The Estonian Generations and Gender Survey 2020: Experience with large-scale web-based data collection

ALLAN PUUR[1]
LUULE SAKKEUS[1]
LIILI ABULADZE[1,2]
MARK GORTFELDER[1]
MARTIN KLESMENT[1]
LEEN RAHNU[1]
TIINA TAMBAUM[1]

1 Tallinn University, Estonian Institute for Population Studies
2 Väestöliitto, Population Research Institute

Abstract

In Estonia, the Generations and Gender Survey 2020 (GGS-II) is the third large-scale demographic survey that collects data on family and fertility dynamics. As the country participates in the Survey of Health, Ageing and Retirement in Europe, the GGS-II opted for a shorter age range of the sample (18–59). The questionnaire in the GGS-II in Estonia follows the GGS-II wave 1 baseline questionnaire. The questionnaire also includes the Global Uncertainties' module developed by the Nordic countries, a battery of questions on the perceived impact of COVID-19, and several country-specific items. The GGS-II in Estonia was implemented using only computer-assisted web interviewing (CAWI). In this article, we present a concise overview of the sampling and data collection process, analyse representativeness and response rates, and briefly assess the data quality. We conclude that despite low response rates, the GGS-II provides a good basis for the analysis of fertility and family dynamics.

Introduction

In Estonia, the Generations and Gender Survey 2020 (GGS-II) is the third large-scale demographic survey that collects data on family and fertility dynamics in the country. The GGS-II was preceded by a survey conducted in 1994 (women) and 1997 (men) in the context of the Family and Fertility Survey programme coordinated by the United Nations Economic Commission for Europe. Estonia also participated in the first round of the Generations and Gender Survey (GGS-I), coordinated since 2009 by the Netherlands Interdisciplinary Demographic Institute (NIDI). In Estonia, the GGS-I took place in 2004 and 2005; due to limited resources, the survey was limited to one wave. Publications on these earlier surveys include methodology reports (EKDK, 1995a; 1999), standard tabulations (EKDK, 1995b; 2008; 2009), analytical books (Katus et al., 2000; Katus et al., 2002), and numerous research articles.

In Estonia, preparations for GGS-II were initiated by the Estonian Institute for Population Studies (EIPS) at Tallinn University. In 2019, a consortium based on EIPS, family sociology researchers at University of Tartu, and the Ministry of Social Affairs was formed and the GGS-II was included in the national roadmap of research infrastructures (ETAG, 2019). For the technical preparation of the survey, the EIPS formed a working group, the members of which are the authors of this paper. Estonia also joined the consortium of the international Generations and Gender Programme (GGP).

In April 2020, funding for the data collection of the first wave of GGS-II in Estonia was obtained from Tallinn University for the period of three years. In 2021–2022, activities related to GGS-II received support from the Estonian Research Council (grant TT7). From 2022, activities of the GGS-II in Estonia are supported by Horizon2020 project GGP-5D.

This article provides a concise overview of data collection and quality of the data of the first panel wave of GGS-II in Estonia. Following the introduction, the second section briefly explains sampling design, techniques and frame. The third section provides information about the questionnaire and data collection procedures. The fourth section presents response and non-response rates. The fifth section discusses representativeness and data quality, followed by a summary.

Sampling design, techniques and frame

The standard age range of the GGS-II sample is 18–79 at the beginning of the first panel wave, but countries can opt for a shorter age range, if a survey on older adults exists in the country. As Estonia participates in the Survey of Health, Ageing and Retirement in Europe (SHARE), the working group tasked with the preparation of the survey opted for a shorter age range (18–59) in Estonia.

The target population of the GGS-II in Estonia consisted of all residents aged 18–59 on January 1, 2021. In terms of birth cohorts, this corresponds to individuals born in from 1961 to 2002. One individual in the sample was born on January 1, 2003 and be-

longs to the 2003 birth cohort. The target population comprised residents living both in non-institutional and institutional households. At the beginning of 2021, the size of the target population within this age range was 715,361 individuals, according to population estimates (Statistics Estonia, 2023).

The sample frame for the GGS-II is the population register that contains information about all legal residents in the country. The sample was drawn in August 2021 by the Information Technology and Development Centre (SMIT) of the Ministry of Interior, which is the government institution running the population register in Estonia. SMIT provides services for survey agencies that want to use population-based probability samples in their surveys.

The sampling unit was an individual. When drawing the probability sample from the population register, a one-stage procedure was applied with respondents being drawn from the register without first selecting higher-order sampling units. For practical reasons, the individuals who belonged to the sample of the pilot survey of the GGS-II were excluded from the sample frame before drawing the sample of the main survey. For unit selection, a random number generator was used. No stratification by gender, age, regions or other demographic characteristics was employed when drawing the sample.

The sample drawn from the population register comprised 32,003 individuals. The sample size was based on the expected response rate (25%) and the expected number of 8000 respondents in the first wave of the GGS-II. The expected response rate was based on the experience of other countries who had used a web-based data collection in their GGS-II surveys and the experience gained from the pilot survey. The register data on respondents came with names, contact information (e-mail addresses, phone numbers, and postal addresses) and core demographic characteristics such as gender, date of birth, marital status, education, and place of residence. Table 1 presents the breakdown of the sample by gender and five-year age groups.

