 motorcycles, while PAPI only allowed responses for 3 vehicles in total
mob6i1_flag, mob6i2_flag	03_wave1a_2025	Values of mob6i1_gen and mob6i2_gen were imputed (see Section 4.7.3 for specific data editing decisons on mob6)
mob10_flag	03_wave1a_2025	Inconsistency: Respondent indicated having taken private flights (mob10 = 1), but reported zero flights for all flight distance categories (mob11i1–mob11i4)
sde15_flag	03_wave1a_2025	Inconsistency PAPI/CAWI: The response option 0 *I have never been employed* was available in CAWI mode, but not in PAPI mode. Cases in which 0 was selected and cases with intmode = 2 *PAPI* were flagged
hlt12i1o_flag, hlt12i2o_flag, hlt12i3o_flag, job10_flag, kid2n1_flag, kid2n2_flag, kid2n3_flag	05_wave1b_2025	Implausible value in the flagged variable (see Section 4.2 for detailed information on value checks)
home2_kid4_flag	05_wave1b_2025	Contradictory entry: Living alone (home2) but also living with a child or children (kid4)
home2_rel3_flag	05_wave1b_2025	Contradictory entry: Living alone (home2) but also living with partner (rel3)

Continued on next page...

Continued from previous page

Flag Variable	Dataset(s)	Description
sde11_sde17_flag	05_wave1b_2025	Inconsistency: More than one degree obtained in the last 6 or 12 months, where one of them is a prerequisite for the other; Note that this could be possible, e.g., through the recognition of a previously earned degree
sde13i1_job10_flag	05_wave1b_2025	Inconsistency: Full-time employment, but weekly working hours below 35

2.1.6 Labels in German and English

Although all GLEN surveys were administered in German, the Stata survey datasets feature bilingual (de and en) variable and value labels, supported by codebooks in both languages. Additionally, the release package includes separate German and English SPSS datasets for each survey. This allows international researchers to reference the original German formulations alongside the English translations to ensure nuances are not lost.

Stata users can toggle between languages using the commands `label language de` for German and `label language en` for English. Note that the Stata datasets are delivered with English as the active language, ensuring English labels are accessible via standard import functions. To access the German labels outside of Stata or SPSS, users can either import the German SPSS files directly or use R's `readstata13` package. This R-based workflow is also an applicable workaround for Python users.

Please note that other datasets, such as the gross data or vignette datasets, contain English labels only.

2.2 Inclusion and exclusion of cases in survey datasets

Unlike the gross sample (invitation sample, see Section 2.3), the survey datasets include only those who participated. Furthermore, individuals may be excluded from the survey dataset for several additional reasons. The definition of the remaining cases primarily depends on identity checks, item response, refusals during field work, and verification of belonging to GLEN's target population. The exact definition per survey is documented in the following section.

2.2.1 Recruitment Survey (2024)

The final sample in the survey dataset is derived from the invitation sample by excluding (subtracting) respondents assigned to steps (1) to (4), as shown in Table 11. Each step is described in detail below.

Table 11: Cases in survey dataset, Recruitment Survey (2024)

	N_0	**of invitation sample**
(1)	– N_1	of non-participants
(2)	– N_2	of participants with `dispcode_rec`=*break-off*
(3)	– N_3	of participants with other self-reported drop-out reason
(4)	– N_4	of participants who failed identity check
	= N	**of survey dataset**

- **(1)**: Individuals who did not participate in the survey are not part of the survey datasets.
- **(2)**: In the Recruitment Survey, an interview was counted as a *break-off* if a person participated online (CAWI), but left the survey before reaching the panel consent question `consent2`. Respondents who took part in PAPI mode were not excluded from the survey dataset (regardless of whether or not they reached the consent question). See Section 2.3 for the definition of `dispcode_rec`.
- **(3)**: Participants have the possibility of informing the GLEN project team about changes of their place of residence, refusals, etc. via various channels (i.e., hotline, email, or via the questionnaire itself). Whenever individuals wish their survey data to be removed from the panel, inform the GLEN team that another person is participating on their behalf, or no longer fit into the target population (e.g., because they live abroad or in a nursing home), they are removed from the dataset. Some cases could not be deleted immediately due to internal processes and are therefore still included in the survey dataset. These cases are anonymized by assigning the value -11 *Not available* to all variables.
- **(4)**: To ensure that only the individuals that were randomly selected participate in the survey, an identity check was conducted by comparing information provided by respondents in the survey with official register data. Respondents were classified as non-target individuals if their reported sex, nationality, and age (differing by at least two years) all deviated from the register data simultaneously, and were consequently excluded from future surveys. This is important to note,

as in these cases the recorded panel consent likely came from a non-target individual.

