DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 32, ARTICLE 24, PAGES 723–774
PUBLISHED 13 MARCH 2015
http://www.demographic-research.org/Volumes/Vol32/24/
DOI: 10.4054/DemRes.2015.32.24

Research Article

Quality of demographic data in GGS Wave 1

Jorik Vergauwen

Jonas Wood

David De Wachter

Karel Neels

This publication is part of the Special Collection on “Data Quality Issues in the First Wave of the Generations and Gender Survey,” organized by Guest Editors Aart C. Liefbroer and Joop Hox.

Table of Contents

Demographic Research: Volume 32, Article 24
Research Article

Quality of demographic data in GGS Wave 1

Jorik Vergauwen[1]

Jonas Wood[2]

David De Wachter[3]

Karel Neels[4]

Abstract

BACKGROUND
A key feature of the Generations & Gender Programme (GGP) is that longitudinal micro-data from the Generations and Gender Surveys (GGS) can be combined with indicators from the Contextual Database (CDB) that provide information on the macro-level context in which people live. This allows researchers to consider the impact of socio-cultural, economic, and policy contexts on changing demographic behaviour since the 1970s. The validity of longitudinal analyses combining individual-level and contextual data depends, however, on whether the micro-data give a correct account of demographic trends after 1970.

OBJECTIVE
This article provides information on the quality of retrospective longitudinal data on first marriage and fertility in the first wave of the GGS.

METHODS
Using the union and fertility histories recorded in the GGS, we compare period indicators of women's nuptiality and fertility behaviour for the period 1970–2005 and cohort indicators of nuptiality and fertility for women born after 1925 to population statistics.

[1] Corresponding author. University of Antwerp, Belgium. E-Mail: jorik.vergauwen@uantwerpen.be.
[2] University of Antwerp, Belgium.
[3] University of Antwerp, Belgium.
[4] Corresponding author and principal investigator. University of Antwerp, Belgium.
E-Mail: Karel.Neels@uantwerpen.be.

RESULTS
Results suggest that, in general, period indicators estimated retrospectively from the GGS are fairly accurate from the 1970s onwards, allowing exceptions for specific indicators in specific countries. Cohort indicators, however, were found to be less accurate for cohorts born before 1945, suggesting caution when using the GGS to study patterns of union and family formation in these older cohorts.

CONCLUSIONS
The assessment of the validity of demographic data in the GGS provides country-specific information on time periods and birth cohorts for which GGS estimates deviate from population statistics. Researchers may use this information to decide on the observation period or cohorts to include in their analysis, or use the results as a starting point for a more detailed analysis of item nonresponse in union and fertility histories, which may further improve the quality of GGS estimates, particularly for these earlier periods and older birth cohorts.

COMMENTS
Detailed country-specific results are included in an appendix to this paper, available for download from the additional material section.

1. Introduction

In 2000 the Population Activity Unit (PAU) of the United Nations Economic Commission for Europe (UNECE) launched the Generations and Gender Programme (GGP) to enhance understanding of the causes and consequences of demographic change in developed countries (Vikat et al. 2007). International comparability is a key feature of the GGP and several, mainly European, countries have become highly committed to the implementation of the programme. The GGP consists of two pillars. The first pillar is a set of *Generations and Gender Surveys* (GGS)[5]. The GGS is a panel survey that collects longitudinal micro-level data on a representative sample of non-institutionalized residents aged 18 to 79 years in each of the participating countries. The first wave of the GGS collects detailed data on partnership histories and (non-)resident children, making it possible to reconstruct changes in union formation and fertility in recent decades and link these to covariates at the individual, household, and contextual levels. To overcome the limitations associated with the retrospective design of the Fertility and Family Surveys (FFS) – the immediate predecessor of the GGP – the GGS

[5] See http://www.unece.org/pau/ggp/welcome.html and http://www.ggp-i.org.

combines elements of a retrospective setup with a prospective panel design (Vikat et al. 2007). The prospective design makes it possible to assess the impact of characteristics recorded in each wave (e.g., values and intentions) on subsequent behaviours, thus contributing to an enhanced understanding of the dynamic nature of demographic behaviour and the life-course. The *Contextual Database* (CDB)[6] (see the contribution of Caporali et al. in this special volume) is the second main pillar of the GGP and provides aggregate indicators at the meso (regional) and macro (national) levels. The CDB contains 200 indicators organized around 16 domains, covering both quantitative and qualitative information at the aggregate level for each of the participating countries, mostly from the 1970s onwards (Spielauer 2007). The possibility of combining longitudinal micro-level data from the GGS with contextual data from the CDB using multilevel (hazard) models allows for the assessment of the effect of various contexts (e.g., cultural, policy, and economic) on changing demographic and social behaviours (Simard and Franklin 2005; Vikat et al. 2007).

The contextual indicators in the CDB are generally population statistics drawn from different standardized national and international sources (e.g., vital statistics). As a result, the validity of analyses that combine contextual indicators with longitudinal micro-level data critically depends on the quality of the retrospective data on union formation and fertility collected in the GGS. To the extent that partnership and fertility histories provide a biased account of past trends, this may also yield biased estimates of the effect of various contextual factors on demographic behaviour. This also applies to explanatory analyses using GGS micro-data to examine associations between individual-level predictors and partnership and fertility histories. In this paper we use the total survey error approach to discuss potential sources of error in surveys such as the GGS, with a specific focus on bias in retrospectively collected event-histories (Blossfeld and Rohwer 2002; Blossfeld, Golsch, and Rohwer 2007; Weisberg 2009). In line with the GGS sample design guidelines, the paper subsequently aims to assess the joint effect of different sources of error on the validity of demographic data in the GGS by comparing indicators of first marriage and fertility estimated retrospectively from the GGS to population statistics (i.e., time-series drawn from vital registration, census, and register data).

[6] See http://www.ggp-i.org/contextual-database.html

2. Theoretical framework: the total survey error approach

Retrospective surveys have become the main source of data on partnership and childbearing histories (Beckett et al. 2001; Murphy 2009; Kreyenfeld et al. 2010). We use the total survey error approach to classify sources of error in surveys (Weisberg 2009; Groves and Lyber 2010). In this approach the term 'error' refers to the difference between the value obtained from a survey and the true value, usually the true value for the larger population of interest (Weisberg 2009). A characteristic feature of demographic surveys is that the true value for the population (at least for a number of characteristics) is frequently available from population statistics. Although population statistics can be subject to error and have been shown[7] to be biased in specific cases (e.g., Velkoff and Miller 1995; Morgan et al. 1999; Aleshina and Redmond 2005), we will consider them as a standard against which the total survey error in the Generations and Gender Surveys can be measured. We consider GGS estimates valid if they reflect the population value without systematic bias (Weisberg 2009).

Differences between population statistics and demographic indicators calculated retrospectively from the GGS may result from different sources of error, which can be classified into three larger types. Section 2.1 discusses the first type of error, which concerns issues related to selecting respondents for a survey. This includes sampling error, coverage error, and error associated with nonresponse at the unit level. Section 2.2 discusses the second type of error, which concerns issues related to response accuracy. This section distinguishes between nonresponse error at the item level, measurement error due to respondents, and measurement error due to interviewers. Finally, section 2.3 discusses the third type of error, which concerns various issues related to survey administration, including post-survey error, mode effects, and comparability effects (Weisberg 2009).

[7] Morgan and colleagues (1999), considering diverging findings on racial differences in first-birth timing in the United States, find that estimates based on vital registration perform worse than estimates using Current Population Survey (CPS) data. Among other factors, the racial bias in vital registration on first-birth timing is related to the fact that information on the number of births is taken from vital registration whereas estimates on the population at risk are drawn from the census. Aleshina and Redmond (2005) document that infant mortality and births in Central and Eastern European states are underreported in official registers due to a less rigorous definition of a live birth compared to countries using the WHO definition. Similarly, Velkoff and Miller (1995) found that misreporting infant deaths as deaths of extremely premature infants or even stillbirths was advantageous to hospitals in the Soviet Union, as infant mortality was used as a criterion to evaluate hospitals.

2.1 Issues related to selecting respondents

2.1.1 Sampling error

Sampling error is the error that occurs when a sample is surveyed rather than the entire population. Systematic bias may result when non-probability sampling is used, whereas probability sampling permits mathematical derivation of the sampling error (Weisberg 2009). Non-probability samples are likely to generate selectivity in the sample to the extent that inclusion probabilities depend on at-home availability patterns of different groups in the population. The random route procedure implemented in the German GGS has, for instance, been shown to induce selectivity (Sauer, Ruckdeschel, and Naderi 2012). However, as most countries implemented probability sampling in line with the GGS sample design guidelines (Simard and Franklin 2005 and Fokkema et al. in this special volume), the next sections focus on the potential impact of coverage error and nonresponse error at the unit level.

2.1.2 Coverage error

Coverage error occurs when the sampling frame does not correspond with the population of interest (Weisberg 2009). A typical issue of coverage error in retrospective surveys such as the GGS is that the collection of event-histories relies on survivors, while these may represent a nonrandom subgroup of their birth cohort with specific characteristics due to selective institutionalization, mortality, or out-migration (Neels 2006; Kreyenfeld et al. 2010)[8]. Selective institutionalization has been found to affect maternity histories because married women are more likely to have children and less likely to enter a retirement home (Sauer, Ruckdeschel, and Naderi 2012). This may result in an underrepresentation of childless women in non-institutionalized populations at older ages, which in turn can bias fertility levels upward for these cohorts. Similarly, mortality is found to be lower among married and cohabiting individuals than among singles (Gadeyne and Deboosere 2000; Sauer, Ruckdeschel, and Naderi 2012), whereas higher mortality is found for both childless and multiparous women (Kvåle, Heuch, and Nilssen 1994; Gadeyne and Deboosere 2000). This may introduce additional bias in nuptiality and fertility indicators for older cohorts. Finally, immigration and emigration are generally considered to be strongly correlated with patterns of nuptiality and family

[8] The resident non-institutionalized population aged 18–79 constitutes the target population of the GGS (Simard and Franklin 2005). For GGS various sampling frames were used depending on the country considered (see Fokkema et al. in this special volume).

formation (Neels 2006; Murphy 2000; 2009; Kreyenfeld et al. 2010; 2011; Sauer, Ruckdeschel, and Naderi 2012). Underrepresentation of migrant groups in surveys is frequently found – e.g., in the German GGS (Kreyenfeld et al. 2010) – and this will lead to an underestimation of nuptiality and fertility when migrant groups have higher marriage or childbearing rates.

2.1.3 Nonresponse at the unit level

Unit nonresponse occurs when a designated respondent does not participate in the survey, thereby limiting how representative the actual respondents are of the population of interest (Weisberg, 2009). Since nonresponse is found to vary significantly between social groups, under or overrepresentation of specific groups may affect the validity of retrospective estimates to the extent that these characteristics are correlated with nuptiality and fertility patterns. Determinants of unit nonresponse frequently discussed in the literature are age, gender, socio-economic position, geographical region, ethnicity, household (com)position, and parity. Unfortunately, the information available in the sampling frame (e.g., registers) is generally too limited to restore the representativeness of the sample with respect to such characteristics via post-stratification (see Fokkema et al. in this special volume on the calculation of post-stratification weights).

Considering age, younger people are repeatedly found to have lower participation rates in surveys (Van Loon et al. 2003; Neels 2006), whereas other sources mention lower participation rates for both the oldest and youngest age groups (Groves and Couper 1998; Goldberg et al. 2001; Nicoletti and Peracchi 2005). Similarly, Régnier-Loilier (2011), studying attrition between waves 1–2 and waves 2–3 of the French GGS, finds higher attrition for people younger than 30 and for people aged 60 and older. With respect to gender, lower response rates are generally found for men (Van Loon et al. 2003; Nicoletti and Peracchi 2005). Also in the French GGS, higher attrition has been found for men, especially among respondents younger than 50 (Régnier-Loilier 2011).

