

# **DeZIM.panel**

## **Data Release 4.0.0**

### **Method & Data Report**

Version June 10th 2024



## Publisher

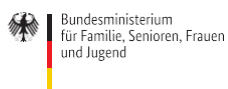
**Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM) e.V.**

Forschungsdatenzentrum DeZIM.fdz | Mauerstraße 76 | 10117 Berlin

+49 (0)30 804 928 93 | [fdz@dezim-institut.de](mailto:fdz@dezim-institut.de)

## Citation

Lietz, Almuth; Schmälzle, Michaela; Zimmermann, Stefan; Köhler, Jonas; Siegel, Madeleine; Li, Ida; Chouaibi, Doreen; Dollmann, Jörg; Jacobsen, Jannes (2024). DeZIM.panel Data Release 4.0.0. Method & Data Report, Version: 10th June 2024. Berlin: Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM).



## Inhalt

1	Introduction .....	1
2	Overview .....	2
3	Survey instrument .....	3
4	Field phase .....	5
4.1	Field phases .....	5
4.2	Survey Duration .....	6
5	Population and response .....	6
5.1	Population .....	6
5.2	Return bw1 .....	8
5.3	Return bs1 .....	9
5.4	Return bw2 .....	10
5.5	Return bw3 .....	11
5.6	Return bw4 .....	12
5.7	Return cw1 .....	13
5.8	Return cw2 .....	14
5.9	Return cw3 .....	15
5.10	Return cw4 .....	16
5.11	Weighting .....	17
6	Data curation and anonymization .....	19
6.1	Editions of scientific use files (SUF) and data access .....	19
6.2	Data structure .....	20
6.3	Survey wave naming scheme .....	20
6.4	Variable naming scheme .....	21
6.5	Open coded variables .....	22
6.6	Aggregated and ex-post coded variables ("_g") .....	22
6.7	Versionized ("_v") and harmonized ("_h") variables .....	22

6.8	Example: Versionizing, harmonization, aggregation.....	23
6.9	Missing conventions.....	24
	Literature.....	27

## List of tables

Table 1: Project overview .....	2
Table 2: Core modules across waves and survey period.....	5
Table 3: Survey duration by wave (in minutes).....	6
Table 4: Return bw1 after delivery.....	8
Table 5: Return bw1 according to overrepresented groups .....	8
Table 6: Return bs1 after delivery .....	9
Table 7: Return bs1 according to overrepresented groups.....	9
Table 8: Return bw2 after delivery.....	10
Table 9: Return bw2 according to overrepresented groups .....	10
Table 10: Return bw3 after delivery.....	11
Table 11: Return bw3 according to overrepresented groups .....	11
Table 12: Return bw4 after delivery.....	12
Table 13: Return bw4 according to overrepresented groups .....	12
Table 14: Return cw1 after delivery .....	13
Table 15: Return cw1 according to overrepresented groups.....	13
Table 16: Return cw2 after delivery .....	14
Table 17: Return cw2 according to overrepresented groups.....	14
Table 18: Return cw3 after delivery .....	15
Table 19: Return cw3 according to overrepresented groups.....	15
Table 20: Return cw4 after delivery .....	16
Table 21: Return cw4 according to overrepresented groups.....	16
Table 22: Comparison of unweighted cell percentages by gender, education, and age group.....	18
Table 23: Missing codes.....	25

## 1 Introduction

To better reflect the post-migrant society in Germany, the DeZIM.panel was set up in 2021, which allows a representative, longitudinal survey of the population (Dollmann et al. 2022). The DeZIM.panel is a survey infrastructure recruited by letter that can be used to determine current moods and trends among people living in Germany who were born between 1941 and 2004. For this purpose, people without a migration history as well as people with a migration history are interviewed. The latter groups are overrepresented in the DeZIM.panel, especially people from Turkey, from other Muslim-majority countries, from states with which West Germany concluded recruitment agreements between 1955 and 1973, from the former USSR and Romania as well as the descendants of these people. As a result, the DeZIM.panel provides enough cases for comparisons and analyses within these groups – making it a unique infrastructure in the German survey landscape. As a multi-topic online access panel, the DeZIM.panel aims to track the impact of external events as well as short- and long-term changes in public opinion, with a particular focus on immigrants and their descendants. In this way, developments can be recorded, and data can be provided that is of particular interest to researchers in the fields of integration, migration, and racism. Integration indicators, experiences of discrimination, xenophobic and racist attitudes as well as national and ethnic identities are recorded as part of the respective surveys.

The DeZIM.panel data is prepared, anonymized, and documented within the framework of the DeZIM.fdz, the research data center of the DeZIM Institute. This data is made available as Scientific Use Files (SUF) for the scientific community to facilitate scientific data reuse. The SUF are offered via various access methods: via download, via remote access and on-site on the premises of the DeZIM Institute. Access to DeZIM data is regulated by law. The prerequisites for using a SUF are a scientific research purpose, employment at a scientific institution and the conclusion of a data usage contract. In addition to the data sets, documentation materials for the data sets are provided.

In the following chapter, the central data and working modalities of the studies are summarized in tabular form. The survey instruments used are discussed in chapter three, the field phases in chapter four and the population, response, and the weighting concept in chapter five. A description of the individual steps of data preparation and anonymization can be found in chapter six.

## 2 Overview

For orientation, the report is preceded by a brief overview of the most important data and working modalities of the study. Please note that the current data release contains data from the recruitment wave<sup>1</sup> conducted in 2021 until wave cw4 (8<sup>th</sup> regular wave). All of these waves are regular surveys with one exception: A special survey was conducted in 2022. This short survey (wave abbreviation: bs1) was intended to measure reactions to the war in Ukraine shortly after the events.

**Table 1: Project overview**

Title	DeZIM.panel - Wave 1 – Wave 8
Responsible institution	DeZIM Institute
Survey institute	DeZIM Institute
Funding	Federal Ministry from Family Affairs, Senior Citizens, Women and Youth (BMFSFJ)
Population	All persons living in Germany, with or without a migration history, born between 1941 and 2004, with an oversampling of people from Turkey and from other majority Muslim countries, from states with guest worker agreements and re-settlers from the East.
Survey method	Online survey via Tivian EFS
Survey period	bw1 <sup>2</sup> : 02.12.2021 – 31.01.2022 bs1: 28.02.2022 – 06.03.2022 bw2: 31.03.2022 – 22.05.2022 bw3: 27.06.2022 – 07.08.2022 bw4: 26.09.2022 – 06.11.2022 cw1: 12.12.2022 – 31.01.2023 cw2: 28.03.2023 – 14.05.2023 cw3: 19.06.2023 – 31.07.2023 cw4: 18.09.2023 – 30.10.2023
Gross sample ("Einsatzstichprobe"):	$n_{\text{wave bw1}} = 6,719$
Panel-ready individuals <sup>3</sup>	$n_{\text{wave bs1}} = 4,855$ $n_{\text{wave bw2}} = 6,655$ $n_{\text{wave bw3}} = 6,557$ $n_{\text{wave bw4}} = 6,523$ $n_{\text{wave cw1}} = 6,500$ $n_{\text{wave cw2}} = 6,439$ $n_{\text{wave cw3}} = 6,368$ $n_{\text{wave cw4}} = 6,293$

<sup>1</sup> For information on the recruitment wave, please refer to the method report provided by Ruland/Sandbrink (2022) (available in German only).