2.2.2 Subsequent surveys after Recruitment Survey (2024)

From the Starter Survey onward, the procedure differs in two respects (see Table 12). First, identity checks based on register data were no longer carried out. Second, the definition of *break-off* in step **(2)** changed: *break-off* is now determined by the proportion of essential questions answered (item response). Participants who responded to less than 50% of essential questions are not included in the survey datasets. However, such cases remain in the gross dataset and these individuals remain panel members, meaning they will be invited to participate in future surveys.[3] Regardless of item response, participants who completed a survey in PAPI mode and returned it to the research agency are not excluded from the survey dataset (in the GLEN Release 1.0, this applies to Wave 1a). While it is assumed that the likelihood of obtaining valid responses is low when fewer than 50% of the items have been answered online, incomplete paper-based questionnaires with response rates below 50% are considered valid, given the additional effort involved in returning them by post.

Table 12: Cases in survey datasets after the Recruitment Survey

		N_0	**of invitation sample**
(1)	–	N_1	of non-participants
(2)	–	N_2	of participants with `dispcode`=*break-off*
(3)	–	N_3	of participants with other self-reported drop-out reason
	=	**N**	**of survey dataset**

2.3 Gross datasets

For each survey, a specific gross dataset is available (`*_gross.dta`) that encompasses the full invitation sample for the respective survey, including all individuals eligible for the survey regardless of their final response status.

[3] For respondents who did not complete the questionnaire, it is not possible to determine which filter questions they would have been routed to. The following approach is therefore applied when calculating their item response rate: First, the last question presented to the respondent prior to break off is identified. Next, the minimum number of questions that non-break off respondents were required to answer from that point onward is determined. This minimum value is then used as the basis for calculating the item response rate of the respondent who did not complete the survey (in addition to the questions that the respondent had already been presented with).

These datasets provide a record of the survey-specific contact process and field work outcomes. For a list of available variables in the gross datasets, see Table 13.

Key components

- **Contact history:** Variables on the mode and timing of the initial invitation, up to three reminders, and information on the total number of contact attempts.
- **Recruitment & background data:** Register-based information (year of birth and sex) and panel consent status for the recruitment sample.
- **Invitation experiments:** Documentation of invitation-related experiments (see variable `inviteexpX`) that cannot be inferred from the questionnaire itself. Note that questionnaire splits remain located within the survey dataset. See Section 3 and the survey-specific codebooks for more details.
- **Field work disposition:** Each case is assigned a *disposition code* summarizing the completion degree (i.e., complete, partial, break-off, or nonresponse).

Note: The definitions of disposition codes vary between the Recruitment Survey and subsequent surveys, as well as by interview mode.

Gross dataset variables

Table 13 lists the variables available in the gross datasets.

Table 13: List of contact and field work variables in each gross dataset

Variable	Explanation
`id`	Respondent identification number (available only if participated in the Recruitment Survey)
`survey`	Survey indicator (for details see Table 4)
`invitemode`	Mode of the initial invitation (1 = Email, 2 = Postal)

`rem1mode`
`rem2mode`
`rem3mode`
	Mode of 1st reminder (1 = Email, 2 = Postal)
	Mode of 2nd reminder (1 = Email, 2 = Postal)
	Mode of 3rd reminder (1 = Email, 2 = Postal)

`invitedate`	Date of the initial invitation

`rem1date`
`rem2date`
`rem3date`
	Date of 1st reminder
	Date of 2nd reminder
	Date of 3rd reminder

`fieldstart` `fieldend`	Date of field start (CAWI) Date of field end (CAWI)
`doby_reg`	Year of birth from register data

Continued on next page...

Continued from previous page

Variable	Explanation
sex_reg	Sex from register data
inviteexp1	Indicator for invitation experiment 1: "illustrated character" in Starter Survey 2025; See Section 3 for further details
inviteexp2	Indicator for invitation experiment 2: "10% exclusion" in Inter-Wave Survey 2025a; See Section 3 for further details
dispcode_rec	Disposition status for Recruitment Survey 1 Complete: item response [80%, 100%] (PAPI) 1 Complete: item response [80%, 100%] & seen consent2 (CAWI) 2 Partial: item response [0%, 80%[(PAPI) 2 Partial: item response [0%, 80%[& seen consent2 (CAWI) 3 Break-off: did not reach consent2 (only CAWI) 4 No interview
dispcode	Disposition status since Starter Survey 2025 1 Complete: item response [80%, 100%] (CAWI & PAPI) 2 Partial: item response CAWI [50%, 80%[; PAPI [0%, 80%[3 Break-off: item response [0%, 50%[(CAWI) 4 No interview
suf	Case included in SUF of current survey (1 = yes, 0 = no)