Despite considerable variety in classifications and operationalizations, several sources report differential unit nonresponse rates in terms of socio-economic position, with higher-SES groups generally being overrepresented in surveys (Goldberg et al. 2001). Literature shows that higher-educated groups are more willing to participate in surveys (Fitzgerald, Gottschalk, and Moffitt 1998; Pickery, Loosveldt, and Carton 2001; Kreyenfeld et al. 2011; Sauer, Ruckdeschel, and Naderi 2012). Similarly, studies that consider alternative indicators of socio-economic context, such as labour market position (Goldberg et al. 2001; Van Loon et al. 2003), housing conditions (Nicoletti and Peracchi 2005; Abraham, Maitland, and Bianchi 2006; Lauwereys, Neels, and De

Winter 2011), or income characteristics (Bergstrand et al. 1983; Fitzgerald, Gottschalk, and Moffitt 1998) mostly find higher response rates for higher socio-economic groups. Also in the case of the German GGS, differential response rates by socio-economic position have been found to lead to an overrepresentation of higher social strata (Kreyenfeld et al. 2010). For the French GGS, Régnier-Loilier (2011) suggests that attrition is highest among lower-educated individuals and unemployed persons. Similarly, Burkimsher (2009) states that the underestimation of fertility for older cohorts in the Bulgarian, Hungarian, and Georgian GGS may be due to the underrepresentation of older women with low SES, since lower education in these cohorts is correlated with having more children.

Geographical region plays an important role, as response rates are often found to vary strongly between regions (Goldberg et al. 2001; Neels 2006), with response rates generally being particularly low in urban regions (Groves and Couper 1998; Abraham, Maitland, and Bianchi 2006). For the French GGS, Régnier-Loilier (2011) has found that attrition is highest in urban regions. For the GGS in Central and Eastern European countries it has been suggested that rural residence may be correlated with both higher fertility and lower participation rates and thus may provide a partial explanation of underestimation of fertility (Burkimsher 2009).

Ethnicity is another characteristic related to lower response rates that may induce bias in nuptiality and fertility indicators. Migrant populations are generally underrepresented in surveys, particularly when the availability of the questionnaire in different languages is limited (Festy and Prioux 2002; Kreyenfeld et al. 2010). Foreigners are frequently found to show higher degrees of unit nonresponse (Bergstrand et al. 1983; Neels 2006), resulting in an underrepresentation of these groups (see Kreyenfeld and colleagues (2010) with respect to the German GGS). Several contributions have discussed the confounding role of migration (e.g., the role of East-West migration in Germany in higher fertility (Kreyenfeld et al. 2010)) or ethnic minorities in the GGS (e.g., underrepresentation of Roma populations with considerably higher fertility in Central and Eastern Europe (Burkimsher 2009)). Also, results for the French GGS (Régnier-Loilier 2011) indicate that attrition is higher for migrant groups.

Unit nonresponse has also been shown to vary in terms of household composition. Nicoletti and Peracchi (2005) claim that contact rates are inversely related to mobility of selected households. As a result, socio-demographic characteristics associated with higher mobility will also induce higher unit nonresponse. Literature shows that length of residence at the current address but also co-residence (Nicoletti and Peracchi 2005) and being married (Bergstrand et al. 1983; Goldberg et al. 2001; Neels 2006) correlate with higher participation in surveys, whereas being a single person household or having an unstable marriage or migration history entails lower participation rates (Fitzgerald, Gottschalk, and Moffitt 1998; Abraham, Maitland, and Bianchi 2006). Similarly, in the context of a household-based sample, Groves and Couper (1998) find a strong negative correlation between the number of household members and the number of contact attempts required to reach designated individuals in households.

Finally, higher parity has been shown to be associated with higher participation in surveys (Nicoletti and Peracchi 2005; Neels 2006), whereas unit nonresponse is higher in single child or childless households (Groves and Couper 1998; Abraham, Maitland, and Bianchi 2006). Fertility research provides ample evidence of the relative underrepresentation of childless women. Because mothers (and particularly mothers of young children) stay at home more often, they are easier to contact (Lievesley 1988; Festy and Prioux 2002). This may result in an overestimation of fertility in younger cohorts, especially in countries where young mothers tend to stay at home (Kreyenfeld et al. 2011; Sauer, Ruckdeschel, and Naderi 2012). This effect of ‘family bias’ was assumed by Festy and Prioux (2002) in the context of the FFS and Kreyenfeld and colleagues (2011) have also referred to it in relation to the GGS. In addition, it has been suggested that people with young children are particularly interested in participating in surveys (Groves and Couper 1998), particularly family surveys, since it is assumed they feel highly involved (Kreyenfeld et al. 2011). This is in line with findings indicating that high involvement or interest in the topic of the study (‘topic saliency’) induces higher participation rates (Krosnick 1991; Couper, Singer, and Kulka 1997; Pickery, Loosveldt, and Carton 2001; De Wulf, Van Kenhove, and Wijnen 2003; Sauer, Ruckdeschel, and Naderi 2012). Buber (2013), analysing attrition patterns between subsequent waves of the Austrian GGS, finds that pregnancy and more traditional attitudes towards marriage are associated with lower attrition.

Besides the relevance of the aforementioned characteristics of the designated respondent, unit nonresponse is also related to interviewer behaviour. The degree of unit nonresponse may be affected by the contact procedure adopted by the interviewer, the number of contact attempts, and the reaction of targeted respondents to interviewer characteristics (Campanelli, Sturgis, and Purdon 1997; Pickery, Loosveldt, and Carton 2001).

2.2 Issues related to response accuracy

2.2.1 Item nonresponse error

Item nonresponse error occurs when the respondent participates in the survey but skips some of the questions, with “don’t know” answers constituting an important item nonresponse problem (Weisberg 2009). To the extent that item nonresponse is non-random it may induce additional bias in sample estimates. Recall problems constitute an important cause of item nonresponse in the retrospective collection of life histories when respondents fail to remember the occurrence or timing of particular events (Blossfeld and Rohwer 2002). For event-history data in particular, both the event and the date of the event are important, as imputation of histories in case of incomplete information potentially introduces additional bias (Zabel 2009). Often the timing of the event proves most cumbersome to remember (Peters 1988; Belli 1998; Klein and Fischer-Kerli 2000; Wu, Martin, and Long 2001; Hayford and Morgan 2008). Well documented recall problems are heaping (i.e., discrete reporting of time since event) and forward telescoping (events being falsely reported as having occurred more recently). Whether or not recall problems occur often depends on the recall period. Underreporting of events is found to increase as time since the event elapses (Sudman and Bradburn 1973; 1974). Particularly for the oldest birth cohorts and the earliest periods of calendar time included in the analysis, recall errors may thus affect the quality of retrospectively estimated indicators (Neels 2006; Kreyenfeld et al. 2010; Sauer, Ruckdeschel, and Naderi 2012).

Apart from the recall period, the occurrence of recall problems depends on both event and respondent characteristics. Landmark events such as childbirth and marriage are often considered as reliable recollections, which can be recorded retrospectively with limited bias (Neels 2006; Neels and Gadeyne 2010; Klein and Fischer-Kerli 2000; Beckett et al. 2001; Hill 2005; Hayford and Morgan 2008; Kreyenfeld et al. 2011). Literature refers to the concept of salience, which relates to emotional involvement with the event considered and the concept of rehearsal of the dates of particular events (e.g., celebrating a child's birthday implies remembering it every year) (Wagenaar 1986). Both emotional involvement and the amount of rehearsal can be assumed to be high in the case of marriage or childbirth, supporting the argument of valid retrospective information on childbearing and partnership histories (Wu, Martin, and Long 2001). However, recall problems are also found for landmark events such as childbirth and union formation (Murphy 2009; Kreyenfeld et al. 2010; Ní Bhrolcháin, Beaujouan, and Murphy 2011).

With respect to respondent characteristics, gender differences in reporting vital events are found (Auriat 1991; Poulain, Riandey, and Firdion 1992; Klijzing and Cairns 2000; O'Connell 2007; Mitchell 2010). Bias often depends on whether men's or women's reports are used (Rendall et al. 1999). Men show a higher tendency to underreport children, particularly non-resident children [9] (Cherlin, Griffith, and McCarthy 1983; Sorensen 1997; Juby and Le Bourdais 1999; Rendall et al. 1999;

[9] Juby and Le Bourdais (1999), assessing declarations of births in the General Social Survey in Canada, find that the number of births declared by partners still living together is almost identical for mothers and fathers. However, underreporting of non-resident children by men was found.

Joyner et al. 2012), while women are considered more accurate in reporting both the occurrence and the correct dates of (vital) events in the domain of union and family formation. Literature not only suggests that women may be more communicative in this respect, which again entails a higher degree of rehearsal (Klijzing and Cairns 2000), but also that in some cases men may not be aware of the birth of their children (Juby and Le Bourdais 1999; Joyner et al. 2012). However contradictory evidence is also found (e.g., Hertrich 1998). With respect to age, previous results suggest that older people tend to underreport children that have left the parental home during periods more distant from the survey (Kreyenfeld et al. 2010). Also, education has an effect on recall of life-course events, with higher-educated women being found to be more reliable in reporting marital histories and divorce dates, childbearing desires, and methods of contraception (Peters 1988; Smith and Thomas 2003; Mitchell 2010).

2.2.2 Measurement error due to the respondent

Measurement error occurs when the measure obtained is not an accurate measure of what was intended to be measured. Measurement error due to the respondent is when the respondent gives an inaccurate answer to the question, which is often really a matter of how well the researcher worded the survey question (Weisberg 2009). The ability to remember is no guarantee that facts are correctly reported. Respondents may choose to edit their answers as a result of learning effects and fatigue (Ní Bhrolcháin, Beaujouan, and Murphy 2011), particularly when respondents lose interest and motivation to participate during the interview. After learning that answering questions in a particular way may prevent further questions, respondents may use this knowledge to shorten the interview (see Ruckdeschel et al. in this special volume). People may opt not to report additional children in order to end the interview as soon as possible (Piacentini, Jensen, and Schaffer 1991; Kessler, Little, and Groves 1995; Lucas et al. 1999; Duan et al. 2007). Also, social desirability may introduce bias. People often want to put themselves in a more favourable light and may give responses they think interviewers want to get. A typical example of social desirability in demographic research that may introduce varying bias by birth cohort is the increased social acceptance and the changing prevalence of cohabitation and non-marital childbirth (Hayford and Morgan 2008).

As argued by Weisberg (2009), measurement error is closely related to question wording and questionnaire construction. When evaluating the quality of survey data it is important to consider the characteristics of the questionnaire (e.g., complexity, question wording). The complexity of the questionnaire is found to be an important determinant of bias in data, as survey complexity may increase interview fatigue (Kreyenfeld et al. 2010; Sauer, Ruckdeschel, and Naderi 2012). Ní Bhrolcháin and colleagues (2011), for instance, found that the main reason for missing births, especially for older cohorts of women (Murphy 2009), was an inaccuracy in the questionnaire, as most of these ‘missing’ children had been reported as household members but had not been repeated when providing birth histories. Beaujouan (2013) identifies both the complex filtering in the GGS as well as inconsistent implementation across countries (e.g., question filters, pre-codes) as a potential threat to data quality.

