Generations and Gender Programme

USER GUIDE FOR GGS TEACHING DATASET

Zsuzsanna Makay [1], *Lívia Murinkó* [1],

Olga Grünwald [2], *Aisling Connolly* [2]

[1] Hungarian Demographic Research Institute

[2] Netherlands Interdisciplinary Demographic Institute

Acknowledgement: This data brief is supported by European Union's Horizon Europe HORIZON-INFRA-2021- DEV-02 grant agreement No 101079357, project GGP-5D (The Generations and Gender Programme Preparatory Phase Project).

Please cite as: Makay, Z., Murinkó, L., Grünwald O., & Connolly, A. (2024). "User Guide for GGS Teaching Dataset" GGP Technical Paper Series. https://doi.org/10.5281/zenodo.13859966

Funded by the European Union

Summary

The Generations and Gender Survey (GGS) is the core component of the Generations and Gender Programme (GGP) – an international research infrastructure providing high quality, open access data on population and family dynamics. The GGS is a cross-national panel survey on life course and family dynamics that tracks the experiences and changes that individuals go through in their personal lives.

The GGS Teaching Datasets are based on the second round of the GGS (GGS-II), which was launched in 2020 and data collection is ongoing. The GGS Teaching Datasets are extracts from the already available data, containing a selection of variables and a subset of respondents from several countries.

The aim of the Teaching Datasets is to simplify and facilitate the use of the rich and complex GGS data. By bringing GGS into the classroom, students can learn statistical analysis using easy-to-use real survey data. The GGS Teaching Datasets can also be a useful resource for researchers with limited experience in quantitative analysis of complex survey data. The dataset has been designed to support teaching and learning of variable management, descriptive statistics, multivariate methods, survival analysis and cross-country comparisons. The variables included allow for the study of topics related to family formation, partnership quality, well-being, work-life balance, family and gender attitudes.

This User Guide describes how the Teaching Datasets have been created, how to access and use them, and provides an overview of the data structure and the variables included.

Table of contents

1. Introduction

1.1. What are GGP and GGS?

The Generations and Gender Programme (GGP) is an international research infrastructure providing high quality, open access data on population and family dynamics. The Generations and Gender Survey (GGS) is the core component of the GGP. It is a cross-national panel survey on life course and family dynamics that tracks the experiences and changes that individuals go through in their personal lives, such as leaving the parental home, union formation and dissolution, parenthood, and the opportunities and challenges that individuals face in their lives. GGS applies a panel design – collecting information on the same persons at three-year intervals (also called 'waves' of data collection) – to allow the examination of causes and consequences of inequalities between genders and generations.

The Teaching Dataset is based on the second round of the GGS (GGS-II), which was launched in 2020 with a revised questionnaire, improved survey design and refreshed samples (Gauthier et al. 2023). GGS-II data collection is ongoing. Data from several countries have already been released. The microdata files are freely available to researchers for non-commercial use after registering in the GGP User Space and signing the relevant data agreement.

1.2. What is GGS Teaching Dataset and who is it for?

The GGS Teaching Dataset is an extract from Wave 1 of the GGS-II data, containing a selection of variables and a subset of respondents from several countries.

The GGS Teaching Dataset has been created with the aim of simplifying and facilitating the use of the rich and complex GGS data. By bringing GGS into the classroom, students can learn statistical analysis using easy-to-use real survey data. To support lecturers and teachers, we have created datasets specifically designed for use by students and new data users. The GGS Teaching Dataset can also be a useful resource for researchers with limited experience in quantitative analysis of complex survey data. The dataset has been developed to support teaching and learning of variable management, descriptive statistics, multivariate methods, survival analysis and cross-country comparisons. The variables included allow for the study of topics related to family formation, partnership quality, well-being, work-life balance, family or gender attitudes.

The GGS Teaching Dataset is intended for student training and teaching purposes. For scientific publications and theses, the original GGS datasets are more appropriate, as the Teaching Dataset contains several simplifications and additional anonymization that make it unsuitable for scientific use. The main datasets are available from the GGP website.

The Teaching Dataset is available for different data management and analysis platforms and can be opened directly by the most common software packages: STATA, SPSS, and Excel.

1.3. Two datasets: starter and advanced

Currently there are two versions of the teaching dataset available: the starter and the advanced dataset. The starter dataset is for introductory statistics classes and the advanced one is for more advanced analysis.

Both datasets are cross-sectional. They cover the same countries and the same number of respondents, but the advanced dataset has more variables.

The main difference between them is that only the advanced dataset includes information on the timing of life events. By including life histories (e.g. fertility history, partnership history), retrospective data analysis techniques can be practiced. The extended dataset is particularly useful for learning and teaching survival/event history analysis.

1.4. Types of skills and techniques that students can learn

Both the starter and the advanced datasets are useful for teaching, learning and practicing the following skills and techniques using real survey data:

- Basic use of a statistical software (describe data and variables, frequencies, dealing with different types of variables, missing data etc.)
- Manipulation of variables: recoding, renaming, creating scales, etc.
- Using a 'weight' variable
- Cross tabulation (contingency tables)
- Linear regression models
- Logistic regression models

In addition to the above statistical methods and data management skills, the advanced dataset offers additional opportunities for users to learn and practice:

- Working with dates
- Survival analysis techniques

2. Data access

2.1. How to download

The data is available for download via the following link: https://ggp.colectica.org/Item/int.ggp/0540f022-9fb7-470c-be95-9d183a28abcb/2 with a one-click procedure. Each of the two data files – the starter and the advanced – has a dedicated page where users can browse the data and download the data file. Each data file is downloaded in a zip format with the data in SPSS, Stata, and Excel formats.