<sup>2</sup> Please note: We use wave abbreviations to refer to the waves. See section 6.4 for an explanation of the wave abbreviations.

<sup>3</sup> The gross and net samples differ in part from the previous releases due to a revised data processing procedure.

Realized interviews	$n_{\text{wave bw1, finished}} = 3,686$ ; $n_{\text{wave bw1, interrupted}} = 99$ $n_{\text{wave bs1, finished}} = 2,626$ ; $n_{\text{wave bs1, interrupted}} = 38$ $n_{\text{wave bw2, finished}} = 3,571$ ; $n_{\text{wave bw2, interrupted}} = 126$ $n_{\text{wave bw3, finished}} = 3,698$ ; $n_{\text{wave bw3, interrupted}} = 39$ $n_{\text{wave bw4, finished}} = 3,608$ ; $n_{\text{wave bw4, interrupted}} = 28$ $n_{\text{wave cw1, finished}} = 3,677$ ; $n_{\text{wave cw1, interrupted}} = 43$ $n_{\text{wave cw2, finished}} = 3,598$ ; $n_{\text{wave cw2, interrupted}} = 86$ $n_{\text{wave cw3, finished}} = 3,584$ ; $n_{\text{wave cw3, interrupted}} = 37$ $n_{\text{wave cw4, finished}} = 3,530$ ; $n_{\text{wave cw4, interrupted}} = 44$
Survey duration (only finished)	bw1: 22 minutes (median), 26 minutes (mean) bs1: 8 minutes (median), 10 minutes (mean) bw2: 27 minutes (median), 32 minutes (mean) bw3: 18 minutes (median), 21 minutes (mean) bw4: 17 minutes (median), 20 minutes (mean) cw1: 23 minutes (median), 26 minutes (mean) cw2: 27 minutes (median), 32 minutes (mean) cw3: 19 minutes (median), 22 minutes (mean) cw4: 21 minutes (median), 25 minutes (mean)
Survey languages	German (source questionnaire), Arabic (not in bs1), English, Russian (not in bs1), Turkish (not in bs1), Portuguese (only bw1 to cw3)
Citation of data	Dollmann, Jörg; Jacobsen, Jannes; Mayer, Sabrina J.; Lietz, Almuth; Köhler, Jonas; Schmälzle, Michaela; Siegel, Madeleine; Zimmermann, Stefan; Chouaibi, Doreen; Kalter, Frank (2024). DeZIM.panel Data Release 4.0.0. Dataset. Version: 4.0.0. Berlin: Deutsches Zentrum für Integrations- und Migrationsforschung (DeZIM).
DOI	10.34882/dezim.panel.download.4.0.0 10.34882/dezim.panel.remote.4.0.0 10.34882/dezim.panel.onsite.4.0.0

### 3 Survey instrument

A total of four survey waves are carried out every year. Standardized online questionnaires are used as survey instruments. The questionnaires are initially created in German and then translated into English, Arabic, Russian, and Turkish. Wave bs1 (short survey on reactions to the war in Ukraine) was only translated into English due to the short preparation time for this survey. Moreover, Portuguese is available as survey language in the following waves: bw1, bw2, bw3, bw4, cw1, cw2, and cw3. To guarantee the quality of the survey translations, the initial translation is reviewed by a second translator.

Half of the regular surveys consist of various core modules (8 minutes) as well as questions on current topics (2 minutes). The core modules focus on political institutions and attitudes as well as political behaviour (Module 1), on social values and attitudes (Module 2), on health and well-being (Module 3) and on work and education

as well as experiences of discrimination (Module 4). Questions about subjective well-being and the assessment of the individual life situation are asked in each wave. Table 2 shows which core module was surveyed in which wave. The specific content of each wave's questionnaire can be found in the questionnaires and codebooks included in the documentation.

The general idea behind the core modules is to collect central constructs from various thematic areas that can be included by a variety of researchers as independent variables, moderators, or dependent variables. Since we intend to ask largely the same questions in the core modules each year, this allows developments to be tracked over time. For current topics, two minutes of survey time are reserved in each wave. All questions were either taken from other established surveys or tested in advance and are generally asked to the entire panel. The core modules and questions on current topics are created by the DeZIM.panel team in consultation with the scientific heads of the DeZIM Institute. The other half of the survey time is reserved for employees of the DeZIM Institute, the DeZIM Research Community and the BMFSFJ (Federal Ministry from Family Affairs, Senior Citizens, Women and Youth), who can submit suggestions and thus have the opportunity to collect cross-sectional and longitudinal data with a total of 10 minutes of survey time per regular survey wave. In contrast to the recurring modules, these questions can be based on current political events. Here too, all questions should, if possible, always be able to be asked to the entire panel and should have been tested in advance. Some of the questions of wave bs1 were also included and published in the Consortium for the Social, Behavioural, Educational and Economic Sciences's standard questionnaire for collecting socio-demographic and crisis-related variables (RatSWD 2023).



## 4 Field phase

### 4.1 Field phases

The field phases can be found in Table 2. Those invited online are always informed that the field phase ends on a specific date, to those invited by letter no end of the field phase is communicated. All responses received up to the start of the next field phase are included in the dataset. Those invited by email receive up to six reminders in addition to the initial invitation. At the end of the field phase, respondents are additionally motivated to take part in the survey by a raffle. During each raffle week, from Monday to Friday, one person is randomly drawn from all those who have taken part up to that point and could win €200 in the form of a desired voucher. People who did not provide an email address or provided an incorrect email address in the recruitment wave are invited to the online survey by letter with a link to the survey and receive another postal reminder letter. All participants receive an incentive after participation (postpaid) in the form of a voucher (Amazon, Zalando or bucher.de) worth of €10. From wave 2, it is possible to have the €10 transferred to a provided bank account.

Please note the following special features regarding the field phase of the short survey on reactions to the war in Ukraine: In contrast to the regular survey waves, in which people are invited to the survey by letter and email, only people who provided us with their email address could be invited to the short survey due to the limited time frame. In addition to the first invitation on February 28th, those invited by email received a reminder on March 2nd, 2022. For this short survey, a raffle was used as an incentive. Participants were informed in the first invitation that, as a reward for their participation, all participants would be entered into a raffle for 20 x €100 vouchers<sup>4</sup>.

