The Generations & Gender Programme: Evaluate, Plan, Initiate

Deliverable 2.1: GGP Technical Case and E-Needs

Den Haag, October 13th 2018

Tom Emery, Susana Cabaco, Peter Lugtig, Vera Toepoel,

Detlev Lück, Robert Naderi, Martin Bujard & Almut Schumann

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 739511.

www.ggp-i.org [@GGP_I](https://twitter.com/GGP_I)

The Generations and Gender Programme (GGP) was launched in 2000 by the Population Unit of the United Nations Economic Commission for Europe (UNECE) and has been coordinated by the Netherlands Interdisciplinary Demographic Institute since 2009. It is a social science Research Infrastructure that provides harmonized, large-scale, longitudinal, cross-national panel data on individual life courses and family dynamics. Over time, the GGP follows respondents through relationships, marriages, parenthood, divorces, deaths and many of the opportunities and challenges that people face along the way. It then tracks the causes and consequences of these events at the individual and societal levels.

In 2016, the GGP received funding for the Horizon 2020-funded Generations & Gender Programme: Evaluate, Plan and Initiate (GGP EPI) to further pursue the programme's aim to continue to expand its potential for scientific advancement, innovation and societal impact (2016-2019) (Grant Agreement No 739511). The GGP EPI project provides the GGP with the means to institutionalize and further formalize key processes required to position GGP for a successful application for inclusion in the ESFRI Roadmap 2020 as a sustainable, world-class research infrastructure accessible to all researchers in Europe and beyond. Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as:

GGP (2018) GGP Technical Case and E-Needs, Deliverable 2.1 of the GGP-EPI project funded under the European Union's Horizon 2020 research and innovation programme GA No: 739511. Available at: www.ggp-i.org/ggpepi

Contents

Preface	4
1. Introduction.....	4
2. Previous fieldwork operations in the GGP	4
3. Fieldwork in other Social Science Infrastructures	6
4. The Push to Web and the GGS.....	8
5. Piloting a new fieldwork model.....	11
6. Technical description of the experiment	14
7. Summary.....	22
8. Bibliography	23

Preface

This deliverable is designed to evaluate the current technical state of the GGP infrastructure and possible amendments and developments that can help increase its performance as part of its infrastructural development. It outlines a technical evaluation that will be used to adapt the fieldwork and implementation guidelines and enable the initiation of a new round of GGP. This technical evaluation will also help inform planning elsewhere such as the requirements for financing and governance structures within the GGP, as well as the capacity of the GGP to increase its socio-economic and scientific impact.

1. Introduction

As a social science infrastructure, the GGP finds itself in a rapidly evolving and developing field that is adapting to fundamental shifts in the data that sustains it. It is the responsibility of research infrastructures to anticipate and manage these fundamental shifts and ensure that scientists can incorporate new data technologies into research and thus accelerate the scientific process. In this report we focus on the data challenges facing the social sciences and how the GGP can adapt to better serve the research scientists examining family dynamics and relationships.

We start with a historical overview of the GGP infrastructure and the community of population scientists and statistical offices out of which it evolved. The GGP in some form or other has existed for several decades yet the way it collects, and processes data has changed dramatically in that time. We then consider the social science landscape of today and how the GGP fits in the landscape of other RI's in the social sciences. We consider how surveys such as the Survey for Health, Ageing and Retirement in Europe (SHARE), the European Social Survey (ESS) and the European Value Study (EVS) have evolved in their operations and their strategies for adapting to changes in data science. The GGP has been cooperating with these RI's on these questions as part of SERISS (Synergies in European Research Infrastructures for the Social Sciences). As part of this project, ESS has led work on examining a 'Push to Web' framework. This framework, called CRONOS, seeks to explore the possibility of collecting some data from respondents via web interviews. It is this challenge which the GGP is also aiming to address. In the fourth section of this report we consider the fundamental advantages and disadvantages of a push to web and use this to outline the need for a piloting of a multimode GGP in a cross-national environment. This is followed by the design of the pilot and the technical description of its implementation. The experiment will take place in January 2018 and will inform the implementation of a new round of data collection in the GGP in 2020.

2. Previous fieldwork operations in the GGP

The Generations and Gender Programme evolved out of a long-standing community. In the 1970's and 1980's statistical offices from across Europe sought greater standardization in the way they collected data on basic demographic processes such as marriages, divorces and births. This demand for standardization was intensified by the observation that fertility was falling, and marital instability was rising rapidly (Van De Kaa 1987). There was an obvious requirement for data that could help establish the causes and consequences of these developments in a cross-national context. National statistical offices of the time were already conducting national surveys on these issues but the United Nations Economic Commission for Europe (UNECE) sought to coordinate these efforts and ensure international comparability (Festy et al. 2002).

On a technical level the Families and Fertility Survey (FFS) focused on the collection of core survey items that were of common interest across countries facing declining fertility. This primarily focused on fertility histories and basic demographics of those aged 18-49 that would enable cross-national analysis of fertility postponement. This was a highly decentralized effort and relied on the cooperation of national statistical offices in the implementation of the survey. Fieldwork operations were entirely the responsibility of the statistical offices and there was limited academic involvement. Once collected the data were harmonized and standardized. The primary product of the FFS was initially standardized tables, common to demography and

not the distribution of micro-data. Over time micro-data was made available after the signing of multiple data agreements which were specific to each national statistical office. These data were then distributed via physical disk through the post. Today, the FFS can be downloaded from the GGS website.

Demographic changes intensified over the fieldwork period of the FFS with further developments in family formation but also rapid extensions in life expectancy, potentially altering intergenerational relations. There was also a significant scientific limitation with the FFS data in that it was cross-sectional and the evolving theories describing changing demographic behavior required longitudinal data (Francesco C. Billari, Philipov, and Testa 2009). Therefore, in 2000 the UNECE created the Generations and Gender Programme as an successor of the FFS. This included extending the data collection to the whole life course (18-79) and was designed to be a panel study with 3 waves with 3 years between each wave (Fokkema et al. 2016). In this design there was to be more ex-ante harmonization of survey instruments. Yet this was difficult given the success of the FFS project itself.

Despite this, several countries continued with their own family and fertility surveys instead of opting for the GGS. From a technical perspective, the FFS has also increased national capacity in fieldwork operations to the point at which they were operating independently of the central coordination efforts. In countries such as Spain, Canada, the Netherlands, Uruguay and Italy the FFS continued to be the model for data collection and the adapted GGP design was not adopted universally. Nevertheless, 19 countries did field the Generations and Gender Survey and 13 of these then fielded a follow up second wave. Unfortunately, the financial crisis struck just at the point in time when the GGS was reaching maturity and subsequently only five countries were able to complete the full three waves. This has brought the costs of the GGS into sharp focus. A decentralized design, large sample sizes, and a requirement for face to face interviews mean that the GGS is a relatively expensive survey to run.

To compound these issues, the GGS extension to older generations was undermined by the development of SHARE. As a survey explicitly designed for the studying of ageing, SHARE was able to fill the data needs of researchers on ageing and thus undermine the arguments for extending the GGP to older generations (Börsch-Supan et al. 2013). The extension of the GGP sample from 18-49 to 18-79 doubled the required sample size from 4,000-5,000 to 8,000-10,000. The GGP had extended its thematic scope and increased the cost to national statistical offices, but this only served to reduce compliance and over extend national teams. A final issue that exacerbated the problems was the move of the central coordination team from the UNECE to NIDI in The Hague in 2009. This was done due to an inability of the UNECE to participate as a lead in a European Commission funded project and the conflicting aim of integrating the GGP within the emerging European Research Area (ESFRI 2006). At a time when the GGP was in-need of strong strategic leadership to help establish the new data collection, it had to simultaneously relocate and reestablish its infrastructure in an entirely new location.