All the countries are appended in the same files, so that the data is ready for cross-country comparison and there is no need for additional post-harmonization. However, users need to be aware that there are a few variables that are not available in all countries (see Table A1 in the Appendix).

The GGS Teaching Dataset is intended for teaching and training purposes only, not suggested for scientific publications, and any form of commercial use is not permitted. For scientific publications and theses, we recommend using the original GGS micro data, which are available from the GGP Data Portal after registration, filing a request and signing the data agreement.

2.2. Supporting materials

There is a dedicated page on the GGP website for the teaching datasets (https://www.ggp-i.org/ggs-teaching-datasets/). This page provides a detailed explanation of the teaching datasets and includes video tutorials that guide users on how to use the data effectively. In addition, the page includes a link to an online codebook where users can browse detailed variable-level information about the dataset.

3. The structure of the data

3.1. Basic structure of the data

The current version of the data (v2.0) covers 12 countries and over 76,000 respondents. The number of variables is 88 in the starter dataset and 124 in the advanced dataset. In order to have the same age range for each country, respondents aged 18–49 are included.

All countries are merged into a single harmonised file. Variables are ordered thematically in the data. The advanced dataset includes all the variables of the starter dataset and some additional variables.

3.2. Selection of variables

The following general principles guided variable selection:

- Few missing cases
- No complex filtering
- Cross-country comparability (no country-specific variables or coding)
- Only cross-sectional information
- Keep it simple but suitable for statistical analysis

Data is based on the Generations and Gender Survey Baseline Questionnaire Version 3.1.1 (Gauthier et al. 2021). Both original and constructed variables are included. All variables have been renamed and some labels also differ from the original GGS datasets.

Basic respondent and household information is included and can be used as independent variables in analyses. Variables of interest (potential dependent variables) are centred on thematic focal points. These include basic demographics, partnership quality, division of household work, fertility intentions and ideals, mental health, working

conditions, family and gender attitudes. Both datasets include a weighting variable and the advanced dataset contains dates of past events (notably, fertility and partnership histories).

3.3. Countries in the dataset

Data collection for GGS-II is underway. Several countries have already implemented GGS-II, while others are preparing for data collection. Other countries will be added as fieldwork is completed and data are available.

Version v2.0 of the Teaching Dataset includes 12 countries. Fieldwork took place between 2020 and 2023 (Table 1).

Please note that – due to the large variation in the number of respondents per country – we recommend that the "country" variable is always included in the analysis.

Table 1: Countries included in the GGS Teaching Dataset v2.0

Country	Period of data collection	Mode	N
Argentina (Buenos Aires)	Aug 2022 – Dec 2022	CAPI	1,335
Austria	Oct 2022 – Mar 2023	CAWI	6,027
Croatia	May 2023 – Jul 2023	CAWI	6,351
Czech Republic	Oct 2020 – Jul 2022	CAWI, CAPI	3,647
Denmark	Mar 2021 – Jul 2021	CAWI	8,142
Estonia	Oct 2021 – Feb 2022	CAWI	6,838
Finland	Oct 2021 – Mar 2022	CAWI	3,062
Germany	2021–2022	CAWI, PAPI	21,579
Moldova	Jan 2020 – Dec 2020	CAPI	4,175
Netherlands	Oct 2022 – Nov 2023	CAWI	5,640
United Kingdom	Aug 2022 – Jan 2023	CAWI	5,683
Uruguay	Oct 2021 – Oct 2022	CAPI, CAWI	3,641

Notes: CAPI = computer-assisted personal interview, CAWI = computer-assisted web interview, PAPI = paper and pencil personal interview

3.4. Anonymization

A number of depersonalisation and anonymization procedures have been applied to the data to minimise the risk of identification (based on Baron et al. 2022).

1. All direct identifiers (names, addresses, telephone numbers, etc.) have been removed from the data.
2. Country is the only available geographical information on place of residence (information on region, settlement etc. have been removed).
3. The original personal identifier of the respondents has been removed and a new identifier variable has been created so that the Teaching Dataset cannot be merged with the GGS main country files.

4. The weight variable has been rounded to 2 decimal places so that it is not possible to identify which respondents have the same combination of values on the weighting variables (age, gender, region, education, and marital status).
5. Continuous objective variables have been aggregated (e.g. age of respondent, age of children) or truncated at the top (e.g. number of children, sisters or brothers).
6. Plain text variables (responses to open questions) are not included.
7. Response modalities with low (<5) frequencies have been identified, systematically reviewed, and removed or recoded. This includes the transformation of some continuous variables into categorical ones (e.g. age into age groups), reducing the number of values of certain categorical variables (e.g. marital status) or removing certain rare events from the life histories (e.g. dates of higher order births).
8. Random noise has been added to the dates of events, while preserving the order of events for each person. This means that the age of the respondent and variables recording age at certain events are also randomly modified.

3.5. Missing values

As with all survey data, the GGS has some missing values. (The only exceptions are gender and age, where respondents with no valid answers have been removed from the dataset.) The reason for a missing value may be refusal, don't know, not applicable (filtering), incomplete survey (when a respondent quits the web survey before reaching the final question), or the fact that the question was not asked in that particular country or for a subset of respondents.