Little information is provided on the translation procedures that countries have used to adapt the questionnaires to language- and country-specific contexts. However, the country-specific data documentation suggests that some countries (e.g., Bulgaria) have validated their translation by re-translating their country-specific questionnaire back to English. Other countries have set up pilot studies to assess the quality of the translation work (e.g., France) (Generations and Gender Programme 2014). We assume that the questions of interest are less country- or language-specific (e.g., birth year of biological child).

Finally, the presence of others than the respondent and the interviewer during the interview possibly generates further bias in the outcomes of the survey (Régnier-Loilier 2007) (e.g., underreporting of children with previous partners due to presence of current partner).

2.2.3 Measurement error due to the interviewer

Measurement error due to the interviewer occurs when effects associated with the interviewers lead to inaccurate measures (Weisberg 2009). The interview situation and interviewer characteristics may also affect the data quality collected in surveys (Davis et al. 2010), even when standardized methods are used (Childs and Landreth 2006). Whenever interaction between a respondent and an interviewer occurs, interviewer bias becomes a realistic problem. Interviewers may unintentionally miscode data, skip questions, or deviate from the prescribed interviewing procedures. Learning effects on the part of the interviewer include intentionally introduced bias as a result of making up responses, misreporting deliberately, falsifying disposition codes, miscoding data or deviating from prescribed interviewing procedures (Sauer, Ruckdeschel, and Naderi 2012). Standardization of the interview situation and instructions for interviewers (e.g., standard responses, show cards) may mitigate the interviewer’s influence on data collection. In the GGS, fieldwork guidelines were provided rather than strict prescriptions concerning fieldwork practices. As a result, fieldwork practices vary greatly by country (see Fokkema et al. in this special volume).

2.3 Survey administration issues

2.3.1 Mode effects

Mode effects refer to the effects that the choice between face-to-face interviewing, phone surveys, mail questionnaires, and other survey modes has on the results that are obtained (Weisberg 2009). In case of the GGS, the use of register data can be considered as a specific mode of data collection where responses were imputed from available registers in some countries (e.g., GGS Norway). This may give rise to additional issues of measurement error when the official (de jure) living arrangement does not reflect the de facto living arrangement of respondents.

2.3.2 Post-survey error

Post-survey error is the error that occurs in processing and analysing survey data, and does not constitute error from the survey process itself (Weisberg 2009). An important post-survey process is data editing, aimed at trying to locate and correct the individual errors in survey data, which in turn raises concern about the possibility of over-editing. In several countries computer-assisted personal interviewing (CAPI) was used during data collection for the GGS, which allows moving part of the data editing process to the interview itself by catching implausible combinations of answers and having the interviewer correct one or more entries, or asking respondents for additional clarification in case of apparent inconsistencies (Weisberg 2009). The lack in the GGS questionnaire of control questions on current civil status or parity would seem to provide limited possibilities for field editing, but it is unclear whether and to what extent such controls were implemented in different countries using CAPI. Similarly, it is unclear for most countries whether information available from administrative registers was used to impute missing data on household composition, partnership histories, and/or maternity histories in the stage of data cleaning.

2.3.3 Comparability effects

Finally, comparability effects refer to differences between survey results obtained by different survey organizations or in different nations or at different points in time (Weisberg 2009). In the case of the GGS, the harmonization of questions that have been adapted to the national context constitutes an additional step in the process of data editing. The harmonization of the GGS is performed at the Netherlands

Interdisciplinary Demographic Institute (NIDI), but a detailed log documenting the changes and corrections implemented in subsequent versions of the harmonized data is hitherto unavailable.

2.4 Assessing total survey error

Studies looking into the quality of retrospective data have come up with mixed results, depending on the topic studied (Dex and McCulloch 1998; Jacobs 2002; Ayhan and Işiksal 2004; Shachar and Eckstein 2007; Gibson and Kim 2010). Although fertility and nuptiality histories are mostly regarded as ‘hard facts’ which respondents are keen to report (Swicegood, Morgan, and Rindfuss 1984; Wu, Martin, and Long 2001), several sources of error may affect the quality of retrospective data and introduce bias in GGS estimates of nuptiality and fertility (Blossfeld and Rohwer 2002; Blossfeld, Golsch, and Rohwer 2007).

In the case of the GGS, the survey guidelines predominantly focus on dealing with selection bias, with the target response rate being set at 80% (Simard and Franklin 2005). Similarly, replacement of non-respondents with other respondents is not advised. To correct for selective nonresponse, the use of nonresponse weight adjustments and the examination of non-random nonresponse patterns are recommended as solutions. To analyse nonresponse it is essential that all individuals sampled in each wave are included in the final dataset with their final disposition codes (Kveder 2005). In addition, the calculation of response rates is standardized, following procedures outlined by Lynn and colleagues (2001), the American Association for Public Opinion Research (2011), and Kveder (2005) (Fokkema et al. in this special volume). Although standardized definitions of final disposition codes and response rates give information on a specific source of survey error (Lynn et al. 2001; Kveder 2005), standardised response rates do not provide any information on the joint impact of various potential sources of survey error on the quality of the longitudinal demographic data in the GGS. As a result, the GGS sample design guidelines recommend that design weights which are adjusted to account for nonresponse or which make use of auxiliary data (e.g., post-stratification) are validated by comparing weighted GGS estimates to other sources (e.g., vital statistics) to verify whether the GGS estimates are accurate (Simard and Franklin 2005). Assessments of the quality of demographic data are available in the literature for the FFS, but only a limited number of contributions have considered the quality of demographic data in the GGS. Festy and Prioux (2002) have compared period total fertility rates calculated retrospectively from the FFS with vital statistics for the 24 participating countries. An overestimation of fertility was found, which the authors ascribe to an overrepresentation of married women with children, who were easier to interview. Also, the Austrian GGS shows overestimations of fertility in line with findings by Festy and Prioux, due to similar mechanisms (Kreyenfeld et al. 2011). Additionally, assessments on the quality of fertility and nuptiality data have been made for the German GGS (Kreyenfeld et al. 2010; Kreyenfeld et al. 2011; Sauer, Ruckdeschel, and Naderi 2012), indicating that fertility and nuptiality are underestimated for older birth cohorts and overestimated for younger birth cohorts. Additional efforts to assess data quality have been made by Burkimsher (2009) for the Bulgarian, Hungarian, and Georgian GGS, finding underestimations of cohort fertility for older cohorts and major validity problems in the case of Bulgaria. Many of the aforementioned contributions draw attention to the importance of weight adjustments and particularly to relevant factors that should be considered in calculating post-stratification weights (which is often an arbitrary decision) (Festy and Prioux 2002). For Austria, special weights were developed by the Vienna Institute of Demography that include fertility information on parity in their computation. It has been suggested that these weighted data perform considerably better (Buber 2010; Kreyenfeld et al. 2011).

In this paper we assess the validity of final estimation weights by comparing GGS estimates of period and cohort indicators of first marriage and fertility as well as mean ages at first marriage and childbearing to population statistics. To our knowledge this is the only contribution assessing the quality of demographic data for all countries where data from GGS wave 1 are currently available (Table 1). The added value is self-evident for countries where the quality of retrospectively estimated demographic indicators has not yet been assessed. Also, for countries for which partial results are available, the consistent approach of our analysis has the advantage that results can be compared across countries. The aim of this contribution is to provide users of the GGS data with information that may help them to decide on the temporal and geographical scope of their analysis or guide further research on item nonresponse in the GGS. We also make suggestions on how data quality may be improved in future surveys.

3. Data & methods

Section 3.1 discusses the Harmonized Data Files used for the analysis, whereas section 3.2 documents the retrospective estimation of indicators of first marriage and fertility from the GGS. Sections 3.3–3.7 reflect on all different indicators that are calculated in the analyses as well as on the results that will be presented.

3.1 Harmonized data files

The analyses use data on 14 countries for which GGS Wave 1 data are currently available (see Table 1 for country-specific information on the harmonized data files used). The analysis is restricted to data for women, as population statistics typically provide information on female rates only (Council of Europe 2005). As a result, our results provide no information on the validity of retrospectively collected demographic data for men. For each country, Table 1 documents the period and cohorts considered in the validation (see below). The time period or birth cohorts for which we assess the quality of indicators on first marriage is subject to variation between countries and is determined by the starting year in which the interviews took place and the age range of respondents included in the cross-sectional sample (see section 3.2). The scope of the validation is determined by the requirement to consider cohort and period indicators of fertility and nuptiality between ages 15 and 49[10], although the results section also includes period and cohort indicators based on the age range 15–39. The GGS estimates were compared to population statistics drawn from various sources. In this paper we mainly use population statistics drawn from the GGP Contextual Database (CDB), although alternative sources have been considered when they provided longer time-series (e.g., Eurostat, Council of Europe, Human Fertility Database).

GGS estimates were calculated using both weighted and non-weighted data[11]. The following indicators of first marriage were estimated retrospectively from the GGS: i) age-specific female first-marriage rates (ASFFMR), ii) the period total female first-marriage rate (period TFFMR), iii) the cohort total female first-marriage rate (cohort TFFMR), iv) the period mean age at female first marriage (period MAFFM), and v) the cohort mean age at female first marriage (cohort MAFFM). In calculating first-marriage rates we considered the earliest of three possible dates. The first date concerns the year of marriage to the current co-resident partner. Co-resident partners are included in the

[10] Age-specific rates at age reached in year *t* are calculated retrospectively considering the respondent's date of birth.

[11] For the Netherlands two different sets of weights were used. The first set of weights corrects for the sample design and takes into account that a random sample of households was taken instead of a random sample of individuals. The second set of weights additionally corrects for the underrepresentation of certain groups (e.g., men, young adults, inhabitants of highly urban and highly rural areas…). Since the GGP sample design guidelines are unclear on the recommended weighting procedure to be used (Simard and Franklin 2005), most countries have implemented their own solution. In general, GGS countries other than the Netherlands provide a single standardized weight variable based on post-stratification of the sample for a limited number of characteristics (frequently age, sex, and either region or urbanization) (see Fokkema et al. in this special volume).

household grid of the GGS[12] and the section on partnerships provides additional information on whether the respondent is married to this partner (in which case month and year of marriage are recorded). A second possibility concerns the date of marriage to the current non-resident partner, as the section on non-resident partners probes whether respondents have a living-apart-together marriage[13]. Finally, the earliest date of marriage in the partnership history of the respondent is considered as, for each partnership, respondents were asked whether and when they married this partner[14].

Table 1: Descriptives for GGS Wave 1 data included in the analyses

						Time perspective:	
Country	**HDF**	**Age range**	**Birth cohorts**	**Year of interview**	**N**	**Period measures**	**Cohort measures**
Australia	4.0	16–99	1906–1990	2005–2006	3944	1970–2005	1920–1957
Austria[1]	3.0	18–46	1963–1990	2008–2009	3001	–	–
Belgium	3.0	18–82	1928–1990	2008–2010	3728	1978–2005	1928–1960
Bulgaria	4.0	18–82	1922–1986	2004	6983	1977–2001	1927–1954
Estonia	3.0	21–81	1924–1983	2004–2005	5034	1974–1998	1924–1955
France	3.0	18–79	1926–1987	2005	5708	1976–2002	1926–1955
Georgia	3.0	18–80	1926–1988	2006	5595	1976–2003	1926–1956
Germany	3.0	17–85	1920–1988	2005	5407	1975–2002	1925–1955
Hungary	3.0	21–79	1926–1983	2004–2005	7517	1976–1998	1926–1955
Italy	4.0	18–64	1939–1985	2003	5115	1989–2000	1939–1953
Netherlands	3.0	18–80	1923–1985	2002–2004	4741	1973–2000	1923–1954
Norway	3.0	19–81	1927–1988	2007–2008	7541	1977–2003	1927–1958
Romania	4.0	18–80	1925–1987	2005	6009	1976–2002	1926–1955
Russia	3.0	17–81	1923–1987	2004	7038	1974–2001	1924–1954

Note: [1] Period and cohort measures cannot be estimated up to age 49 given the limited age range of respondents in the cross-sectional sample.
[2] HDF: Harmonized Data File Release.