Table 1. *Sample size by gender and five-year age groups, Estonian GGS-II*

Age group	Men	Women	Both
18–24	2,004	1,893	3,897
25–29	1,767	1,621	3,388
30–34	2,402	2,146	4,548
35–39	2,257	2,073	4,330
40–44	2,112	1,981	4,093
45–49	2,089	2,007	4,096
50–54	1,889	1,936	3,825
55–59	1,822	2,004	3,826
Total	16,342	15,661	32,003

Source: Estonian GGS-II, authors' calculations

Questionnaire, lessons from the pilot survey and fieldwork

The questionnaire of the GGS-II in Estonia follows the GGS-II wave 1 baseline questionnaire prepared by the international GGP consortium (Gauthier et al., 2021). It consisted of nine modules, with questions on demographics, partnership and fertility histories, family planning, household, intergenerational relations, well-being, work, income, and attitudes. The questionnaire also includes the Global Uncertainties' module developed and implemented by the Nordic countries, a battery of questions on the perceived impact of COVID-19 (similar to that used in GGS-II in Moldova), and several country-specific items, such as ethnicity and language, previous partners' country of birth and education, experience of intimate partner violence, hours of social media use, and other themes. A complete list of country-specific items included in the Estonian GGS-II questionnaire is available from the international GGP website (GGP, 2023).

The Estonian GGS-II working group prepared the questionnaire in two languages (Estonian and Russian). Most of the questions were translated from English by two expert translators and afterwards checked by the Estonian GGS-II working group members. The digital versions of the questionnaires were programmed by the GGP Central Hub. The pdf-files of questionnaires in both languages are available from the survey website (Estonian GGS-II, 2021a). The GGP Central Hub also provided the individual web links for all respondents. The answers provided by the respondents were stored directly in the database administered by the Central Hub.

Given the budget constraints, the GGS-II in Estonia was implemented using only computer-assisted web interviewing (CAWI). To test the digital questionnaires and survey procedures, a pilot survey was implemented in the period from April to June 2021. The target population, sampling frame, and methods were identical to those used in the main survey described above. The sample size was 12,165 in the pilot survey. The survey agency Kantar Emor carried out the data collection both in the pilot and the main survey of the GGS-II in Estonia. In the pilot survey, the overall response rate was 27.4%, the response based on completed interviews was 18.1%. The relatively large difference between the two rates may be due to the fact that in the pilot survey the respondents with incomplete answers were not specifically reminded to complete all modules. Data from the pilot survey are available to users through the Central Hub.

An important objective of the pilot survey was to obtain information about the effectiveness of incentives in increasing response rate. To this end, a test of different unconditional and conditional incentives for respondents was included in the pilot survey. The sample was randomly divided into three groups of equal size with slightly more than 4,000 respondents in each group. All members of Group 1 were offered an option of a free two-month access to online media (12 different newspaper and journals for up to five readers). Group 2 was given an option to participate in a lottery of 55 shopping gift cards with a value of 50 euros for responding to the survey. Members of Group 3 participated in a donation to two NGOs dealing with mental health of children and young people (unconditional incentive). The donation (1,300 euros) on behalf of the respondents was made

by the survey agency. Members of the third group who had responded to the survey also participated in a lottery of 32 shopping gift cards (conditional incentive).

The analysis found no evidence of a strong advantage of one incentive scheme over the others. However, more respondents in Groups 2 and 3 responded quickly, before receiving any reminders (Tambaum & Klesment, 2021). Based on results of the pilot survey and cost considerations, the working group chose a scheme combining both unconditional and conditional incentives for the main survey. Unconditional incentives included a joint donation (6,000 euros) to three NGOs dealing with child and mental health. Conditional incentives consisted of a lottery. Participants who responded to the survey had an opportunity to win one of four main prizes (à 2,000 euros), one of 20 shopping gift cards (à 300 euros), or one of 215 shopping gift cards (à 50 euros). In the pilot survey, the cost of incentives per potential respondent was 4.1 euros in Group 1, 0.40 euros in Group 2, and 0.72 euros in Group 3.