Most missing values do not have specific codes in the Teaching Dataset, partly because of the difficulty of harmonizing them across countries. However, for some subjective variables (questions on fertility intentions and attitudes) the 'Don't know' missing code has been retained so that students can practise working with it.

4. Contents of the dataset: an overview

The dataset contains variables from almost all sections of the main GGS questionnaire, except for work history. The aim was to cover a wide range of topics in order to provide teachers and students with several options for analysis.

The following table gives an overview of the topics and variables included in the starter and the advanced datasets (Table 2). More details on each variable are available in the next section.

Table 2: Overview of variables

Variable name(s)	Label	Included in the…	
		starter dataset	advanced dataset
BASIC INFORMATION			
id	Respondent ID	1	1
country	Country of residence	1	1
weight	Weight	1	1
int_m, int_y	Interview date	0	1
gender	Gender	1	1
birth_m, birth_y	Month/year of birth of respondent	0	1
agegr5	Age group (5 years)	1	1
countrybirth	Born in country	1	1
EDUCATION AND WORK			
edu	Highest level of education	1	1
wrk_employment	Employment status	1	1
wrk_activity	Activity status	1	1
wrk_dummy	In paid work last week	1	1
wrk_occupation	Occupation ISCO major groups	1	1
wrk_time	Work time	1	1
wrk_hours	Hours worked per week	1	1
wrk_home	Working from home	1	1
wrk_evening	Evening work	1	1
wrk_weekend	Weekend work	1	1
wrk_life_bal1-4	Work-life balance	1	1
CHILDREN AND FERTILITY			
ch_nb	Number of children of respondent	1	1

Variable name(s)	Label	Included in the…	
		starter dataset	advanced dataset
ch_dummy	Respondent has children (dummy)	1	1
chx_m, chx_y	Month/Year of birth of child x	0	1
chx_ager	Age group of respondent at the birth of child x	1	1
ch_youngest	Age of the youngest child	1	1
fer_int1	Intention to have a child in next 3y	1	1
fer_int2	Intention to have a child at all	1	1
fer_int3	Total number of children intended	1	1
fer_int4	General ideal family size	1	1
fer_int5	Personal ideal family size	1	1
fer_int_d	Intends to have children	1	1
ch_impact1 – ch_impact8	Child impact	1	1
PARTNERSHIP			
par_stat	Partnership status	1	1
par_dummy	Respondent has a coresident partner/spouse	1	1
par_met	Place first met current partner	1	1
mar_stat	Marital status	1	1
mar_dummy	Respondent is married	1	1
par_sat	Satisfaction with relationship	1	1
par_age	Age group at first cohabitation or marriage	1	1
pstart_m, pstart_y	Date started living with current partner/spouse (month/year)	0	1
pmar_m, pmar_y	Date of marriage with current spouse (month/year)	0	1
par_nb	Number of past partnerships (marriage or cohabitation)	0	1
pstart_mx, pstart_yx	Date of starting living together with partner x (month/year)	0	1
pmar_x	Married to partner x	0	1
pend_mx, pend_yx	End of relationship with partner x (month/year)	0	1
HOUSEHOLD			
hh_type	Household type	1	1
hh_alone	Living alone in household	1	1
hw1 – hw6	Division of household tasks	1	1
hw_sat	Household tasks satisfaction	1	1

Variable name(s)	Label	Included in the… starter dataset	Included in the… advanced dataset
GENERATIONS			
parents_mar	Biological parents ever got married	1	1
parents_mar_m, parents_mar_y	Date of parents' marriage (month/year)	0	1
parents_end	Biological parents ever broke up	1	1
parents_end_m, parents_end_y	Date of parents' (first) break-up (month/year)	0	1
brothers	Number of brothers ever born	1	1
sisters	Number of sisters ever born	1	1
alive_m	Biological mother alive	1	1
alive_f	Biological father alive	1	1
alive_gp	Number of grandparents alive	1	1
WELLBEING AND ATTITUDES			
lonely1 – lonely6	Loneliness	1	1
happiness	Happiness scale	1	1
subinc	Subjective income	1	1
att_fam1 – att_fam8	Family values	1	1
att_gender1 – att_gender5	Gender importance	1	1
att_work1, att_work2	Ideal work hours	1	1
relig	Religiosity	1	1

5. Description of variables

This section provides a brief description of all the variables in the Teaching Dataset. Variable names and labels, value labels (if there are any), question wording (as the question appeared in the main GGS questionnaire), filtering and any additional information on missing values or variable construction are listed below.

Variables are listed in the order in which they appear in the datasets. We also indicate whether a variable is available in both datasets or only in the advanced dataset. Table A1 in the Appendix show the availability of each variable by country.

5.1. Identifiers and basic demographic information

id	Respondent ID	Both datasets

An integer – running from 1 – that is unique to each respondent. This is different from the ID variable that the main GGS data files use.

country	Country of residence	Both datasets

The country where the interview was fielded.