With respect to fertility, GGS estimates for the following indicators were compared to population statistics: i) age-specific female fertility rates (ASFR), ii) the period total fertility rate (period TFR), iii) the cohort total fertility rate (cohort TFR), iv) the period mean age at childbearing (period MAC), and v) the cohort mean age at childbearing (cohort MAC). The fertility indicators only use data on biological children of the respondent: we do not consider adopted, foster, or stepchildren. Dates of birth of biological children were drawn from different sections of the GGS questionnaire. First,

[12] Respondents are asked to fill out the household grid at the start of the interview. 'To begin, I would like to ask you about all persons who live in this household. Who are they? To help me keep track of your answers, please tell me their first names and how they are related to you.'
[13] Are you currently having an intimate (couple) relationship with someone you're not living with? This may also be your spouse if he/she does not live together with you.'
[14] (Apart from your current partnership or marriage,) have you ever before lived together with someone as a couple or have you ever been married?'

the household grid provides the dates of birth of biological children residing in the household of the respondent. Second, dates of birth of non-resident and deceased biological children were drawn from a separate section in the questionnaire on non-resident children[15].

3.2 Retrospective estimation of age-specific rates

The retrospective estimation of demographic indicators from the GGS follows the procedure outlined by Neels (2006) for the validation of period and cohort indicators of first marriage and fertility, estimated retrospectively from the Belgian censuses of 1991 and 2001 (Neels 2006; Neels and Gadeyne 2010). Using the earliest date of marriage in the partnership history, age-specific female first-marriage rates ($ASFFMR_t$) are calculated by calendar year and age reached during the year (see left panel of Figure 1):

$$ASFFMR_i = \frac{\text{first marriages in year } t \text{ to women who reach age } i \text{ during year } t}{\text{number of women born in year } t-i} \tag{1}$$

where the $ASFFMR_t$ for age *i* reached in year *t* relates the number of first marriages in year *t* to women born in year *t* - *i* to the number of person-years lived by these women in year *t* (the number of person-years lived is equal to the cohort size in the sample, as there is no mortality or emigration in the retrospective estimation of age-specific rates). Similarly, using the dates of birth of biological children, age-specific fertility rates ($ASFR_t$) are calculated by calendar year and age reached during the year:

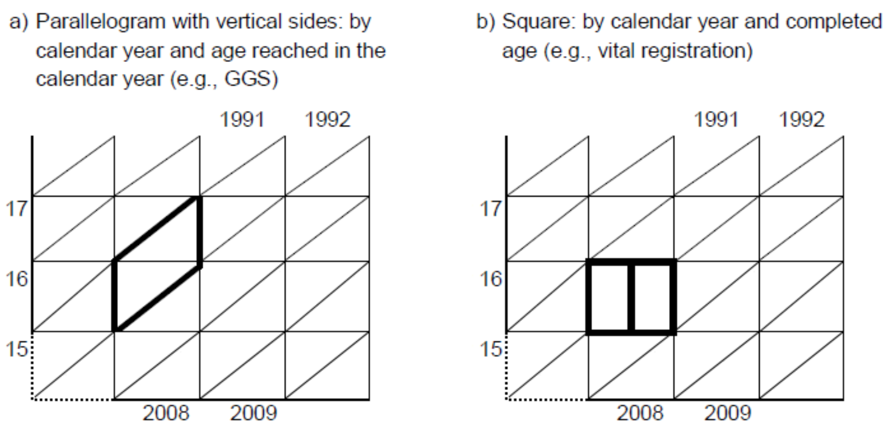
$$ASFR_i = \frac{\text{births in year } t \text{ to women who reach age } i \text{ during year } t}{\text{number of women born in year } t-i} \tag{2}$$

In contrast to the age-specific rates estimated from the GGS, age-specific rates drawn from population statistics are routinely calculated by calendar year and completed age, where events to women aged *i* in completed years are related to the midyear population of women aged *i* on their last birthday (right panel of Figure 1)

[15] 'We already talked about those children who currently live in your household. In addition to them, have you given birth to/fathered any other children or have you ever adopted any other children? Do not include stepchildren, that is, children belonging to your current or prior partner/spouse. I will ask you about those children later.'

(Wunsch and Termote 1978; Calot 1984; Neels 2006). The GGS estimates of the $ASFFMR_t$ and $ASFR_t$ by age reached in year *t* refer to events in adjacent age groups and are centred at exact age *i* (left panel of Figure 1), whereas the $ASFFMR_t$ and $ASFR_t$ drawn from population statistics are typically centred at age (*i* + 0.5). To obtain a closer approximation between age schedules of first marriage and fertility, estimated from the GGS and corresponding schedules drawn from population statistics, the age-specific rates estimated from the GGS were averaged over adjacent ages. The age-specific rate at completed age *i* drawn from population statistics is thus approximated by the mean of the age-specific rates at ages reached *i* and (*i* + 1), estimated retrospectively from the GGS (Wunsch and Termote 1978; Calot 1984; Neels 2006).

Figure 1: Calculation of age-specific fertility rates



The time periods or cohorts considered in the analysis depend on the year in which fieldwork started and the age range included in the cross-sectional samples. For period indicators the earliest year that can be considered in the analyses is the year in which the oldest female birth cohort (which includes at least 20 women) reaches its 50th birthday (see Table 1), or, alternatively, the earliest year for which all age-specific rates can be estimated between completed ages 15 and 49. Similarly, the most recent year considered is the year in which the youngest cohort in the sample turns 15. Given that in most countries the GGS sampled individuals aged 18 and older, period indicators can typically be estimated up to 3 years before the year in which the fieldwork started. Cohort indicators reflect the proportion ever marrying by age 49 or completed fertility by age 49 in the cohorts considered. In contrast to the estimation of the period measures, the age-specific rates by age reached estimated from the GGS were not averaged over adjacent age-groups for the estimation of cohort indicators. For cohort indicators, the assessment starts with the oldest cohort (having at least 20 respondents) included in the survey and includes the cohort of women turning 49 in the year preceding the start of the fieldwork.

3.3 Period total female first-marriage rate and period total fertility rate

From the ASFFMR, the period TFFMR in year *t* is obtained as:

$$Period \quad TFFMR_t = \sum_{i=15}^{49} \frac{1}{2}(g_i^t + g_{i+1}^t) \tag{3}$$

where g_i^t and g_{i+1}^t are the GGS estimates of the ASFFMR by ages *i* and (*i* + 1) reached in year *t*. Similarly, the period TFR in year *t* is obtained as:

$$Period \quad TFR_t = \sum_{i=15}^{49} \frac{1}{2}(f_i^t + f_{i+1}^t) \tag{4}$$

where f_i^t and f_{i+1}^t are the GGS estimates of the ASFR by ages *i* and (*i* + 1) reached in year *t*.

3.4 Period mean age at female first marriage and period mean age at childbearing

The calculation of the period mean ages is based on the age-specific and total rates. The period MAFFM and the period MAC are obtained as:

$$Period \quad MAFFM_t = \frac{\sum_{i=15}^{49} \frac{1}{2}(g_i^t + g_{i+1}^t) * (i + 0.5)}{Period \quad TFFMR_t} \tag{5}$$

$$Period \quad MAC_t = \frac{\sum_{i=15}^{49} \frac{1}{2}(f_i^t + f_{i+1}^t) * (i + 0.5)}{Period \quad TFR_t} \tag{6}$$

where g_i^t, g_{i+1}^t and f_i^t, f_{i+1}^t in equations 5 and 6 respectively represent the ASFFMR and ASFR by age reached *i* and (*i* + 1) in year *t* calculated retrospectively from the GGS.

3.5 Cohort total female first-marriage rate and cohort total fertility rate

Using the age-specific rates by age reached, the cohort TFFMR is obtained as follows:

$$Cohort\ \ TFFMR_t = \sum_{i=15}^{49} (g_i^t) \tag{7}$$

where g_i^t are the ASFFMR by age reached *i* in year *t* drawn from the GGS for a cohort of women born in year *t*. Similarly, the cohort TFR is obtained as:

$$Cohort\ \ TFR_t = \sum_{i=15}^{49} (f_i^t) \tag{8}$$

where f_i^t are the ASFR by age reached *i* in year *t* drawn from the GGS for a cohort of women born in year *t*.

3.6 Cohort mean age at female first marriage and cohort mean age at childbearing

Cohort mean ages at first marriage and childbearing are derived from the corresponding cohort schedules. The cohort MAFFM and MAC for women born in year *t* are obtained as follows:

$$Cohort\ \ MAFFM_t = \frac{\sum_{i=15}^{49} i * g_i^t}{Cohort\ \ TFFMR_t} \tag{9}$$

$$Cohort\ \ MAC_t = \frac{\sum_{i=15}^{49} i * f_i^t}{Cohort\ \ TFR_t} \tag{10}$$

where g_i^t and f_i^t in equations 9 and 10 respectively represent the ASFFMR and ASFR by age reached *i* in year *t* estimated from the GGS.

3.7 Indicators of survey error

Following Weisberg (2009), three indicators of survey error are reported. First, we calculate the mean difference between the GGS estimates and the corresponding population values throughout the observation period or range of cohorts considered, as well as for subsequent five-year periods or birth cohorts. The mean difference provides an indication of the systematic error or bias in GGS estimates. If the systematic error is constant across the period or range of cohorts considered, the variance of the GGS estimate may still be correct. Also, adding a constant to a variable does not affect correlations with other variables (e.g., time-series of relevant contextual variables) or regression coefficients, so these statistics may still be correct (Weisberg 2009). Second, for each indicator we calculate the standard deviation between the GGS estimates and the corresponding population statistics (the square root of the mean squared error which is the average of the squared deviations between the GGS estimates and the population values), which is taken as an indication of the random error in GGS estimates. Assuming random error has a mean of zero, it does not affect the mean of a variable, but does increase the variance of a variable and therefore directly affects correlations with other variables (e.g., time-series of relevant contextual variables), which are reduced in magnitude or attenuated (Weisberg 2009). Similarly, random error attenuates regression coefficients and reduces the statistical power of hypothesis tests compared to the situation where random error is absent. Third, to illustrate the joint impact of random error and changing systematic bias over time, we calculate the zero-order correlation between the GGS estimates and the population values for each demographic indicator considered.

4. Results

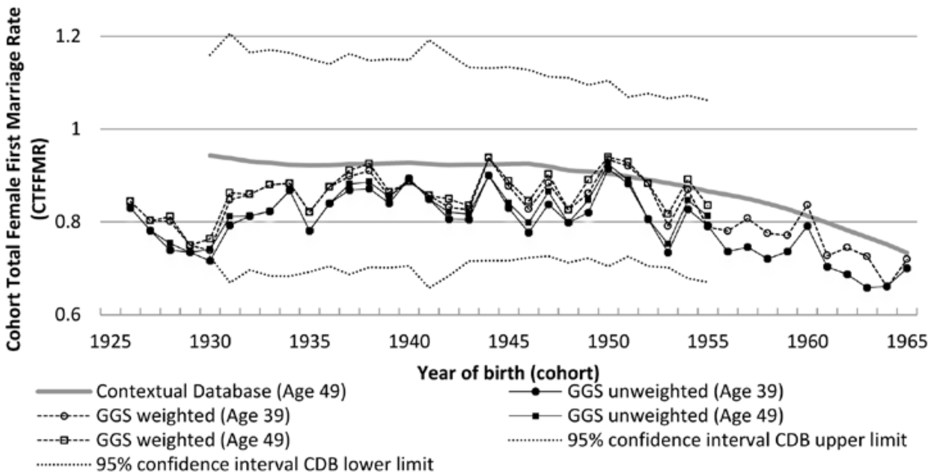
For each indicator of first marriage and fertility, graphs for a limited number of countries have been included in the text in order to illustrate variation in data. Detailed country-specific figures comparing estimates based on the GGS (and their 95% confidence intervals) against population statistics are included in the appendix.