Regarding fieldwork, this had significant implications as national teams lacked oversight and leadership from the central coordination team, leading to high attrition at the microlevel but also the failure to secure sustainable funding. In terms of operations, national teams were provided with a core questionnaire and fieldwork guidelines and told to implement the survey with limited oversight of sampling, translation, processing and fieldwork. This led to large deviations over the course of the data collection process and limited capacity for recourse amongst the central coordination team. It also led to a considerable slow-down in the processing of micro level data that was collected. Some of the data collected in 2004 went unprocessed at the central offices for a period of 10 years and was only released once the central coordination team had re-established itself and acquired sufficient resources to complete its responsibilities.

Despite the evident problems in the data collection for the Generations and Gender Survey, the data has had a very large impact on the field of social demography and led to significant insights that have been used to inform policy (Kotowska, Matysiak, and Mynarska 2006; Buber-Ennsner and Neuwirth 2013). The demand for the data is high and the thematic issues it covers remain prescient (King 2011). At the core of the FFS and GGS lie data on partnership and fertility histories that help social demographers monitor developments in

demographic behavior at the micro-level but with the ability to capture cross-national differences. Surveys such as the ESS and SHARE have shown what is achievable through a centralized survey model and the production of high quality, comparable socio-demographic data is of high importance and a priority for the future of the European Research Area.

3. Fieldwork in other Social Science Infrastructures

The GGP's decentralized and organic origins help understand its current structure and technical design. However, they do not reflect the changing capacities and functions of a survey organization that have brought about by technological development. The FFS predates the World Wide Web and this was still evident when field began on the GGS in 2004. Since the GGP was founded there have been two European Survey Research Infrastructures emerge that have fundamentally reshaped how a cross-national survey can be conducted. Yet, the European Social Survey (ESS) and the Survey of Health, Ageing and Retirement in Europe (SHARE) have distinct approaches and philosophies.

The European Social Survey was established in 2002, two years after the GGP, and was focused on lifting the general quality of survey processes across Europe and measuring social attitudes and values. The ESS does deliver high quality scientific data on important topics but the aim of the infrastructure is far broader than this. Firstly, the ESS is aimed at in country capacity building. The ESS maintains strong national teams and requires a high degree of national level coordination as one of its founding aims was to increase standards in survey research and share best practices. This has been evident in many countries where the fieldwork procedures established by the ESS have bled through into other surveys and commercial practices, raising the quality of social data considerably. The second broader function of the ESS is the access to high quality comparative data for educational purposes. The ESS makes its data feely available to researchers and students at the click of a button and this has led to the infrastructure accruing more than 100,000 users, the clear majority of whom are students. At a time when demand for quantitative skills is increasingly rapidly, the ESS provides a simple yet high quality tool for teaching and learning of social statistics.

These broader aims are reflected in the structure of ESS. Its primary technical investments in the past 15 years have been on fieldwork processing issues to increase quality. The ESS work on translation and fieldwork management has set new standards in the field and been used to raise standards elsewhere. The ESS is designed for high and intensive investments in methodological issues¹ over and above delivering scientifically cutting-edge data in the most efficient manner (Jowell et al. 2007; Stoop et al. 2010). For the GGP, the ESS provides many benefits. On many occasions fieldwork agencies that the GGP employs have worked on the ESS and having done so is considered a mark of quality and trust. It ensures that fieldwork agencies are aware of modern standards and capable of producing data that meets the needs of scientists. This is vital given the dominance of commercial data collection amongst fieldwork agencies. For the Central Coordination Team, the ESS provides guidance and best practices on a wide range of issues for which it would be inefficient for the GGP to develop itself. Nevertheless, the ESS operational model differs from the GGP in so much as it deliberately avoids high centralization. This is because to do so would undermine its national capacity building efforts. It therefore does not make use of the latest technologies in cross-national implementation. It does not use a central server for the processing of data and nor does it have primary control over national level activities such as tendering and fieldwork.

SHARE, in contrast to the ESS, is much closer to the GGP in terms of its technical functions. The aim of SHARE is to deliver high quality data for scientific research on a predefined and socially pressing topic, with an emphasis on personal finances and health. In the absence of an ageing survey for Europe in the late 1990's, it is no coincidence that the GGP saw expansion into older generations as a simple extension of existing operations, such is the symmetry in their operations. Nevertheless, SHARE has evolved into a survey research

¹ For an overview of the ESS activities related to data quality assessment see http://www.europeansocialsurvey.org/methodology/ess_methodology/data_quality.html.

infrastructure that is specifically targeted at monitoring the health and finances of those aged over 50. It is a well suited infrastructure for this task because SHARE samples the whole population over 50. This is a difficult task as persons over 80 bring about fieldwork challenges that require specialized techniques. For example, the population over 80 invariably live in specialist accommodation which poses problems for sampling. They often have cognitive issues which make the completion of a survey complex from both a practical and ethical perspective. Mortality is also much higher in this age group and this requires specific panel maintenance procedures that are less prevalent in other surveys. SHARE has developed specialist techniques and technologies to address these challenges. It is a global leader in ageing research and the GGP benefits immeasurably from SHARE's contribution to survey research generally.

In contrast to the ESS, SHARE has no impediments to centralization and it has sought efficiencies through centralization where possible. The SHARE survey instrument is developed centrally, the translation into national languages is overseen by the center, the training of interviewers is directed centrally, interview trainers themselves are trained at the central office, the data from the field goes directly to a central office and harmonization, access and documentation are all managed centrally (Börsch-Supan, Hank, and Jürges 2005). This has led to high quality data that is processed relatively quickly. At the core of this system is the centrally controlled survey instrument which is deployed through Blaise software (<https://www.blaise.com>). On top of this software, SHARE have built many survey innovations that have improved data quality and opened new areas of research. This is no small achievement given that SHARE is now operational in 26 European Countries and requires the questionnaire to be translated into dozens of languages. The standards developed in centralization by SHARE over the past 10 years will be deployed in the new round of GGP data collection, not as an innovation but simply as a matter of course.

A further survey infrastructure with many close parallels to the GGP is the European Value Study (EVS). This study is conducted once every nine years with a cross-sectional sample. The long gap in between waves and its lack of a long term central coordination team that is dedicated to the infrastructure mean that the EVS faces many of the challenges that the GGP has faced. With low centralization and the inertia that comes with a long-standing project, EVS has not been integrated into the ESFRI roadmap process like ESS and SHARE have. As projects that were established in the last decade, SHARE and ESS could tailor their institutional development to meet the requirements of ESFRI, whilst EVS and GGP have relied on long standing institutional commitments, bringing into question their long-term sustainability. Nevertheless, this structure in many ways suits the scientific aims of the EVS. There is no requirement for a panel design and the target of the studies (values) generally change slowly over time. The hibernation cycle that EVS deploys is therefore exceptionally cost-efficient in many respects.