2.4 Vignette datasets

The vignette datasets (*_vig.dta) provide information on (multi-)factorial survey experiments. The central variables are: 1) the vignette identification number vigXid, which indicates the unique vignette number drawn for each respondent, 2) experimental treatments are captured by the variables vigXdimY, and 3) the outcome variables vigXqZ. Note that the outcome variables are included in both the survey datasets and, for the sake of completeness, the vignette datasets. While Section 3 includes a short summary of the vignette experiments, comprehensive documentation is available in the GLEN Technical Report No. 1 (Bozoyan et al., 2026) and No. 3 (Wolter et al., 2026). Please note that labels in the vignette datasets are provided exclusively in English. To access the original German wording and gain a more thorough understanding of the design, researchers should refer to the aforementioned technical reports.

2.5 Geospatial and regional information

The GLEN survey data can be linked to geospatial information to incorporate objective indicators of the respondents' living environment and surrounding infrastructure. This may involve spatial data at various aggregation levels. For example, information

at the municipal level, such as population density, can be linked to the survey data using municipal codes. Furthermore, geocodes of the respondents' place of residence (longitude and latitude geocodes) can be used to perform more fine-grained analyses (e.g., green space around the place of residence). It is also possible for guest researchers to merge their own georeferenced data with the survey data.
However, the use of spatial data requires appropriate data protection measures. For this reason, access to spatial data is restricted to guest workstations in the secure data center in Kaiserslautern. More detailed information on access to the secure data center is published on the GLEN website.

2.5.1 Available regional data

A limited number of regional variables was linked to the GLEN survey data in GLEN Release 1.0 (available in the survey datasets). These include:

- **Federal state** (`state`): Federal state in which the respondent lives. Due to data protection reasons, Rhineland-Palatinate & Saarland as well as Lower Saxony & Bremen were combined.
- **RegioStaR4** (`regiostar4`): Differentiated regional statistical regional type in which the respondent lives (metropolitan urban region, regiopolitan urban region, rural region close to an urban region, or peripheral rural region) (BMVI, 2018). This assignment is done at the municipal level.

2.5.2 Updating places of residence

In order to assign correct regional and geodata to the respondent data, panel participants are asked to notify the GLEN project team of any changes of address. They can enter a new address directly at the end of the questionnaires or contact the survey institute by phone or email. Furthermore, an address search is carried out if postal invitations could not be delivered.
The survey institute Verian provides updated addresses at the beginning of the panel waves only (shortly before the invitations for the next panel waves are sent out), not at the beginning of the inter-wave surveys. The addresses are stored as anonymized geocodes and are separated from the survey datasets. The geocodes used to add regional information (`state`, `regiostar4`) to a given survey dataset are based on the updated geocodes provided for the subsequent panel wave. For example, the regional information for Wave 1a and Inter-Wave Survey 2025a correspond to the updated geocodes delivered at the beginning of Wave 1b.

In rare cases for which no updated geocodes are available (e.g., because a person has withdrawn from the panel between Wave 1a and Wave 1b), the information from the previous panel wave (here: Wave 1a) is used.
Note: The variable `newad1` only reflects cases in which respondents reported a move at the end of the questionnaire. It does not account for moves that were communicated via email, hotline, or address search, and is therefore not sufficient to fully capture residential moves.

3 Experiments

Several types of experiments are implemented within GLEN:

- methodological design experiments *conducted prior* to the survey,
- methodological design experiments *embedded within* surveys,
- split experiments regarding content, and
- factorial survey experiments.

All experiments are fully available to users for analyses. Substantive experiments are included only in the Starter Survey and the inter-wave surveys; the Recruitment Survey and the panel waves do not contain experiments. All experiments concerning questionnaire content are described in the respective codebooks. Table 14 lists the experiments conducted for each survey and identifies the outcome variables affected by the experimental treatments in GLEN Release 1.0.

Table 14: Experiments overview of GLEN Release 1.0

Experiment	Datasets	Outcomes	Treatments
Illustrated character split	`02_starter_2025_gross`		`inviteexp1`
Symbolic scale split	`02_starter_2025`	`area1i1 - area1i4`	`area1_split1`
Environmental policies split	`02_starter_2025`	`pol6i1 - pol6i6`	`pol6i1_split1 - pol6i6_split1`
Sugar policy split	`02_starter_2025`	`pol7`	`pol7_split1; pol7_split2`
Transport policies vignettes	`02_starter_2025_vig`	`vig1q1`	

Continued on next page...