4.1 Cohort total female first marriage rate and mean age at female first marriage

The assessment of the quality of the cohort total female first marriage rate (cohort TFFMR) is limited to countries for which population statistics on cohort nuptiality patterns are available (Austria, Belgium, Bulgaria, France, Germany, Hungary, Italy, Netherlands, Norway, and Romania). In most of the countries considered, the GGS underestimates the cohort TFFMR, particularly for the older birth cohorts included in the survey. The underestimation is particularly articulated in Belgium and Norway (Figure 5) up to the 1945 cohort and in Germany (Figure 3) up to the cohort born in 1955. Compared to population statistics, the GGS-based TFFMR is about 0.15 first marriages per woman too low in Belgium and Norway in the oldest cohorts, ranging up to an underestimation of 0.20 marriages per woman in Germany (Table 2). In Germany GGS-based TFFMR for cohorts born before 1940 is no longer included in the 95% confidence intervals around the population parameter. Also in Bulgaria, France (Figure 2), and the Netherlands, the GGS underestimate the cohort TFFMR for all birth cohorts considered. In the Netherlands, the underestimation ranges up to 0.50 first marriages per woman for the oldest cohorts. In Italy, Romania, and Hungary, no cohorts are found where the GGS estimate differs substantially from vital statistics. For Austria the analysis is only possible for a limited period, due to the limited age range covered in the cross-sectional sample. For these cohorts, however, the results do not show substantial deviations between GGS estimates and population statistics. In Hungary the difference between the GGS estimates and population statistics is generally below -0.10 for cohorts born between 1940 up to the mid-1960s, with deviations between GGS and population statistics being even smaller for cohorts born before 1940.

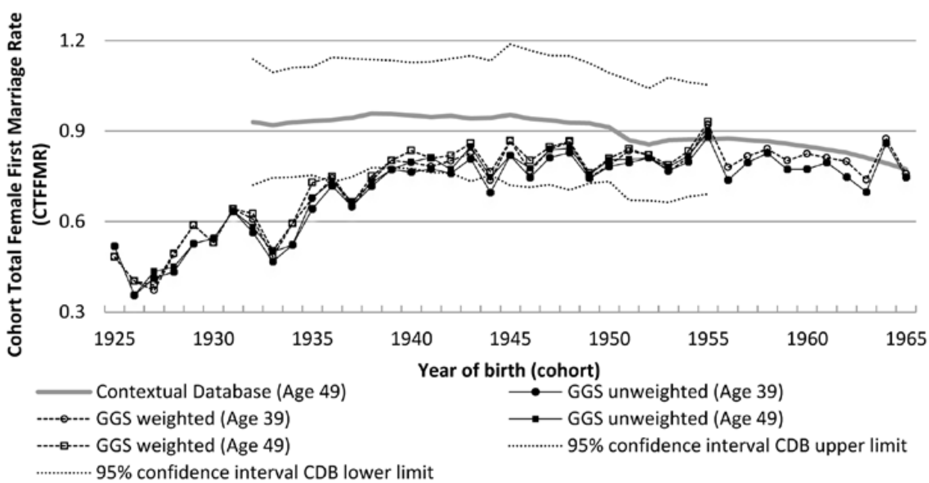
In sum, the outcomes show that compared to population statistics the GGS generally underestimate the cohort TFFMR, particularly for the older birth cohorts. The mean deviation between GGS estimates and populations statistics is negative in all countries considered (Table 3), indicating a general underestimation of cohort first-marriage intensities. The changing nature of the systematic bias over subsequent birth cohorts in tandem with substantial year-to-year variation in the GGS estimates in general gives rise to low zero-order correlations between TFFMR series estimated from the GGS and population statistics (Table 3), suggesting that long-term trends in cohort nuptiality patterns are often poorly reflected by the GGS (in Belgium a significant negative correlation was found). However, if zero-order correlations are calculated separately for younger birth cohorts (results not included), the quality of the cohort nuptiality data is clearly higher (resulting in positive cross-correlations) for the more recent birth cohorts.

Figure 2: Cohort total female first marriage rate, France, 1926–1965



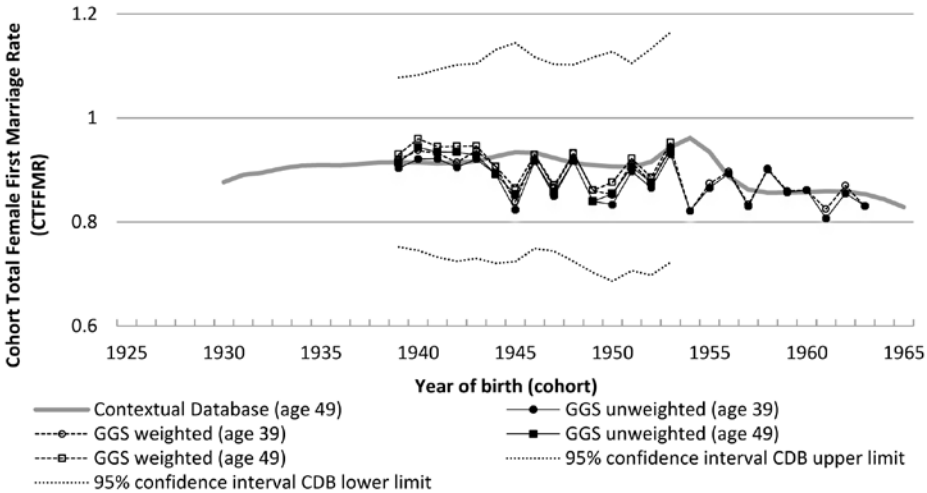
Source: Generations & Gender Survey France (Wave 1), Calculations by authors.

Figure 3: Cohort total female first marriage rate, Germany, 1925–1965



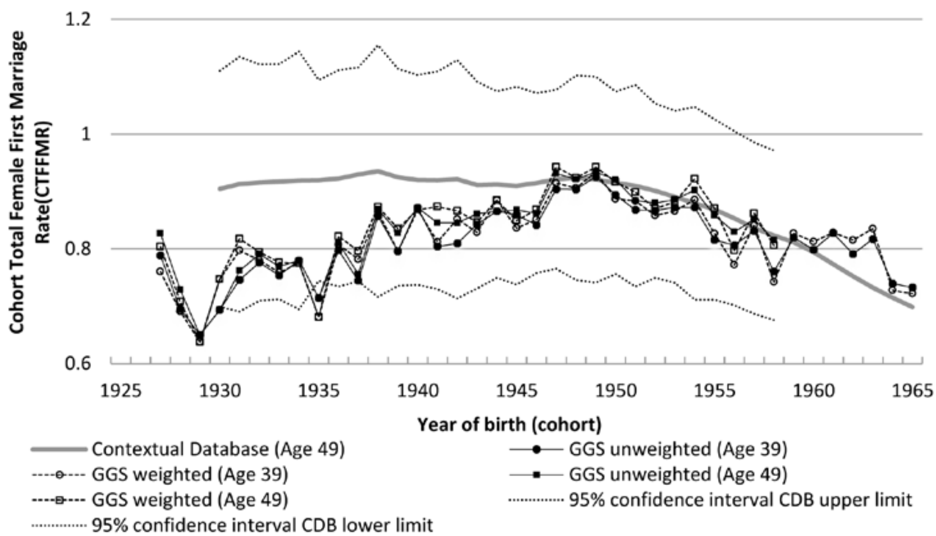
Source: Generations & Gender Survey Germany (Wave 1), Calculations by authors.

Figure 4: Cohort total female first marriage rate, Italy, 1931–1965



Source: Generations & Gender Survey Italy (Wave 1), Calculations by authors.

Figure 5: Cohort total female first marriage rate, Norway, 1927–1965



Source: Generations & Gender Survey Norway (Wave 1), Calculations by authors.

Table 2: Cohort TFFMR (calculated up to age 49): mean difference between GGS estimates and population statistics for subsequent 5-year birth cohorts

	1930–1934	1935–1939	1940–1944	1945–1949	1950–1954	1955–1959
Australia	–	–	–	–	–	–
Austria	–	–	–	–	–	–
Belgium	–	-0.1374	-0.1305	-0.0402	0.0055	0.0030
Bulgaria	–	–	–	-0.0611	-0.0765	–
Estonia	–	–	–	–	–	–
France	-0.0825	-0.0446	-0.0512	-0.0476	0.0016	–
Georgia	–	–	–	–	–	–
Germany	–	-0.2077	-0.1293	-0.1081	-0.0583	–
Hungary	-0.0077	-0.0135	-0.0463	-0.0547	-0.0574	–
Italy	–	–	0.0239	-0.0310	–	–
Netherlands (w0)	-0.4529	-0.2594	-0.2175	-0.1886	-0.1764	–
Netherlands (w1)	-0.4977	-0.2908	-0.2437	-0.2173	-0.2018	–
Norway	-0.1317	-0.1250	-0.0489	-0.0128	-0.0014	–
Romania	–	–	0.0204	0.0083	-0.0869	–
Russia	–	–	–	–	–	–

Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 3: Cohort TTFMR and cohort MAFFM: mean difference, standard deviation of difference and zero-order correlation between GGS estimates and populations statistics

Country	Cohorts	CTFFMR			CMAFFM		
		Correlation	Mean differ.	S.D. differ.	Correlation	Mean diff.	S.D. differ.
Australia	–	–	–	–	–	–	–
Austria	–	–	–	–	–	–	–
Belgium	1933–1960	-0.509**	-0.063	0.079	0.453*	0.649	0.663
Bulgaria	1942–1954	0.002	-0.077	0.034	-0.116	0.435	0.657
Estonia	–	–	–	–	–	–	–
France	1930–1955	-0.091	-0.044	0.047	0.240	0.096	0.615
Georgia	–	–	–	–	–	–	–
Germany	1932–1955	-0.185	-0.146	0.108	0.574**	0.689	0.600
Hungary	1930–1955	0.494*	-0.037	0.032	0.087	-0.056	0.392
Italy	1939–1953	0.064	-0.004	0.036	0.659**	-0.347	0.378
Netherlands (w0)	1930–1954	-0.204	-0.259	0.116	0.575**	1.496	0.922
Netherlands (w1)	1930–1954	-0.229	-0.290	0.121	0.558**	1.474	0.938
Norway	1930–1958	0.051	-0.057	0.065	0.536**	0.474	0.701
Romania	1938–1955	0.130	-0.020	0.060	0.386	0.011	0.396
Russia	–	–	–	–	–	–	–

Note: Significance levels: p < .05 (*), p < .01 (**).
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 4: Cohort MAFFM: mean difference between GGS estimates and population statistics in subsequent 5-year birth cohorts

	1930–1934	1935–1939	1940–1944	1945–1949	1950–1954	1955–1959
Australia	–	–	–	–	–	–
Austria	–	–	–	–	–	–
Belgium	–	0.4853	0.9144	0.3817	0.8732	0.3916
Bulgaria	–	–	–	0.8486	-0.0260	–
Estonia	–	–	–	–	–	–
France	-0.2020	-0.1221	0.0725	0.1207	0.3607	–
Georgia	–	–	–	–	–	–
Germany	–	0.7315	1.1081	0.5484	0.6031	–
Hungary	-0.5623	-0.0917	0.2482	0.0787	0.0685	–
Italy	–	–	-0.4450	-0.4750	–	–
Netherlands (w0)	1.7968	0.9351	1.4779	1.2412	2.0268	–
Netherlands (w1)	1.7303	0.9294	1.5014	1.1221	2.0841	–
Norway	-0.1227	0.8346	0.5081	0.5358	0.2166	–
Romania	–	–	-0.0085	-0.0385	0.0975	–
Russia	–	–	–	–	–	–

Source: Generations & Gender Survey (Wave 1), Calculations by authors.