From a fieldwork perspective, EVS inherits a survey that has been fielded since the 1980's and the nature of its instrument, like the GGP, reflects this. If the EVS was designed from scratch today, it would likely look very different and take full advantage of the adaptive interviewing capacities in CAPI. Yet, the need to maintain the historical measures for comparability limit EVS' ability to rapidly adapt new technologies in ways that mirror the GGP's own challenges. Nevertheless, the EVS has made progress where possible, adopting the centralized translation process from SHARE and fieldwork management tools used by ESS. Yet given the content of the EVS, it has yet to make significant advances into web interviewing which would severely reduce costs and increase sustainability. EVS has been unable to adopt this approach for several reasons. First, it is cyclical infrastructure and it is exceptionally difficult for it to make long term investment in innovations itself. To test and pilot an online EVS would require significant funds and the central hub of EVS is not sufficiently large to accommodate such a project. Secondly the EVS, unlike SHARE and ESS, does not prioritize EU member states. Like the GGP it takes a broad definition of Europe and thus operates in more countries with sub-optimal fieldwork conditions. The adoption of a web approach in these contexts is exceptionally difficult, especially in ensuring comparability across countries and time.

Of the three infrastructures that have been discussed here, all of them are in the process of adapting to new technological and practical realities. Yet, none of them is well positioned to address the primary challenge in social research today which is the transition of interviews to an internet mode. The ESS has made tentative

advances in this direction with the CRONOS panel in SERISS but the ESS remains committed to implementing the gold standard in survey research which for the time being remains face to face. SHARE faces difficulty in a mode transition to Web for two reasons. First, their sample is old and internet penetration rates amongst this demographic, whilst increasing, remain low. Second, SHARE is very reliant on objective health measures that require a face to face interview such as cognitive and physical tests. The inability of SHARE to deliver a web survey is highlighted by the experiments being conducted in the Netherlands where an online survey was used in its place but was deemed insufficiently compliant with the core SHARE study. Finally, EVS doesn't have the centralized infrastructure to deliver a survey online and overcome the challenges posed by broad country coverage. The GGP is largely free of these constraints and is well placed to take advantage of a 'Push to Web'.

4. The Push to Web and the GGS

The broad advantages and disadvantages of a Push to Web strategy are well known and well documented (Millar and Dillman 2011; Lugtig and Toepoel 2016). By conducting interviews online, costs are reduced considerably as they remove the costliest component of fieldwork which is the interviewer. Internet surveys, on some measures, can also be preferable as they allow respondents to answer discreetly without interviewer interference (Jäckle et al. 2015). Total Survey Error approaches have shown that interviewers and fieldwork operations constitute a significant proportion of survey error and internet surveys circumvent these specific sources (Groves and Lyberg 2010). Revilla (2013) analysed measurement invariance and quality of composite scores in a face-to-face and a web survey and concluded that even using different modes of data collection, "as long as the exact same wording and scales are used and the samples are drawn randomly from the population, equivalent measurements can be obtained" (2013: 26). Internet surveys also open a survey up to more interactive and adaptive survey items and instruments such as the integration of social media, cognitive tests or iterative and modularized content (de Bruijne and Wijnant 2013). This flexibility allows for the possibility of immediate feedback during web interviews. Across six web survey experiments, F. Conrad and colleagues found evidence of the advantages of using prompts and interactive interventions in reducing speeding and straightlining (Conrad et al. 2017). In addition, the vast paradata generated during data collection can provide insights about potential issues affecting fieldwork and the quality of the data collected, both at a technical level (for example break offs, device used, software issues) and user level, allowing the possibility to learn more about user behaviour and missing data (Bluemke et al. 2017; Bruton-Smith and Tarling 2017). Of course, such advantages also come with risks. Sampling Frames in internet surveys are hard to establish and converting respondents to a web design replaces interviewer error with a new set of issues and complications (Bethlehem 2010). A recent study comparing ten web and face-to-face surveys from Africa, Asia and South America found that web samples are usually younger, more often male, higher educated than face-to-face samples – although the differences are smaller in countries with higher internet usage (Tijdens and Steinmetz 2016). They also severely limit objective measures, particularly in the field of health or interviewer observations which in the past have been a fruitful source of data for social surveys. On what regards measurement error, it is important to have in mind that respondents in web surveys can complete the questionnaires in different devices and a study has shown that measurement errors can differ between devices – larger on tablets and smartphones than on PCs (Lugtig and Toepoel 2016). Finally, whilst they are generally cheaper than face to face interviews, they are by no means free and the gains from internet surveys are often marginal rather than a punctuated equilibrium in the data collection process (Bianchi, Biffignandi, and Lynn 2017).

In the context of these advantages and disadvantages, the GGP is well placed. Regarding sampling issues, the GGP includes the population 18-49 which is more predisposed to internet surveys and amongst which there is a higher internet penetration rate (Poushter and Stewart 2017). It doesn't include objective health measures like those in SHARE, instead focusing on items such as sexual health, personal attitudes, beliefs and histories which have been shown to be better measured via the web (Vikat et al. 2007). The GGP is also a survey that is under high cost pressures and any attempts to ease those pressures should be considered. Finally, the GGP is

constantly seeking innovations in the nature of the data it collects. In the fields of demography and sociology, there is an increasing pursuit of innovative and creative operationalisations to get at concepts that previously lay outside the capacity of research infrastructures to measure. For example, detailed geo referenced data, social media data, internet behavioural data and broader technological paradata from the 'internet of things' (Bluemke et al. 2017). These data are reshaping theories and concepts in social theory and the GGP needs to be a platform upon which these can be delivered. A survey with the capacity for online interviews better facilitates this trend in the field with more reactive and adaptive methodologies that can better service the scientific community. The research cycle and initial costs of a face to face are prohibitive for such innovations. It is therefore the intention of the GGP to explore a Push to Web strategy and examine its feasibility.

Work Package 2 (WP2) of the Evaluate, Plan, Initiate project focuses on ways in which the fieldwork process of the Generations and Gender Survey (GGS) can be further centralized and standardized. Whilst this will involve the creation of a central instrument, the main goal of WP2 is to work out guidelines on how this instrument is to be developed and fielded in a diverse range of countries at the end of the project in 2019. This will likely require the instrument to migrate between modes (i.e. face to face, web, phone, mobile) which brings about technical and methodological issues with regards to comparability. In order to inform the decisions about future rounds of fieldwork, an experiment on fieldwork strategies forms part of WP2. To ensure that the work conducted in this work package supports the new round of fieldwork, it is necessary to summarise the key developments and trends that are expected in the design of the GGS after 2019:

- 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, programmed and provided centrally. The key reasons for doing this are:
 - a) to ensure that the questionnaire is administered in the same way across countries, eliminating the need for post-harmonization
 - b) to make sure that GGS data can be prepared for data release quickly and efficiently
 - c) to better monitor fieldwork in each country.
- 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, meaning that data collection will take place in as much the same way as possible across countries.
- The fieldwork procedures should be flexible. 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.

The 4 principles above guided the practical choices in fieldwork procedures for future GGS rounds as well as the following design considerations:

- In some countries, the GGS may use individual 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 others' 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 paper-and-pencil or internet 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 5-50% depending on e.g. the topic of the survey, the number of reminders sent, and level of incentives offered.
- Using F2F as a follow-up mode for non-respondents to an online-survey can increase response rates, and reduce sample selectivity. It was decided that such a consecutive mixed-mode design of Web -> F2F (Push to web (P2W)) could form 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:
 - a) In countries with lower internet penetration rates, more interviews will have to be conducted F2F.
 - b) 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.
 - c) 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. email addresses) in wave 1, the proportion of web interviews will likely increase in later waves.
 - d) Over time, also internet penetration rates will change which will also affect the proportion of web interviews.
 - e) 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.
 - f) This implies a risk of measurement 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.
 - g) In order to minimize such measurement effects, the GGS questionnaire should be redesigned and then tested in the experiment. There will be two versions of the GGS questionnaire: a face-to-face and web version. We will try to minimize differences between the two versions (mainly instructions to interviewers) to minimize measurement effects.
 - h) Also, measurement effects may be reduced by conducting sensitive parts of the F2F interview by CASI (that is, self-administered), instead of CAPI.
 - i) Scientists using the GGS data need to be instructed to conduct sensitivity analysis for mode effects when analysing the data.