Table 2: Core modules across waves and survey period

Wave	Field phase	Postal invitation	Postal Reminder	Core Module
bw1	02.12.2021 – 31.01.2022	08.12.2021	13.01.2022	Politics
bs1	28.02.2022 – 06.03.2022	--	--	Reactions to the war in Ukraine
bw2	31.03.2022 – 22.05.2022	05.04.2022	09.05.2022	Attitudes & Values
bw3	27.06.2022 – 07.08.2022	01.07.2022	27.07.2022	Health & Well-being
bw4	26.09.2022 – 06.11.2022	06.10.2022	26.10.2022	Work, Education & Experiences of Discrimination
cw1	12.12.2022 – 31.01.2023	15.12.2022	18.01.2023	Politics
cw2	28.03.2023 – 14.05.2023	31.03.2023	02.05.2023	Attitudes & Values

<sup>4</sup> <https://www.wunschgutschein.de/>

cw3	19.06.2023 – 31.07.2023	19.06.2023	30.08.2023 <sup>5</sup>	Health & Well-being
cw4	18.09.2023 – 30.10.2023	28.09.2023	23.10.2023	Work, Education & Experiences of Discrimination

## 4.2 Survey Duration

The survey duration is always roughly announced in the first email invitation and on the welcome page of the questionnaire. As far as the actual interview duration is concerned, Table 3 shows the arithmetic mean, the standard deviation, the median as well as the interquartile range (IQR). The median is always slightly below the arithmetic mean. As mentioned previously, wave bs1 is a short survey, i.e. it was designed to be substantially shorter than the regular studies. Table 3 also shows the IQR resulting from the first and third quartile. For instance, in wave bw1, this means that 50% of the participants took between 17 and 31 minutes to complete the survey.

Table 3: Survey duration by wave (in minutes)

Wave	Announced Duration	Mean	SD	25%-Quantil	50%-Quantil Median	75%-Quantil
bw1	15-20	26	15	17	22	31
bs1	5	10	7	6	8	12
bw2	25	32	16	21	27	38
bw3	15	21	11	14	18	24
bw4	20	20	11	13	17	24
cw1	20	26	13	17	23	31
cw2	25	32	17	21	27	38
cw3	20	22	12	14	19	25
cw4	20	25	13	16	21	29

## 5 Population and response

### 5.1 Population

The population of the study consists of all people living in Germany with and without a migration history who were born between 1941 and 2004. The sample includes an oversampling of people from Turkey, other predominantly Muslim countries, countries with which West Germany concluded recruitment agreements between 1955 and 1973 (so-called “guest workers”), individuals from the former USSR and Romania (so-called “Spätaussiedler”), as well as the descendants of these people. This is a random sample based on the residents'

<sup>5</sup> Due to a technical error, the postal reminder was sent out later than in other waves. As participants were still able to take part until the field start of the following wave (18.09.2023), this had no effect on the opportunity to take part.

registration offices, with which current moods and trends in the German population can be mapped longitudinally.

The DeZIM.panel sample is based on a two-stage stratified sampling. First, 57 municipalities or 60 Primary Sampling Units (PSU) were randomly selected in a Probability Proportional to Size (PPS) selection. The drawing was carried out with implicit stratification according to federal state and BIK municipality size classes. In the second step, a disproportionate approach was chosen for the selection of people (secondary sampling units) to achieve oversampling in the five subgroups with a migration history. These groups were taken into account with a higher probability of being selected but had to be preclassified onomastically because the sample characteristics provided by the registration offices did not include any information on migration biography. The operational sample for the survey included a total of 37,583 addresses out of the 103,926 drawn addresses. Further information on random sampling can be found in Dollmann et al. (2022) and in Ruland & Sandbrink (2022). A total of 9,168 people took part in the recruitment wave, of which 6,719 agreed to take part in further follow-up surveys.

The following sections present the wave-specific response rates. The first table shows the response rate for participants invited by email or letter. The second table shows the response rate of the five subgroups with a migration history in comparison to participant without a migration history. For the calculation of the response rate, we followed the AAPOR standard definitions (AAPOR 2023).

## 5.2 Return bw1

In total, 6,719 people were invited to the first regular wave of the survey. The majority of respondents (52.2%) were invited via email. People who did not provide an email address in the recruitment wave were invited by letter (45.4%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. In wave bw1, this applied to 2.5% of participants.

**Table 4: Return bw1 after delivery**

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	3,506	3,047	166	6,719
<i>Undeliverable</i>	4	69	0	73
<i>Cleaned up sent invitations</i>	3,502	2,978	166	6,646
Welcome page opened [b]	2,365	1,531	57	3,953
Questionnaire completed [c]	2,182	1,453	51	3,686
AAPOR RR1 [b/a]	67.5%	50.2%	34.3%	58.8 %
AAPOR RR2 [c/a]	62.2%	47.7%	30.7%	54.9 %

**Table 5: Return bw1 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	465	666	565	444	976	3,583	20	6,719
<i>Undeliverable</i>	3	15	7	6	14	27	1	73
<i>Cleaned up sent invitations</i>	462	651	558	438	962	3,556	19	6,646
Welcome page opened [b]	263	327	319	265	567	2,206	6	3,953
Questionnaire completed [c]	245	283	295	249	525	2,084	5	3,686
AAPOR RR1 [b/a]	56.6%	49.1%	56.5%	59.7%	58.1%	61.6%		58.8%
AAPOR RR2 [c/a]	52.7%	42.5%	52.2%	56.1%	53.8%	58.2%		54.9%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

### 5.3 Return bs1

In total, 4,855 people were invited to the first short survey. As mentioned previously, due to the limited field time, only people who had provided us with a email address by February 28, 2022, could be invited to the short survey. Accordingly, all participants were invited by email.

**Table 6: Return bs1 after delivery**

	Total
Sent out Invitations [a]	4,855
<i>Undeliverable</i>	<i>119</i>
<i>Cleaned up sent invitations</i>	<i>4,736</i>
Welcome page opened [b]	2,713
Questionnaire completed [c]	2,626
AAPOR RR1 [b/a]	55.9 %
AAPOR RR2 [c/a]	54.1 %

**Table 7: Return bs1 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	316	459	410	315	716	2,630	9	4,855
<i>Undeliverable</i>	<i>7</i>	<i>16</i>	<i>14</i>	<i>11</i>	<i>23</i>	<i>46</i>	<i>2</i>	<i>119</i>
<i>Cleaned up sent invitations</i>	<i>309</i>	<i>443</i>	<i>396</i>	<i>304</i>	<i>693</i>	<i>2,584</i>	<i>7</i>	<i>4,736</i>
Welcome page opened [b]	161	191	218	188	404	1,549	2	2,713
Questionnaire completed [c]	153	172	214	180	392	1,513	2	2,626
AAPOR RR1 [b/a]	50.9%	41.6%	53.2%	59.7%	56.4%	58.9%		55.9%
AAPOR RR2 [c/a]	48.4%	37.5%	52.2%	57.1%	54.7%	57.5%		54.1%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

#### 5.4 Return bw2

In total, 6,655 people were invited to wave bw2. The majority of respondents (71.4%) were invited via email. People who did not provide an email address or provided a not reachable email address in the previous waves were invited by letter (28.2%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. In wave bw1, this applied to 0.4% of participants.

