



The Generations & Gender Programme

Three-country Experiment on Modes of Data Collection

19 October 2020

Tom Emery, Detlev Lück, Almut Schumann, Susana Cabaço, Peter Lugtig, Vera Toepoel, Martin Bujard,
Robert Naderi, Norbert F. Schneider

ggp@nidi.nl

Please do not cite or circulate without permission

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 739511 and the Federal Institute for Population Studies (BiB).



Contents

1.	General Description of the GGP Experiment on Modes	3
2.	Design of the GGP Experiment on Modes	7
2.1.	Overview of the GGP Fieldwork Experiment	7
2.2.	Fieldwork Specifications and Sampling	8
2.3.	Design Specifications in Germany	9
2.4.	Design Specifications in Croatia	12
2.5.	Design Specifications in Portugal	12
3.	Empirical Findings	13
3.1.	The Common Experiment: Comparing Push-to-Web and Face-to-Face	13
3.2.	The German Experiment: Comparing Incentives	16
3.3.	The Croatian Experiment: Comparing Reminder Strategies	17
3.4.	The Portuguese Experiment: Comparing Respondent Selection Strategies	18
4.	Conclusions and Recommendations	20

Executive Summary

The GGP carried out its first round of data collection during the period 2002-2012. During this period of total of 20 countries collected data. This has resulted in many papers and has supported major advances in knowledge in population and family dynamics. However, behind this scientific excellence, there were several technical challenges.

The first one is that a non-centralized approach was used. What this means is that although there was a centrally approved questionnaire (approved by the GGP consortium board), there was little collaboration between the central hub and national teams in implementing the survey. There were three consequences of this approach: (1) there variations in how the data was collected: some countries opting for pen and paper and others for CAPI; (2) there were large country variations in compliance to the central questionnaire (resulting in incomparability and time-consuming post-harmonisation); and (3) there was a long period of time between the end of the data collection and the release of the data file to users.

The second challenge is that during the period 2002-2012, the context itself of survey data collection has changed. Not only have the response rates been largely declining in many countries, but moreover the actual cost of carrying out face-to-face surveys has increased steeply. These two changes have threatened the sustainability of conventional survey data collection.

In preparing for the new round of data collection in 2020, the GGP has invested in various pilots to address these two technical challenges:

- (1) it has tested a more centralized mode of data collection as well as various pieces of software to standardize fieldwork operation and speed up data processing; and
- (2) it has tested the feasibility of online data collection, to reduce the cost of surveys and hopefully to increase the response rates.

These two technical options were tested as part of a three-country experiment. This document reports on the experiment and its main findings. The three countries chosen for the experiment were selected to deliberately reflect some of the most challenging fieldwork environments for face to face data collection. This report illustrates that, whilst there are many significant challenges involved in centralizing data collection and conducting online data collection, the existing technical and methodological design of the GGP is feasible. It also outlines the limitations to these approaches and mitigation measures in countries which have specific, problematic conditions.

1. General Description of the GGP Pilot

In order to launch a new round of the Generations and Gender Survey (GGS) in 2020, the GGP-Consortium has started a process of reflecting and investigating how such a new panel study can best be designed in terms of survey methodology. The main focus of this technical options analysis was on designing the survey in such a way that, on the one hand, it would be cost-efficient and as many countries as possible would be able to find funding, and, on the other hand, the survey quality in terms of coverage, representativity, and accurate measurement would meet the high quality standards GGS has been aiming for also in the past. This process of reflecting and investigating has been assigned to the work package 2 (WP2) “Methodological & Technical Development” of the EU-funded project GGP-EPI (“Evaluate, Plan, Initiate”). This work package consists of partners from NIDI (Netherlands Interdisciplinary Demographic Institute), BiB (German Federal Institute for Population Research) and the University of Utrecht, who are at the same time the first eight authors of this report.

Following a proposal from the University of Utrecht, WP2 has considered a sequential mixed-mode design as promising, consisting of a combination of CAWI and CAPI modes, also known as “web first”, “web push” or “push-to-web” design. Its aim is to reduce survey costs by interviewing as many respondents as possible through self-administered web-interviews, which is the cheapest mode available. At the same time, the follow-up interviews conducted as interviewer-administered face-to-face interviews are supposed to prevent a selection bias as it would be expected from a pure online survey. The mixed-mode design may even have the advantage of accessing wider range of people than any single mode survey. For these reasons, the push-to-web design has been tested in direct comparison to the CAPI mode, as it has been used so far in GGS, in an experimental pilot study.

Aside from this methodological experiment, the pilot study has tested further potential design innovations. These follow a consideration of the key developments and trends in internationally comparable survey research. According to these considerations, the future GGS data collection will be coordinated centrally. Data servers for storage of survey data and the questionnaire management system will be centralized. All countries will work with the same common digital version of the questionnaire, programed and provided centrally. The key reasons for doing this are (1) to ensure that the questionnaire is administered in the same way across countries, eliminating the need for post-harmonization as far as possible, (2) to make sure that GGS data can be prepared for data release quickly and efficiently, and (3) to better monitor the fieldwork in each country. The strategy of centralization suggests that:

- Sample management will still be coordinated within each GGP country, so that it complies with legal requirements for the storage of personal (identification) data in accordance with national legislations.
- Sampling and fieldwork procedures will be pre-harmonized and coordinated by the central GGS “hub” at NIDI, meaning that data collection will take place in as much the same way as possible across countries.
- Sampling and fieldwork procedures should nevertheless be flexible and adjusted in cooperation between central hub and national team. There are large differences between countries (e.g. in internet penetration rates, availability of sampling frames etc.), and it is likely that with time, these differences will change further (within and between countries). Fieldwork guidelines should therefore be both robust for country differences, and future proof at the same time.

This pilot study has been conducted in the three countries Germany, Croatia and Portugal, in order to gain insights to what extent findings would be generalizable and applicable to other GGS countries. Each of the three countries provides a different context for survey research; none of them provides particularly good conditions. ***The purpose of this selection was to rather under- than over-estimate the chances of the revised design. If the design would be implementable in Portugal, Germany and Croatia it should be realisable in any other GGS country also.***

Practical issues to consider in the design of the study

The three principles above guided the practical choices in fieldwork procedures for future GGS rounds. In addition to these, there were a number of technical, practical and methodological constraints which informed the design of the study:

- In some countries, the GGS may use individual population registers as the basis for drawing samples. In other countries, the GGS will have to rely on address sampling. Using phone registers or Random Digit Dialling as the basis for sampling are not feasible due to low coverage rates of phones, and high nonresponse rates in telephone surveys.
- As a result, the GGS will rely on a combination of individual- and address-based sampling. In some countries, persons will be sampled, while in other households will be sampled, with an additional sampling stage of selecting an individual within a household.
- Because addresses are used, the initial mode of contact with respondents can be made through letters and/or interviewers. Contact modes can be combined (e.g. sending a letter first, after which interviewers visit a house).
- Face-to-face (F2F) interviews are much more expensive than self-administered paper-and-pencil or web interviews. The costs of F2F interviews will likely increase in future, while for other modes, the costs will stay relatively stable.
- The GGS questionnaire is complex. It is long and involves complex routing procedures. Because of this, using paper-and-pencil as an administration mode was ruled out.
- There have been successful cases of other surveys that send letters, and then ask respondents to complete a survey online by visiting a URL or scanning a QR code. Response rates for such web surveys can be anywhere between 5% and 50% depending on e.g. the topic of the survey, the number of reminders sent and the incentives offered.
- Using F2F as a follow-up mode for non-respondents to an online survey can increase the overall response rates and reduce sample selectivity. It was decided that such a consecutive mixed-mode design of web and F2F follow-up – also referred to as “push-to-web” (P2W) – forms the basis of future GGP surveys.
- There will be flexibility in the proportion of interviews that will be conducted by web and F2F, both between countries and over time. Considerations here are:
 - In countries with lower internet penetration rates, more interviews will have to be conducted F2F.
 - Countries may opt to not assign all cases to web-first if there is a large budget or if costs for F2F interviews in the respective country are affordable. In countries with high costs for F2F interviews and limited financial resources, efforts can be made to maximize the proportion of interviews conducted on the web, for example by sending more reminders, or increasing incentives.