Values:

1 Argentina (Buenos Aires)
2 Austria
3 Croatia
4 Czech Republic
5 Denmark
6 Estonia
7 Finland
8 Germany
9 Moldova
10 Netherlands
11 United Kingdom
12 Uruguay

weight	Weight	Both datasets

Post-stratification weigh is produced using Iterative Proportional Fitting based on the most recent and reliable information on population figures provided by the country teams on five items: age, gender, region, level of education, and marital status. This accounts for selectivity in response, making within-country and cross-country-comparative research more reliable. The weights are normalized for each country (e.g. their mean is close to 1).

int_m	Interview date (month)	Advanced dataset
int_y	Interview date (year)	Advanced dataset

These variables contain the month and the year in which the interview took place.

gender	Gender of respondent	Both datasets

Question: "What is your gender?"
Respondents who choose the "other" category and where information was missing have been deleted from the dataset due to low cell sizes.

Values:

1 Male
2 Female

birth_m	Month of birth of respondent	Advanced dataset
birth_y	Year of birth of respondent	Advanced dataset

Question: "When were you born? [MM/YYYY]"

These variables contain the month and the year in which the respondent was born. The few respondents with missing information on the year have been deleted from the dataset. Missing months have been imputed with a random value.

agegr5	Age groups (5 years)	Both datasets

Age of respondent in years, calculated from monthly information on respondent's date of birth and interview date, then recoded to 5-year categories.

Values:

1 18–24
2 25–29
3 30–34
4 35–39
5 40–44
6 45–49

countrybirth	Country of birth (dummy)	Both datasets

Question: "Were you born in this country?"

Values:

1 Yes
2 No

5.2. Education and work

edu	Highest level of education	Both datasets

Question: "What is the highest level of education you have completed?"
Responses are coded according to the International Standard Classification of Education, ISCED 2011, levels of education, aggregate

levels[1]. Low education = ISCED 0–2, medium education = ISCED 3–4, high education = ISCED 5–8.

Values:

1 Low
2 Medium
3 High

wrk_activity Activity status	Both datasets

Questions: "Which of the items best describes your current employment status?"

Values:

1 In education or training
2 Employed or self-employed
3 Unemployed
4 On child related leave
5 Other inactive

wrk_employment Employment status	Both datasets

Questions: "Have you been in paid work in the last week?
'In paid work' refers to someone who either:
a) worked for pay, profit or family gain for at least one hour in the last week or;
b) were not at work during the last week but has a job or business from which they were temporarily absent (i.e. on holiday or some form of leave)."
Further questions related to work are only asked to those who have answered 'In paid work'.

Values:

[1] For more information, see: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)#Background (visited on 20/09/2024).

1 In paid work
2 No in paid work but looking for work
3 Not in paid work and not looking for work

wrk_dummy	In paid work last week	Both datasets

The variable has been constructed from wrk_employment.
Further questions related to work are only asked to those who have answered 'yes' to wrk_dummy.

Values:
0 No
1 Yes

wrk_occupation	Occupation ISCO major groups	Both datasets

Question: "What is your current occupation? […] If you are engaged in two or more jobs or businesses, please consider only the one in which you spend most of your working hours."
ISCO codes for occupation (major groups) are used.

Values:
1 Managers
2 Professionals
3 Technicians and Associate Professionals
4 Clerical Support Workers
5 Service and Sales Workers
6 Skilled Agricultural, Forestry and Fishery Workers
7 Craft and Related Trades Workers
8 Plant and Machine Operators, and Assemblers
9 Elementary Occupations

wrk_time	Work time	Both datasets

Question: "Is your work full time or part time?"

Values:

0 Part time
1 Full time

wrk_hours Hours worked per week	Both datasets

Question: "How many hours per week do you normally work in this job or business including overtime?"

Values:

1 0–20
2 21–40
3 More than 40

wrk_home Work from home	Both datasets

Question: "Thinking about the last four weeks, did you do any work at home, including using internet for professional purpose, checking emails, having professional phone calls?"

Values:

1 Yes, twice or more per week
2 Yes, less than twice per week
3 No

wrk_evening Evening work	Both datasets

Question: "Thinking about the last four weeks, did you ever work for at least 2 hours in the evening or at night (between 8 p.m. and 5 a.m.)?"

Values:

1 Yes, twice or more per week
2 Yes, less than twice per week
3 No

wrk_weekend Weekend work Both datasets

Question: "Thinking about the last four weeks, did you work on Saturdays or Sundays?"

Values:

1 Yes, twice or more in the last four weeks
2 Yes, less than twice in the last four weeks
3 No

wrk_life_bal1 Work balance: Too tired to do chores Both datasets

Question: "How often has each of the following happened to you during the past three months?
I have come home from work too tired to do the chores that need to be done."

Values:

1 Several times a week
2 Several times a month
3 Once or twice a month
4 Never

wrk_life_bal2 Work balance: Difficult to fulfil family responsibilities Both datasets

Question: "How often has each of the following happened to you during the past three months?
It has been difficult for me to fulfil my family responsibilities because of the amount of time I spent on my job."

Values:

1 Several times a week
2 Several times a month
3 Once or twice a month
4 Never

wrk_life_bal3 Work balance: Too tired to function at work Both datasets

Question: "How often has each of the following happened to you during the past three months?
I have arrived at work too tired to function well because of the household work I have done."

Values:

1 Several times a week
2 Several times a month
3 Once or twice a month
4 Never

wrk_life_bal4 Work balance: Hard to concentrate because of family responsibilities Both datasets

Question: "How often has each of the following happened to you during the past three months?
I have found it difficult to concentrate at work because of my family responsibilities."