5. Piloting a new fieldwork model

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 and data quality. This must be achieved by comparing respondents who are approached using a push to web (P2W) approach to those who will be interviewed using the traditional F2F methodology that was used 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 and the results will be differentiated by country. These countries should represent much of the diversity of European countries, particularly regarding availability of register data for drawing samples, the costs for face-to-face interviews and the social acceptance of survey research in its various forms. The choice of countries also should rather over- than under-estimate the difficulties and costs of a new GGS by providing suboptimal conditions. After discussion, the three selected countries are: Germany, Croatia, and Portugal.

For saving costs, in each country a region will be selected for conducting the experiment. This region needs to include urban as well as rural neighbourhoods. And again, it should provide suboptimal conditions for online and F2F research, compared to other regions within the same country. In every country, we will conduct an experiment in which respondents will be randomly assigned to the F2F-only condition, or a push-to-web condition. In each participating country an additional country-specific methodological experiment will be executed to further investigate how a push-to-web design should be implemented. These will be:

- In Germany, different strategies for providing incentives will be tested.
- In Croatia, different strategies for sending out reminders will be tested.
- In Portugal, different strategies for selecting a respondent from the people living at the same address (within random route sampling) will be tested.

Research Questions

The methodological experiment pursues several goals. Above all, the intended new mode for GGS research, that is, a push-to-web design, shall be tested for its practicability and suitability. For some of the goals, country-specific results are desirable; for others only, the pooled data from all three countries will provide sufficient case numbers, so that they will be answered on a transnational level only. The main goals of the experiment are:

		All	Germany	Croatia	Portugal
Primary Research Questions					
1a	What is the total response rate for push-to-web and for face-to-face design?	✓			
1b	What differences are there between the three countries involved?		✓	✓	✓
1c	What differences are there between different strategies for providing incentives?		✓ ¹		
1d	What differences are there between different strategies for sending out reminders?			✓	

1e	What differences are there between different strategies for selecting a respondent from the people living at the same address (within address-based sampling)?				✓
2a	How large is the web response rate (web interviews divided by initial gross sample) in the push-to-web design?	✓			
2b	What differences are there between the three countries involved?		✓	✓	✓
2c	What differences are there between different strategies for providing incentives?		✓		
2d	What differences are there between different strategies for sending out reminders?			✓	
2e	What differences are there between different strategies for selecting a respondent from the people living at the same address (within address-based sampling)?				✓
3a	How large are the F2F response rates (F2F interviews divided by initial gross sample / by adjusted gross sample) in the push-to-web design?	✓			
3b	What differences are there between the three countries involved?		✓	✓	✓
3c	What differences are there between different strategies for providing incentives?		✓ ¹		
3d	What differences are there between different strategies for sending out reminders?			✓	
3e	What differences are there between different strategies for selecting a respondent from the people living at the same address (within address-based sampling)?				✓
4a	How large is the proportion of web interviews and of F2F interviews within the push-to-web design?	✓			
4b	What differences are there between the three countries involved?		✓	✓	✓
4c	What differences are there between different strategies for providing incentives?		✓ ¹		
4d	What differences are there between different strategies for sending out reminders?			✓	
4e	What differences are there between different strategies for selecting a respondent from the people living at the same address (within address-based sampling)?				✓
5a	How does the web take-up translate into costs for conducting GGS survey research?	✓			
5b	What differences are there between the three countries involved?		✓	✓	✓
5c	What differences are there between different strategies for providing incentives?		✓ ¹		

5d	What differences are there between different strategies for sending out reminders?	✓	
5e	What differences are there between different strategies for selecting a respondent from the people living at the same address (within address-based sampling)?		✓
6	How many interviews are abandoned before completion in each mode?	✓	
7	How large are mode effects for sensitive questions (comparing CAPI and CASI mode, randomly assigned)?	✓ ²	

Secondary Research Questions

8a	What kinds of biases (due to unit nonresponse) are expected in push-to-web and in face-to-face design?	✓	
8b	What kinds of biases (due to unit nonresponse) are expected in the subsamples of people interviewed online and those interviewed F2F, within the push-to-web designs?	✓	
9a	What percentages of item nonresponse are expected in push-to-web and in face-to-face design?	✓	
9b	What percentages of item nonresponse are expected in the subsamples of people interviewed online and those interviewed F2F, within the push-to-web designs?	✓	
10a	What kind of response biases or response styles (such as social desirable responding, acquiescence, etc.) are expected in push-to-web and in face-to-face design?	✓	
10b	What kind of response biases or response styles are expected in the subsamples of people interviewed online and those interviewed F2F, within the push-to-web designs?	✓	

Further Research Questions

There are further goals for which pre-testing will be sufficient. However, these need to be predominantly reached as a pre-condition for successfully conducting the GGP experiment. Therefore, the pilot needs to have its own pre-test in which the following research questions are addressed:

- Do the Blaise programmed digital questionnaire and its implementation by national fieldwork agencies work well (also on small screen devices, such as smart phones)?
- Does interrupting and later continuing the self-administered online-interview work well (at any position in the questionnaire)?
- Does the central administration of the fieldwork on NIDI servers work well?
- Does the revised GGS questionnaire (in the three translated versions) work well?
- Does the Push to Web strategy produce high quality demographic data on key indicators (i.e. Fertility Rates, Divorce Rates, etc)

6. Technical description of the experiment

Timeline

We decided that the timeline for the GGP experiment should be as follows:

- In Spring and Summer 2017 the GGP experiment is planned and designed. This involves the clarification of technical as well as of organizational questions.
- During Summer 2017, the technical revision of the GGS questionnaire is executed.
- In Autumn 2017 in all countries participating in the GGP experiment, a fieldwork agency should have been found for executing the fieldwork for the experiment and it should start preparing the fieldwork.
- In Autumn 2017, the revision of the questionnaire and of the Blaise programming should be completed, so that it can be used and tested within the experiment.
- In Autumn of 2017, national teams will be required to translate the questionnaire.
- The experiment should be conducted in Spring 2018.
- Data from the experiment should be analysed from Summer 2018, leaving us enough time to discuss implications from the experiment for the fieldwork guidelines.
- Fieldwork guidelines should be finalized by the end of 2018.
- At that time, preparations for the GGS relaunch need to be intensified by all participating countries.

Questionnaire and Technical Implementation

All interviews in the pilot study use the same GGS questionnaire:

- The experiment uses the latest possible updated version of the GGS questionnaire. Revisions to this questionnaire should be completed by October 2017.
- The online questionnaire, programmed in Blaise, is provided to the fieldwork agencies in the national languages by the research team. The Blaise program will identify if it is a CAPI or CAWI interview automatically and will load the correct version of the questionnaire accordingly. It has a responsive web design that allows good readability on devices with small screens (such as smart phones). The questionnaire will be administered via the GGP server based at NIDI. This can be done using an offline mode using the Blaise App or Data Entry Program (DEP) or completed via a web browser.
- The fieldwork agency should use its own digital environment (e.g. for administering the status of each participant, the numbers of interviews conducted by each interviewer etc.). The necessary interfaces for connecting the Blaise questionnaire to the environment are provided by the research team.
- In face-to-face interviews, the interviewer uses a laptop or tablet computer (CAPI) which she/he can pass on to the respondent for specific sensitive questions, so that the respondents answer in a self-administered way (CASI). The redesign of the questionnaire will identify the sensitive questions that will be used for this.
- The fieldwork in the push-to-web condition will consist of 2 separate phases. In phase 1, all sampled households will receive an invitation letter on the same day (Friday), which marks the start of the web-phase. Once the fieldwork moves into the F2F-phase, the web questionnaire is closed, and no longer available to respondents.