Continued from previous page

Experiment	Datasets	Outcomes	Treatments
Survey invitation split	04_interwave2025a_gross		inviteexp2
Content split 1 (climate scepticism)	04_interwave2025a		content_split1
Content split 2 (carbon pricing)	04_interwave2025a		content_split2
CO_2 pricing policies vignettes	04_interwave2025a_vig	vig2q1 - vig2q3	

3.1 Starter Survey (2025)

Illustrated character split: An invitation experiment was conducted, in which the invitation letter included an illustrated character in a sample of 2,000 randomly selected individuals from the subgroup invited by postal mail (i.e., online panelists who had not provided their email address in the Recruitment Survey or respondents participating on paper). In the dataset 02_starter_2025_gross, the variable inviteexp1 identifies the control (0) and experimental (1) groups.

Rating stars split: In a second methods experiment, a random subsample of 1,000 respondents were invited to answer the question on residential area description using a non-verbalized symbolic scale (rating stars) instead of a verbalized scale. This is indicated by the variable area1_split1 in the dataset 02_starter_2025. To see how the stars are represented, please see the Screenshots.

Environmental policies split: In the first split experiment, the item wording of six policy interventions was varied in a question on environmental policies (pol6). The variables indicating the split are pol6i1_split1 through pol6i6_split1 in the dataset 02_starter_2025. Randomization was conducted independently for each item.

Sugar policy split: The second split experiment was included in a question on policy interventions to reduce sugar consumption in soft drinks (pol7). While the first and fourth response options of this variable remained constant, the second and third were experimentally varied. The split for response option two is recorded in the variable pol7_split1, and the split for response option three in pol7_split2 in the dataset 02_starter_2025.

Transport policies vignettes: Finally, a vignette experiment was included, in which transport policy interventions related to electric vehicles were varied with respect to their implementation timeframe and objective. Further information on the vignette study is provided in the GLEN Technical Report No. 1 (Bozoyan et al., 2026). The variable `vig1q1`, which contains the evaluation of the policy measure, is included in the dataset `02_starter_2025`. The dataset `02_starter_2025_vig` contains the combinations of the vignette dimensions `vig1dim1` to `vig1dim3`.

3.2 Inter-Wave Survey 2025a

Survey invitation split: In a methodological experiment, a random 10% of the eligible panelists were not invited to the survey. In the dataset `04_interwave2025a_gross`, the variable `inviteexp2` distinguishes the control group (0) from the experimental group (1).

Additionally, two experiments were included in which different content modules were tested against each other.

Content split 1 (climate skepticism): In the first content split, a random subsample of 25% of participants (`content_split1 = 1`) were assigned questions on climate change skepticism (`att4`), while the remaining 75% (`content_split1 = 2`) received questions concerning their online behavior (`web3`, `web4`, and `web5`). The variable `content_split1` is available in the dataset `04_interwave2025a`.

Content split 2 (carbon pricing): In the second content split, a random 90% of participants (`content_split2 = 1`) received a module on CO_2 pricing including a factorial survey experiment (`vig2q1, vig2q2, vig2q3, pol14-pol17`), while the remaining 10% (`content_split2 = 2`) were asked questions on cultural participation and control beliefs (`lsr3-lsr6, psy7`). The variable `content_split2` is also included in the dataset `04_interwave2025a`.

CO_2 pricing policies vignettes: Details of the factorial survey experiment are described in the Technical Report No. 3 (Wolter et al., 2026). The vignette dimensions and information treatments are documented in the dataset `04_interwave2025a_vig`. All vignette evaluations are also included in the dataset `04_interwave2025a`.

4 Data editing

There are two main sources for the raw data that undergo editing. CAWI data, collected via SoSci Survey, are edited directly by the GLEN team. In contrast, PAPI questionnaire data undergo initial processing and editing by Verian before being finalized by the GLEN team.
The following sections outline the basic data processing steps, followed by a more detailed description of specific editing procedures and the logic behind key data decisions.

4.1 Missing values

Table 15 provides an overview of the missing value codes applied across the datasets. All missing values are represented by negative integers; notably, Stata's system missing values (e.g., `.`, `.a`, `.b`, etc.) were not used.[4]

Notes on the value -3 *Does not apply*:
As noted in Table 15, the missing value -3 denotes that a question was filtered for a respondent (*Does not apply*). In specific instances, however, *Does not apply* serves as a substantive response category and must not be confused with the standard missing value code -3. For example, in Wave 1b, the items `rel12iX` (partnership labor division) include *Does not apply* as a regular response option with the value 50. In contrast, respondents without a cohabiting partner who were automatically filtered out were assigned the technical missing code -3 *Does not apply* (filtered out).