Table 8: Return bw2 after delivery

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	4,749	1,877	29	6,655
<i>Undeliverable</i>	10	29	5	44
<i>Cleaned up sent invitations</i>	4,739	1,848	24	6,611
Welcome page opened [b]	3,216	554	10	3,780
Questionnaire completed [c]	3,051	510	10	3,571
AAPOR RR1 [b/a]	67.7%	29.5%	34.5%	56.8%
AAPOR RR2 [c/a]	64.2%	27.2%	34.5%	53.7%

Table 9: Return bw2 according to overrepresented groups

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	460	663	560	436	967	3,549	20	6,655
<i>Undeliverable</i>	7	3	3	6	10	15	0	44
<i>Cleaned up sent invitations</i>	453	660	557	430	957	3,534	20	6,611
Welcome page opened [b]	232	312	300	249	549	2,134	4	3,780
Questionnaire completed [c]	222	265	281	234	519	2,046	4	3,571
AAPOR RR1 [b/a]	50.4%	47.1%	53.6%	57.1%	56.8%	60.1%		56.8%
AAPOR RR2 [c/a]	48.3%	40.0%	50.2%	53.7%	53.7%	57.7%		53.7%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

### 5.5 Return bw3

In total, 6,557 people were invited to wave bw3. The majority of respondents (77.2%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (22.3%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. In wave bw3, this applied to 0.5% of participants.

Table 10: Return bw3 after delivery

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	5,064	1,463	30	6,557
<i>Undeliverable</i>	10	3	0	13
<i>Cleaned up sent invitations</i>	5,054	1,461	30	6,545
Welcome page opened [b]	3,485	301	8	3,794
Questionnaire completed [c]	3,404	286	8	3,698
AAPOR RR1 [b/a]	68.8%	20.6%	26.7%	57.9%
AAPOR RR2 [c/a]	67.2%	19.5%	26.7%	56.4%

Table 11: Return bw3 according to overrepresented groups

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	454	653	547	430	948	3,506	19	6,557
<i>Undeliverable</i>	1	3	2	0	1	6	0	13
<i>Cleaned up sent invitations</i>	453	650	545	430	947	3,500	19	6,544
Welcome page opened [b]	236	292	308	243	549	2,162	4	3,794
Questionnaire completed [c]	228	267	298	238	535	2,128	4	3,698
AAPOR RR1 [b/a]	52.0%	44.7%	56.3%	56.5%	57.9%	61.7%		57.9%
AAPOR RR2 [c/a]	50.2%	40.9%	54.5%	55.3%	56.4%	60.7%		56.4%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

## 5.6 Return bw4

In total, 6,523 people were invited to wave bw4. The majority of respondents (80.7%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (19.3%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. In wave bw4, this applied to 0.03% of participants.

**Table 12: Return bw4 after delivery**

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	5,263	1,258	2	6,523
<i>Undeliverable</i>	24	12	0	36
<i>Cleaned up sent invitations</i>	5,239	1,246	2	6,487
Welcome page opened [b]	3,545	153	2	3,700
Questionnaire completed [c]	3,470	137	1	3,608
AAPOR RR1 [b/a]	67.4%	12.2%	100.0%	56.7%
AAPOR RR2 [c/a]	65.9%	10.9%	50.0%	55.3%

**Table 13: Return bw4 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	450	650	544	431	939	3,490	19	6,523
<i>Undeliverable</i>	6	3	3	6	5	13	0	36
<i>Cleaned up sent invitations</i>	444	647	541	425	934	3,477	19	6,487
Welcome page opened [b]	238	298	305	247	533	2,075	4	3,700
Questionnaire completed [c]	231	268	299	245	516	2,045	4	3,608
AAPOR RR1 [b/a]	52.9%	45.8%	56.1%	57.3%	56.8%	59.5%		56.7%
AAPOR RR2 [c/a]	51.3%	41.2%	55.0%	56.8%	55.0%	58.6%		55.3%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).



## 5.7 Return cw1

In total, 6,500 people were invited to wave cw1. The majority of respondents (82.0%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (17.9%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. In wave cw1, this applied to 0.1% of participants.

**Table 14: Return cw1 after delivery**

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	5,331	1,166	3	6,500
<i>Undeliverable</i>	18	15	0	33
<i>Cleaned up sent invitations</i>	5,313	1,151	3	6,467
Welcome page opened [b]	3,640	113	2	3,755
Questionnaire completed [c]	3,571	104	2	3,677
AAPOR RR1 [b/a]	68.3%	9.7%	66.7%	57.8%
AAPOR RR2 [c/a]	67.0%	8.9%	66.7%	56.6%

**Table 15: Return cw1 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	449	647	543	430	937	3,475	19	6,500
<i>Undeliverable</i>	7	5	2	2	6	11	0	33
<i>Cleaned up sent invitations</i>	442	642	541	428	931	3,464	19	6,467
Welcome page opened [b]	237	283	307	255	540	2,126	7	3,755
Questionnaire completed [c]	233	268	300	250	526	2,095	5	3,677
AAPOR RR1 [b/a]	52.8%	43.7%	56.5%	59.3%	57.6%	61.2%		57.8%
AAPOR RR2 [c/a]	51.9%	41.4%	55.2%	58.1%	56.1%	60.3%		56.6%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

## 5.8 Return cw2

In total, 6,439 people were invited to wave cw2. The majority of respondents (72.5%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (16.5%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. Additionally, all participants were invited by letter, who never participated in a survey after the recruitment wave – additionally to being invited by email. In wave cw2, this applied to 11.0% of participants.