There were other lessons learned from the pilot survey. First, a small proportion of respondents (2.5%) had no valid e-mail address. They received invitations to participate in the survey by mail, but very few responded. Out of 300 people who were contacted by letter, only seven responded. Based on this experience, it was decided not to send invitations by mail in the main survey. Second, the working group decided that invitations to participate in the survey will be sent to all participants within one week after the beginning of the data collection. For a more even distribution of responses over the data collection period, in the pilot survey, invitations were sent in two batches: the first batch at the beginning of fieldwork period (in early April 2021) and the second batch one month later (in early May 2021). The analysis showed that in the second batch, the response rate was 5.3 percentage points lower than in the first batch. The concentration of media coverage of the survey during its launch period may explain the difference. Third, given that invitations were sent at the beginning of fieldwork period, it was decided to increase the number of electronic reminders from three in the pilot survey to five in the main survey. Also, two video clips were prepared for the main survey, links to which were attached to reminders. In one video clip, the Minister of Social Affairs addressed the potential respondents, explaining the objectives and importance of the survey. In another video clip, the respondents were addressed by two well-known entertainment artists. Furthermore, it was also agreed that participants belonging to subgroups with lower than average response rates would receive reminders by phone. In the pilot survey, phone reminders were used on a limited scale to test their effectiveness. Results of the test showed that the overall success rate of phone reminders was 17.3%, that is, less than one-fifth of those reminded by phone responded to the survey. Excluding cases where the survey agency failed to reach the respondent by phone (one third of the cases included in the test), the success rate increases to 27.2%. Finally, changes to the digital questionnaire based on experience of the pilot survey were very few.

The data collection of the main survey of GGS-II in Estonia started on October 25, 2021 and ended on March 3, 2022. Within the first week, all respondents with valid e-mail addresses in the population register received invitations from the survey agency. The invitation letter explained the aims and importance of the GGS-II, voluntariness of response, data protection, and incentives for respondents. Invitation letters were prepared

in two languages (Estonian and Russian); the language of letters sent to individual respondents was chosen based on the information available from the population register. Overall, 1,833 respondents (5.7%) had either missing or invalid e-mails in the database. These respondents were contacted by phone. The invitation was not sent to 611 respondents (1.9%) who had asked the survey agency to remove them from the list of respondents in previous surveys carried out by Kantar Emor.

The data collection was supported by a media campaign carried out both in Estonian and Russian languages. Members of the working group wrote articles and gave interviews to national and local newspapers, and presented the survey in several radio and TV programs, mostly in the first month of data collection (Estonian GGS-II, 2021b). During the two last months of the data collection (January and February 2022), radio and TV channels of the Estonian Public Broadcasting broadcasted brief information clips about the survey. There were also three information campaigns on Facebook.

Table 2 presents the monthly dynamics of web interviews during the period of data collection. It shows significant variation in the number of interviews across months. Although the data collection started in the last week of October, a quarter of all questionnaires (25.4%) were completed during this very first week. Nearly a third of questionnaires (30.2%) were completed in November, the first full month of data collection. The first reminders to respondents were sent in mid-November, and the second reminders were sent in early December.

Table 2. *Number of interviews by month of data collection, Estonian GGS-II*

Month	Number of interviews	Share of interviews, %	Cumulative share,%
October 2021	2,320	25.4	25.4
November 2021	2,766	30.2	55.6
December 2021	1,039	11.4	67.0
January 2022	1,737	19.0	86.0
February 2022	1,267	13.9	99.9
March 2022	17	0.2	100.0
Total	9,146	100.0	100.0

Source: Estonian GGS-II, authors' calculations

In December, the number of responses markedly decreased, with only 11.4% of the total interviews taking place in that month. It was around this time that the survey agency began to remind respondents about the survey by phone, in addition to further e-mail reminders, which would be sent to respondents in mid-January, early February, and late February, respectively. Telephone reminders targeted subgroups with lower than average

response rates (men and ethnic groups other than Estonians) and respondents who had answered only the first modules. In total, approximately 6,000 respondents (19%) received reminders by phone. An increase in the number of responses in January and February suggests that these additional efforts had a positive impact. Estonian-speaking women, i.e., the group with the highest response rate, did not receive any more reminders from January 2022. The number of responses in March was very small, as the data collection ended at the beginning of this month.

Response rates and non-response

The total number of responses to GGS-II in Estonia is 9,146 (Table 3). Of those, 8,212 (89.8%) are complete responses, i.e., the respondents answered all modules of the questionnaire. Complete responses may include situations where the respondent has a few questions unanswered (item non-response). The data collection also resulted in 934 partially complete responses, which is 10.2% of the total. Based on the approach adopted by the Central Hub, a response was considered partially complete if the respondent had answered at least the first two modules (demographics module and life history module). In addition, over 2,800 respondents had started the questionnaire but left before answering the first two modules.

The proportion of complete responses among all responses varies moderately across respondent subgroups. It is slightly higher among men (90.5%) than women (89.2%). Older respondents turned out to be more diligent in answering all modules than their younger peers. Among young respondents aged 18–24, the proportion of complete responses is 85.4%; among those aged 55–59, it reaches 94.5%. The variation between age groups likely also explains differences in completeness of responses associated with marital status. The proportion of complete answers ranges from 87.6% among never-married, who are the youngest among marital status groups, to 94.9% among widows/widowers, who are also the oldest. The educational gradient of completeness of responses is moderately positive. The lowest proportion of complete responses (85.6%) is found among respondents with basic or lower education (ISCED 0-2), while respondents with tertiary (ISCED 5–8) education show the highest proportion (91.6%). Among respondents with secondary education (ISCED 3–4), the proportion of complete responses is 89.8%. Differences in completeness of responses across regions and ethnic groups are small.