- In later waves, efforts can be made to “push respondents to web” who had been interviewed personally in the first wave. Because one can collect further information (e.g. e-mail addresses) in wave 1, the proportion of web interviews will likely increase in later waves.
- Over time, also internet penetration rates will change which will also affect the proportion of web interviews.
- For all these reasons, the proportion of interviews conducted in web and F2F will not only differ over countries, but it also will change over time.
- This implies a risk of selection and mode effects: if respondents answer the same question differently depending on the mode being offered (web vs. F2F), differences between countries or over time could be the result of shifts in the proportions of interviews in each mode.
- In order to minimize such mode effects, the GGS questionnaire is redesigned and tested in the experiment. There are two versions of the GGS questionnaire: a face-to-face and web version. We took care to minimize differences between the two versions (mainly instructions to interviewers) to minimize measurement effects.
- Also, measurement effects are reduced by conducting sensitive parts of the F2F interview by CASI (that is, self-administered), instead of CAPI (that is, personal interviewed).
- Scientists using the GGS data need to be instructed to conduct sensitivity analysis for mode effects when analysing the data.

2. Design of the GGP Experiment on Modes

The basic design tested in the fieldwork experiment on modes is a form of a sequential mixed-mode design – the so-called “push-to-web” design. It consists of an initial self-administered web survey, followed by a phase of personal face-to-face interviews with all people who have not participated in the web interview during the initial phase. This design is compared to the “classic” GGS design, which is interviews being conducted in face-to-face mode only. This comparison of an experimental group in push-to-web and the reference group in face-to-face mode is the common part of the experiment in all three countries in order to evaluate the potential new design in the three quite different country contexts. Additionally, country-specific experiments were included in each of the three countries that would allow us to identify an optimal design variation. These included a variation of incentives in Germany, a variation of reminders being sent out in Croatia and a variation of procedures for identifying a target person within a sampled household in Portugal.

2.1. Overview of the GGP Fieldwork Experiment

Based on the factors affecting future fieldwork trends, the principle aim of the fieldwork experiment is to examine the degree to which a push-to-web within the new round of GGS data collection can be achieved without adverse effects on response rates or data quality. This must be achieved by comparing respondents who are asked to use a push-to-web (P2W) approach to those who will be interviewed using the traditional F2F methodology that was utilized in the first round of the GGS.

In order to generate results that serve as an orientation for the GGS relaunch for as many participating countries as possible (e.g. for the calculation of costs), the pilot study will be conducted in three GGP partner countries, in different parts of Europe with different country-specific conditions. The obtained results of the study will then be differentiated by country. These three countries should represent much of the diversity of European countries, particularly regarding the availability of register data for drawing samples, the costs for F2F-interviews and the social acceptance of survey research in its various forms. The choice of countries also should rather over- than underestimate the difficulties and costs of a new GGS. Therefore, countries that provide suboptimal conditions in the variety of the cited challenges for survey research are chosen. After discussion, the three selected countries are: Germany, Croatia, and Portugal. In order to limit costs, each country chooses one region for conducting the experiment. This region needs to include urban as well as rural neighbourhoods. Furthermore, it should provide, as far as possible, suboptimal conditions for online and F2F research, compared to other regions within the same country.

As a baseline, we conduct an experiment in each country in which respondents are randomly assigned either to the F2F-only, or to the push-to-web condition. In addition, each country executes a specific methodological experiment in order to further investigate *how* a push-to-web design should be implemented best. These country-specific methodological experiments are:

- In Germany: different strategies for providing incentives for participation
- In Croatia: different strategies for sending reminders
- In Portugal: different strategies for selecting a respondent from the people living at the same address (within random route sampling)

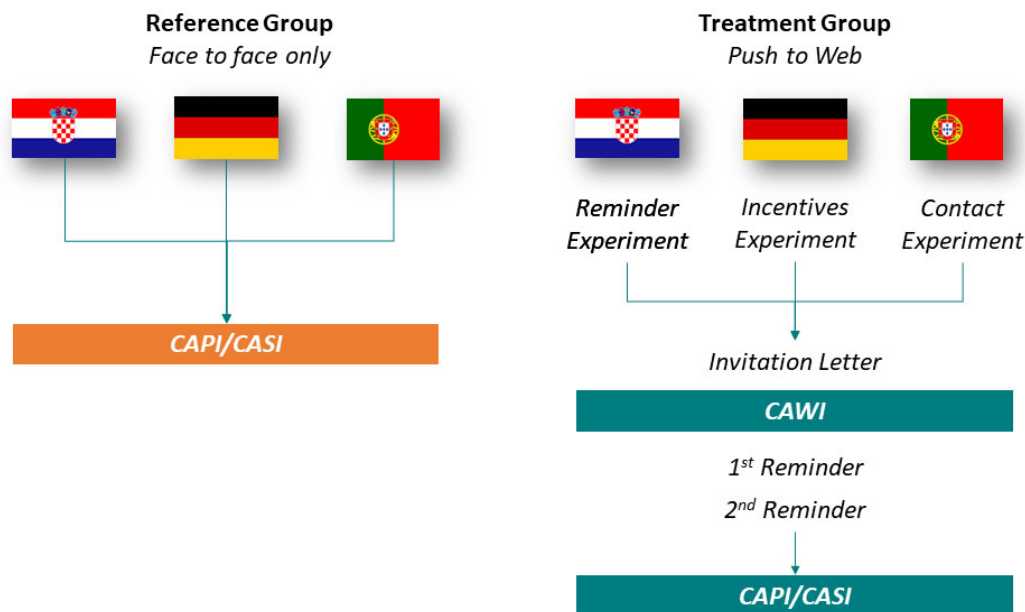


Figure 1 - General design of the GGP Experiment on Modes in Croatia, Germany and Portugal

2.2. Fieldwork Specifications and Sampling

The following technical requirements for fieldwork hold for all countries in the GGP Experiment:

- In each participating country – Germany, Croatia, and Portugal – one reference group is interviewed in a face-to-face (only) mode (CAPI/CASI) and at least two treatment groups are interviewed in a push-to-web design (CAWI or CAPI/CASI).
- All groups – reference as well as treatment – are randomly selected from the same gross sample.
- The gross sample is drawn according to country-specific optimal techniques of drawing representative samples, that is, based on registers in Germany and Croatia, and based on random route in Portugal. In Germany, local registers of 36 communities in Bavaria are employed. In Croatia an election register is used. The Portuguese sample is drawn on address-based sampling.
- In order to reduce costs, the samples are not representative for the countries as a whole but rather for the single region within these three countries. However, these regions include urban as well as rural areas and offer comparatively suboptimal conditions for survey research, for example low provision of high-speed internet. The regions are defined by the local organising team in consultation with the fieldwork agency in a way that costs are reduced (e.g. by picking a region in which the fieldwork agency has many interviewers).
- Furthermore, the samples are restricted to the age range of 18 to 49 years.