Values:

1 Several times a week
2 Several times a month
3 Once or twice a month
4 Never

5.3. Children and fertility

ch_nb	Number of children of respondent	Both datasets

The total number of biological and adopted children the respondent has had.
More than 4 children are recoded to 4+ due to small cell sizes.

Values:

…
4 4 or more

ch_dummy	Respondent has had children	Both datasets

The variable has been constructed from ch_nb.

Values:

0 No
1 Yes

ch'x'_m	Date of birth of child 'x' (month)	Advanced dataset
ch'x'_y	Date of birth of child 'x' (year)	Advanced dataset

Question: "When was the child born? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked to respondents who have had children. Missing months have been imputed with a random value. Implausible years have been set to missing (dates before the birth of the respondent or after the interview). Only the first four children are included due to small cell sizes at higher parities.

ch'x'_ager Age of respondent at the birth of child 'x'	Both datasets

Age of respondent in years at the birth of the first four children. The coding is different for the first child. Ages less than 14 have been recoded to 14.

Values for the first child:

1 14-24
2 25-29
3 30-34
4 35+

Values for the second, third, and fourth children:

1 14-29
2 30-34
3 35+

ch_youngest Age group of the youngest child	Both datasets

Age of the youngest child of the respondent (in years, grouped into 5-year categories).

Values:

1 0-4
2 5-9
3 10-14
4 15-19
5 20+

fer_int1 Intention to have a child in the next 3y	Both datasets

Question: "Do you intend to have a/another child during the next three years? Please take into account only biological children".

Values:

1 Definitely not
2 Probably not
3 Unsure
4 Probably yes
5 Definitely yes
6 Currently expecting a child

fer_int2	Intention to have a child at all	Both datasets

Question: "Supposing you do not have a/another child during the next three years, do you intend to have any (more) children at all?"
This question was not asked to respondents who were currently expecting a child.

Values:

1 Definitely not
2 Probably not
3 Unsure
4 Probably yes
5 Definitely yes

fer_int3	Total number of children intended	Both datasets

Question: "How many more children – including biological and adoptive children – do you intend to have overall? [Not including existing children]"
This question was not asked to respondents who answered "definitely not" to the previous question (fer_int2).

Values:

…
4 4 or more
99 Don't know

fer_int4	General ideal family size	Both datasets

Question: "Generally speaking, what do you think is the ideal number of children for a family?"

Values:

…
4 4 or more
99 Don't know

fer_int5	Personal ideal family size	Both datasets

Question: "For you personally, what would be the ideal number of children you would like to have or would have liked to have had?"

Values:

…
4 4 or more
99 Don't know

fer_int_d	Intention to have children (dummy)	Both datasets

Dummy variable constructed from fer_int3.

Values:

0 No
1 Yes

ch_impact1	Child Impact: Do what you want	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I

would like you to tell me what effect you think this would have on various aspects of your life.
The possibility to do what you want"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact2	Child Impact: Money to spend	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
The amount of money you can spend"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse

5 Much worse
99 Don't know

ch_impact3	Child Impact: Realize other goals	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
The possibility to realize other goals in life"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:
1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact4	Child Impact: Joy from life	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
The joy and satisfaction you get from life"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.

The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact5	Child Impact: Employment opportunities	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
Your employment opportunities"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact6	Child Impact: Partners work opportunity	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
Your partner's employment opportunities"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact7 Child Impact: Security in old age	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
The care and security you may get in old age"
Question was not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

ch_impact8	Child Impact: Closeness with spouse	Both datasets

Question: "Even though you might not intend to have a/another child, we would still want your opinion about this possibility. Suppose that during the next 3 years you were to have a/another child. I would like you to tell me what effect you think this would have on various aspects of your life.
The closeness between you and your partner"
Question not asked to those who said that they definitely didn't intend to have any (more) children.
The items on the impact of having a child were optional and were not asked in a number of countries.

Values:

1 Much better
2 Better
3 Neither better nor worse
4 Worse
5 Much worse
99 Don't know

5.4. Partnership

par_stat	Partnership status	Both datasets

The variable has been constructed from three yes-or-no questions:
"Do you have a partner at the moment?"
"Are you and your partner legally married?"
"Does your partner live with you in the same household?"
Both opposite and same sex relationships are considered. LAT = living apart together relationship.

Values:
1 Living with spouse
2 Cohabitation
3 LAT
4 No partner

par_dummy	Respondent has a coresident partner/spouse	Both datasets

The variable has been constructed from par_stat (those living with spouse or cohabiting have been coded as "yes").

Values:
0 No
1 Yes

par_met	Place first met current partner	Both datasets

Question: "How did you and your partner meet?"
It was asked of respondents living in either marriage, cohabitation or LAT relationship. Answer categories have been simplified.

Values:
1 Through work
2 In education (school, university, college etc.)
3 Online
4 At a bar, nightclub or dance club

5 Through a social organization, health club, gym, church or volunteer group
6 At a private party, social event or trip
7 Through friends or family
8 Other

mar_stat	Marital status	Both datasets

The variable has been constructed using the question "Are you and your partner legally married?" and information on past partnerships (marriages and divorces).
Legal marital status. Registered partnerships are not considered. Divorced and widowed respondents are grouped together in the "previously married" category due to low cell sizes.

Values:
1 Never married
2 Married
3 Previously married

mar_dummy	Respondent is married (dummy)	Both datasets

The variable has been constructed from mar_stat (married respondents have been recoded to "yes").