The overall response rate that considers both complete and partially complete answers is 28.7%. In the first round of the GGS in Estonia, which was carried out in 2004–2005 and used face-to-face interviews, the response rate was 2.4 times higher (70.2%) (EKDK, 2008). Thus, the response rate in the GGS-II resembles the non-response rate in the GGS-I. That said, the response rate in the GGS-II is not exceptionally low, given the long-term decrease in survey participation, the mode of data collection used in the Estonian GGS-II, and the length of the questionnaire. The response rate is comparable to other countries that used web-based data collection in GGS-II (Gauthier et al., 2023).

Table 3. *Number of responses and response rates by socio-demographic characteristics, Estonian GGS-II*

Characteristics	Number			Rate, %		
	All	Complete	Partial	All	Complete	Partial
All	9,146	8,212	934	28.7	25.7	3.0
Gender						
Men	4,002	3,621	381	24.6	22.3	2.3
Women	5,144	4,591	553	32.9	29.3	3.6
Age group						
18–24	1,123	959	164	28.9	24.7	4.2
25–29	1,016	892	124	30.0	26.4	3.6
30–34	1,261	1,105	156	27.8	24.3	3.5
35–39	1,330	1,180	150	30.8	27.3	3.5
40–44	1,174	1,058	116	28.8	25.9	2.9
45–49	1,183	1,079	104	29.0	26.4	2.6
50–54	1,073	1,003	70	28.2	26.3	1.9
55–59	986	936	50	25.9	24.6	1.3
Marital status						
Never-married	4,187	3,669	518	27.5	24.1	3.4
Married	3,686	3,380	306	31.8	29.1	2.7
Widowed	1,006	915	91	24.5	22.3	2.2
Divorced	98	93	5	29.5	28	1.5
Unknown	169	155	14	26.7	24.4	2.3
Education						
Basic	1,259	1,078	181	21.0	18.0	3.0
Secondary	3,184	2,829	355	26.7	23.7	3.0
Tertiary	4,338	3,988	350	34.5	31.8	2.7
Unknown	365	317	48	25.6	22.2	3.4
NUTS3 region						
North	4,408	3,981	427	29.6	26.8	2.8
Central	741	660	81	26.3	23.4	2.9
North-East	791	702	89	27.4	24.3	3.1
South	2,167	1,938	229	28.8	25.8	3.0
West	1,039	931	108	27.4	24.5	2.9
Ethnicity						
Estonians	7,040	6,338	702	30.0	27.1	2.9
Other groups	2,106	1,874	232	24.9	22.1	2.8

Source: Estonian GGS-II, authors' calculations

Table 3 also shows differences in response rates by socio-demographic characteristics. Despite significant efforts made by the survey agency, the overall response rate for men (24.6%) remains clearly lower than that for women (32.9%). By contrast, the efforts to reduce disparity in response rates between age groups have proven to be rather successful. Except for a somewhat lower response rate among 55–59-year-olds, the rate is at the level of 28–31% in all other age groups. However, as noted above, younger respondents had less patience to answer all modules. Differences between marital status groups are similar in magnitude to gender differences. The highest response rate is characteristic of people who are married (31.8%). On the other end, divorced respondents exhibit the lowest response rate (24.5%) among marital status groups.

The largest difference in responses rates is related to education. Of participants with tertiary education, more than a third (34.5%) responded to the survey. Among participants with secondary and basic education, response rates were 26.7 and 21.0, respectively. In contrast to education, regional differences in response rates are relatively small. Finally, regarding ethnicity, 30% of Estonians responded to the survey, while participants from other ethnic groups have noticeably lower response rates (24.9%) despite efforts from the survey agency and the working group.

Table 4 shows survey nonparticipation (unit non-response). This reflects an unwillingness or inability of individuals to share their experiences in response to request from a survey. Overall, 22,757 potential respondents (71.3%) did not respond to the survey. The variation in non-response by subgroups mirrors differences in response rates discussed earlier in this section.

The information from the survey agency allows us to distinguish four categories of non-response. First, 2,882 respondents (9.0% of the sample) refused to participate in the survey. Refusals to participate were received by e-mail or phone. About two fifth of the respondents who refused to participate mentioned a reason of their refusal, such as lack of time or interest, length of the questionnaire, etc. Refusals also include people who had asked to be removed from participant lists after being contacted for previous surveys. Second, 959 potential respondents (3%) are non-contacts. These are the respondents for whom the survey agency was unable to obtain a valid e-mail or telephone contact.

Third, the largest group includes 18,916 potential respondents (59.3%) who did not answer for other reasons or whose reason for nonparticipation remains unspecified. Specific reasons such as lack of access to computer or smartphone, poor digitals skills, or bad health were mentioned by a very small number (87) of individuals in this group. An overwhelming majority of this group (18,829 potential respondents) received invitations and reminders, but the respondents did not react to the invitation in any way; the invitations were not returned for technical reasons.