- Proxy responding is not allowed.
- All experimental groups, including reference and treatment groups, are represented by case numbers that are suitable for answering the research questions formulated beforehand. Hence, a minimum sample size for each group should be $n=200$. As far as possible, the required case number should exceed the minimum sample size!
- Not only the treatment groups which are interviewed in a push-to-web design, but also the two sub-samples in each treatment group of those interviewed in self-administered online-mode respectively those interviewed in F2F-follow-up mode, should each be represented by credible case numbers requiring $n > 200$ respondents.
- In each country, there is one “anchor” treatment group for the push-to-web condition, which follows an internationally standardised and comparable design as listed below, and one treatment group with a further experimental condition.

Table I - Allocation of target net samples across the experiment

<i>minimum case numbers ...</i>	<i>Germany</i>			<i>Croatia</i>			<i>Portugal</i>		
	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W exp</i>	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W exp</i>	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W exp</i>
<i>for all interviews</i>	200	400	400	200	400	400	200	400	400
<i>web interviews</i>	–	200	200	–	200	200	–	200	200
<i>face-to-face</i>	200	200	200	200	200	200	200	200	200

2.3. Design Specifications in Germany

In Germany, the experimental design is realised as follows:

- The gross sample is drawn from the registers of local registration offices.
- In coordination with the fieldwork agency, Bavaria is chosen as the region for conducting the fieldwork experiment. Here, 36 sample points are selected. From each of these sample points either $n=250$ or $n=500$ people are drawn for the gross sample, so that half of the gross sample represents urban and the other half rural regions in Bavaria. This implies that the sample is disproportional, over-representing rural areas.
- In order to test the push-to-web design in various variations, five treatment groups are generated with various combinations of incentives (unconditional / conditional):
 - a “second reference group”, not receiving any incentive (0 / 0),
 - the “anchor group”, receiving 5 Euro unconditional pre-paid incentive (5 / 0),
 - a “small conditional incentive group”, receiving 5 Euro post-paid incentive after a successfully completed interview (0 / 5),
 - a “large unconditional incentive group”, receiving 5 Euro unconditional incentive and further 25 Euro after a successfully completed interview (5 / 25),
 - a “large conditional incentive group”, receiving 30 Euro conditional post-paid

incentive after a successfully completed interview (0 / 30).

Given that F2F-interviews are particularly expensive in Germany, the experiment would not be viable without a limitation of the F2F-interviews conducted. Negotiations with fieldwork agencies came to the insight that, aside from n=200 interviews in the F2F only reference group, another n=200 F2F interviews are financially feasible within two of the push-to-web treatment groups each. For these, the “anchor group” (5/0) and the “large unconditional incentive group” (5/25) are chosen. Hence, we proceed as follows:

- In “the anchor group” (5 / 0) and in the “large unconditional incentive group” (5/25), all participants who do not participate in the web survey, but are instead willing to be interviewed in person, are interviewed face-to-face in the F2F follow-up phase. The number of people contacted is chosen in a way that enough cases can be expected to remain after the web phase, so that minimum case numbers for the F2F follow-up are realistic to be reached.
- In all other treatment groups, no F2F follow-up interviews are conducted (unless unexpected financial resources allow doing so).

Table 2 - Allocation of target net sample across experimental groups in Germany

	group R <i>reference</i>	group 0 <i>2nd reference</i>	group 1 <i>anchor</i>	group 2	group 3	group 4
general interview mode	F2F only	mixed-mode / push-to-web (web + F2F)				
incentives (unconditional / conditional)	(5 / 0)	(0 / 0)	(5 / 0)	(0 / 5)	(5 / 25)	(0 / 30)
A) total gross sample (est.)	~8,000					
B) initial group gross samples (estimated)	~575	~2,000	~1,400	~1,500	~1,200	~1,300
intended mode at 1st contact	face-to-face	web	web	web	web	web
response rate (estimated)	30-40%	10-15%	15-20%	15-20%	15-25%	15-25%
minimum sample size	200	250	250	250	250	250
C) sample size (estimated)	200	250	250	250	250	250
D) adjusted group gross samples (estimated)	(~375)	(~1,750)	~1,150	(~1,250)	~950	(~1,050)
E) random sub-sample		0*		0*	~500*	0*
intended mode after 2nd reminder	—	(face-to-face)	face-to-face	(face-to-face)	face-to-face	(face-to-face)

response rate – based on adjusted g. g. s. (estimated)	—	15-25%	15-25%	15-25%	15-25%	15-25%
minimum sample size	—	0*	200	0*	0*	0*
F) sample size (estimated)		0*	~230	0*	~100*	0*
G) total sample size	~200	~250	~480	~250	~350	~250
total response rate (estimated)	30-40%		25-45%			

* For financial reasons, the F2F follow-ups in the non-anchor treatment groups are conducted only to the extent that financial means allow. Before all other non-anchor treatment groups, these interviews are conducted in the “5/25” group. For this purpose, a random subsample will be drawn.

2.4. Design Specifications in Croatia

In Croatia, the design of the experiment is realised as follows:

- The sample is drawn from the Croatian Electoral register and covers the whole geographical area of Croatia.
- Respondents in the Push to Web group are sent a letter informing them of the study which includes a link to the survey and 40 kuna (approximately €5).
- No phone contacts are allowed, even in case a telephone number is available from the electoral register. In the F2F phase, the interviewer may collect new telephone numbers of respondents as part of the F2F interview (just like outlined in the general guidelines).
- In order to test the push-to-web design in various variations, two treatment groups are formed:
 - An “anchor group”, receiving two reminders in a two-week time interval (14 days and 28 days after the initial contact).
 - A group, receiving two reminders in a one-week time interval (7 days and 14 days after the initial contact).

2.5. Design Specifications in Portugal

In Portugal, the design of the experiment is realised as follows:

Interviews follow the general guidelines for fieldwork as pointed out earlier in this document. A sample is drawn by using a random route design. In a first moment, when the interviewers established the first contact, and were successful in getting access and being able to speak with someone, they had to establish if in a given address there were any eligible respondents residing and, if so, leave the invitation letter and incentive (€5 voucher).

In the case of the face-to-face group, the interviewer would aim at scheduling a date and time to return for the face-to-face interview with the respondent.

The invitation letters provided instructions on how the respondent should be selected in each household. The person to be chosen was the next person to have their birthday who was aged between 18-49. The experiment focused, during the first contact, on the reference, or not, by the interviewer to the fact that there was the incentive in the envelope. It should be noted that the person opening the door and receiving the envelope was not necessarily the selected person to participate in the survey.

- In one group, the interviewer conducted the eligibility check and, if there were eligible respondents in the household, the interviewer would leave the envelope (containing the invitation letter and incentive), mentioning that there was a €5 voucher inside.
- In the other group, the procedure was very similar with a key difference. If there was anyone eligible to participate in the household, the interviewer would leave the envelope (containing the invitation letter and incentive), without any references to the fact that there was a €5 voucher inside.