**Table 16: Return cw2 after delivery**

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	4,665	1,063	711	6,439
<i>Undeliverable</i>	0	18	0	18
<i>Cleaned up sent invitations</i>	4,665	1,045	711	6,421
Welcome page opened [b]	3,498	81	165	3,744
Questionnaire completed [c]	3,395	63	140	3,598
AAPOR RR1 [b/a]	75.0%	7.6%	23.2%	58.1%
AAPOR RR2 [c/a]	72.8%	5.9%	19.7%	55.9%

**Table 17: Return cw2 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	444	639	539	426	925	3,447	19	6,439
<i>Undeliverable</i>	0	6	1	1	2	8	0	18
<i>Cleaned up sent invitations</i>	444	633	538	425	923	3,439	19	6,421
Welcome page opened [b]	243	291	302	255	533	2,114	6	3,744
Questionnaire completed [c]	231	258	292	247	511	2,053	6	3,598
AAPOR RR1 [b/a]	54.7%	45.5%	56.0%	59.9%	57.6%	61.3%		58.1%
AAPOR RR2 [c/a]	52.0%	40.4%	54.2%	58.0%	55.2%	59.6%		55.9%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

### 5.9 Return cw3

In total, 6,368 people were invited to wave cw3. The majority of respondents (75.2%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (16.1%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. Additionally, all participants were invited by letter, who never participated in a survey after the recruitment wave – additionally to being invited by email. In wave cw3, this applied to 8.7% of participants.

**Table 18: Return cw3 after delivery**

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	4,791	1,024	553	6,368
<i>Undeliverable</i>	0	3	0	3
<i>Cleaned up sent invitations</i>	4,791	1,021	553	6,365
Welcome page opened [b]	3,524	76	77	3,677
Questionnaire completed [c]	3,449	67	68	3,584
AAPOR RR1 [b/a]	73.6%	7.4%	13.9%	57.7%
AAPOR RR2 [c/a]	72.0%	6.5%	12.3%	56.3%

**Table 19: Return cw3 according to overrepresented groups**

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	443	621	534	419	911	3,421	19	6,368
<i>Undeliverable</i>	0	2	0	0	0	0	1	3
<i>Cleaned up sent invitations</i>	443	619	534	419	911	3,421	18	6,365
Welcome page opened [b]	239	273	300	251	532	2,076	6	3,677
Questionnaire completed [c]	229	257	295	244	510	2,044	5	3,584
AAPOR RR1 [b/a]	54.0%	44.0%	56.2%	59.9%	58.4%	60.7%		57.7%
AAPOR RR2 [c/a]	51.7%	41.4%	55.2%	58.2%	56.0%	59.7%		56.3%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

### 5.10 Return cw4

In total, 6,293 people were invited to wave cw4. The majority of respondents (76.8%) were invited via email. People who did not provide an email address in the previous waves were invited by letter (15.4%). Participants, who were initially invited online, but their email address was no (longer) valid or bounced for other reasons (full mailbox), those participants were additionally invited by post at the moment of the postal reminder (4 to 6 weeks after field start), if a valid postal address was available. Additionally, all participants were invited by letter, who never participated in a survey after the recruitment wave – additionally to being invited by email. In wave cw4, this applied to 7.8% of participants.

Table 20: Return cw4 after delivery

	Online	Postal	Online + Postal	Total
Sent out Invitations [a]	4,831	971	491	6,293
<i>Undeliverable</i>	2	8	0	10
<i>Cleaned up sent invitations</i>	4,829	963	491	6,283
Welcome page opened [b]	3,490	64	64	3,618
Questionnaire completed [c]	3,422	52	56	3,530
AAPOR RR1 [b/a]	72.2%	6.6%	13.0%	57.5%
AAPOR RR2 [c/a]	70.8%	5.4%	11.4%	56.1%

Table 21: Return cw4 according to overrepresented groups

	[1]	[2]	[3]	[4]	[5]	[6]	missings	Total
Sent out Invitations [a]	440	605	523	417	901	3,389	18	6,293
Undeliverable	1	2	1	0	2	4	0	10
<b>Cleaned up sent invitations [b]</b>	439	603	522	417	899	3,385	18	6,283
Welcome page opened [c]	226	266	299	234	518	2,066	9	3,618
Questionnaire completed [d]	213	248	289	230	509	2,034	7	3,530
AAPOR RR1 [b/a]	51.4%	44.0%	57.2%	56.1%	57.5%	61.0%		57.5%
AAPOR RR2 [c/a]	48.4%	41.0%	55.3%	55.2%	56.5%	60.0%		56.1%

[1] from Turkey, [2] from other Muslim-majority countries, [3] from countries with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

### 5.11 Weighting

The basis of the original recruitment study was a population registration office sample that was qualified using onomastics methods. The sample was used disproportionately according to the results of the onomastics (explicit stratification). With multi-stage PPS sampling (probability proportional to size), the calculation of a design weight can theoretically be omitted since this is constant for all units. Due to the explicit disproportionate stratification, i.e. the separate use of onomastically qualified target persons and the fact that the importance weight is calculated in the first selection stage over the entire population and this community selection is used for all target groups, the gross sample is not a self-weighted sample. In addition, a proportion of municipalities were not able to provide sufficient addresses and not all municipalities took part, which means that there were dropouts among the municipalities. Therefore, despite PPS drawing, the selection probabilities were first determined, and design weights were calculated. The calculation of the design weight and its modification can be found in Ruland & Sandbrink (2022).

To assess possible biases in sample selection and field work, the distribution of our sample with regard to certain central characteristics was compared with the distribution of these characteristics in the German Microcensus (MZ) (see Table 22, see also Dollmann et al. 2022). The results show that from recruitment (W0) to panel consent (W0P) to actual participation in wave 1 (W1), the bias towards higher education and younger participants increases. No clear patterns can be identified with regard to gender. Carrying out a multivariate logistic regression on the probability of not participating in the panel or wave 1 shows that, in addition to age, education and gender, migration status also had a significant positive association. In the recruitment wave, the probability of not participating was 8 percentage points significantly higher for immigrants than for people without a migration history, but no significant difference was found between people with a migration history and the descendants of immigrants. In the first wave, participation was significantly higher for immigrants compared to people without a migration history by 14 percentage points and for children of immigrants by 5 percentage points (all  $p < 0.05$ ) (Dollmann et al. 2022).

These results show the necessity to take non-response characteristics into account in the weighting. Due to the different selection probabilities and the stratification of the groups, the data should be weighted for the analyses. We therefore offer a design weight that takes into account different selection probabilities [dweight], as well as a combined weight that also takes into account unit non-response in the recruitment sample (based on age groups, gender, federal state and size of municipalities [BIK]) [dweight\_adj]. In addition to the adjusted design weight, it is recommended to estimate a retention factor for the analysis sample and multiply this by the adjusted design weight.