Finally, 100 participants turned out to be ineligible. Most of them were living outside the country although according to the register, they were residents of Estonia. This group also includes two potential respondents who had died during the period between the date when the sample was drawn and the beginning of the data collection. Non-eligible participants are not considered when calculating the cause-specific non-response rates; they are not shown in Table 4.

Table 4. *Number of non-responses and non-response rates by socio-demographic characteristics, Estonian GGS-II*

Characteristics	Number				Rate, %			
	All	Refusal	Non-contact	Other or unspe-cified	All	Refusal	Non-contact	Other or unspe-cified
All	22,757	2,882	959	18,916	71.3	9.0	3.0	59.3
Gender								
Men	12,258	2,247	620	9,391	75.4	13.8	3.8	57.8
Women	10,499	635	339	9,525	67.1	4.1	2.2	60.9
Age group								
18–24	2,764	354	79	2,331	71.1	9.1	2.0	60
25–29	2,366	352	111	1,903	70	10.4	3.3	56.3
30–34	3,280	360	193	2,727	72.2	7.9	4.3	60.1
35–39	2,987	370	128	2,489	69.2	8.6	3.0	57.7
40–44	2,904	408	126	2,370	71.2	10	3.1	58.1
45–49	2,899	377	107	2,415	71	9.2	2.6	59.2
50–54	2,734	311	101	2,322	71.8	8.2	2.7	61
55–59	2,823	350	114	2,359	74.1	9.2	3.0	61.9
Marital status								
Never-married	11,042	1,355	487	9,200	72.5	8.9	3.2	60.4
Married	7,921	1,079	285	6,557	68.2	9.3	2.5	56.5
Widowed	3,095	379	158	2,558	75.5	9.2	3.9	62.4
Divorced	234	24	8	202	70.5	7.2	2.4	60.8
Unknown	465	45	21	399	73.3	7.1	3.3	62.9
Education								
Basic	4,729	601	228	3,900	79.0	10.0	3.8	65.1
Secondary	8,749	1,174	350	7,225	73.3	9.8	2.9	60.5
Tertiary	8,219	962	334	6,923	65.5	7.7	2.7	55.1
Unknown	1,060	145	47	868	74.4	10.2	3.3	60.9
NUTS3 region								
North	10,473	1,427	445	8,601	70.4	9.6	3.0	57.8
Central	2,074	205	76	1,793	73.7	7.3	2.7	63.7
North-East	2,096	427	82	1,587	72.6	14.8	2.8	55
South	5,355	567	223	4,565	71.2	7.5	3.0	60.7
West	2,759	256	133	2,370	72.6	6.7	3.5	62.4
Ethnicity								
Estonians	16,390	1,554	671	14,165	70	6.6	2.9	60.5
Other groups	6,367	1,328	288	4,751	75.1	15.7	3.4	56.1

Note: Non-eligible participants (n=100) are not included in the table.
Source: Estonian GGS-II, authors' calculations

Table 4 presents cause-specific non-response rates by socio-demographic groups. Men exhibit higher refusal rates than women; a similar though smaller gender difference is characteristic of non-contacts. By contrast, the rate of non-response for other or unspecified reasons is slightly higher among women. The observed pattern of gender differences may be explained by the fact that men were more intensively contacted by phone. Among educational groups, cause-specific non-response exhibits a distinct pattern. Rates of refusals, non-contacts, and other or unspecified reasons appear the highest among participants with basic education, and the lowest among those with tertiary education. As regards ethnicity, the refusal, non-contact, and non-eligibility rates are higher among survey participants from other ethnic groups and lower among Estonians. However, for non-response rate for other or unspecified reasons, the pattern is reversed. The observed pattern may stem from the fact that participants from other ethnic groups were more intensively contacted by phone. Across regions, the refusal rate is particularly high in the northeast, reflecting a high concentration of Russian-speakers in this region. Cause-specific non-response rates across age and marital status groups do not show any distinct pattern.

In order to compensate for varying non-response rates across subgroups, harmonised weights have been constructed by the GGP Central Hub. Weights were constructed based on five variables: gender, age groups (18–30, 31–45, 46–60), region (five NUTS3 regions), level of education (basic, secondary, tertiary), and marital status (never-married, ever-married). Population data for weights was derived from the population census (December 31, 2021) that fell into the period of data collection of GGS-II.

Representativeness and data quality

Representativeness

Representativeness of the GGS-II in Estonia can be assessed by comparing the resident population aged 18–59 and survey respondents in terms of core socio-demographic characteristics (Table 5).

Examining the data, a noticeable difference between survey respondents and the population emerges with regard to gender. There are fewer men among respondents than in the population (43.8% vs 51.1%). The difference in the gender composition between the population and respondents stems from the lower response rates of men discussed in the previous section. However, the difference between complete and partial responses by gender is not statistically significant (Table 5).

Unlike gender, no sizeable differences can be observed in the age composition of the population and the GGS-II respondents. In line with age-specific response rates, the biggest difference from the total population concerns the oldest five-year age group (55–59), who are slightly underrepresented among respondents (10.8% in the survey vs 12% in the population). Differences in the proportion of other age groups are small and do not exhibit a clear pattern. Age differences between complete and partial respondents are

statistically significant though with younger people having a larger proportion of partial than complete responses.