The analysis of the cohort MAFFM excludes Australia, Estonia, Georgia, and Russia, as population statistics on the cohort MAFFM were not available for these countries. For France, Belgium, Bulgaria, Germany, Hungary, Italy, Norway, and Romania, the GGS estimates of the cohort MAFFM show accurate approximations of population statistics without substantial deviation (<1 year). In the Netherlands the GGS overestimates the cohort MAFFM for nearly all cohorts, with deviations ranging up to 2.08 years (Table 4).

In sum, the GGS estimates of the cohort MAFFM provide a closer approximation of population statistics than is the case for GGS estimates of the cohort TFFMR. As a result, the time-series of cohort MAFFM estimated from the GGS show significant positive correlations with the MAFFM drawn from population statistics (Table 3). Considering the results for both the cohort TFFMR and the cohort MAFFM, we can conclude that the cohort nuptiality data in the GGS are fairly accurate for cohorts born after 1940–1945, notwithstanding exceptions in a limited number of countries (see summary indicators and detailed country results in the appendix).

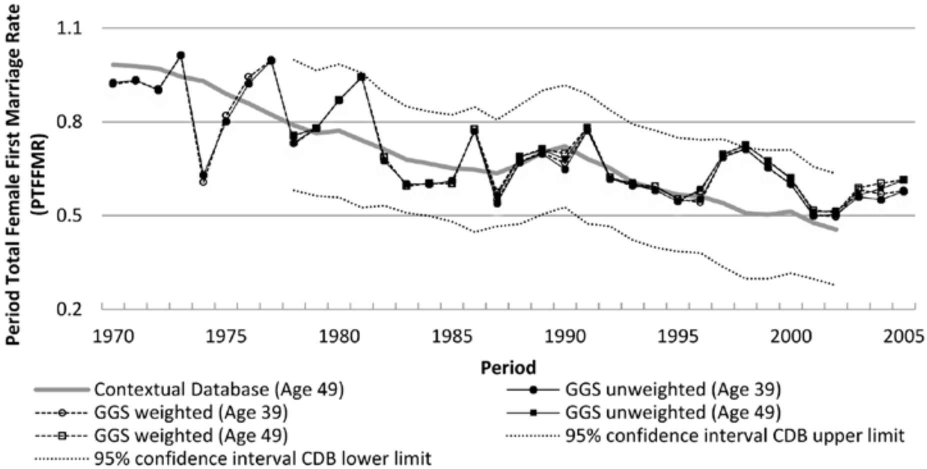
4.2 Period total female first-marriage rate and mean age at female first marriage

The quality of period TFFMR was assessed for all countries where period population statistics on first marriage were available (Austria, Belgium, Bulgaria, Estonia, France, Georgia, Germany, Hungary, Italy, Netherlands, Norway, Romania, and Russia). Compared to cohort indicators of first marriage, the period TFFMR estimated from the

GGS approximate population statistics more closely. Consistent with the findings for the cohort TFFMR, however, an underestimation of first-marriage intensities is found in earlier periods for a number of countries. This underestimation mainly occurs in the late 1970s (the Netherlands and Russia in Table 6) and 1980s (Bulgaria and the Netherlands). For these countries the underestimation of the period TFFMR in earlier periods is generally below -0.05, as shown in Figure 8 for Hungary and Figure 9 for Russia. The Netherlands is exceptional, as the deviation in the period TFFMR is on average -0.16 during the 1970s. In Austria, Belgium, Germany, Italy, and the Netherlands the GGS overestimates period first-marriage intensities in the 1990s and the 2000s, with the bias being particularly pronounced in Germany (ranging up to +0.12 on average in the period 1995–1999). In Belgium and Italy, overestimations of the period TFFMR ranging up to +0.10 on average are found after 1995. Countries where the GGS estimates provide fairly accurate approximations of population statistics are Estonia (Figure 7), Georgia, Norway, and Romania. Despite variation in data quality across countries, the GGS-based period TFFMR in most countries shows significant positive correlations with time-series drawn from population statistics, indicating that the GGS accurately represent period trends in first-marriage intensities in these countries.

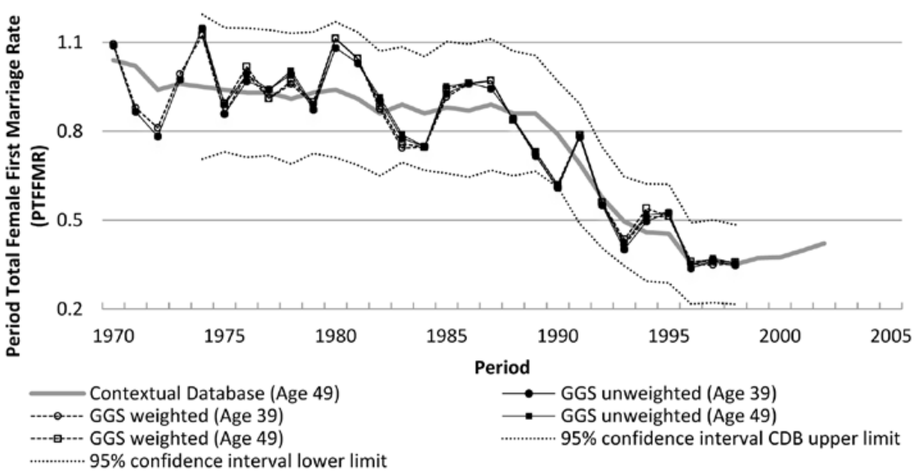
The quality of period MAFFM could be assessed for all countries, except Austria - due to a lack of comparability between population statistics (mean ages based on ASFFMR up to age 49) and the GGS (mean ages based on ASFFMR up to age 34). For most countries (Belgium, Bulgaria, Estonia, France, Germany, Hungary, Italy, Norway, Romania, and Russia) the period MAFFMs estimated from the GGS do not deviate substantially from population statistics in the period considered (deviation <1 year). In Australia the GGS overestimates the period MAFFM by more than one year between 1980 and 1984. In Georgia the GGS underestimates the period MAFFM throughout the observation period, with deviations ranging up to 3 years in specific years. The Netherlands is again exceptional, as the GGS overestimates the period MAFFM in most periods, with deviations ranging up to +1.75 years in subsequent 5-year intervals. For most countries, however, the estimates of the period MAFFM based on the GGS show significant positive correlations with population statistics, illustrating that the GGS adequately represent period trends in terms of the MAFFM in these countries (Table 5).

Figure 6: Period total female first-marriage rate, Belgium, 1970–2005



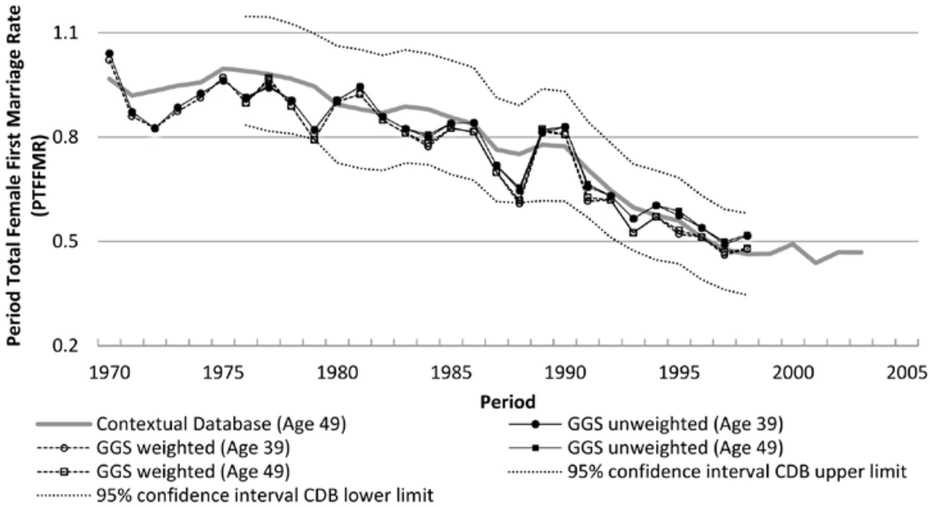
Source: Generations & Gender Survey Belgium (Wave 1), Calculations by authors.

Figure 7: Period total female first-marriage rate, Estonia, 1970–2002



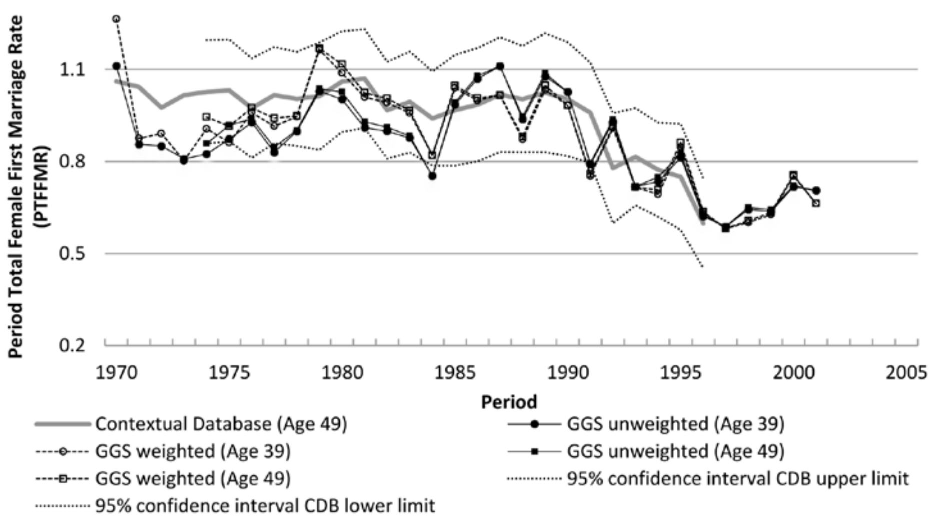
Source: Generations & Gender Survey Estonia (Wave 1), Calculations by authors

Figure 8: Period total female first-marriage rate, Hungary, 1970–2003



Source: Generations & Gender Survey Hungary (Wave 1), Calculations by authors.