3. Empirical Findings

The empirical findings focus on the comparison between the traditional Face-to-Face (F2F) survey mode, used in former waves of the GGS, and the Push-to-Web (P2W)-design, tested in the GGP pilot regarding the response rate and indicators of data quality. The F2F reference group and the P2W reference group have the same conditions of receiving a €5 prepaid incentive and two reminder letters with an interval of two weeks. These groups only differ in the mode of data collection: While the F2F reference target persons were interviewed in CAPI mode by an interviewer, the P2W reference target persons were first asked to participate online in the web and if they did not after two reminders, an interviewer came to their home to conduct the CAPI interview. Section 3.1. focuses on the comparison between the two reference groups across the three countries to receive insights into how P2W performs compared to F2F. The following sections (3.2 to 3.4) are dedicated to each individual country and consider the response rates and data quality for the respective experimental groups in the country experiment.

3.1. The Common Experiment: Comparing Push-to-Web and Face-to-Face

As a first assessment for the performance of P2W, we look at the data across all three countries under study. In doing so, we exclude the experimental groups in each country and look at the reference F2F and P2W groups in each country to examine the performance of P2W relative to F2F under the common fieldwork procedures.

Response Rates

For calculating response rates, the gross sample of units was divided into:

- a) completed interviews,
- b) those which were eligible and didn't complete an interview and
- c) those which were not eligible.

This information is stored in the contact protocols from each specific country. For person-based samples (like for Germany and Croatia), all cases of the gross sample should be eligible since every target person belongs to the target population based on information included in the official sample frame list of Croatia and Germany. In Germany some target persons were not fully or not finally processed by an interviewer due to time and capacity restrictions of the fieldwork agency. Since this was the case, these sample units were not included for calculating the response rate - they were indicated as "not eligible". In Portugal, the sample is based on households because there is no person register sample frame available. The interviewer had to find the right household and it could be that the person opening the door and receiving the envelope was not necessarily the selected person to participate in the survey. As a consequence of this design, Portugal had a higher proportion of not eligible persons who were contacted by an interviewer. For households which could not be contacted (e.g. because five times no person was found in household) a proportion of eligible and not eligible persons has to be calculated manually¹. Due to this, one has to be aware that the response rate for Portugal is partly based on assumptions of eligibility within the gross sample. This is also the reason why no Chi-square-Test is calculated and reported for Portugal.

¹ If "no contact" is recorded in the Portuguese contact protocols, the interviewer did not know whether the selected person is eligible (as target population = 18-49 years old). We had to calculate the proportion of cases which were contacted and not eligible. This proportion then needs to be extrapolated to the not contacted cases ("no contact"). This proportion was calculated for each of the three experimental groups in Portugal – the amount of not eligible persons from the not contacted persons varies between 30-40 % depending on the group.

Response rates are calculated as follows:

$$\text{Response Rate} = \text{Number of complete interviews} / \text{Number of eligible cases}$$

Table 1 shows the distribution of complete interviews mentioned in the contact protocols for all three countries together comparing only the ratios for the F2F reference group and the P2W reference group.

Table 3 - Response Rate from Face-to-Face and Push-to-Web- group across all countries

	F2F ref.	P2W ref.
Complete interviews (%)	19.00	24.73
Total number of eligible cases (N)	2,985	5,119

%= relative frequencies. Source: GGP Pilot, own calculations

The results in table 3 illustrate that the response rates in F2F are much below the F2F rate identified in gold standard studies overall. However, the extent of this and the reasoning behind this is varied across countries and so we will address this issue in the country specific reporting. Nevertheless, it should also be remembered that the three countries in the study were chosen because of their challenging fieldwork conditions and where the future of F2F interviewing looks questionable. F2F response rates in the study were generally low though and fieldwork agencies reported that this was for several reasons.

Firstly, the tender for fieldwork provided for a very small F2F operation of just 200 completed interviews. This complicated fieldwork management and greatly increased overheads as interviewers needed to be trained and fieldwork management procedures implemented for a relatively small number of interviews. For example, most fieldwork agencies were utilizing only 6-10 interviewers in the experiment which meant that the withdrawal of interviewers in the early stages of fieldwork was proportionally more disruptive. The withdrawal of interviewers in the early stages of fieldwork is commonplace and over time fieldwork agencies focus investment and management on high performing and committed interviewers. With such a low number of cases however, this became very difficult to manage.

The second issue was that the fieldwork window for F2F was unreasonably short. As the study was designed as Push to Web experiment, the timetable for fieldwork was also structured around fieldwork running to a CAWI timetable. For example, in Croatia the fieldwork agency was given less than 3 months to complete all necessary contacts with the sampled individuals, including the follow ups to non-respondents in CAWI. In retrospect, this was highly impractical and placed too great a demand on the fieldwork agency. Typically, fieldwork operations can last from 6-9 months. The fieldwork for the latest round for which data is available in the ESS shows fieldwork lasting 6 months in Germany, 8 months in Portugal and 6 months in Croatia. Given that the fieldwork team is reduced to accommodate the smaller sample size, it cannot be expected that the smaller number of cases would lead to a shorter fieldwork window. It is therefore highly likely that the fieldwork operations were stopped before the sample could be fully exhausted and before the fieldwork team reached optimal efficiency.

Given the deficiencies identified in the baseline F2F response rates, we will also benchmark the P2W response rates against those observed in the latest round of the ESS. The ESS places great emphasis on achieving high response rates and full and transparent reporting of fieldwork operations. The questionnaire for the ESS and the technical protocols for the ESS differ from those of the GGS and therefore the comparison is not as ideal as the controlled experimental conditions

in the design. It does however still represent a gold standard in survey research. The general nature of the ESS survey, its length and the mature nature of its contact protocols suggest that the response rates achieved by the ESS should be higher than those that would be achieved should the F2F component of the experiment have been conducted under better conditions. The latest response rates which are available for the ESS for are as follows:

- Germany (2019, Round 9): 27.6%
- Portugal (2017, Round 8): 45.0%
- Croatia (2015, Round 7): 45.7%

In all cases these are higher than the F2F response rate observed in the GGS experimental study. Therefore, whilst the results show that the response rate is higher in P2W than F2F, considerable caution is required in the evaluation of this result.

In table 2 the differentiation of counties is taken into account.

Table 4 - Response Rate from Face-to-Face and Push-to-Web- group for each country

	Germany		Croatia		Portugal	
	F2F ref	P2W ref.	F2F ref.	P2W ref.	F2F ref.	P2W ref.
Complete interviews (%)	29,49	23,66	27,67	49,45	11,70	9,81
Total number of eligible cases (N)	685	1.365	600	1.450	1.700	2.304
Chi-squared-test (χ^2)	8.13**		82.11***		Not feasible	

*%= relative frequencies; χ^2 = Chi-squared-test; Level of significance: * $p < 0.05$ | ** $p < 0.01$ | *** $p < 0.001$; Source: GGP Pilot, own calculations*

If you look at the countries separately, P2W only works best for Croatia. In this instance the P2W baseline group actually outperformed the latest ESS F2F response rate (49.5% v 45.7%). Due to cultural and structural conditions it may be that people in Croatia are more willing to participate in surveys in general and because of a higher degree of digitization in the country they are not sceptical about web surveys. It could also be that incentives were higher relative to average incomes.

Germany and Portugal held moderately higher response rates in F2F. In Germany, people tend to participate more often in CAPI mode with an interviewer than in a self-administered CAWI mode. Maybe it is because it seems more serious, more official and more difficult to cancel when an interviewer visits you at your home than to follow an invitation for a web survey. It should also be noted that the response rate in the experiment for F2F was higher than in the ESS which ran concurrently to the study (29.9% v 27.6%). This would suggest that the fieldwork operations in Germany achieved the best practice standards in country and therefore should not be discounted as an adequate benchmark within the study.