- The data collected in all interviews is transmitted to a NIDI server and stored there.
- Fieldwork agencies will maintain a database in which a log is kept of fieldwork efforts. By the end of data collection, an anonymized version (only containing an anonymous ID that can be linked to the substantive data) as well as interview outcome data, will be sent to NIDI. This fieldwork log will at the minimum contain the fieldwork codes as listed in Appendix A. The respondent ID's to be used will be provided to fieldwork teams at the same time as the questionnaire and must be used in the fieldwork process.
- Within three weeks after the end of the fieldwork, the fieldwork agency also provides the database with all collected contact addresses and IDs of the participants in a separate post. These data will need to be protected with a key, and send through a secure connection.
- Once a week, the national organising team as well as the fieldwork agency are updated by NIDI with the current state of the dataset.
- Sensors will be used to collect passive data as meta or para-data for auditing purposes, particularly time markers measuring when an interview was begun and ended as well as how long each module took, but also the GPS code measuring the position of the device used for filling out the questionnaire. This will be done by the Blaise program without the fieldwork agency needing to be involved.

Pre-Test

Before the start of the actual experiment, the fieldwork agency conducts a pre-test:

- The pre-test has at least 20 participants. These do not need to be selected in any particular way. The sampling is left to the fieldwork agency.
- At least 5 participants are interviewed personally in a face-to-face interview. At least 15 participants fill out the online questionnaire. Among the latter a variation of devices shall be used, including desktop, laptop, tablet computer and a variety of common brands and models of smartphones (incl. Apple iPhone and Samsung Galaxy).
- Participants and interviewers as well as researchers pay attention to whether:
 - any technical problems occur with the questionnaire programming (Is the questionnaire correctly displayed? Does the operation of the questionnaire work well? ...),
 - any problems occur with display of the questionnaire on small devices,
 - any technical problems occurring through interruptions of the self-administered CAWI interview (at any position in the questionnaire)?
 - any technical problems occur with the questionnaire-environment interface (Can the fieldwork agency monitor the status of each participant? ...),
 - any technical problems occur with the transmission and storage of data on NIDI servers,
 - any problems occur with the understandability of the translation of the questionnaire.
- At the end of the pre-test, the fieldwork agency provides a short and informal report in English language regarding the results to the questions mentioned above.
- Based on the report, researchers and fieldwork agency discuss the problems and necessary measures for their solution and implement these measures to enable the successful completion of the experiment.

Fieldwork Specifications and Sampling

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

- In each participating country – Germany, Croatia, and Portugal – one reference group will be interviewed in a face-to-face (only) mode (CAPI/CASI) and at least two treatment groups will be 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, a local register is used for a region that still needs to be identified. In Croatia, a health insurance register is used. In Portugal, address based sampling is used.
- To save costs, the samples are not representative for the countries but only for one region each within these countries. This region will include urban as well as rural areas. And it will offer comparatively suboptimal conditions for survey research, such as low provision of high speed internet. This region can be defined by the local organising team in consultation with the fieldwork agency in a way that reduces costs (e.g. by picking a region in which the fieldwork agency has many interviewers).
- Samples are furthermore restricted to the age range of 18 to 49 years.
- No proxy responding is allowed.
- All groups need to be represented by case numbers that are suitable for answering the research questions formulated before. A minimum sample size for each group should be $n = 200$. If possible, case numbers should clearly exceed this lower limit!!
- For the treatment groups (interviewed in push-to-web design), also the two subsamples of those who were interviewed in self-administered online-mode and those interviewed in face-to-face mode, both should be represented by credible case numbers ($n > 200$).
- In each country, there is one anchor group for P2W which will follow an internationally standardised and comparable design (see below).

	Germany			Croatia			Portugal		
Minimum case numbers	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W variant</i>	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W variant</i>	<i>F2F only</i>	<i>P2W anchor</i>	<i>P2W variant</i>
all interviews	200	400	200	200	400	400	200	400	400
web interviews	–	200	200	–	200	200	–	200	200
face-to-face	200	200	–	200	200	200	200	200	200

Contact protocol (P2W Anchor)

- As a first contact, people are contacted by mail. They receive a letter through the post, inviting them to participate in an online survey. The design of the envelopes and mail are up to the national teams but guidelines for the content of the invitation letter are to be found in Appendix B.

- In Portugal, as a country with address-based sampling, the invitation is sent to the household rather than the individual. Here, the interviewer follows the protocol for selecting the right respondent (see later section on within household selections).
- The invitation letter offers a short URL as well as a QR code, linking to the same website, and a password that prevents that any other person, finding the website by chance, from filling in the questionnaire.
- The letter also comes with 5 Euro in cash (if possible) or in vouchers as an unconditional pre-paid incentive. In Portugal and Croatia, where monetary incentives are not allowed by law vouchers are used.²
- In Croatia, where the currency is Kuna instead of Euro, a round amount close to 5 Euro (37 Kuna) is defined, to be as comparable as possible. If available, a 35 or 40 Kuna voucher is used.
- The letter briefly explains the study, its purpose, the institutions being responsible for it as well as all necessary information on data protection and legal background (such as participation being voluntary). It emphasizes the scientific, political and societal relevance of the project and expresses a warm invitation to participate. It also mentions the incentive, as an unconditional gift. The letter looks professional, clearly structured and appealing. The text is straight to the point and avoids lengthy formulations.
- All letters (initial invitation and reminders) should be designed to arrive on a Friday.
- People who do not fill in the online questionnaire within 14 days after the first contact receive a first reminder, again by post.
- People who do not fill in the online questionnaire within 14 days after the first reminder receive a second reminder. In this letter they are informed that, unless they fill in the questionnaire within the forthcoming 5 days, an interviewer will visit them at home, so that they can have a face-to-face rather than an online interview. It is also noted that it will not be possible to complete the web survey after this 5-day period.
- After 5 further days, the online questionnaires are deactivated.
- Shortly after this date, an interviewer visits people at home who still haven't participated.
- With people who then are willing to participate but do not have the time right away, the interviewer agrees upon a date for the interview. The interviewer can take a phone number and/or e-mail address (at the door) at this stage and call or e-mail the respondent to make an appointment.
- For sampled persons who are not at home at the interviewer's visit, a letter is left, saying that he/she is trying to get into contact and offering possibilities to suggest a date for an interview by e-mail, phone or post.
- If a person does not offer any date for an interview, the interviewer will try a second and later a third time to visit. Visits occur on different weekdays, including at least one weekend day, as well as at different hours. At least one visit is made in the early evening hours.
- If after three visits the interviewer has not been in contact with the selected person, the outcome code is set to "non-contact". No further contact attempt is made, unless the sampled person contacts the research agency/interviewer.

² Sending monetary bills by mail may not be allowed in some countries. Research has shown that vouchers are however not as effective as real money, so we need to be aware of that while analyzing the data.