Table 15: Missing codes

Code	Meaning	Notes
-1	*Don't know*	Where stipulated by the questionnaire design, this represents the residual category (comprising responses such as *Don't know* or *No opinion*)
-2	*No answer*	Assigned if a respondent has seen the question page but provided no information

Continued on next page...

[4]The only exception is the `id` variable in the gross data of Recruitment Survey, where a system missing value was assigned to individuals who did not participate.

Continued from previous page

Code	Meaning	Notes
-3	*Does not apply*	Indicates the question was filtered out; Skip patterns are listed as entry filters below the variable description in the respective codebook
-4	*Filter error / Incorrect entry*	Denotes filter errors; For further details, please refer to Section 4.4
-5	*Invalid entry*	Indicates false entries in PAPI mode, e.g., misinterpreting open entry fields as checkboxes
-6	*Unreadable answer*	Mostly for PAPI responses
-7	*Incomplete data*	Non-generatable: This code is assigned to generated variables when one or more component variables required for the calculation contain missing information
-8	*Inadmissible multiple responses*	The respondent ticked multiple boxes for a single-choice item in the paper questionnaire (PAPI only)
-9	*Only in CAWI*	Variable is not captured in PAPI mode
-10	*Only in PAPI*	Variable is not captured in CAWI mode
-11	*Not available*	Assigned when data has been redacted for anonymization or removed due to data deletion requests
-33	*Break-off*	Assigned if a respondent did not answer the question due to interview dropout (i.e., ended interview at this or previous CAWI page)

4.2 Value checks

Value checks are conducted to ensure that all entries fall within the expected range for each variable and are consistent with the questionnaire. In addition, the plausibility of variables with open-ended text or numerical entries is manually assessed.

For variables with numerical open entries, a plausibility check is performed. Based on a subjective evaluation, flags are assigned to values that appear highly implausible. These flags are intended to guide data users who may wish to exclude such values

from their analyses, as they are likely the result of incorrect respondent input. The flag variables are dummy coded, with 0 indicating *Plausible value* and 1 indicating *Implausible value*. For example, if a respondent's height is entered in an open field, values below 55 cm or above 250 cm are flagged as implausible. This procedure is applied to all relevant variables. Plausible value ranges are determined through brief inspection, and any values outside these ranges are flagged accordingly. It is left up to the data user to exclude these cases based on the flag provided. Further information on flag variables can be found in Section 2.1.5.

4.3 Consistency checks

A range of consistency checks are conducted to identify logically impossible or empirically implausible combinations of values across two variables or across PAPI and CAWI mode. In general, the originally reported values are not edited. Instead, a corresponding flag variable is generated to indicate the detected inconsistency. For these flag variables, value 0 indicates that no inconsistency was identified, while value 1 indicates that an inconsistency is present. See Section 2.1.5 for more information and a full overview of all flag variables.

Consistency within a survey
Inconsistent and/or illogical responses provided by a respondent within the same questionnaire are flagged. For example, if a respondent reports an advanced degree without the required prerequisite qualification (combinations of `sde3iX`), such as a doctorate without a Bachelor's degree, this combination is flagged (`sde3_flag`).

Consistency across survey modes
Differences between PAPI and CAWI questionnaires are flagged when they may lead to inconsistencies and are not explicitly documented in the respective codebook. For example, unlike in PAPI, selecting `sde2i1` (no vocational training yet) together with one or more additional vocational qualifications was not possible by design in CAWI mode. Therefore, combinations of `sde2i1` with `sde2i2-sde2i7` were flagged (`sde2_flag`).

4.4 Filter checks

In the codebooks, the full filter condition for each variable is documented in both descriptive text and formal logic. In CAWI, it defines the display logic determining whether a question is shown to the respondent, whereas in PAPI, it represents the intended routing path through the questionnaire.
Deviations from the filter conditions can occur in two ways depending on survey mode.

In CAWI mode, filter inconsistencies arise when respondents use the back button to change earlier answers, causing subsequent responses that no longer match the updated filter logic to remain in the data. In PAPI mode, filter inconsistencies occur when respondents overlook or misapply routing instructions on paper, leading them to answer questions that should have been skipped or skip those that should have been answered.

For *CAWI data*, data cleaning checks for such inconsistencies and filter-inconsistent values are recoded to -4 *Filter error / incorrect entry*. For *PAPI data*, filter inconsistencies remain unresolved in the dataset. As researchers cannot verify which participant response is factual post-hoc, no automated corrections or manual imputations were applied. This preserves data integrity by avoiding subjective assumptions, leaving the final decision on how to handle these cases (e.g., through exclusion or recoding) to the scientific end-user.