In Portugal the overall response rate for both groups is on a very low level, but there is a slightly higher participation in the F2F reference group. This could be due the fielding conditions in Portugal were rather challenging (e.g. no sample of persons, many holiday houses etc). Portugal is a notoriously difficult country to conduct fieldwork in for established cross-national surveys. Fieldwork in the European Social Survey lasts a very long time. For Round 8, fieldwork lasted 238 days and resulted in 1,270 interviews. This implies a rate of just 5 interviews per day which is the same as the rate achieved in the GGP between the fieldwork dates of 14th May and 15th September

2018. Given that the fieldwork operations in Portugal also included the contact attempts for the CAWI sample, the issues affecting F2F response rates would have also affected the CAWI response rates. The contact protocols developed for CAWI in Portugal didn't mitigate the inherent obstacles to fieldwork operations there and subsequently the results for all groups were disappointing.

3.2. The German Experiment: Comparing Incentives

Process and Outcomes of the Fieldwork

Due to the delays in starting and processing fieldwork, the CAPI follow-up phase was not completed. Kantar had intermediately planned to complete fieldwork by the end of September. When approaching this date, suggested on short notice to indeed end fieldwork, without having all target persons contacted four times. Arguments were the very low response rates within the follow-up, the accordingly low motivation of interviewers (being paid per realised interview), and the long timespan that was to be expected for processing the last cases. Based on the misunderstanding that only few cases would be missing, this proposal was initially supported. Accordingly, Kantar interrupted the fieldwork by the end of September 2018. However, the clarification of the numbers led to the insight that 718 (36.3%) of the 1,979 cases that were supposed to be contacted within the CAPI follow-up hadn't been processed. Based on this information, BiB insisted to reactivate the field. This was realised only by early November.

From November to mid of December a second CAPI follow-up phase was conducted. Here, the agreement was to still not process every case. The reason was that interviewers were still extremely demotivated, and partly reluctant, to proceed with the task, which involved for them a very bad ratio of payment on the one side and invested time, travel as well as effort on the other side. For the same reason, Kantar insisted to give interviewers the option to offer a post-paid incentive of 10 Euro, so that cases from the second follow-up phase haven't received the same incentives as in the first follow-up phase. Only sample points with a relevant number of unprocessed addresses were reactivated. Among the 718 unprocessed cases, 250 weren't pursued any further. Another 81 cases couldn't be contacted four times by interviewers until mid of December, so that in the end altogether 331 cases (16.7% of all cases in the CAPI follow-up phase) remained unprocessed. 387 cases were processed during the second follow-up phase, so that in the end 1,648 cases (83,3% of the 1,979 cases in the follow-up) were fully processed. 26 interviews were realised during the second follow-up.

The final codes in the contact protocols deviate between the first and the second follow-up phase: While response rates remain similarly low (around 4% for the lower incentive and around 7% for the higher incentive), there were fewer strong refusals and more target persons that couldn't be contacted within field time. This may partly be due to the shorter time span of the second follow-up phase, or due to selection effects. Partly, it must be assumed that demotivated interviewers may have declared a target person as not at home, without actually trying four times to contact this person.

Response Rates

The experiment in Germany includes four further experimental groups using different combinations of unconditional and conditional incentives. Only in one experimental group (P2W 5/25) a face to face follow-up was carried out. The other three groups (CAWI 0/0, CAWI 0/5, CAWI 0/30) can be regarded as pure web surveys.

Table 5 - Response Rates for Germany

	F2F ref. (5/0)	P2W ref. (5/0)	CAWI (0/0)	CAWI (0/5)	P2W (5/25)	CAWI (0/30)
Complete interviews (%)	29.49	23.66	9.80	13.31	39.37	24.90
Total number of eligible cases (N)	685	1,365	2,592	1,728	1,496	1,008
Chi-squared-test (χ^2)	610.15***					

%= relative frequencies; χ^2 = Chi-squared-test; Level of significance: * $p < 0.05$ | ** $p < 0.01$ | *** $p < 0.001$; Source: GGP Pilot, own calculations

Primary findings from the German incentives experiment:

- Comparison between F2F and P2W reference groups with 5 Euro unconditional incentive show higher response rates for the F2F group, indicating that, ceteris paribus, people are more likely to participate if an interviewer is present.
- An incentive (unconditional or conditional) appears to increase the response rate, as the experimental group CAWI (0/0) without any incentive reveals the lowest response rate compared to all other experimental groups.
- The highest incentive groups P2W (5/25) and CAWI (0/30) with each 30 Euro incentive rank among the highest response rates, however, it can be assessed that an unconditional incentive preceding the unconditional incentive enhances the response rate. This could be due to generated trust as people might be more confident to receive the promised monetary reward when they already received a share. The face-to-face follow up in the P2W (5/25) group might additionally increase the response rate.
- For the same amount of money, prepaid incentives work better than post-paid incentives (as for the group P2W ref. (5/0) and CAWI (0/5).

Overall the displayed response rates point in the direction that incentives increase participation, that unconditional incentives work comparatively better than conditional incentives and that a mixture of unconditional and conditional incentives with a high monetary reward of about 30€ as in group P2W (5/25) motivates people the most to participate in the study. Under these conditions, CAWI outperforms F2F, the European Social Survey (27.6%) and the recruitment survey for the GESIS online panel (31.56%)².

3.3. The Croatian Experiment: Comparing Reminder Strategies

Fieldwork Processes and Outcomes

The fieldwork in Croatia was conducted by IPSOS. The processing of the CAWI invites and sending of reminders was done in singular batches and led to notable surges in activities around the receipt of letters with almost all CAWI interviews conducted within the 6-week period, compared to around 10 weeks in Germany where CAWI completions were staggered. There were a considerable number of complaints and inquiries with regards to the financial incentives given that they were delivered in the form of cash, in the enveloped alongside.

² <https://dbk.gesis.org/dbksearch/download.asp?db=E&id=54261>

Face to Face fieldwork in Croatia was relatively poor in performance given that the sample was national rather than geographically clustered which disadvantaged the fieldwork team whose travel time and logistical constraints increased substantially. As in Portugal, contact rates were low and there was a high no-contact rate within the F2F fieldwork component. There were substantial fears at the outset that the low internet penetration rate of around 72% would require significant F2F follow ups in Croatia, this was not realized. CAWI responses were high and the resulting follow ups were very difficult for interviewers to complete as respondents were very demotivated and not inclined to cooperate. This is despite of follow ups being conducted close to the finish of the online component. Interviewers reported that, when a F2F interviewer requested an interview with a CAWI non-respondent, a common response was to either say they had just completed the online survey or would do so shortly rather than have a face to face interview. This suggested a strong preference for the online mode.

Response Rates

The Croatian experiment was carried out with one further experimental group where people receive the 5 Euro prepaid incentive like in the reference group but vary in the timing of receiving the reminder. Those in the experimental group received their reminders one week apart rather than two.

Table 6 - Response Rates for Croatia

	F2F ref. (5/0)	P2W ref. (5/0)	P2W (5/0, weekly)
Complete interviews (%)	27.67	49.45	51.93
Total number of eligible cases (N)	600	1,450	1,450
Chi-squared-test (χ^2)	107.65***		

%= relative frequencies; χ^2 = Chi-squared-test; Level of significance: * $p < 0.05$ | ** $p < 0.01$ | *** $p < 0.001$; Source: GGP Pilot, own calculations

In Croatia the response rates show that (unlike in Germany) with the same unconditional incentive (5/0) people are more likely to participate in the P2W reference mode than in the F2F reference mode. This is higher also than the ESS response rate of 45.7% recorded in 2015. A shortage of the timeframe in between reminder letters from two weeks to one week only slightly increases the response rate (51.9%). A possible explanation could be that people who do not want to participate ignore reminder letters after one as likely as after two weeks and the span in between letters does not affect their decision that much. One important consideration in this regard however is the time between the first letter and the engagement of fieldwork interviewers with face to face follow ups of online non-respondents. Compressing the time between reminders, subsequently reduces the time between the initial letter and the engagement by the interviewer. In Germany, this was reported as an obstacle to the engagement of such non-respondents and so compressed reminders could be efficient in this regard.