- When a contact is made with the sampled person, an appointment is attempted. If the respondent does not want to make an appointment, two further contact attempts are required. If at the third attempt, a sampled person offers a soft refusal (“not time now”, “I am busy”, etc.) a fourth and eventually a fifth contact attempt is made. After three to five unsuccessful contact attempts, the final interview outcome is set to “refusal”. If a respondent offers a hard refusal (“I don’t want to participate”) at any attempt, no further attempts are made, and the outcome is set to “refusal”.
- If a respondent does not speak the language of the GGS questionnaire adequately, or has health problems (hearing, infirmity), the interview outcome is set to ‘other non-interview’, with a specification of why no interview was held.

Contact protocol (F2F only reference group)

- As a first contact, people are contacted by post. They receive a letter, announcing that in within the forthcoming 10 days, an interviewer will be visiting them at home to conduct a personal interview. The design of the envelopes and mail are up to the national teams but guidelines for the content of the invitation letter are to be found in Appendix B.
- In Portugal, as a country with address-based sampling, the invitation is sent to the household rather than the individual. Here, the interviewer follows the protocol for selecting the right respondent (see later section on within household selections).
- The letter also comes with 5 Euro in cash (if possible) or in vouchers as an unconditional pre-paid incentive. In Portugal and Croatia, where monetary incentives are not allowed by law vouchers are used.
- In Croatia, where the currency is Kuna instead of Euro, a round amount close to 5 Euro (37 Kuna) is defined, to be as comparable as possible. If available, a 35 or 40 Kuna voucher is used.
- The letter briefly explains the study, its purpose, the institutions being responsible for it as well as all necessary information on data protection and legal background (such as participation being voluntary). It emphasizes the scientific, political and societal relevance of the project and expresses a warm invitation to participate. It also mentions the incentive, as an unconditional gift. The letter looks professional, clearly structured and appealing. The text is straight to the point and avoids lengthy formulations.
- After 3 to 9 days, an interviewer visits people for conducting the interview.
- When a contact is made with the sampled person, an appointment is attempted. If no contact is made at the household level, two further contact attempts are made. If after three attempts, no contact is made with the sampled household, it is set to noncontact, and no further interview attempt is made, unless the sampled person contacts the research agency/interviewer.
- If contact is made at the household level, the interviewer identifies the sampled respondent (in Croatia and Germany), and attempts an interview (or makes an appointment). A total of 5 attempts are made to interview the sampled respondent, if contact has been made at the household level within 3 attempts.
- If before or at the third contact attempt, a sampled person offers a soft refusal (“no time now”, “I am busy”, etc.) a fourth and eventually a fifth contact attempt is made. After three to five unsuccessful contact attempts, the final interview outcome is set to “refusal”. If a respondent offers a hard refusal (“I don’t want to participate”) at any attempt, no further attempts are made, and the outcome is set to “refusal”. See appendix A for detailed contact outcome codes.

- With people who are willing to participate but do not have the time right away, the interviewer agrees upon a date for the interview. The interviewer can take a phone number and/or e-mail address (at the door) at this stage and call or e-mail the respondent to make an appointment
- For sampled persons who are not at home at the interviewer's visit, a letter is left, saying that he/she is trying to get into contact and offering possibilities to suggest a date for an interview by e-mail, phone or post.
- If a person does not offer any date for an interview, the interviewer will try a second and later a third time to visit. Visits occur on different weekdays, including at least one weekend day, as well as at different hours. At least one visit should be made in the early evening hours.
- If a respondent does not speak the language of the GGS questionnaire adequately, or has health problems (hearing, infirmity), the interview outcome is set to 'other non-interview', with a specification of why no interview was held.

Deliverables and further issues

- Interviewers need to be trained before the start of the (main) fieldwork. One detail of the interviewer training is the randomly assigned variation of CAPI and CASI mode (among F2F interviews) for questions. Interviewers will be randomly assigned to either use CASI or CAPI for specific sections of the questionnaire containing sensitive questions. Interviewers should be trained accordingly.
- The fieldwork is monitored by the fieldwork agencies. Measures for an effective control of the accurate realisation of interviews, of the compliance of interviewer instructions and of the data quality shall be taken (such as the analysis of times and durations of all interviews of the same interviewer). It is particularly important to check which respondent completes the questionnaires. To assess this, it is necessary that a subsample of web respondents is contacted by test calls, that their identity is verified and that it is checked whether they personally completed the survey.
- Within three weeks after the end of the fieldwork, the fieldwork agency provides a short report in English language regarding the technical and organisational performance of the fieldwork, including response rates, the reasons for non-participation (refusal, no contact, ...) and incomplete interviews, compliance of interviewer instructions and technical functioning.
- Within three weeks after the end of the fieldwork, the fieldwork agency provides a database in SPSS or Stata format with all contacted people – anonymised by an identification number – informing about the number of contact attempts, the days and hours of each contact attempt, as well as the final status regarding their (non-) participation ("interview completed", "interview abandoned without completion", "refusal", "no contact", "other non-interview", and specification of 'other' reason not to have participated) (see Appendix A).
- Within eight weeks after the end of the fieldwork, the fieldwork agency provides a financial report in English language. This report explains the (factually incurred) costs (not necessarily the costs the researchers are charged for), distinguished in fixed costs (e.g. for the set-up of an interview mode) and the costs per interview in the various modes. It further gives estimates on what a large survey with n=10,000 would cost in each of the tested modes (face-to-face only, push-to-web in each of the groups).

Country-specific experiments

Aside from the anchor group, as described above, there shall be further treatment groups in each country, varying one condition each for further country-specific experiments. The conditions that are varied are:

- the strategy for providing incentives – in Germany,

- the strategy for sending out reminder letters – in Croatia, and
- the strategy for selecting a respondent from the people living at the same address (within address-based sampling) – in Portugal.

The document now further discusses country-specific requirements for the study.

Design Specifications in Germany

The design of the experiment will be realised as follows in Germany:

- The gross sample is drawn from the registers of local residents' registration offices.
- A region will be chosen, in coordination with the fieldwork agency, in which a small number of sample points (about 12) will be selected. These will partly be urban and partly rural.
- For testing the push-to-web design in various variations, five treatment groups will be formed with various combinations of incentives (unconditional / conditional):
 - a "second reference group", not receiving any incentive at all (0 / 0),
 - the "anchor group", receiving 5 Euro unconditional pre-paid incentive (5 / 0),
 - the "small conditional incentive group", receiving 5 Euro post-paid incentive after a successfully completed interview (0 / 5),
 - the "large unconditional incentive group", receiving 5 Euro unconditional incentive and further 25 Euro after a successfully completed interview (5 / 25),
 - the "large conditional incentive group", receiving 30 Euro conditional post-paid incentive after a successfully completed interview (0 / 30).