Married individuals are overrepresented among respondents in relation to their share in the population (40.3% vs 34.2%). This likely reflects a greater interest in the theme of the survey in this group. The proportion of never-married and divorced respondents is lower than that in the population, while the proportion of widows and widowers appears closely similar to the latter. Differences between complete and partial responses by marital status are also statistically significant with never-married having a higher proportion of partial than complete responses.

Relatively large differences between GGS-II respondents and the population are observed with regard to education. Due to higher response rate, among the respondents the highly educated make up a much larger share than in the population (47.4% vs 37.4%). The proportion of people with secondary education shows the opposite pattern (34.8% vs 45%). People with basic education are also underrepresented among the respondents, but the size of the difference from the population is smaller. Educational differences between complete and partial responses are statistically significant with those with up to and including secondary education having higher proportion of partial than complete responses.

Given their higher response rate, the proportion of Estonians among the respondents is higher than that in the population (77% vs 68.8%). By contrast, people from other ethnic groups are underrepresented among the respondents (23% vs 31.2%). Among the characteristics examined, differences between the survey respondents and the population are smallest by region. In none of the NUTS-3 regions does the difference in the proportion of the region's residents exceed 0.6 percentage points. The differences between complete and partial responses by ethnicity as well as by the NUTS-3 region are not statistically significant.

In sum, the representativeness analysis shows noticeable differences between the respondents and the total population by gender, marital status, education, and ethnicity, caused by variation in response rates. For age groups and regions, the composition of respondents and the population is closely similar. For most of the characteristics examined, differences from the total population do not follow precisely the same pattern among the respondents who completed the survey wholly and partly. The relatively small share of the latter group among all respondents means that representativeness is mainly shaped by the respondents who gave complete answers. However, there are significant differences between complete and partial responses by age, marital status as well as by education.

Table 5. *Population and responses by socio-demographic characteristics, Estonian GGS-II and official statistics*

Characteristics	Frequency distribution, %				
	Population (official statistics)	All responses	Complete responses	Partial responses	χ² test between complete and partial responses
Gender					
Men	51.1	43.8	44.1	40.8	0.054
Women	48.9	56.2	55.9	59.2	
Age group					
18–24	12.2	12.3	11.7	17.6	0.000
25–29	10.6	11.1	10.9	13.3	
30–34	14.2	13.8	13.5	16.7	
35–39	13.5	14.5	14.4	16.1	
40–44	12.8	12.8	12.9	12.4	
45–49	12.8	12.9	13.1	11.1	
50–54	12.0	11.7	12.2	7.5	
55–59	12.0	10.8	11.4	5.4	
Marital status					
Never-married	48.8	45.8	44.7	55.5	0.000
Married	34.2	40.3	41.2	32.8	
Widowed	12.6	11.0	11.1	9.7	
Divorced	1.0	1.1	1.1	0.5	
Unknown	3.4	1.8	1.9	1.5	
Education					
Basic	14.9	13.8	13.1	19.4	0.000
Secondary	45.0	34.8	34.4	38.0	
Tertiary	37.4	47.4	48.6	37.5	
Unknown	2.7	4.0	3.9	5.1	
NUTS3 region					
North	47.7	48.2	48.5	45.7	0.556
Central	8.7	8.1	8.0	8.7	
North-East	9.0	8.6	8.5	9.5	
South	23.3	23.7	23.6	24.5	
West	10.8	11.4	11.3	11.6	
Ethnicity					
Estonians	68.8	77.0	77.2	75.2	0.615
Other groups	31.2	23.0	22.8	24.8	

Source: Estonian GGS-II and Statistics Estonia, authors' calculations

Quality check: fertility indicators

To assess validity of survey results, age-specific fertility rates of women aged 20–49 were calculated from the GGS-II for 2020/2021. The survey-based estimates of fertility rates were then compared to those based on the population register and reported in official statistics (Statistics Estonia, 2023).

The comparison shows that the difference between the estimates obtained from the survey and official statistics does not follow a uniform pattern across age groups (Table 6). Among young women aged 20–24, the survey moderately underestimates the actual fertility level. However, in age groups 25–29, 30–34, and 35–39, the differences go in the opposite direction, as the survey-based fertility rates exceed those based on the register. The overestimation of actual fertility level is most evident among women in their 30s, reaching 11–12% in relative terms. Among older women, the pattern reverses again, with the survey-based fertility rates being somewhat lower than those reported in official statistics.

The sum of age-specific fertility rates, shown at the bottom of Table 6, approximates the total fertility rate. Given that the overestimation of the actual fertility levels occurs in prime childbearing age, whereas the underestimation occurs among younger and older women, it is not surprising that the GGS-II in Estonia overestimates the current overall fertility level. However, the overestimate of the level of period fertility is not large. In relative terms, the survey overestimates the level of period fertility in 2020/2021 by 3.2%.