Figure 9: Period total female first-marriage rate, Russia, 1970–2001



Source: Generations & Gender Survey Russia (Wave 1), Calculations by authors:

Table 5: Period TTFMR & period MAFFM: mean differences, standard deviation of difference and zero-order correlation between GGS estimates and population statistics

Country	Periods	Period TFFMR			Period MAFFM		
		Correlation	Mean of differ.	S.D. of differ.	Correlation	Mean diff.	S.D. differ.
Australia	1977–2005	–	–	–	0.922**	0.739	0.761
Austria[1]	1988–2003	0.672**	0.434	0.057	–	–	
Belgium	1978–2002[2]	0.643**	0.039	0.086	0.841**	0.215	0.933
Bulgaria	1977–2001	0.914**	-0.028	0.080	0.832**	0.037	0.523
Estonia	1974–1998	0.924**	0.016	0.093	0.803**	0.035	0.592
France	1976–2002	0.743**	0.017	0.067	0.902**	-0.073	0.872
Georgia	1990–2001[3]	0.832**	0.090	0.119	0.696**	-1.591	0.676
Germany	1975–2001[4]	0.133	0.086	0.134	0.857**	0.028	0.926
Hungary	1976–1998	0.946**	-0.037	0.054	0.886**	-0.185	0.414
Italy	1989–2000	0.317	0.080	0.069	0.742**	0.213	0.450
Netherlands (w0)	1973–2000	0.124	-0.024	0.151	0.897**	1.410	0.968
Netherlands (w1)	1973–2000	0.104	-0.048	0.149	0.891**	1.415	1.001
Norway	1977–2003	0.596**	0.042	0.067	0.890**	0.253	0.941
Romania	1976–2002	0.621**	0.027	0.124	0.718**	-0.315	0.457
Russia	1974–1996	0.747**	-0.016	0.090	0.812**	-0.146	0.504

Note: Significance levels: not significant (–), p < .05 (*), p < .01 (**).
[1] for Austria rates between the ages 15 and 24 based on the GGS are compared to rates from vital statistics for the same age range.
[2] due to availability of external data the analysis period of the Period MAFFM deviates (1978–2004).
[3] due to availability of external data the analysis period of the Period MAFFM deviates (1990–2003).
[4] due to availability of external data the analysis period of the Period MAFFM deviates (1975–2002).
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 6: Period total female first-marriage rate (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year intervals

	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999
Australia	–	–	–	–	–
Austria[1]	–	–	–	0.0674	-0.0101
Belgium	–	0.0258	0.0108	0.0097	0.1094
Bulgaria	–	-0.0281	-0.0519	0.0376	0.0281
Estonia	0.0094	0.0173	0.0166	-0.0141	–
France	–	0.0608	0.0243	-0.0061	0.0636
Georgia	–	–	–	0.1109	0.0604
Germany	0.0115	0.0616	0.0480	0.0799	0.1225
Hungary	–	-0.0291	-0.0414	-0.0303	–
Italy	–	–	–	0.0725	0.1016
Netherlands (w0)	-0.1385	-0.0638	-0.0865	0.0031	0.1806
Netherlands (w1)	-0.1630	-0.0843	-0.1023	-0.0116	0.1392
Norway	–	0.0310	0.0708	0.0370	0.0411
Romania	–	-0.0225	0.1047	0.0480	0.0775
Russia	-0.0184	-0.0199	0.0000	-0.0497	–

Note: [1] for Austria rates between the ages 15 and 24 based on the GGS are compared to rates from vital statistics for the same age range.
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 7: Period MAFFM (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year intervals.

	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004
Australia	–	1.1019	0.6069	0.6981	0.3458	0.3745
Austria	–	–	–	–	–	–
Belgium	–	0.2028	0.1659	0.5753	0.4546	-0.6016
Bulgaria	–	0.0580	0.1850	0.0676	-0.0145	–
Estonia	-0.0731	0.1498	-0.3210	0.2147	–	–
France	–	0.2421	-0.0449	-0.1674	-0.0039	–
Georgia	–	–	–	-1.1827	-1.7041	–
Germany	0.7909	0.4403	0.3959	-0.7325	-0.2378	–
Hungary	–	-0.3008	0.0309	-0.4795	–	–
Italy	–	–	–	0.2997	0.2317	–
Netherlands (w0)	0.8585	1.7272	1.7396	1.2330	1.4224	–
Netherlands (w1)	0.8995	1.6888	1.7553	1.2378	1.4522	–
Norway	–	0.1932	0.7968	0.1261	0.0788	–
Romania	–	0.0821	-0.2834	-0.5071	-0.5336	–
Russia	0.1960	-0.0491	-0.4125	-0.4909	–	–

Source: Generations & Gender Survey (Wave 1), Calculations by authors

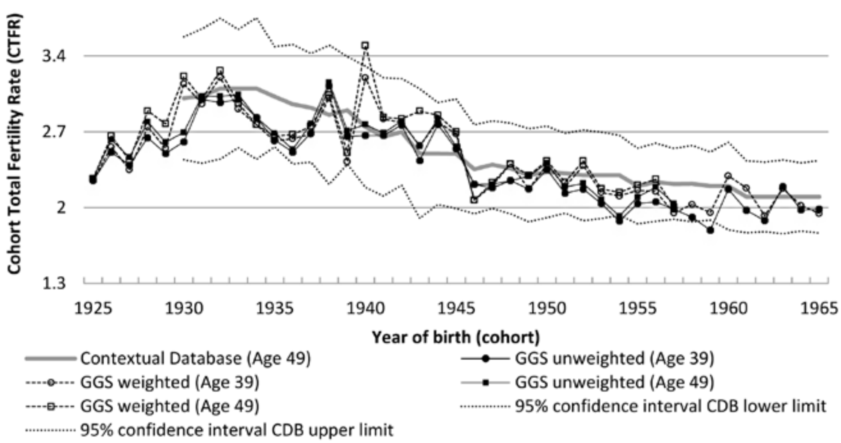
4.3 Cohort total fertility rate and mean age at childbearing

The cohort TFR estimated from the GGS was compared to population statistics for all 14 countries included in the analysis. Again, the temporal scope of the analysis is limited for Austria, given the limited age range in the cross-sectional sample. The results for the cohort TFR generally correspond with the conclusions drawn from the cohort nuptiality data. Consistent with the results found for the cohort TFFMR, the GGS underestimates the cohort TFR (Table 9) for women born in the 1930s with deviations ranging up to -0.43 children per woman in Bulgaria, -0.46 children per woman in Belgium, and -0.59 children per woman in Romania (Figure 12). Also, in other countries the GGS substantially underestimates the cohort TFR for cohorts born up to the 1940s (Russia, Figure 13) or even the 1950s (Germany and Hungary). In Bulgaria, Georgia, and Italy deviations are generally below -0.40 children per woman. It should be noted, however, that the underestimation of cohort fertility rates does not occur in all countries. In France the GGS estimates seem to reflect fertility patterns accurately for cohorts during the 1930s, while underestimations are found for cohorts born at the end of the 1940s and the beginning of the 1950s. In Australia (Figure 10), Estonia, the Netherlands (Figure 11), and Norway, the cohort TFRs estimated from the GGS are found to be fairly accurate across the whole range of cohorts considered. Table 8 shows the correlations between GGS-based cohort TFRs and population statistics. In most countries the GGS series reflect the trends emerging from population statistics.

For Belgium, Bulgaria, Estonia, Russia, and Romania similar conclusions can only be drawn for later cohorts (e.g., after 1945).

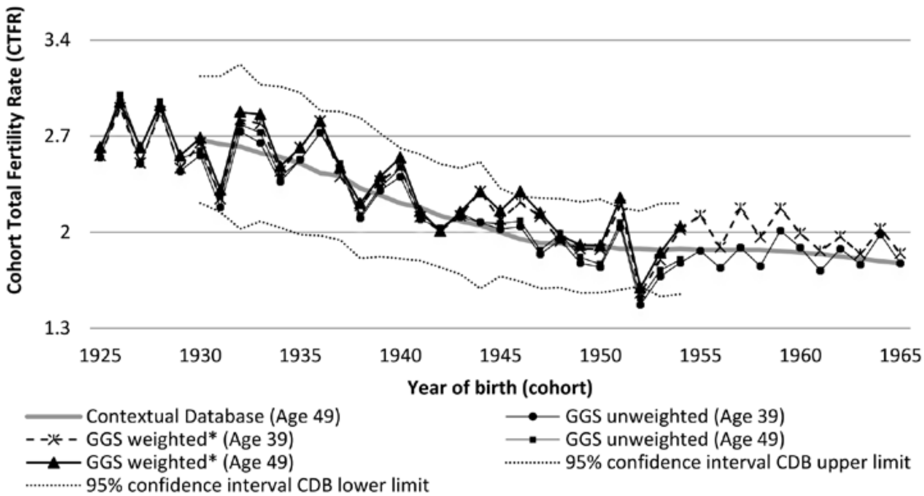
Population statistics on the cohort MAC are available for Australia, Belgium, Bulgaria, Estonia, France, Germany, Hungary, Italy, the Netherlands, Norway, Romania, and Russia. No results can be presented for Austria and Georgia. For Bulgaria, Estonia, France, Hungary, Norway, Romania, and Russia, the GGS estimates of the cohort MAC do not deviate substantially from population statistics (Table 10). For Belgium and Germany the GGS estimates overestimate the cohort MAC up to 1 year for the birth cohorts between 1945–1950 and 1945–1955. In Australia and the Netherlands the GGS overestimate mean ages at childbearing for all cohorts considered. In Australia the deviation between both series ranges up to 3 years in specific years, whereas deviations are generally smaller in the Netherlands, with deviations generally being limited to 1 year. In Italy the GGS substantially underestimates the cohort MAC for cohorts born in the 1940s. In sum, the GGS estimates provide a better approximation of long-term trends in the cohort MAC than is the case for the cohort TFR. As a result, for most countries significant positive correlations are found between GGS estimates of the cohort MAC and population statistics (Table 8). This is not the case, however, for the entire range of cohorts in Estonia, Italy, and Russia. In these countries, positive and significant correlations only emerge for cohorts born after 1945, indicating that the GGS more accurately reflect trends in the cohort MAC for the more recent birth cohorts.

Figure 10: Cohort total fertility rate, Australia, 1925–1965



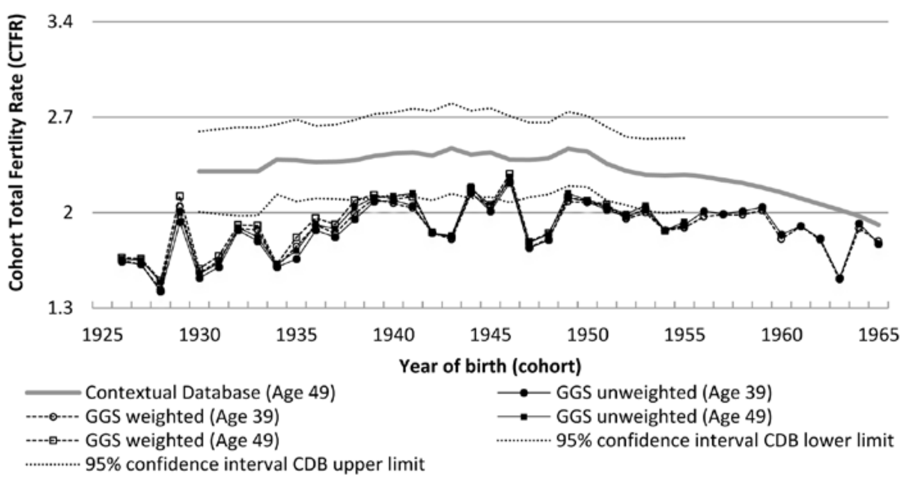
Source: Generations & Gender Survey Australia (Wave 1), Calculations by authors.