Given that face-to-face interviews in Germany are particularly expensive, the experiment would not be realisable without a limitation of the face-to-face interviews that need to be conducted. Negotiations with fieldwork agencies came to the insight that, aside from n=200 interviews in the face-to-face only reference group, another n=200 face-to-face interviews within the push-to-web design can be financed. This is how we will proceed:

- In the anchor group (5 / 0), all participants who did not participate in the web survey but are ready to be interviewed personally are interviewed in follow-up face-to-face interviews, without limitation. The number of people being contacted in the anchor group is chosen in a way that the number of face-to-face interviews will only slightly exceed n=200.
- In the group with 5 Euro unconditional and 25 Euro conditional incentive, as many follow-up face-to-face interviews will be conducted as financial means allow. For this purpose, among the participants who did not participate in the web survey, a random subsample will be drawn.
- In all other treatment groups face to face follow-up interviews will not be conducted (unless unexpected financial resources allow to do so).

	group R reference	group 0 2 nd reference	group 1 anchor	group 2	group 3	group 4
<i>general interview mode</i>	<i>F2F only</i>		<i>mixed-mode / push-to-web</i>			
<i>incentives</i>	(5 / 0)	(0 / 0)	(5 / 0)	(0 / 5)	(5 / 25)	(0 / 30)
<i>(unconditional / conditional)</i>						

A) total gross sample	~8,000					
B) initial group gross samples	~575	~2000	~1400	~1500	~1200	~1300
intended mode at 1 st 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	200-250	200-250	200-250	200-250	200-250
C) sample size	200	250	250	250	250	250
D) adjusted group gross samples	(~375)	(~1750)	~1150	(~1250)	~950	(~1050)
E) random sub-sample		0*		0*	~500*	0*
intended mode after 2 nd reminder	—	(face-to-face)	face-to-face	(face-to-face)	(face-to-face)	(face-to-face)
response rate – based on adjusted g. s.	—	15-25%	15-25%	15-25%	15-25%	15-25%
minimum sample size	—	0*	200-250	0*	0*	0*
F) sample size		0*	~230	0*	~100*	0*
G) total sample size	~200	~250	~480	~250	~350	~250
total response rate	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 will be conducted in the “5/25” group. For this purpose, a random subsample will be drawn.

- In the case of an unexpectedly very low response rate for web interviews, the design is adjusted: The anchor group is split in two groups, by random selection: Only in group 1a, all participants are followed up by face-to-face interviews. Here, a number of people is contacted that leads to about n=200 face-to-face interviews, even if this doesn't lead to 200 web interviews. In group 1b, only web interviews are conducted. Here as many people are contacted as necessary to reach the minimum number of web interviews. For analyses (except those estimating the response rates and web/F2F ratios), both groups are merged.

Design Specifications in Croatia

The design of the experiment will be realised as follows in Croatia:

- For testing the push-to-web design in various variations, two treatment groups will be formed:
 - the “anchor group”, receiving two reminders (14 days and 28 days after the initial contact). No phone contacts are allowed, even when a telephone number is available from the health

register. When in the F2F phase, the interviewer may record new telephone numbers of respondents as part of the F2F interview (just like outlined in the general guidelines).

- a group, receiving three reminders (14 days, 28 days and 35 days after the initial contact). In this final group, respondents for whom a known telephone number is available through the health records, will additionally be called between days 28-42 to remind them of the survey. Three contact attempts are made by telephone to reach a respondent. If a respondent is contacted, the interviewer reminds the respondent of the web survey, and informs the respondent of the time left (anywhere between 1 and 14 days) to complete the web survey. If a respondent actively brings up the possibility of doing the interview F2F, the interviewer is allowed to make an appointment, for the period after day 42. The interviewer cannot actively bring up the possibility of F2F interview himself, and should try to make sure the respondent completes the web survey. Once this group moves into the F2F phase (after day 42), the interviewer can call the respondent again to make an appointment for a F2F interview before visiting the respondent at home.

Design Specifications in Portugal

The design of the experiment will be realised as follows in Portugal:

Interviews will follow the general guidelines for fieldwork as pointed out earlier in this document. A sample will be drawn using a random route design. As such, the invitation letters will be addressed to households instead of individuals, and an experiment will focus on how to select individuals from within those households.

- Two thirds (2/3) of respondents will be randomly assigned to the “traditional invitation method” condition. The invitation letter will contain a household grid and a subsequent explanation of who the survey is intended for (i.e. the next person to have their birthday who is aged between 18-49)
- One third (1/3) of respondents will be randomly assigned to a condition in which respondent selection is actually conducted on the web. The web survey starts with completing the household grid (this is part of the regular survey). The computer will automatically select one eligible respondent from the household grid, and then ask the respondent to either
 - make sure that this person continues with the survey or
 - complete the e-mail address of this person, so that this person will receive a separate invitation for the survey. The survey for this person will continue after the household grid.

7. Summary

To conclude, it is important to reflect on the added value that the pilot study brings in terms of the broader field of survey methodology and data collection. It is clear that, in the near future, social science research infrastructures have to adapt to the changes that advancements in technology present, finding solutions to challenges that ‘push to web’ strategies entail (in particular, mode effects, coverage, non-response). Despite the opportunities that web surveys can offer, it remains to be seen how the implementation of a Push to web (mixed-mode) design will perform in the context of the GGS, with a complex and relatively long questionnaire, in sub-optimal conditions (including areas with low internet penetration). In addition, there are also important lessons to be learnt from the survey paradata and fieldwork reports. The results of the pilot study will inform the model and guidelines for future data collection. Part of the vision and ambition for the next round will then include the incorporation of improvements in all stages of the survey to achieve higher efficiency, lower fieldwork costs, quicker data processing and better microdata.

Transitioning to a survey future online is a significant challenge but the potential benefits are considerable. Not only will a web based survey be cheaper and more efficient in its operation, it also opens the possibility

for greater flexibility and complexity in the deployment of the infrastructure. Social Survey Infrastructures have yet to match the hard sciences in their ability to fully open up and allow active control of the infrastructure by scientists. The exception to this would be the European Social Survey which reserves space in every round for groups of scientists to field modules. However, this is very limited 'beam time' and is dominated by well-established and senior teams. What is more, access to the infrastructure is limited to once every two years. If the GGP is able to successfully shift to a survey future online, the ability to field modules after a very short development phase, out of the regular cycle of the survey would open the opportunity for smaller projects and less senior researchers to also utilise the infrastructure. The GGP will be able to target specific sub-populations, run innovative experiments in specific countries and even field modules in response to social, economic and political events as they unfold. The shortening of the research life cycle brought about by a push to web opens up a new range of opportunities for science.

Finally, a push to web approach reflects a broader trend within social demography which seeks to incorporate more diverse forms of data into social science research. The internet is not just a tool for social scientists but also a fundamental object of study, one that is fundamentally reshaping human behaviours (F C Billari, Giuntella, and Stella 2017). A future online means that the GGP is better placed to capture internet related data, especially in the form of paradata. The speed of typing during a web survey, the type of browser used, the model of tablet and the speed of internet connection all give researchers tell-tale insights into the respondent themselves. It also provides a portal to integrate social media data and other web data that the respondent has (with consent) and therefore allowing researchers to understand an increasingly significant part of our social world. By utilizing the scientific principles of a probability based sample and a long standing survey structure alongside the power, diversity and depth of new forms of data a survey future online could form the basis of a revolution in the social sciences (Francesco C Billari and Zagheni 2017).