Table 6. *Age-specific fertility rates and total fertility rate for women in 2020/2021, Estonian GGS-II and official statistics*

Age group	Age-specific fertility rate, per 1000		Difference (GGS-II – official statistics)
	Estonian GGS-II	Official statistics	
20–24	34.5	41.1	-6.6
25–29	99.7	96.4	3.3
30–34	111.1	100.2	10.9
35–39	65.1	58.3	6.8
40–44	13.9	16.2	-2.3
45–49	0.6	1.3	-0.7
Total fertility rate	1.62	1.57	0.05

Note: The total fertility rate shown in the table excludes women aged below 20.
Source: Estonian GGS-II and Statistics Estonia, authors' calculations

We also compared the average number of ever-born children among GGS-II respondents to the 2021 population census. Similar to quality check based on age-specific fertility rates, the comparison is limited to women, because the population census did not ask men about the number of children they have ever had. Unlike the results obtained from age-specific fertility rates, Figure 1 shows that the GGS-II slightly underestimates the lifetime fertility of women. On average, the number of children ever born reported by the survey is 4.8% lower than what the census shows. Further examination of parity-specific data indicated that the observed difference is due to an overestimation of childlessness in the survey. Among the GGS-II respondents, the proportion of childless women is 30.1%, while in the census, the proportion is 26.5%. The proportion of women with one, two, and three or more children is slightly lower in the survey than in the census, respectively. Further evidence on the quality of GGS-II fertility data is available from Leocádio et al. (2023). The study focused on the fertility data among women of five countries where the survey was implemented via the web, and the data processing has been completed (Estonia, Norway, Finland, Denmark, and Sweden). For Estonia, the study found no systematic deviation for the cohort fertility indicators (cohort total fertility rate, mean age at childbearing, and parity distribution) compared to estimates from the Human Fertility Database. In line with our results, Leocádio et al. (2023) also reported an overestimation of the period fertility level for Estonia, with the difference from the UN statistics being within the confidence intervals.
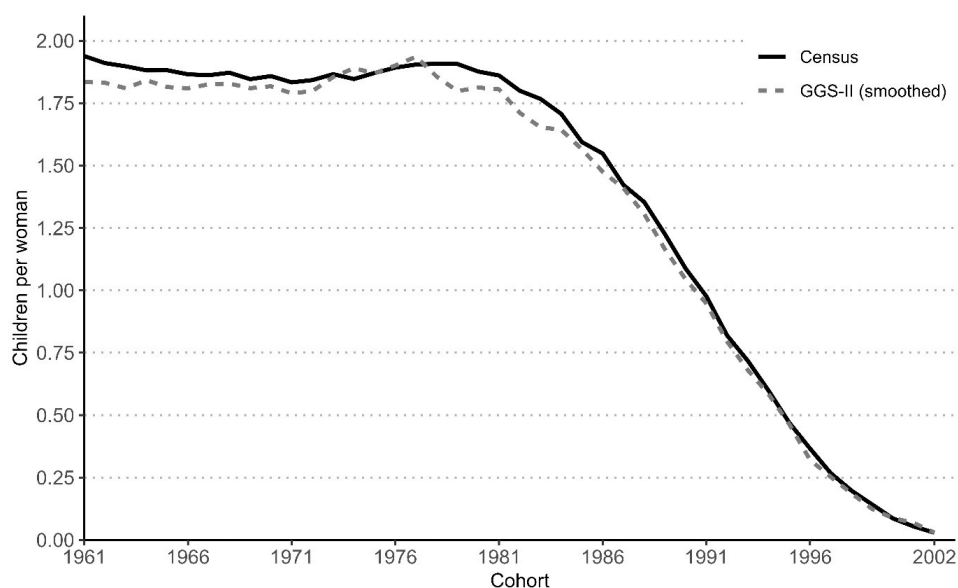

Figure 1. *Average number of ever-born children, cohorts of women born in 1961–2002, Estonian GGS-II and census 2021*

Summary

This article provided a concise overview of data collection and quality of the data of the first panel wave of GGS-II in Estonia. Unlike previous surveys on family and dynamics implemented in Estonia in the context of Family and Fertility Survey and Generations and Gender programmes, the data of the GGS-II were collected using web-based interviewing. The main survey was implemented in the period of 18 weeks from the end of October 2021 to the beginning of March 2022. The survey used a combination of conditional (a lottery of cash prizes and shopping gift cards) and unconditional incentives (donation to NGOs dealing with mental well-being and health).

The overall response rate in the survey was 28.7%. Nine-tenths of responses were complete (all modules answered) and one-tenth were partially complete (at least demographics and life history modules completed). Lower than average response rates were observed among men, participants with secondary or basic education, divorced participants, and participants from ethnic minority groups. Response rates were higher among women, participants with tertiary education, married participants, and the ethnic majority group.