3.4. The Portuguese Experiment: Comparing Respondent Selection Strategies

As reported in section 3.1, fieldwork in Portugal was very problematic. The fieldwork agency recruited around 10 interviewers to conduct the fieldwork, but this proved to be insufficient given the level of contacts required also for the CAWI component of the experiment. In Portugal,

between 1st May 2018 and 1st November 2018, the fieldwork agency therefore recorded 9,643 contact attempts. This suggests a contact rate of approximately 78 houses per day. The fieldwork company indicated that from the initial recruitment of interviewers, only 8 remained actively engaged with the project and subsequently they were contacting around 10 households per day. This is a credible rate of contacts, but the interviewers reported persistent problems with access to properties, particularly apartment blocks and those residences which were ultimately uninhabited. This both slowed contact attempt rates and decreased the success rate of contacts. Ultimately, the fieldwork agency underestimated the work required for CAWI recruitment under the technical guidelines and the level of staff needed to deliver this. The tendering guidelines were also unrealistic on the time required to conduct such levels of contacts, especially given that fieldwork is largely dormant for July and August each year.

Response Rates

The additional experimental group in Portugal was a variation how the interviewer left the envelope that means he made no reference to the fact that there was a 5 Euro voucher inside. As already mentioned, and explained in section 3.1, to calculate the response rate for the experimental groups we had to estimate the proportions of “eligible” and “not eligible” sample units. The findings presented should be interpreted with care.

Table 7 - Response Rates for Portugal

	F2F ref. (5/0)	P2W ref. (5/0)	P2W (5/0, voucher hidden)
Complete interviews (%)	11.70	9.81	7.89
Total number of eligible cases (N)	1,700	2,304	2,015

%= relative frequencies; Source: GGP Pilot, own calculations

In Portugal only moderate differences between the three experimental groups are noticeable since the overall response rate is very low in Portugal. The F2F reference group shows a higher response rate than the P2W groups. Comparing the two P2W groups reveals that making no reference to the 5 Euro voucher in the handed envelope to the household P2W (5/0, no reference to voucher) leads to the lowest response rate. When people are unaware of the incentive, the motivation to pass on the invitation letter to the selected household member is lower and the completion rate is therefore much less.

4. Conclusions and Recommendations

Opening GGS to web mode: Goals and design of the pilot study

The pilot study of the Generations and Gender Survey (GGS) had the purpose of first testing the revised questionnaire and second assessing ways in which the fieldwork process of the GGS can be improved in terms of survey methodology. It focused on implementing a CAWI mode in a mixed-mode design as well as on a centralisation of fieldwork organisation in terms of questionnaire programming and data hosting. The pilot was funded by the “Generations and Gender Programme: Evaluate, Plan, Initiate (GGP-EPI)” and the Federal Institution of Population Research Germany (BiB).

The background of the mixed-mode design assessed by the pilot study, was the increasing possibilities provided by the internet for survey research and proceedings in CAWI methodology. While the GGS questionnaire was originally designed for personal face-to-face interviewing, a main goal of the pilot was to test how well this questionnaire can be applied in CAWI mode, at least in a cross-sectional perspective. Additionally, ways of centralizing fieldwork organisation promised to make survey research more efficient, faster and more comparable between participating countries. Before we summarise findings and draw conclusions regarding which modes are recommendable for a future round of GGS data collection we restate what the experimental study has assessed and what it hasn't, since this has implications regarding the limitations of any conclusions we can draw.

What the pilot study has investigated: Technical Options Analysis

All of the designs compared in the pilot study used a common version of the GGS questionnaire: a version that represents the state of revision in early 2018, which was somewhat close to the final version but is not identical with it. They all used a questionnaire programming in Blaise that was provided centrally by the Central Coordination Team at NIDI and that fieldwork institutes were required to use. All web respondents had the option of interrupting an interview whenever and as often as they decided to and to continue at any point in time within the next days (during the phase of web interviewing). All data collected in all interviews was uploaded to a central server at NIDI in The Hague where it was hosted. Interviewers used an App called “Data Entry Program (DEP)” by Blaise on their computers for downloading and operating the questionnaire as well as for uploading the data to the NIDI server. Online respondents used an online questionnaire in their country language, hosted by NIDI.

Our pilot study has also compared two modes of data collection: The reference mode was the for GGS customary mode of personal face-to-face interviewing. The alternative was a so-called “push-to-web” design: a sequential mixed-mode design, consisting of a self-administered web interview, combined with a follow-up by personal face-to-face interviews among those who had not participated online. This was done in an experimental design in the three countries Germany, Croatia and Portugal. Each target person received an unconditional incentive of 5€ (or 40 Croatian Kuna), already with the invitation letter or (in Portugal) with the interviewer's first visit. For “push-to-web” groups, two reminder letters were sent out after the first invitation, the first one 14 days, the second one 28 days after the initial invitation. Some days after the second reminder an interviewer would start contacting the remaining target persons who, until then, had not participated.

Within each country one additional country-specific experiment was conducted, each varying a condition for the data-collection in mixed-mode. In Germany, four alternative ways of offering incentives were tested, varying the amount as well as the timing of the donation (no incentive, 5€

conditionally post-paid, a combination of 5€ pre-paid and 25€ post-paid, and 30€ post-paid). In Croatia, a weekly time interval for the reminder letters was tested (instead of the biweekly time interval). In Portugal, an alternative for defining the target person within a selected household was tested.

What the pilot study hasn't investigated: Limitations of the design

The pilot study did not assess the following:

- The pilot study did not test any longitudinal design. There was no second or third wave conducted, as the GGS pilot study design is demanding, and accordingly no evidence collected regarding issues such as attrition. The only evidence collected is a question asked in Germany (at the end of the interview) regarding the respondent's consent to store contact information for a potential later re-contact.
- The pilot study did not test modes in any other country than the three countries mentioned. Given the country-specific findings found for Germany, Croatia, and Portugal, it must be assumed that any fourth country comes with, again, different conditions, presumably leading to country-specific effects. The three countries have been chosen in a way that much of the heterogeneity of Europe would be captured; however, none of the three countries can offer insights for any fourth country with certainty.
- The pilot study focused on the age range 18 to 49, which excludes elderly people aged 50 to 79, which have been included in earlier rounds of data-collection. These younger generations are presumably more mobile and accordingly harder to reach at home, but, on the other hand, mainly familiar with the internet and web interviewing. For elder generations, the results are not transferable.
- The pilot study did not test any other design than the ones mentioned. It did not, for example, assess a concurrent mix-mode design or a mixed-mode design, combining web interviews and paper-and-pencil interviews. It did not test other incentives or reminder strategies than those mentioned.
- The pilot study did not use the final version of the GGS questionnaire, as it will be used in 2020 in the new round of data-collection. It used the state of questionnaire revision in early 2018, which represented a significant improvement and shortening, compared to the earlier GGS; however, it also was still longer and more complicated than the final version.
- The pilot study did not test any data-collection under the condition that national fieldwork institutes would program the questionnaire or host the decentralized data on their own. Issues such as data cleaning or post-harmonisation of data for comparability are probably (and were supposed to be) less present than they used to be in former GGS surveys. On the other hand, technical problems linked to the transmission of data from fieldwork institutes to NIDI or linked to the necessary adaptation of fieldwork institutes to Blaise software as well as organisational problems, such as institutes not participating in the tendering process due to the demanding technical requirements, may affect the outcome of the pilot study in ways that were not an issue in previous designs for GGS data-collection.