4.5 Variable consolidation

Data editing involves refining variables to ensure they are presented in a user-friendly format. In some questions, if several elements are combined, data are stored by the survey software in auxiliary variables. In the data preparation process, these variables are recoded to easy-to-use variables that retain their essential information. A primary example is the measurement of the number of children in a household. In CAWI mode, this is captured using two different components (see Figure 1 for graphical representation during CAWI):

- `kid1_aux`: Radio button (1 = *children present*, 2 = *no children*).
- `kid1_dd_aux`: Dropdown menu used to specify the exact number of children (if `kid1_aux` = 1).

Figure 1: Visualization of auxiliary variables for `kid1` in SoSci Survey (Wave 1b)

In the SUF, these are fused into a single generated variable, `kid1`, which indicates the child count directly. The logic is as follows:

- `kid1=0` if `kid1_aux=2`
- `kid1=kid1_dd_aux` if `kid1_aux=1` & `kid1_dd_aux!=missing`
- If a respondent selects 1 (*children present*) on the radio button but fails to select a value from the drop- down menu, the exact count remains unknown. In the SUF, this results in a value of -2 *No answer* for the fused variable `kid1`, even though the presence of (an unknown number of) children was indicated.

Consequently, users may observe discrepancies between the fused variable and subsequent survey routing. For example, the variable `kid2n1` (number of children under 14) is filtered based on both the presence of children (`kid1_aux = 1`) OR the number of children (`kid1_dd_aux > 0`). If a respondent clicked the radio button indicating the presence of children but left the dropdown blank, the question for `kid2n1` was still posed and will appear in the dataset, even though the parent variable `kid1` is coded as -2 *No answer*. In such cases, the filter logic in the survey instrument diverges from the simplified filter notation (e.g., `kid1 > 0`) typically found in the respective codebooks.

Note: The auxiliary variables are not part of the SUF.

4.6 PAPI editing

The initial stage of PAPI data editing was conducted by Verian. During this process, unreadable answers—including instances where open-entry fields were misused as checkboxes or radio buttons—were coded as -5 *Invalid entry*. Additionally, any inadmissible multiple responses were assigned a value of -8.
In a subsequent step, the GLEN team harmonized the PAPI data to align with the final CAWI logic. For variable sets involving multiple responses, the following rule applies: any contradictory or inadmissible response triggers a value of -8 for all variables within that specific set to ensure internal consistency.

4.7 Specific data editing decisions

4.7.1 `sde4*` - Citizenship(s)

The processing of citizenship entries followed a ”first-entry” priority rule to ensure data consistency across modes:
PAPI: In cases where respondents provided multiple entries for citizenship, only the first valid entry was retained. CAWI: While the survey instrument allows for up to two distinct citizenships, some respondents entered the same citizenship twice. In these

instances of redundant information, the duplicate was removed and only the first entry was used.

4.7.2 `home12*` - Heating system at home

The question concerning heating system types at home does not include a "None of the above" option, meaning that respondents could select any possible combination of the 16 checkboxes. The following rules were applied sequentially to the `home12*` variables, in the order listed below:

1. If only *Don't know* is selected in a category (main/additional heating), all 7 remaining variables in that category are set to -1 *Don't know*.
2. If nothing is selected in both categories, all 16 variables are set to -2 *No answer*.
3. If nothing is selected for the main heating system, but at least one entry is given for an additional heating system, the main heating system is set to -2 *No answer*.
4. If at least one checkbox was selected for the main heating system, but none for the additional heating system, the additional heating system is not set to -2 *No answer*, but to 0 *Not mentioned*.

4.7.3 `mob6*` - Number of motor vehicles (PAPI)

In PAPI mode, respondents occasionally marked an open entry field as if it were a checkbox (see Figure 2). These cases were coded as -5. For the variables `mob6i1` and `mob6i2` (representing the number of cars and motorcycles), the number of overall vehicles was subsequently imputed using responses to the follow-up questions on vehicle type for up to three vehicles (`mob7n1`, `mob7n2`, `mob7n3`). The imputed variables are `mob6i1_gen` and `mob6i2_gen`.
The indicators `mob6i1_flag` and `mob6i2_flag` serve to document data adjustments by flagging values that have been imputed. It should be noted that the true number of vehicles cannot always be recovered, as `mob7n*` records information for at most three vehicles.

4.7.4 `mob10` - Private flights in last 12 months (PAPI)

In PAPI mode, the *Yes* value for variable `mob10` was derived from flight entries recorded in the follow-up questions (`mob11iX`).