Values:
0 No
1 Yes

par_sat	Satisfaction with relationship	Both datasets

Question: "How satisfied are you with your relationship with your partner? On a scale from 0 to 10 where 0 means 'not at all satisfied' and 10 means 'completely satisfied' and 5 means 'about average',

what number best represents your satisfaction with your relationship?"
It has been asked of respondents living in either marriage, cohabitation or LAT relationship.

Values:

0 Not at all satisfied
…
5 About average
…
10 Completely satisfied

par_age	Age at first cohabitation or marriage	Both datasets

Age of respondent (in years, grouped into 9 categories) at the beginning of the first cohabitation or marriage. Ages less than 14 have been recoded to 14.

Values:

1 14-19
2 20-21
3 22-23
4 24-25
5 26-27
6 28-29
7 30-34
8 35-39
9 40-49

pstart_m	Date started living with current partner/spouse (month)	
pstart_y	Date started living with current partner/spouse (year)	Advanced dataset

Question: "When did you and he/she first start living together?

[MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of respondents who lived in marriage or cohabitation.
Missing months have been imputed with a random value. Implausible years have been set to missing (dates before the birth of the respondent or after the interview).

pmar_m	Date of marriage with current spouse (month)
pmar_y	Date of marriage with current spouse (year)
	Advanced dataset

Question: "When did you marry? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of married respondents.
Missing months have been imputed with a random value. Implausible years have been set to missing (dates before the birth of the respondent or after the interview).

par_nb	Number of past partnerships (marriage or cohabitation)
	Advanced dataset

Question: "Not including your current relationship, how many partnerships did you have where you lived together?"
The question was asked of respondents who lived in marriage or cohabitation.

pstart_m'x' Date of starting living together with partner 'x' (month)
pstart_y'x' Date of starting living together with partner 'x' (year)
Advanced dataset

Question: "When did you start living together with …? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of respondents who had lived in marriage or cohabitation before the current partnership.
Missing months have been imputed with a random value. Implausible years have been set to missing (dates before the birth of the respondent or after the interview).
Only the first three partners are included due to small cell sizes at higher order unions.

pmar_'x' Married to partner 'x' Advanced dataset

Question: "Were you and … legally married?"
The question was asked of respondents who had lived in marriage or cohabitation before the current partnership.
Only the first three partners are included due to small cell sizes at higher order unions.

pend_m'x' End of relationship with partner 'x' (month) Advanced dataset
pend_y'x' End of relationship with partner 'x' (year)Advanced dataset

Question: "When did that happen? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of respondents who had lived in marriage or cohabitation before the current partnership.
Missing months have been imputed with a random value. Implausible years have been set to missing (dates before the birth of the respondent or after the interview).

Only the first three partners are included due to small cell sizes at higher order unions.

5.5. Household and housework

hh_type	Household type	Both datasets

Constructed variable describing the type of household of the respondent.

Values:

1 Living alone
2 Single parent
3 Couple with no children
4 Couple with children
5 Other

hh_alone	Living alone in household	Both datasets

The variable has been constructed from hh_type.

Values:

0 No
1 Yes

hw_1	Housework: Preparing meals	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Preparing daily meals"
Question only asked if there was a coresident partner.

Values:

1 Always me

2 Usually me
3 Equally me and partner
4 Usually partner
5 Always partner
6 Always or usually someone else

hw_2	Housework: Vacuuming	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Vacuum cleaning the house"
Question was only asked if there was a coresident partner.

Values:
1 Always me
2 Usually me
3 Equally me and partner
4 Usually partner
5 Always partner
6 Always or usually someone else

hw_3	Housework: Doing laundry	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Doing the laundry"
Question wasconly asked if there was a coresident partner.

Values:
1 Always me
2 Usually me
3 Equally me and partner

4 Usually partner
5 Always partner
6 Always or usually someone else

hw_4	Housework: Small repairs	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Doing small repairs in and around the house"
Question was only asked if there was a coresident partner.

Values:
1 Always me
2 Usually me
3 Equally me and partner
4 Usually partner
5 Always partner
6 Always or usually someone else

hw_5	Housework: Finances	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Paying bills and keeping financial records"
Question was only asked if there was a coresident partner.

Values:
1 Always me
2 Usually me
3 Equally me and partner
4 Usually partner
5 Always partner
6 Always or usually someone else

hw_6	Housework: Social	Both datasets

Question: "The next questions are about who does what in your household. Please indicate who does the following tasks in your household.
Organising joint social activities"
Question was only asked if there was a coresident partner.

Values:

1 Always me
2 Usually me
3 Equally me and partner
4 Usually partner
5 Always partner
6 Always or usually someone else

hw_sat	Satisfaction with housework	Both datasets

Question: ""How satisfied are you with the division of household tasks between you and your partner? On a scale from 0 to 10 where 0 means 'not at all satisfied' and 10 means 'completely satisfied' and 5 means 'about average', what number best represents your satisfaction with the division of household tasks?"
Question was only asked if there was a coresident partner.

Values:

0 Not at all satisfied
…
5 About average
…
10 Completely satisfied

5.6. Generations

parents_mar	Biological parents ever got married	Both datasets

Question: "Did your biological parents ever get married?"