Opportunities and limitations of involving web interviews as a mode for the GGS data-collection

In a lot of ways, the pilot study supports the consideration that web mode, as a component within a mixed-mode design, is an interesting alternative to be used for future data-collection in GGS. A central advantage of the web mode compared to personal interviews is the cost reduction. However, this must be balanced against the administration of a second mode which also introduces extra fixed costs, these are not negligible.

All in all, the data quality provided by a “push-to-web” design is decent. It is not decreased in ways that would rule out a mixed-mode design. In most aspects, it is actually very comparable to face-to-face mode or even improved. It is comparable, for example, regarding coverage and selection biases as well as regarding item non-response and measurement biases. The web mode allows for a better representation of the young and mobile generation in the GGS because they have a much higher internet penetration and their response rates are lower in face-to-face than for older cohorts. The response rates in a push-to-web design are overall slightly reduced. However, they can be even higher than in face-to-face if the push-to-web is combined with a higher incentive in the right combination of pre-paid and post-paid or if the right structural and cultural context is in place such as in Croatia. Increasing the incentive would easily be possible since it would still – or even more so – make push-to-web more cost efficient.

The period of fieldwork can be similarly long as in a pure face-to-face survey – potentially shorter but potentially also longer. It could be shorter since the data-collection in web mode is faster so that after the web phase relatively few target persons remain for being interviewed. It could be longer because the target persons remaining after the web phase for follow-up face-to-face interviewing are partly reluctant to be interviewed at all, and they may be annoyed by the repeated invitations and reminders, so that the response rate during the follow-up is clearly reduced. The crucial questions that decide upon the length of the fieldwork period in push-to-web survey is whether there are enough interviewers in the follow-up phase and whether they are active and motivated, despite the lower response rates. It should be considered, whether the marginal added value of follow-up face-to-face interviewing is cost efficient and contributes significantly to response rates and data quality.

Despite the, in many ways, encouraging findings, outlined above, the pilot study has also revealed a few problems related to CAWI as a mode of data-collection that need to be taken into account, when considering a mixed-mode design involving web interviews.

Firstly, a problem that the pilot study only finds hints for is the **risk of attrition**. As emphasized before, the pilot has neither conducted a second wave nor has it carried out a re-contact with the invitation to participate in such a second panel wave. Accordingly, there is no empirical evidence how high the attrition rate would be for a GGS conducted in web mode or in a push-to-web design. However, already the fact that there is no empirical evidence emphasizes the risk that the attrition could be quite high: Whereas the involvement of interviewers in the “classic” face-to-face mode has proven to tend towards reasonable attrition rates and offers the chance to limit attrition further by training and motivating interviewers accordingly, this is not the case for the self-administered web mode. The absence of an interviewer and of the identification of the survey through a personal contact makes it likely that attrition would be increased. This assumption is further emphasized by an empirical finding of the pilot in Germany: The ratio of respondents giving consent by the end of a web-based interview that their contact data may be stored for a re-contact is much lower than after a face-to-face interview: Between 27% and 41% of respondents (depending on the incentive offered) give their consent in web mode, while 66% do so when asked by an interviewer. Of course, the timing when and the way how this consent is asked for could be modified in a way that more people agree to be re-contacted. (For example, respondents could be asked by the beginning of the interview.). The improvements of the questionnaire design and questionnaire programming, as described above, might increase the consent rate. Nevertheless, the comparison of the consent rates between the modes supports the assumption that the attrition for people interviewed online would be clearly higher than for those interviewed personally. This, again, bears the risk that in a push-to-web design, the second and third panel wave would be affected by selective drop-out. And it bears the risk that the consecutive waves would have very

low case numbers, potentially even too low for any longitudinal analyses or for continuing the panel after wave two. To examine this issue further, adequate benchmarks should also be established that reflect the compound effect on representivity of both non-response and attrition. Secondly, the pilot reveals that the approach of conducting GGS as a push-to-web survey has led to dissatisfying results in Portugal. The reason seems to be that web interviews can only be conducted effectively in a country with a **sampling frame that allows individual sampling** by mail or e-mail. If such a sampling frame does not exist and address-based sampling must be applied neither the cost saving advantages of web mode nor satisfying response rates are likely to be achieved. In such a case, as experienced by the pilot study in Portugal, interviewers need to visit each household in person – where a household member may or may not be at home to answer the door –, identify a target person within each household – who may or may not be present during this visit – and then deliver the invitation letters for participating online personally. This procedure has shown to be neither efficient nor effective for motivating people to participate in a web survey. Accordingly, also a push-to-web design is not recommendable in such a country context.

A problem of web mode that is also documented by the pilot study is a substantial share of **break-offs** in the web interviews. Since personal interviews are rarely abandoned, the break-off rates are clearly higher in web mode than in face-to-face. This increase can be reduced by a decent post-paid incentive, in particular when offered in combination with a small unconditional incentive upfront. But the break-off rates still remain quite considerable. They are the outcome of a questionnaire design oriented at face-to-face, implying a much longer interview and more complicated question wordings than are advised for web surveys. CAWI interviews are usually not supposed to be longer than 20 minutes while the GGS (pilot) questionnaire has a duration of 52-70 minutes (concluding from Kazakhstan and Germany). It can be conceded that the questionnaire has been further shortened and improved after the pilot study, regarding the weariness of its loop-based structure as well as regarding the annoyance of some of its contents and wordings. Also, the technical side of programming and formatting has been improved. Accordingly, there is reason to hope that the experience of a self-administered interview will be less annoying and more pleasant in the future rounds of data-collection than it has been in the pilot, so that break-off rates will be lower. However, this is only a presumption. Nevertheless, it still must be assumed that the share of people abandoning the interview before its end would be high in a GGS web interview. Similarly to low response rates, high break-offs are problematic in particular because they bear the risk of increasing biases: It is quite plausible that the ones who abandon the interview before its end are a selective group, leading to an under-representation of this group in the net sample – in particular if the loop-based questionnaire structure increases the length and weariness of an interview for a particular group of people, such as respondents with many children and complex family biographies. More than a low response rate, a high break-off rate must raise the concern that there may be a strong bias towards the under-representation of these social groups with family biographies that are particularly interesting for family-demographers. It has to be ensured through the questionnaire design that respondents cannot shorten the interview by under-reporting the number of their children or similar ways of misreporting.

Further risks of the push-to-web design that have been observed during the pilot study rather require awareness and according measures or fieldwork monitoring than affecting the principle choice for or against push-to-web. One such risk is related to the **performance of interviewers in the follow-up**: Given that only those cases are to be contacted by the interviewers who have not participated in web mode the expectable response rates for this follow-up are much lower than in a fresh sample in a pure face-to-face survey. Also, the fact that these target persons have already been contacted three times will presumably lead to a lower motivation, if not to significant

annoyance. Accordingly, the face-to-face follow-up is less motivating for the interviewers than other assignments. And interviewers working as freelancers on two or more assignments at the same time may decide to invest no or little activity into the follow-up. This must be avoided by all means for example by avoiding overlap with other survey, by according payment schemes for interviewers or by fieldwork monitoring.