Table 22: Comparison of unweighted cell percentages by gender, education, and age group

Personen ...	Gender: male				Education: high school diploma				Age group: 18-30 years old			
	MZ	W0	W0P	W1	MZ	W0	W0P	W1	MZ	W0	W0P	W1
[1]	51,8	47,2	44,5	42,3	27,7	35,4	38,8	43,5	34,6	31,0	35,8	39,3
[2]	60,0	57,9	59,1	58,8	47,2	45,1	47,5	51,1	44,1	33,3	35,1	35,7
[3]	47,6	38,9	41,8	39,0	37,8	43,1	48,5	53,8	28,7	18,5	22,1	24,8
[4]	54,1	47,8	50,1	46,2	35,0	46,8	51,6	61,3	26,4	23,1	26,6	30,7
[5]	50,6	41,8	43,2	41,3	48,5	54,9	58,4	66,2	22,5	19,8	23,0	24,7
[6]	50,1	45,8	46,7	47,7	40,8	45,8	49,5	54,3	27,2	20,1	22,0	23,9

[1] from Turkey, [2] from other Muslim-majority countries, [3] from states with which West Germany concluded recruitment agreements between 1955 and 1973, so-called “guest workers,” [4] from the former USSR as well as Romania and their descendants, [5] with a different migration history, [6] without a migration history. For group composition, see Dollmann et al. (2022).

MZ: Microcensus 2017, W0P: People who took part in the recruitment wave and gave their panel consent. Abitur: excluding people with other/ongoing training. Gender: excluding people who identify as diverse.

## 6 Data curation and anonymization

The following sections provide an overview of the available data editions and data access, the data structure as well as the wave and variable naming scheme including information about the procedure for aggregation, versioning, harmonization, and the missing convention used.

### 6.1 Editions of scientific use files (SUF) and data access

DeZIM.panel data is made available exclusively for scientific research purposes, i.e. it is only made available to data users in the scientific community with a specific research purpose. Researchers are granted access to the DeZIM.panel data if they have signed a data usage contract with DeZIM. The various data packages, the so-called editions, differ according to the amount of information they contain, the level of data protection and the type of data access. With data release 4.0.0, the DeZIM.panel offers three editions (see Figure 1).

Figure 1: Editions of DeZIM.panel release 4.0.0



Data users can apply for all three editions. The editions have a different information level, which is represented by the "restriction level". A lower data restriction level makes it possible to provide more information. The **Onsite Edition** is the least restrictive data version and allows access to sensitive personal variables (e.g. "wrk0026x02 Postal code of current employer"). On request, it is also possible to use open-coded variables, which are first checked from a data protection perspective. To use the Onsite Edition, researchers can submit a request to the Research Data Center. In a secure working environment, researchers can use the complete data of the Onsite Edition at DeZIM's institute and take their results with them after they have been checked by the Research Data Center. The **Remote Edition** of the DeZIM.panel data also provides access to information classified as sensitive, with researchers working online in an environment hosted by DeZIM. When applying for the Remote Edition, data users receive access to the data. As a third edition, DeZIM offers the data in the **Download Edition**. Access is provided via an encrypted download and the data is strictly anonymized. However, it should be noted at this point that the download version is still sufficient for the majority of analyses. Variables that have been anonymized can be used by requesting remote or onsite data via the above-mentioned access

methods. In addition, syntax created with the Download Edition can also be transferred to the Remote Edition as far as possible.

Which variables are anonymized or fully available in which data editions is documented in the codebook. The "Aggregation level" column in the sheet "Variables" documents the anonymization level according to the access path of the data: "Download" means that the variable is accessible in the Download Edition. "Remote" means that the variable is available in the Remote Edition but is anonymized in the Download Edition. "Onsite" means that the variable is accessible in the Onsite Edition but is anonymized in the Download and Remote Edition. The category "Onsite on demand" refers to (semi-)open information that can be accessed Onsite on demand at the DeZIM Institute after a specific variable request.

## 6.2 Data structure

With the release 4.0.0 the data of the DeZIM.panel is provided in longitudinal data structure. Release 4.0.0 contains the data from the recruitment wave (aw0, 2021), the first eight regular survey waves (bw1-bw4 (2022), cw1-cw4 (2023)) and the short survey on reactions to the war in Ukraine (bs1, 2021). Data sets in long format offer a compressed and user-friendly data set structure for longitudinal analyses. Here, each person (lfd) has one row for each survey wave (wave). This means that the DeZIM.panel does not release several data sets of cross-sectional data for the different waves, but one data set in which all survey waves are included, and all identical variables have the same variable name across waves. A respondent can appear several times in such a dataset, but only once for each survey wave.

The following three variables are key variables for using the longitudinal dataset:

- **lfd**: The variable lfd is an individual permanent identifier for all participants ever registered in DeZIM.panel.
- **wave**: The variable wave is the wave identifier. For each wave, information for all participants invited in the corresponding survey wave is included in the dataset.
- **par0011**: The variable par0011 (Final outcome of contact) documents whether the survey is completed and if this is not the case, it documents the reasons (e.g., interruption, undeliverable invitation, non-participation, etc.).

## 6.3 Survey wave naming scheme

The DeZIM.panel is surveyed four times a year. The following name scheme is used to distinguish between the waves:

- Distinction between survey years: a = 2021, b = 2022, c = 2023, etc.
- Distinction between regular waves [w] and short surveys [s]

According to this wave name scheme, the recruitment wave of 2021 represents wave aw0, the four regular waves in 2022 are named wave bw1, bw2, bw3, bw4 and the short survey on reactions to the war in Ukraine



conducted in 2022 is named wave bs1. The four regular waves conducted in 2023 are named cw1, cw2, cw3, and cw4.

**Please note:** Each year, the first wave starts in the middle of December of the year before and continues until March. Since most of the field time is between January and March, the new year is used as prefix for the wave abbreviations. The year of the field start is documented in the variable year.

#### 6.4 Variable naming scheme

With the release version 4.0.0 of the DeZIM.panel data, the variable naming concept was changed. Except for special variables such as identifiers or weighting variables, all variable names in the survey data are defined according to the naming convention presented in Figure 2:

Figure 2: Variable naming scheme

Topic	Topic number	{x}	{Item number}	{_o}	{_v*}	{_h} / {_g}
<ul style="list-style-type: none"> <li>• soc</li> <li>• edu</li> <li>• wrk</li> <li>• pol</li> <li>• val</li> <li>• dcr</li> <li>• hlt</li> <li>• lif</li> <li>• var</li> <li>• par</li> <li>• svy</li> </ul>	<ul style="list-style-type: none"> <li>• 0001</li> <li>• ...</li> <li>• 9999</li> </ul>	<ul style="list-style-type: none"> <li>• x</li> </ul>	<ul style="list-style-type: none"> <li>• 01</li> <li>• ...</li> <li>• 99</li> </ul>	<ul style="list-style-type: none"> <li>• _o</li> </ul>	<ul style="list-style-type: none"> <li>• _v1</li> <li>• ...</li> <li>• _v99</li> </ul>	<ul style="list-style-type: none"> <li>• _h</li> <li>• _g</li> </ul>

A variable name always consists of at least one topic and a number from 0001-9999 (topic number), which is usually assigned chronologically. In total, the variable names are divided into eleven topics:

- "pol" (Political topics such as elections, organisations, political engagement, reactions/opinions on political events)
- "val" (Attitudes/values and personality)
- "hlt" (Health topics such as the healthcare system, mental and physical health or care)
- "wrk" (Work and income)
- "edu" (Education and vocational training)
- "dcr" (Experiences with discrimination)
- "lif" (Life, family/friends/social networks, leisure activities)
- "soc" (Demography, home, wealth)
- "svy" (Survey variables such as interview language and survey evaluation)
- "var" (Random/recoding triggers, filter variables for vignettes, preloads)
- "par" (Paradata such as interview duration, time of survey start)

The "topic number" refers to a question in the questionnaire. If a question consists of several items (e.g., matrix question with a standardized scale or question with multiple answers), these items are integrated into the variable names using an "x" as the item separator and a two-digit "item number".