Values:
0 No
1 Yes

parents_mar_m	Date of marriage of biological parents (month)	
parents_mar_y	Date of marriage of biological parents (year)	
		Advanced dataset

Question: "When did they get married? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of respondents whose biological parents ever got married to each other (who answered "yes" to parents_mar).
Missing months have been imputed with a random value. Implausible years have been set to missing (dates after the interview).

parents_end	Biological parents ever broke up	Both datasets

Question: "Did your biological parents ever break up?"

Values:
0 No
1 Yes
8 Not applicable, never together

parents_end_m	Date of biological parents' (first) break-up (month)	
parents_end_y	Date of biological parents' (first) break-up (year)	
		Advanced dataset

Question: "When did that first happen? [MM/YYYY] If you are unsure of the precise date, please provide your best estimate."
The question was asked of respondents whose biological parents ever broke up (who answered "yes" to parents_end).
Missing months have been imputed with a random value. Implausible years have been set to missing (dates after the interview).

brothers	Number of brothers ever born	Both datasets

Question: "How many brothers do you have? Including those who are deceased."
More than 4 brothers are recoded to 4 due to small cell sizes.

Values:

…
4 4 or more

sisters	Number of sisters ever born	Both datasets

Question: "How many sisters do you have? Including those who are deceased."
More than 4 sisters are recoded to 4 due to small cell sizes.

Values:

…
4 4 or more

alive_m	Biological mother alive	Both datasets

Question: "Is your biological mother still alive?"

Values:

0 No, not alive anymore
1 Yes, still alive
99 Don't know

alive_f	Biological father alive	Both datasets

Question: "Is your biological father still alive?"

Values:
0 No, not alive anymore
1 Yes, still alive
99 Don't know

alive_gp	Number of grandparents alive	Both datasets

Question: "How many of your grandparents are alive?"
More than 4 grandparents are recoded to 4 due to small cell sizes.

Values:
4 4 or more
99 Don't know

5.7. Wellbeing and attitudes

lonely1	Loneliness: People to lean on	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently.
There are plenty of people I can rely on when I have problems."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:
1 Yes
2 More or less
3 No

lonely2	Loneliness: General sense of emptiness	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently. […]
I experience a general sense of emptiness."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:

1 Yes
2 More or less
3 No

lonely3	Loneliness: Miss having people around	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently. […]
I miss having people around."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:

1 Yes
2 More or less
3 No

lonely4	Loneliness: May people I can trust	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently. […]
There are many people I can trust completely."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:

1 Yes
2 More or less
3 No

lonely5	Loneliness: Feel rejected	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently. […]
Often, I feel rejected."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:
1 Yes
2 More or less
3 No

lonely6	Loneliness: Enough people I feel close	Both datasets

Question: "The next six statements are about your current experiences. Please indicate for each of them to what extent they have applied to you recently. […]
There are enough people that I feel close to."
Variables lonely1 to lonely6 can be used to create a loneliness scale.

Values:
1 Yes
2 More or less
3 No

happiness	Happiness scale	Both datasets

Question: "Taking all things together, how happy would you say you are? Please note that 0 means 'extremely unhappy' and 10 means 'extremely happy'"

Values:

0 Extremely unhappy

…

10 Extremely happy

subinc	Subjective income	Both datasets

Question: "A household may have different sources of income and more than one household member may contribute to it. Thinking of your household's total monthly income, is your household able to make ends meet with great difficulty, with difficulty, with some difficulty, fairly easily, easily or very easily?"

Values:

1 With great difficulty
2 With difficulty
3 With some difficulty
4 Fairly easily
5 Easily
6 Very easily

att_fam1	Values: Marriage outdated	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? Marriage is an outdated institution."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam2	Values: Unmarried cohabitation	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? It is alright for a couple to live together without getting married."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam3	Values: Divorce is permissible	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? It is all right for a couple with an unhappy marriage to get a divorce even if they have children."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam4	Values: Women need children	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? A woman has to have children in order to be

fulfilled."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam5	Values: Child needs a father and mother	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? A child needs a home with both a father and a mother to grow up happily."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam6	Values: Men need children	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? A man has to have children in order to be fulfilled."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree

99 Don't know

att_fam7	Values: Homosexual couple rights	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? Homosexual couples should have the same rights as heterosexual couples do."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_fam8	Values: Child suffers if mother works	Both datasets

Question: "To what extent do you agree or disagree with each of the following statements? A pre-school child is likely to suffer if his/her mother works."

Values:

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
99 Don't know

att_gender1	Gender importance: Political leaders	Both datasets

Question: "The next questions are about the roles of men and women. On the whole, who would make better political leaders, men

or women?"

Values:

1 Men definitely
2 Men slightly
3 Both sexes equally
4 Women slightly
5 Women definitely
99 Don't know

att_gender2	Gender importance: University	Both datasets

Question: "For whom is a university education more important, men or women?"

Values:

1 Men definitely
2 Men slightly
3 Both sexes equally
4 Women slightly
5 Women definitely
99 Don't know

att_gender3	Gender importance: Job	Both datasets

Question: "For whom is having a job more important, men or women?"

Values:

1 Men definitely
2 Men slightly
3 Both sexes equally
4 Women slightly
5 Women definitely

99 Don't know

att_gender4	Gender importance: Childcare	Both datasets

Question: "For whom is looking after the home and children more important, men or women?"

Values:

1 Men definitely
2 Men slightly
3 Both sexes equally
4 Women slightly
5 Women definitely
99 Don't know

att_gender5	Gender importance: Small children	Both datasets

Question: "Who are better at caring for small children, men or women?"