A second such risk is the **correct entry of the ID number** by web respondents. If a respondent failed to type in her or his ID number, which in Germany was a nine digit alphanumeric ID, into the online questionnaire this did not only cause difficulties in later on identifying the case and linking it to the correct contact protocol record, potentially leading to a loss of data. It also caused the risk that this respondent would not be able to continue with her or his interview after deciding to interrupt it. In that case, the respondent, when returning to the interview (then presumably with the correct ID) would either need to start the interview all over or abandon the interview, which is more likely. Therefore, the risk of mistyped IDs needs to be avoided by programming a filter in the Blaise questionnaire that only allows valid ID numbers to be entered.

The optimal way of designing a GGS in push-to-web design

Irrespective of the risk of attrition for the longitudinal design of the GGS and other potential downsides of involving web mode into GGS data-collection, the pilot allows conclusions to be drawn for an optimal design for a push-to-web survey if push-to-web should be implemented:

- Incentives should be paid both unconditionally pre and conditionally post-paid. The pre-paid incentive should be rather low (e.g. 5 €) in order to efficiently improve the target person's general attitude towards the survey. The conditional post-paid incentive should be more generous, offering a good reason to remain patient and complete the interview. The optimal amounts will certainly differ between European countries and depend on factors such as survey saturation, cultural factors and purchasing power. Also, institutional and legal issues have to be considered: In some countries it is, for example, not allowed to send money in cash by mail.
- Reminder letters should be sent out one week after the initial invitation and again one week later.
- High motivation and activity level of interviewers in the follow-up phase needs to be ensured – by taking according measures, as discussed above.
- It has to be ensured that the respondent is technically able and encouraged to interrupt the interview any time he/she wants or needs to and continue at any time in the near future, as often as she/he wishes.
- The programming of the questionnaire should have a filter allowing only 'valid', that is existing IDs to be typed in, so that respondents mistyping their ID are warned and forced to correct it.
- The design and programming of the questionnaire should ensure that the biographical information is collected in a way that neither respondents nor interviewers have a reason to misreport (for example to under-report the number of their children in order to reduce the duration of the interview) and that the biographical information provided by the respondent cannot be corrected by him later on.

For further details, independently from the mode in which GGS is conducted, the GGS fieldwork guidelines need to be considered. These include aspects such as wording and formatting of invitation and reminder letters, sending out letters shortly before the weekend, high importance of panel maintenance between the waves, appropriate translations of the questionnaire into national

languages etc.

Centralised or decentralised organisation of fieldwork

The pilot has tested a much more centralised organisation of the fieldwork: The programming of the questionnaire was provided by NIDI. Interview data was uploaded to and stored at a server at NIDI, centrally. These measures were considered because it was expected that they would make fieldwork more efficient, because it would allow a more efficient quality control of fieldwork and because it would make data from various counties more comparable and reduce the need for post-harmonisation. These expectations were in many ways confirmed. However, also risks of a centralised fieldwork were identified:

- The wording and structure of the questionnaire can and should in general be centralised. However, it cannot be fully identical for every country; instead, functional equivalent wordings and operationalizations need to be found. This does not only require deviations from a word-by-word translation. It may also imply modifications in the structure of the questionnaire, such as an additional question or dropping a question. This can be necessary, for example, if a particular national educational system requires an extra question for defining the ISCED level of a respondent or if a particular topic is violating cultural norms in a way that keeping the question would make interview break-offs very likely.
- The central programming of the questionnaire in Blaise has strong advantages in terms of efficiency and of comparability of the national datasets and their formatting, and in terms of reducing the need for post-harmonisation. However, this central programming is a time-consuming, challenging and highly important task. The centralisation makes it central to the success of the project since any mistake in the programming would contaminate data in all national surveys at the same time. This task is the responsibility of the Central Coordination Team at NIDI – in its current size – bears the risk that the identification and correction of mistakes may not succeed in all cases due to lacking personnel capacities, in particular when several countries start fieldwork within a short period of time and require small national adaptations simultaneously. If this plan is pursued, a relevant personnel reinforcement of the Central Coordination Team needs to be considered.
- Also, the central hosting of data has at least two sides to it: On the one hand it enables a quite effective central monitoring of fieldwork activities as well as a fast processing of the collected data. On the other hand, the organisation and coordination of the fieldwork activities can become more complicated, slow and error prone. This affects the fieldwork institutions in charge of this organisation, but also interviewers and national teams. Examples include technical problems of fieldwork institutes and interviewers when operating Blaise on their devices, technical problems when transmitting data, time lags and the need for extra communication for identifying which cases have actually been interviewed (so that interviewers need to be paid for) and which haven't (and need to receive reminder letters) etc.

Recommendations

Work Package 2 of the EU funded project “Generations and Gender Programme: Evaluate, Plan, Initiate (GGP-EPI)” had the assignment to identify potential technical and methodological innovations and improvements for future rounds of data-collection in the “Generations and Gender Survey”. Based on the state of research in survey methodology and on theoretical reflections of the team, several innovations, as described above, were considered and accordingly tested in the here presented pilot study. After collecting experience while conducting the pilot study as well as after analysing the data collected through the pilot and assessing its findings, the team of Work Package

2 wishes to share its insights with the Consortium Board of the “Generations and Gender Programme” as well as with the project partners of GGP-EPI. These are subsumed in the following recommendations.

Regarding the centralisation of fieldwork activities, we recommend:

- The wording and structure of the questionnaire should be centralised. However, wordings and operationalisations need to be functional equivalent instead of standardised and identical across countries. There must be room for few minor well-reasoned country-specific adjustments.
- The central programming of the questionnaire in Blaise should be undertaken but linked to a relevant personnel reinforcement of the Central Coordination Team at NIDI.
- Also, the central hosting of data should be offered to national teams, without being mandatory and alternative arrangements should be available to countries without compromising the centralized nature of the operations.

Regarding the use of online data collection:

- Online data collection can be considered for reducing survey costs – and eventually for improving coverage. Web mode should particularly be considered in countries where costs for face-to-face interviews are high and according funding is unavailable.
- Web mode should not be implemented if there is no sampling frame that allows individual based sampling and an initial contact by mail or e-mail, without further testing. In countries where no such sampling frame exists, further piloting and testing of push to web strategies is required.
- A sequential mixed mode design is only recommendable if a part of the cost savings is reinvested into more generous incentives. Generally, all advice given above regarding “The optimal way of designing a GGS in push-to-web design” should be taken into account.
- It must also be stressed that face-to-face interviews are very expensive, at least in many European countries, and it is imaginable that some partners will not be able to find sufficient funding for a data collection in CAPI mode. In such cases, a data collection in a mixed mode design, including online interviews, also in the first wave, or even telephone interviews, can and should certainly be considered. In these cases, however, alternative strategies should be found for keeping panel attrition low, such as asking consent at the beginning of the interview rather than at the end. Also, a larger net sample should be considered.
- A further recommendation for keeping panel attrition low, independently from the choice of mode, are intensive panel maintenance activities, in particular maintaining frequent contact with respondents between the waves. Given that GGS waves are three years apart, this is more important in GGS than in other panel studies. Approximately every six months a reason should be found for contacting respondents (by post or by e-mail), refreshing their memory of the GGS and ensuring that they haven’t relocated, for example by sending Christmas cards, brochure with exemplary GGS results or by conducting very short interim surveys.