### 6.5 Open coded variables

If the variable type of an entire question or an individual item is an open string (string/character), the variable name is given an additional "\_o" to indicate string variables in the variable name.

### 6.6 Aggregated and ex-post coded variables ("\_g")

The data set contains variables whose variable names end with a "\_g". These variables are defined as aggregated or ex-post-coded variables. Aggregated variables are relevant if the Download Edition is used, as not all information is included there for data protection reasons (see section 6.1). These aggregated variables are provided for variables that contain particularly sensitive information. These aggregated variables aggregate particularly sensitive information into larger response categories than in the original variable. The aggregated variables have the same variable name (with "\_g") as the original variable (without "\_g") so that aggregated and original variables can be easily associated. In the Download Edition, the sensitive variable is anonymized, and the aggregated variable is provided with less information. In the Remote and Onsite Editions of the dataset, the complete information is available to data users; the aggregated variables should be understood as an offer for users of the Download Edition. In a few cases, "\_g" in the variable name also describes an ex-post-coding. In these ex-post-codings, open or semi-open information is categorized.

**Example:** The variable edu0001\_v1 (highest general school-leaving degree) contains the response option "Other school leaving certificate, namely:" Where possible, the respondents' open answers were integrated into the response scale in the variable edu0001\_g ("Highest general school degree - recoded").

### 6.7 Versionized ("\_v") and harmonized ("\_h") variables

Across the survey waves of the DeZIM.panel, questions may change in content or be measured in different ways. To enable data users to use the DeZIM.panel easily and as consistently as possible, some variables are versionized and harmonized. Versionized variables inform users that very similar constructs have been measured differently over time. These versionized variables are intended to document changes in data collection over time. The harmonized variables aim to make it easier for data users to conduct analyses over time. For this purpose, versionized variables are harmonized ex-post in order to harmonize the question constructs and measurement differences between the survey waves so that the content is consistent and comparable across the survey waves.

A question construct can differ in terms of the following dimensions: Question text, instruction text, answer categories (wording, quantity) or variable type (metric vs. ordinal, single vs. multiple responses, survey with or

without decimals). To decide whether a variable should be versionized, these dimensions were compared between the survey waves. If differences<sup>6</sup> were found in the described dimensions for questions that were surveyed in several waves, the variable was versionized and the variable name was extended with "\_v\*". Due to the changed measurement of a question construct, a change in response behavior can be assumed and the comparability of the question construct between waves might be limited. Data users can use the codebook to understand the wave-specific wording of a question and its answer options in order to decide to what extent comparability is given for the respective analysis.

For a few central variables, the DeZIM.panel offers harmonized variables that harmonize the version variables into a common variable "\_h" in order to enable the most panel-consistent usability. These variables are intended solely as a harmonization proposal and cannot be used universally for every research demand. Therefore, we would like to encourage data users to check the version variables also contained in the dataset and, if necessary, to harmonize them according to the research interest.

**Please note:** One central variable to many analyses, namely gender, is not provided as a harmonized variable in release 4.0.0. The dataset contains several variables: soc0001\_v1 (Gender) (self-administered), soc001\_v2 (EWO: gender of target person) (information provided by the residents' registration office and used in the sampling strategy), soc0001\_v3 (Gender from birth certificate) (self-administered). In some cases, the information differs between the variables. This might reflect an actual change in gender status or signify an inconsistency or a correction. No harmonization decision was taken here. Users can decide whether they want to exclude these cases in their analysis or do plausibility checks.

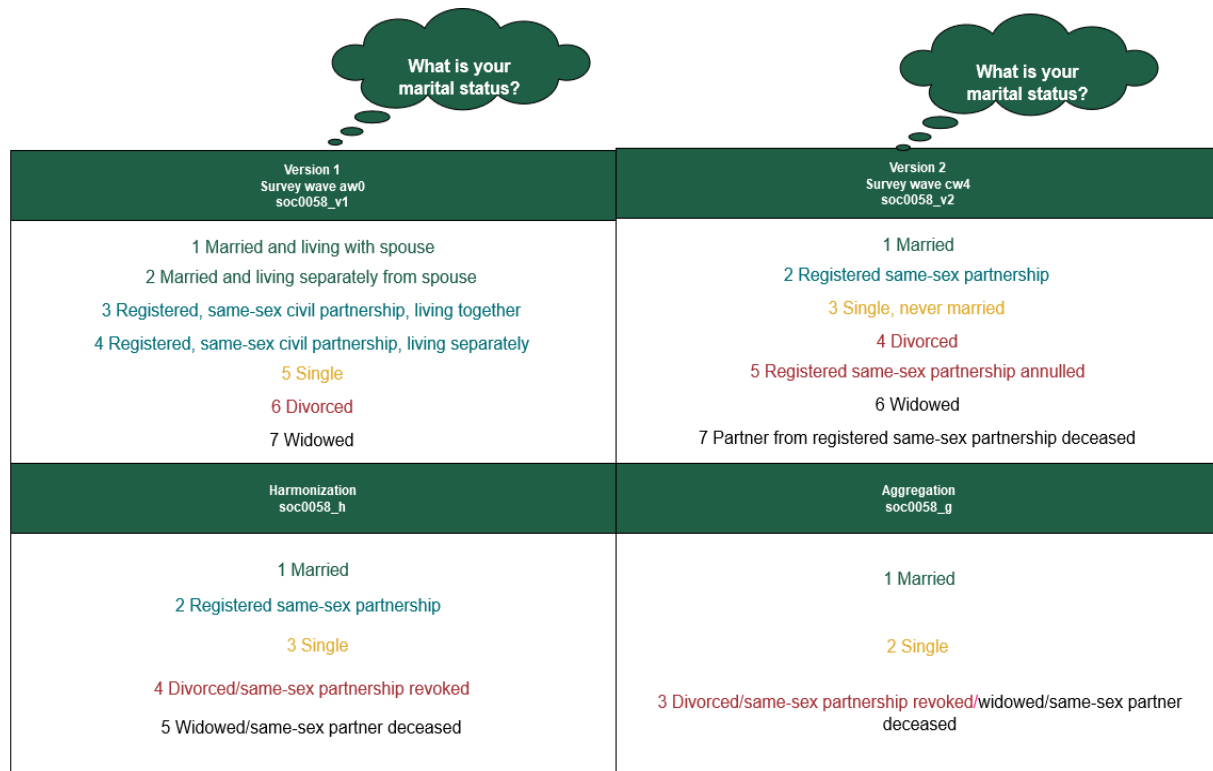
## 6.8 Example: Versionizing, harmonization, aggregation

An example of a variable from data release 4.0.0 that has been versionized, harmonized and aggregated is marital status (see Figure 3).