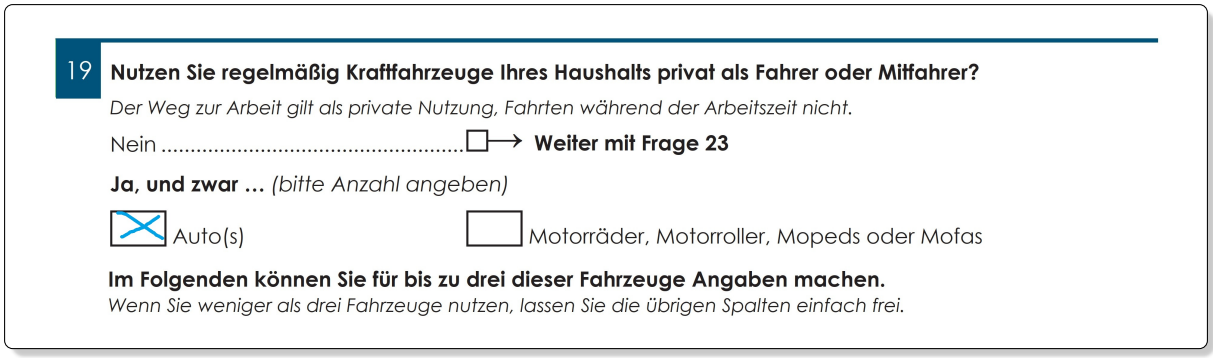

Figure 2: Illustrative fictional example of an open-ended field (`mob6i1`) misinterpreted as a checkbox in the PAPI questionnaire (Wave 1a).

4.7.5 `mob11iX` - Number of private flights

In Wave 1a for the question on the number of flights taken within the last 12 months (`mob11i1` – `mob11i4`), some respondents did not answer the `mob11iX` items in the standard way, as illustrated in Figure 3.

Figure 3: Illustrative example of inconsistent response pattern for `mob11iX` in CAWI mode (Wave 1a).

The standard response format for each of the four flight types is either

- selecting the upper bullet point and entering a number of flights, or

- selecting the lower bullet point *Gar nicht* (English: *Not at all*), which indicates zero flights.

Deviations from this format were handled as follows:

- If no bullet point was selected but at least one other flight type contained a number, the response was set to the value 0, for zero flights.
- If the bullet point for flights was selected but no number was provided, the response was coded as -2 *No answer*.

4.8 Anonymization

Table 16 provides an overview of the anonymization measures applied to variables included in the SUF. Anonymization follows a graded approach based on the level of identification risk associated with different types of variables, including direct identifiers, special categories of personal data, demographic characteristics, regional information, paradata, and open-ended responses. For each variable group, Table 16 documents the procedure used.

Table 16: Anonymization procedures by variable

Variable	Dataset(s)	Procedure
Direct identifiers		
`consent4`, `consent6`	`01_recruitment_2024`, `02_starter_2025`, `03_wave1a_2025`, `05_wave1b_2025`	Binarized (1 *Mentioned*).
`dobm`	`01_recruitment_2024`, `02_starter_2025`, `03_wave1a_2025`	Binarized (1 *Mentioned*)
`doby`	`01_recruitment_2024`, `02_starter_2025`, `03_wave1a_2025`	Bottom-coded ($\leq$ *1934* becomes *1934*)
`doby_gen`	`All`	Bottom-coded ($\leq$ *1934* becomes *1934*)
Sensitive characteristics		
`sex`	`01_recruitment_2024`, `02_starter_2025`	Category 3 *Diverse* randomly assigned to other categories (N in Recruitment Survey = 115)

Continued on next page...

Continued from previous page

Variable	Dataset(s)	Procedure
sex_gen	All	Category 3 *Diverse* randomly assigned to other categories
sde4o31	01_recruitment_2024	Rare nationalities are aggregated; in cases of aggregation, replaced by sde4o32 if available
sde4o32	01_recruitment_2024	Binarized (1 *Mentioned*)
sde4o2	01_recruitment_2024	Rare nationalities are aggregated to broader regional categories
sde6	01_recruitment_2024	Rare countries of origin are aggregated to broader regional categories
bmi_gen	03_wave1a_2025	Top-coded ($\geq$ *40* becomes *40*) and bottom-coded ($\leq$ *18* becomes *18*)
eng1i2, eng1i2o	03_wave1a_2025	Recoded to -11 *Not available*; including subsequent adjustment for eng1i2o
hlt8, hlt9, bmi_gen	03_wave1a_2025	Binarized (1 *Mentioned*); BMI calculated
rel4	05_wave1b_2025	Category 3 *Diverse* of partner's sex randomly assigned to other categories (N in Wave 1b = 25)
eng2i8	05_wave1b_2025	Recoded to -11 *Not available*
hlt10i1 - hlt10i9	05_wave1b_2025	hlt10i1 and hlt10i7 are not anonymized; hlt10_gen indicator for at least one other disease mentioned (0 *No*, 1 *Yes*)
Demographic characteristics		
isco08	03_wave1a_2025, 04_interwave2025a, 05_wave1b_2025	Reduced from 4 to 2 digits; rare categories are anonymized
kldb2010	03_wave1a_2025, 04_interwave2025a, 05_wave1b_2025	Reduced from 4 to 2 digits; rare categories are anonymized
isco08rel	05_wave1b_2025	Reduced from 4 to 2 digits; rare categories are anonymized
kldb2010rel	05_wave1b_2025	Reduced from 4 to 2 digits; rare categories are anonymized

Continued on next page...