Values:

1 Men definitely
2 Men slightly
3 Both sexes equally
4 Women slightly
5 Women definitely
99 Don't know

att_work1	Ideal work hours for mothers	Both datasets

Question: "Consider a family with a mother, father and a two-year old child. How many hours a week should the mother work?"

Values:

0 Not at all
…
999 Don't know

att_work2	Ideal work hours for fathers	Both datasets

Question: "Consider a family with a mother, father and a two-year old child. How many hours a week should the father work?"

Values:
0 Not at all
…
999 Don't know

relig	Religiosity	Both datasets

Question: "Regardless of whether you belong to a particular religion, how religious would you say you are? Please express your religiosity on a scale of 0 to 10 where 0 means 'Not at all religious' and 10 means 'Very religious'."

Values:
0 Not at all religious
…
10 Very religious
99 Don't know

6. Bibliography

Baron, J., Lenoir, J., Caporali, A. and Merly-Alpa, T. (2022): *Survey data anonymisation and pseudononymisation*. Aubervilliers, Ined Datalab. http://hdl.handle.net/20.500.12204/AX_gorcGBnm4X3q6Cn49

Gauthier, A. H., Liefbroer, A., Ajzen, I., Aassve, A., Beets, G., Billari, F., Bühler, C., Bujard, M., Cabaço, S., Corijn, M., Désesquelles, A., Dommermuth, L., Dykstra, P., Emery, T., Fadel, L., Fokkema, T., Hansen, T., Hlebec, V., Hoem, J., Klobas, J., Kogovšek, T., Koops, J. C., Kveder, A., Lappegård, T., Lück, D., Lugtig, P., MacDonald, A., Macura, M., Makay, Z., Mills, M. C., Murinkó, L., Mynarska, M., Neyer, G., Pailhé, A., Petrič, G., Pinnelli, A., Ratikainen, J., Rayboud, A., Rijken, A., Slagsvold, B., Solaz, A., Spéder, Z., Thévenon, O. and Vikat, A. (2021). *Generations and Gender Survey Baseline Questionnaire 3.1.1.* Netherlands Interdisciplinary Demographic Institute [distributor]. https://www.ggp-i.org/wp-content/uploads/2022/07/BaselineQuestionnaire_3.1.1.pdf

Gauthier, A. H., Kong, S., Grünwald, O., Bujard, M., Caporali, A., Deimantas, V. J., Emery, T., Jablonski, W., Koops, J., Rijken, A. and Schumann, A. (2023). *Data Brief: The Generations and Gender Survey second round (GGS-II).* https://hal.science/hal-04384243

7. Appendix

Table A1: Variable availability (v2.0)

Variable name(s)	Argentina/ Buenos Aires	Austria	Croatia	Czech Republic	Denmark	Estonia	Finland	Germany	Moldova	Netherlands	Uruguay	UK
id												
country												
weight												
int_m, int_y												
gender												
birth_m, birth_y												
agegr5												
countrybirth						p						NA
edu												
wrk_activity												
wrk_employment												
wrk_dummy												
wrk_occupation												
wrk_time												NA
wrk_hours												
wrk_home												
wrk_evening												
wrk_weekend			NA									
wrk_life_bal1-4												
ch_nb												
ch_dummy												
chx_m, chx_y												
ch1_ager – ch4_ager												
ch_youngest												
fer_int1												
fer_int2												
fer_int3												
fer_int4												

Variable name(s)	Argentina/ Buenos Aires	Austria	Croatia	Czech Republic	Denmark	Estonia	Finland	Germany	Moldova	Netherlands	Uruguay	UK
fer_int5												
fer_int_d												
ch_impact1		NA			NA	NA	NA	NA	NA			NA
ch_impact2		NA			NA	NA	NA	NA	NA			NA
ch_impact3		NA			NA	NA	NA	NA	NA			NA
ch_impact4		NA			NA	NA	NA	NA	NA			NA
ch_impact5					NA	NA	NA	NA	NA			NA
ch_impact6					NA	NA	NA	NA	NA			NA
ch_impact7		NA			NA	NA	NA	NA	NA		NA	NA
ch_impact8		NA			NA	NA	NA	NA	NA		NA	NA
par_stat												
par_dummy												
par_met												
mar_stat												
mar_dummy												
par_sat												
par_age												
pstart_m, pstart_y												
pmar_m, pmar_y												
par_nb												
pstart_mx, pstart_yx												
pmar_x												
pend_mx, pend_yx												
hh_type												
hh_alone												
hw1 – hw6												
hw_sat												
parents_mar												
parents_mar_m, parents_mar_y												
parents_end												
parents_end_m, parents_end_y												

Variable name(s)	Argentina/ Buenos Aires	Austria	Croatia	Czech Republic	Denmark	Estonia	Finland	Germany	Moldova	Netherlands	Uruguay	UK
brothers												
sisters												
alive_m												
alive_f												
alive_gp												
lonely1 – lonely6												
happiness												
subinc												
att_fam1												
att_fam2												
att_fam3												
att_fam4												
att_fam5												NA
att_fam6												
att_fam7												
att_fam8												
att_gender1												
att_gender2												NA
att_gender3												
att_gender4												
att_gender5												
att_work1, att_work2												
relig												

Notes: NA = not available