---

<sup>6</sup> Grammatical or orthographic adjustments in questions, instructions or answers are not considered to require versionizing in the DeZIM.panel. A modified filter structure is also not currently considered to require versionizing, although distribution disparities are very likely.

Figure 3: Versionizing, harmonization, aggregation of the variable marital status



Version 1 Survey wave aw0 soc0058_v1	Version 2 Survey wave cw4 soc0058_v2
1 Married and living with spouse 2 Married and living separately from spouse 3 Registered, same-sex civil partnership, living together 4 Registered, same-sex civil partnership, living separately 5 Single 6 Divorced 7 Widowed	1 Married 2 Registered same-sex partnership 3 Single, never married 4 Divorced 5 Registered same-sex partnership annulled 6 Widowed 7 Partner from registered same-sex partnership deceased
Harmonization soc0058_h	Aggregation soc0058_g
1 Married 2 Registered same-sex partnership 3 Single 4 Divorced/same-sex partnership revoked 5 Widowed/same-sex partner deceased	1 Married 2 Single 3 Divorced/same-sex partnership revoked/widowed/same-sex partner deceased

The construct of marital status was asked in the aw0 recruitment wave and in the cw4 wave. Both questions were asked identically, but different response options were possible. As the response options cannot be integrated in the longitudinal data format (different coding of the valid values), the variables were stored in two variable versions (soc0058\_v1 and soc0058\_v2) using the same variable name soc0058. Researchers can use both versions independently for their analyses. However, data users who wish to use the variable in a longitudinal perspective may be able to avoid data preparation effort and use the harmonized variable provided by the Research Data Centre, which recodes and integrates the variable versions into the harmonized variable (soc0058\_h) shown above. The response options are highlighted to make the recoding and integration steps visually visible. As the question on marital status also provides information on sensitive information such as same-sex partnerships, an aggregated variable (soc0058\_g) is provided that can be used for the Download Edition. In the Download Edition, the variables soc0058\_v1, soc0058\_v2 and soc0058\_h were anonymized and only the variable soc0058\_g is released for researchers. Full access to all marital status variables is available if the Remote or Onsite Edition is used.

## 6.9 Missing conventions

Each variable in the DeZIM.panel has ten missing codes with data release 4.0.0. All missing codes have negative values and are defined from -999 to -900 (see Table 23).

Table 23: Missing codes

Value	Value label	Definition
-999	No answer	Answer category in questionnaire
-998	Don't know	Answer category in questionnaire
-997	Does not apply	Answer category in questionnaire / xxx
-996	Anonymized	Anonymization carried out
-971	Filter Missing	Missing due to filter guidance, split questions
-969	Not determinable or implausible value	Information that cannot be determined (e.g. for open information) or implausible values
-950	Technical error	Technical error during survey
-902	Interrupted	Participants who did not complete all the survey
-901	No participation in this wave	Participants who did not participate in this wave
-900	Question not part of the survey program this wave	Question was not part of the survey program in this wave

The values "-999 No answer" and "-998 Don't know" are defined as response options for almost every question in the DeZIM.panel and are direct responses from the respondents. Sometimes "-997 Does not apply" is also defined as an answer in the questionnaire. However, the code "-997 Does not apply" is also assigned during data preparation if a question is not skipped due to a filter condition, but because a question was not defined as a required response and the respondent deliberately skipped the question. Whether the answer option "Does not apply" was a predefined response in the question program or was assigned in the data preparation process can be checked in the questionnaires. A variable is coded with the code "-996 Anonymized" if the information is classified as sensitive. For example, in the Download Edition of the dataset, some personal and region-specific variables that are available in the less restrictive Remote and Onsite Editions have been anonymized. Likewise, open coded variables in the Download and Remote Editions are only provided with the value "-996 Anonymized". However, these can be used Onsite on request (see section 6.1). If questions with filter routing are asked in order to exclude certain respondents from answering the question, these respondents receive the value "-971 Filter Missing" for the filtered variable in the dataset. The missing code "-950 Technical error" refers to the following events:

- a) A technical error occurred due to the user's browser navigation behavior: Participants can return to previous questions in the survey context using a browser back button. If participants used the back button in the browser instead of the back button in the survey context, it could lead to new requests sent to the database. This could result in one page arriving with a delay while another is processed first (so-called race condition). As a result, information provided by participants was overwritten, resulting in invalid values.
- b) There were errors in the questionnaire programming: For example, the filter routing was incorrectly programmed or a question was not programmed as a required question, so that people could move on

to the next question without providing valid information. All variables with this missing code are documented in the codebook ("Variables" sheet, "Description" column).

In addition, the missing category "-969 Not determinable or implausible value" is used when the collected data could not be assigned to any value in the coding lists. The coding of this missing category is also documented in the codebook at variable level. Respondents who only partially completed the survey, are coded "-902 Interrupted" for these variables without valid information. People who (a) did not accept the invitation to participate in the survey in the respective wave, (b) only visited the welcome page or (c) only completed the first required question on life satisfaction are coded as "-901 Did not participate in current wave". The missing code "-900 question not part of the question program in this wave" describes for the data in longitudinal format whether a question was asked within the survey wave. If a question was not asked in a survey wave, no information can be available for the respondent and the value "-900 question not part of the question program in this wave" is assigned.

## Literature

AAPOR (2023). Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. 10. Auflage. The American Association for Public Opinion Research, Lenexa, Kansas.

Dollmann, J.; Mayer, S. J.; Lietz, A.; Siegel, M.; Köhler, J. (2022). DeZIM.panel – Data for Germany's Post-Migrant Society. Jahrbücher für Nationalökonomie und Statistik. <https://doi.org/10.1515/jbnst-2022-0025>

RatSWD (Rat für Sozial- und Wirtschaftsdaten). (2023). Standardfragenkatalog zur Erhebung soziodemographischer und krisenbezogener Variablen. (RatSWD Output Series, 7. Berufungsperiode Nr. 4b). Berlin. <https://doi.org/10.17620/02671.76>

Ruland, Michael; Sandbrink, Katharina (2022). Gesellschaftliches Zusammenleben in Deutschland. Rekrutierungsstudie DeZIM.panel. Methodenbericht, Version: 1.0.0. Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.