8. Bibliography

- Bethlehem, Jelke. 2010. "Selection Bias in Web Surveys." *International Statistical Review* 78 (2). Blackwell Publishing Ltd: 161–88. doi:10.1111/j.1751-5823.2010.00112.x.
- Bianchi, Annamaria, Silvia Biffignandi, and Peter Lynn. 2017. "Web-Face-to-Face Mixed-Mode Design in a Longitudinal Survey: Effects on Participation Rates, Sample Composition, and Costs." *Journal of Official Statistics* 33 (2). De Gruyter Open: 385–408. doi:10.1515/jos-2017-0019.
- Billari, F C, O Giuntella, and L Stella. 2017. *Does Broadband Internet Affect Fertility?* papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029806.
- Billari, Francesco C., Dimiter Philipov, and Maria Rita Testa. 2009. "Attitudes, Norms and Perceived Behavioural Control: Explaining Fertility Intentions in Bulgaria." *European Journal of Population* 25 (4): 439–65. doi:10.1007/s10680-009-9187-9.
- Billari, Francesco C, and Emilio Zagheni. 2017. "Big Data and Population Processes: A Revolution?," no. June (January). SocArXiv: 1–12. doi:10.17605/OSF.IO/F9VZP.
- Bluemke, Matthias, Matthias Bluemke, Bernd Resch, Clemens Lechner, René Westerholt, and Jan-Philipp Kolb. 2017. "Integrating Geographic Information into Survey Research: Current Applications, Challenges and Future Avenues." *Survey Research Methods* 11 (3): 307–27. doi:10.18148/srm/2017.v11i3.6733.
- Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42 (4). Oxford University Press: 992–1001. doi:10.1093/ije/dyto88.
- Börsch-Supan, Axel, Karsten Hank, and Hendrik Jürges. 2005. "A New Comprehensive and International View on Ageing: Introducing the 'Survey of Health, Ageing and Retirement in Europe.'" *European Journal of Ageing* 2 (4): 245–53. doi:10.1007/s10433-005-0014-9.

- Bruijne, Marika de, and Arnaud Wijnant. 2013. "Comparing Survey Results Obtained via Mobile Devices and Computers." *Social Science Computer Review* 31 (4). SAGE PublicationsSage CA: Los Angeles, CA: 482–504. doi:10.1177/0894439313483976.
- Buber-Ennsner, Isabella, and Norbert Neuwirth. 2013. "Families in Austria." Vienna. doi:10.4135/9781412975957.n9.
- Conrad, Frederick, Roger Tourangeau, Mick Couper and Chan Zhang. 2017. "Reducing speeding in web surveys by providing immediate feedback." *Survey Research Methods* 11(1): 45-61. doi:10.18148/srm/2017.v11i1.6304
- ESFRI. 2006. "ESFRI European Roadmap for Research Infrastructures." Brussels. https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf.
- Festy, Patrick., France. Prioux-Marchal, United Nations. Economic Commission for Europe., United Nations Population Fund., and Institut national d'études démographiques (France). 2002. *An Evaluation of the Fertility and Family Surveys Project*. United Nations. [https://books.google.co.uk/books?hl=en&lr=&id=jtnoPjiCsxAC&oi=fnd&pg=PR3&dq=Festy,+Patrick+and+France+Prioux+\(2002\),+%3FAn+evaluation+of+the+Fertility+and+Family+Surveys+Project%3F,+New+York+and+Geneva:+United+Nations+Publication,+accessed+via&ots=DPVPdNu](https://books.google.co.uk/books?hl=en&lr=&id=jtnoPjiCsxAC&oi=fnd&pg=PR3&dq=Festy,+Patrick+and+France+Prioux+(2002),+%3FAn+evaluation+of+the+Fertility+and+Family+Surveys+Project%3F,+New+York+and+Geneva:+United+Nations+Publication,+accessed+via&ots=DPVPdNu).
- Fokkema, Tineke, Andrej Kveder, Nicole Hiekel, Tom Emery, and Aart C. Liefbroer. 2016. "Generations and Gender Programme Wave 1 Data Collection: An Overview and Assessment of Sampling and Fieldwork Methods, Weighting Procedures, and Cross-Sectional Representativeness." *Demographic Research* 34 (1): 499–524. doi:10.4054/DemRes.2016.34.18.
- Groves, R. M., and L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74 (5). Oxford University Press: 849–79. doi:10.1093/poq/nfq065.
- Jäckle, Annette, Peter Lynn, Jon Burton, and Jon Burton. 2015. "Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response." *Doi.org* 9 (1). European Survey Research Association. doi:10.18148/srm/2015.v9i1.5475.
- Jowell, Roger., Rory Fitzgerald, Caroline Roberts, and Gillian Eva. 2007. *Measuring Attitudes Cross-Nationally : Lessons from the European Social Survey*. SAGE. <https://books.google.co.uk/books?hl=en&lr=&id=NQrOpbvbyzWC&oi=fnd&pg=PP1&dq=the+european+social+survey&ots=nkkVeJa4B3&sig=YbaUFYVaFAIMNLBYtUSRag13Yo#v=onepage&q=the+european+social+survey&f=false>.
- Kaa, D J Van De. 1987. "Europe's Second Demographic Transition." *Population Bulletin* 42 (1). Population Reference Bureau: 1–59. <http://www.ncbi.nlm.nih.gov/pubmed/12268395>.
- King, Gary. 2011. "Ensuring the Data-Rich Future of the Social Sciences." *Science (New York, N.Y.)* 331 (6018). American Association for the Advancement of Science: 719–21. doi:10.1126/science.1197872.
- Kotowska, Irena E, Anna Matysiak, and Monika Mynarska. 2006. "The Life of Poles: From Leaving the Parental Home to Retirement. Insights from the Generations and Gender Survey (GGs-PL)." Warsaw. http://www.ggp-i.org/wp-content/uploads/2017/10/ggp_Life_of_Poles_eng_fin.pdf.
- Lugtig, Peter, and Vera Toepoel. 2016. "The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey." *Social Science Computer Review* 34 (1): 78–94. doi:10.1177/0894439315574248.
- Millar, M. M., and D. A. Dillman. 2011. "Improving Response to Web and Mixed-Mode Surveys." *Public Opinion Quarterly* 75 (2). Oxford University Press: 249–69. doi:10.1093/poq/nfro03.
- Poushter, Jacob, and Rhonda Stewart. 2017. "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies But Advanced Economies Still Have Higher Rates of Technology Use." Accessed November 10. www.pewresearch.org.

- Revilla, Melanie. 2013. "Measurement Invariance and Quality of Composite Scores in a Face-to-face and a Web Survey." *Survey Research Methods* 7(1): 17-28. doi: <http://dx.doi.org/10.18148/srm/2013.v7i1.5098>.
- Stoop, Ineke A. L., Jacques Billiet, Achim Koch, and Rory Fitzgerald. 2010. *Improving Survey Response : Lessons Learned from the European Social Survey*. Wiley.
<https://books.google.co.uk/books?hl=en&lr=&id=8MlxJtXrbokC&oi=fnd&pg=PR5&dq=the+european+social+survey&ots=YpQTZJJuKz&sig=TkKd86GxCWIY776QwxHnOwJAj4#v=onepage&q=the+european+social+survey&f=false>.
- Tijdens, Kea and Stephanie Steinmetz. 2016. "Is the web a promising tool for data collection in developing countries? An analysis of the sample bias of 10 web and face-to-face surveys from Africa, Asia, and South America." *International Journal of Social Research Methodology* 19(4): 461-479. doi: 10.1080/13645579.2015.1035875.
- Vikat, Andres, Zsolt Spéder, Gijs Beets, Francesco C Billari, Christoph Bühler, Aline Désesquelles, Tineke Fokkema, et al. 2007. "Generations and Gender Survey: Concept and Design." In *Generations & Gender Programme: Concepts and Guidelines*, edited by United Nations Economic Commission for Europe, 35–113. United Nations.
<http://content.epnet.com/ContentServer.asp?T=P&P=AN&K=28743147&EbscoContent=dGJyMMTo5oSepq84v%2BvLOLCmrIcepq5Ss664TLaWxWXS&ContentCustomer=dGJyMPGpsU22rbdOuePfgeyx%2BEu3q64A&D=sih>.