The representativeness analysis showed noticeable differences between the respondents and the total population by gender, marital status, education, and ethnicity. This highlights the importance of employing weights when using the GGS-II data. Although we did not analyse item non-response in this article, data exploration suggests that, like unit non-response, item non-response is also higher in the GGS-II than in the previous round of the survey. On a positive note, we conclude that notwithstanding the low response rate, the GGS-II provides a fairly accurate account of fertility patterns in Estonia.

Acknowledgements

Authors' contributions

AP participated in developing the concept of the article, planned and conducted the empirical analysis, interpreted the results, drafted and revised the manuscript. LS and TT provided study materials. LA, MG and LR participated in developing the concept of and in revising the article. MK maintained the database and provided study materials. All authors read and approved the final version of the manuscript.

References

EKDK (Eesti Kõrgkoolidevaheline Demouuringute Keskus) (1995a). *Eesti Pere- ja Sündimusuuring. Metodoloogiaülevaade. Estonian Family and Fertility Survey. Methodological report.* RU, Series A, No. 39. Tallinn: EKDK. Available at: https://www.digar.ee/arhiiv/et/books/115183

EKDK (Eesti Kõrgkoolidevaheline Demouuringute Keskus) (1995b). *Eesti Pere- ja Sündimusuuring. Standardtabelid. Estonian Family and Fertility Survey. Standard tabulations.* RU, Series C, No. 6. Tallinn: EKDK. Available at: https://www.digar.ee/arhiiv/et/raamatud/118105

EKDK (Eesti Kõrgkoolidevaheline Demouuringute Keskus) (1999). *Eesti Pere- ja Sündimusuuring: meesküsitluse metodoloogiaülevaade. Estonian Family and Fertility Survey: Methodological report of male survey.* RU, Series A, No. 40. Tallinn: EKDK. Available at: https://www.digar.ee/arhiiv/et/raamatud/115184

EKDK (Eesti Kõrgkoolidevaheline Demouuringute Keskus) (2008). Eesti Pere- ja Sündimusuuring. Teine ring. Standardtabelid. *Estonian Family and Fertility Survey. Second round. Standard tabulations*. RU, Series C, No. 26. Tallinn: EKDK. Available at: https://www.digar.ee/arhiiv/et/books/118104

EKDK (Eesti Kõrgkoolidevaheline Demouuringute Keskus) (2009). Eesti Pere- ja Sündimusuuring. Teine ring. Standardtabelid: välispäritolu rahvastik. *Estonian Family and Fertility Survey. Second round. Standard tabulations: Foreign-origin population* RU, Series C, No. 27. Tallinn: EKDK. Available at: https://www.digar.ee/arhiiv/et/raamatud/118106

ETAG (Eesti Teadusagentuur) (2019). *Estonian Research Infrastructure Roadmap 2019*. Tartu: Estonian Research Council. Available at: https://etag.ee/wp-content/uploads/2019/06/ETAg_Research_Infrastructure_Roadmap_2019.pdf

Estonian GGS-II (2021a). *Estonian Generation and Gender Survey-II. Wave 1 questionnaire*. Available at: https://ggp2020eesti.ee/kusitlusinstrumentaariumid/

Estonian GGS-II (2021b). *Estonian Generation and Gender Survey-II. Media coverage during data collection*. https://ggp2020eesti.ee/kajastus-meedias/

GGP (Generations and Gender Programme) (2023). *An overview of country-specific questions in GGP-II*. Available at: https://www.ggp-i.org/ggs-round-ii/

Gauthier, A., Kong, S., Grünwald, O., Bujard, M., Caporali, A., Deimantas, V., Emery, T., Jablonski, W., Koops, J., Rijken, A., & Schumann, A. (2023). *Data Brief: The Generations and Gender Survey second round (GGS-II)*. GGP Technical Paper Series. https://doi.org/10.5281/zenodo.10220746

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., & Vikat, A. (2021). *Generations and Gender Survey Baseline Questionnaire 3.1.1*. The Hague: Netherlands Interdisciplinary Demographic Institute. Available at: https://www.ggp-i.org/wp-content/uploads/2022/07/BaselineQuestionnaire_3.1.1.pdf

Katus, K., Puur, A., & Sakkeus, L. (2000). F*ertility and Family Surveys in Countries of the ECE Region. Standard Country Report. Estonia*. New York and Geneva: United Nations.

Katus, K., Puur, A., & Põldma, A. (2002). *Eesti põlvkondlik rahvastikuareng. Cohort population development in Estonia*. RU, Series D, No. 2. Tallinn: Eesti Kõrgkoolidevaheline Demouuringute Keskus.

Leocádio, V., Gauthier, A. H., Mynarska, M., & Costa, R. (2023). The quality of fertility data in the web-based Generations and Gender Survey. *Demographic Research* 49(3): 31–46. doi: 10.4054/DemRes.2023.49.3

Statistics Estonia (2023). *Database*. Available at: https://andmed.stat.ee/en/stat

Tambaum, T., & Klesment, M. (2021). Organisation of the Estonian GGS-II pilot survey. GGP connect seminar 6, 26 October 2021. Available at: https://www.ggp-i.org/ggp-connect-seminar-series/#toc15