Continued from previous page

Variable	Dataset(s)	Procedure
Geographic information		
`state`	`All`	*Rhineland-Palatinate* and *Saarland*, as well as *Bremen* and *Lower Saxony*, are each grouped into a single category
Open-ended responses		
`Open-ended responses`	`All`	All open-ended answers binarized (1 *Mentioned*)

Direct identifiers are removed or reduced. Special categories of personal data and demographic characteristics are only retained in detailed form if this characteristic is sufficiently common in the population; otherwise, they are recoded or aggregated into broader categories. Regional information is released only in aggregated form, and only selected paradata relevant for analyses are included in the SUF. These decisions are informed by population-level frequency checks based on Microcensus reference data, where available, and by additional empirical checks within the GLEN data, including combinations of potentially identifying characteristics where necessary.
Open-ended responses are excluded from the SUF due to high disclosure risk. Instead, binary indicators are provided to reflect whether a response was given. For further information on the possibility of using open-ended responses or other anonymized variables, please contact the GLEN team directly.

4.9 Weighting

Due to noncontact, selective participation, and panel attrition, the GLEN sample does not perfectly represent the target population (i.e., all adults with a permanent primary residence in Germany). To better approximate the composition of the target population, GLEN provides adjustment weights for each of the five surveys included in GLEN Release 1.0. Using raking, the marginal distributions of key sociodemographic characteristics in the survey data are calibrated to the corresponding distributions from official reference data (Microcensus 2024 and population update data 2024). Weights are available both for CAWI and PAPI participants combined (applies to the Recruitment Survey and Wave 1a, variable `cweight1`) and for CAWI participants only (variable `cweight2`). A detailed description of both the construction and the characteristics of the weights can be found in Leßner et al. (2026).

References

BMVI. (2018). *RegioStaR: Regional Statistical Spatial Typology for Mobility and Transport Research*. Federal Ministry of Transport and Digital Infrastructure. Retrieved March 5, 2026, from https://www.bmv.de/SharedDocs/DE/Anlage/G/regiostar-raumtypologie-englisch.pdf?__blob=publicationFile

Bozoyan, C., Schmiedeberg, C., & Best, H. (2026). *Acceptance of transport policies: A factorial survey experiment in the German Longitudinal Environmental Study (GLEN)*. GLEN Technical Report No. 1. https://doi.org/10.31235/osf.io/dyz7c_v3

Brix, J., & Steinacker, G. (2026). *Methodenbericht GLEN – Umweltpanel. Welle 1a*.

DESTATIS. (2026a). *Hintergrundpapier für die Veröffentlichung von Ergebnissen zur Bevölkerung nach Einwanderungsgeschichte*. Statistisches Bundesamt. Retrieved May 21, 2026, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Methoden/Erlauterungen/einwanderungsgeschichte-hintergrundpapier.html

DESTATIS. (2026b). *Migrationshintergrund*. Statistisches Bundesamt. Retrieved May 21, 2026, from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Glossar/migrationshintergrund.html

EUROSTAT. (2026). *International Standard Classification of Education (ISCED)*. Eurostat statistical office. Retrieved May 21, 2026, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)

ILO. (2023). *The International Standard Classification of Occupations (ISCO-08) companion guide*. International Labour Organization.

Leiner, D. J. (2026). SoSci Survey. https://www.soscisurvey.de

Leßner, L., Bozoyan, C., & Schmiedeberg, C. (2026). *Weights for the German Longitudinal Environmental Study (GLEN)*. GLEN Technical Report No. 2. https://doi.org/10.31235/osf.io/r5zbm_v1

Schmiedeberg, C., Königer, L., & Emmer, J. (2026). *GLEN Design and Cumulative Field Report*. Release 1.0. GLEN Documentation.

Wolter, F., Auspurg, K., & Diekmann, A. (2026). *Public Support for Different CO_2 Pricing Policies: A Factorial Survey Experiment in the German Longitudinal Environmental Study (GLEN)*. GLEN Technical Report No. 3. https://doi.org/10.31235/osf.io/utyzj_v1