Figure 11: Cohort total fertility rate, Netherlands, 1925–1965



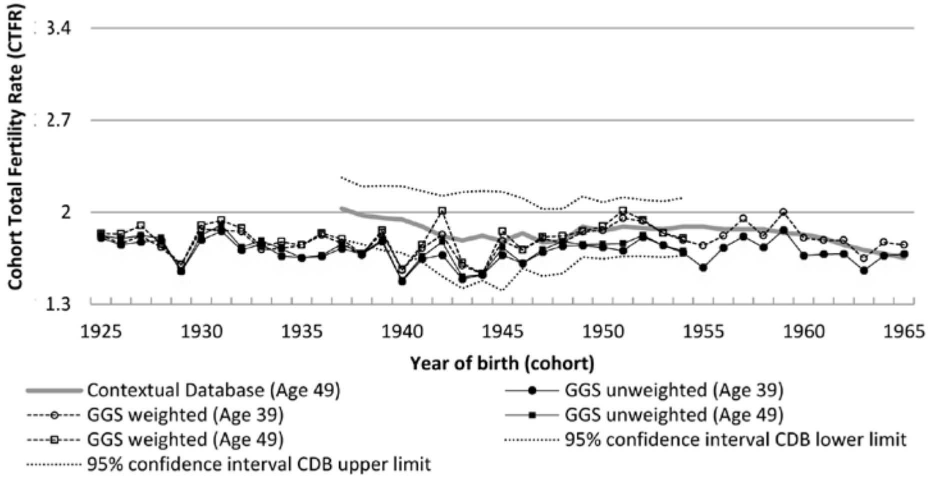
Source: Generations & Gender Survey Netherlands (Wave 1), Calculations by authors.

Figure 12: Cohort total fertility rate, Romania, 1926–1965



Source: Generations & Gender Survey Romania (Wave 1), Calculations by authors.

Figure 13: Cohort total fertility rate, Russia, 1925–1965



Source: Generations & Gender Survey Russia (Wave 1), Calculations by authors.

Table 8: Cohort TFR & Cohort MAC: mean difference, and standard deviation of difference and zero-order correlation between GGS estimates and population statistics

Country	Cohorts	Cohort TFR			Cohort MAC		
		Correlation	Mean Diff.	St. Dev. Diff.	Correlation	Mean Diff.	St. Dev. Diff.
Australia	1930–1957	0.759**	0.004	0.258	0.393**	1.277	0.914
Austria	–	–	–	–	–	–	–
Belgium	1930–1960	0.106	-0.227	0.286	0.689**	0.497	0.596
Bulgaria	1930–1954	-0.307	-0.340	0.136	0.667**	0.458	0.479
Estonia	1944–1955	0.227	0.038	0.117	-0.072	0.489	0.491
France	1930–1955	0.808**	-0.106	0.217	0.667**	0.008	0.473
Georgia	1935–1953	0.480*	-0.250	0.124	–	–	–
Germany	1930–1955	0.600**	-0.239	0.182	0.598**	0.484	0.714
Hungary	1930–1955	-0.100	-0.157	0.110	0.419*	0.029	0.420
Italy	1939–1953	0.753**	-0.196	0.107	0.290	-0.282	0.555
Netherlands (w0)	1930–1954	0.816**	0.113	0.189	0.783**	0.731	0.579
Netherlands (w1)	1930–1954	0.819**	0.090	0.191	0.814**	0.668	0.534
Norway	1930–1958	0.792**	-0.088	0.132	0.832**	0.251	0.504
Romania[1]	1930–1955[1]	0.385	-0.406	0.159	0.812**	0.199	0.446
Russia	1937–1954	-0.046	-0.073	0.158	0.111	0.200	0.796

Note: Significance levels: not significant (–), p < .05 (*), p < .01 (**).
[1] due to limited availability of data from external sources, the data quality assesment of the cohort MAC refers to the period 1934−1958.
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 9: Cohort TFR (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year birth cohorts

	1930–1934	1935–1939	1940–1944	1945–1949	1950–1954	1955–1959
Australia	-0.0259	-0.2051	0.3599	-0.0422	-0.0270	–
Austria	–	–	–	–	–	–
Belgium	-0.4259	-0.4618	-0.2543	-0.2036	-0.0606	0.0111
Bulgaria	-0.4324	-0.3907	-0.3471	-0.2767	-0.2509	–
Estonia	–	–	–	0.0330	0.0768	–
France	0.1468	-0.0238	-0.1090	-0.0632	-0.0891	–
Georgia	–	-0.3127	-0.2344	-0.2755	–	–
Germany	-0.4446	-0.2209	-0.1724	-0.2721	-0.1428	–
Hungary	-0.2553	-0.2485	-0.0849	-0.1290	-0.0797	–
Italy	–	–	-0.1728	-0.2200	-0.2295	–
Netherlands (w0)	0.0521	0.1281	0.1012	0.2060	0.0781	–
Netherlands (w1)	0.0290	0.1165	0.1020	0.1588	0.0459	–
Norway	-0.0864	-0.1651	-0.1890	-0.0592	-0.0172	–
Romania	-0.5984	-0.3821	-0.4324	-0.3956	-0.3212	–
Russia	–	–	-0.1582	-0.0012	0.0196	–

Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 10: Cohort MAC (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year birth-cohorts

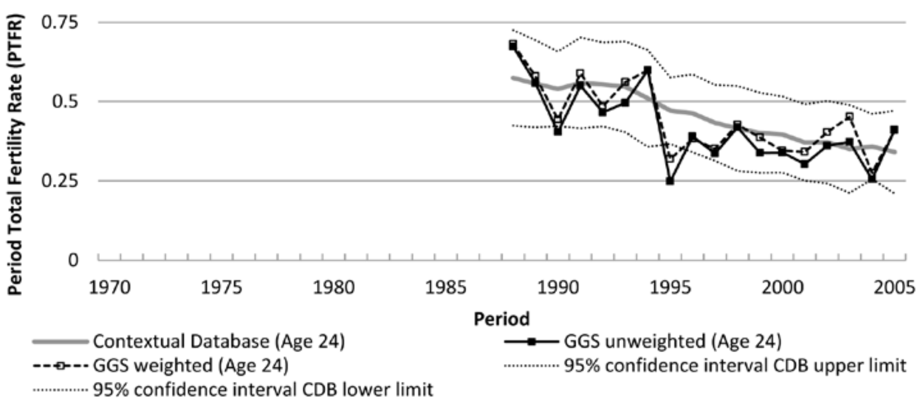
	1930–1934	1935–1939	1940–1944	1945–1949	1950–1954	1955–1959
Australia	1.1359	1.2771	1.0566	0.8828	–	–
Austria	–	–	–	–	–	–
Belgium	0.4856	0.3273	0.4750	0.8843	0.3859	0.3865
Bulgaria	0.6129	0.2964	0.4847	0.6221	0.2717	–
Estonia	–	–	–	0.3680	0.6088	–
France	-0.1607	-0.0227	0.1258	0.0389	0.0274	–
Georgia	–	–	–	–	–	–
Germany	0.1056	0.2455	0.2518	0.9129	0.7283	–
Hungary	0.0037	-0.3247	0.4168	0.0989	-0.0964	–
Italy	–	–	-0.7394	-0.2280	0.4742	–
Netherlands (w0)	0.7505	0.6043	0.5188	1.1507	0.6314	–
Netherlands (w1)	0.7016	0.5799	0.5217	0.9406	0.5943	–
Norway	0.2858	0.1020	0.0259	0.5467	0.0833	0.4468
Romania	–	0.2573	-0.0588	0.2536	0.5243	–
Russia	–	–	0.4205	0.0804	0.3021	–

Source: Generations & Gender Survey (Wave 1), Calculations by authors.

4.4 Period total fertility rate and mean age at childbearing

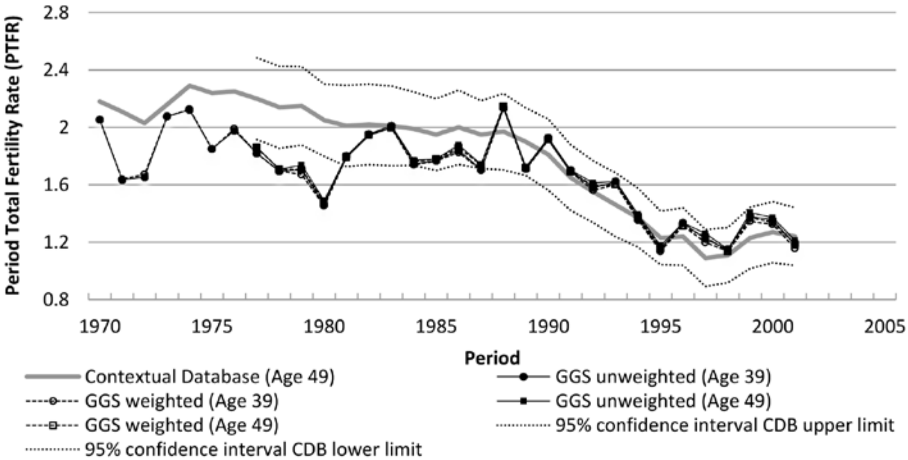
For all countries included in the analysis, population statistics provide time series of the period TFR, making it possible to assess the quality of GGS estimates. The results for the GGS-based period TFR show that, similar to the results for nuptiality, period indicators for fertility show smaller deviations from population statistics than cohort fertility indicators. Consistent with findings regarding cohort indicators, which particularly show underestimations of fertility in older cohorts, underestimations of the period TFR typically occur in the earliest periods considered. In the late 1970s and the 1980s, Bulgaria, Georgia, Estonia, the Netherlands, and Romania show clear underestimations of the period TFR (Table 12). Results for Bulgaria in Figure 15 and Georgia in Figure 17 are examples in this respect. In more recent periods the GGS tends to overestimate the period TFR in some countries, similar to overestimations found for period TFFMR. This overestimation of period fertility indicators is particularly noticeable during the 1990s in Georgia and Germany (up to +0.5 children per woman) and at the turn of the millennium in Belgium (up to +0.2 children per woman in selected years). Countries that do not show substantial periods of under or overestimation of fertility are Australia, Belgium, France (Figure 16), Hungary, Italy, Norway, and Russia. Also, the period TFR drawn from the Austrian GGS data (Figure 14) shows no substantial deviations, notwithstanding the limited time scope of the analysis.

Figure 14: Period total fertility rate, Austria, 1988–2005



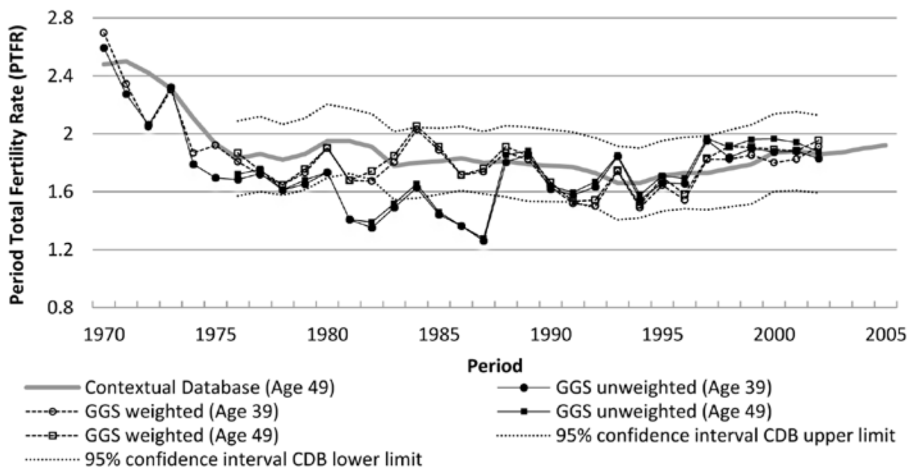
Source: Generations & Gender Survey Austria (Wave 1), Calculations by authors.

Figure 15: Period total fertility rate, Bulgaria, 1970–2001



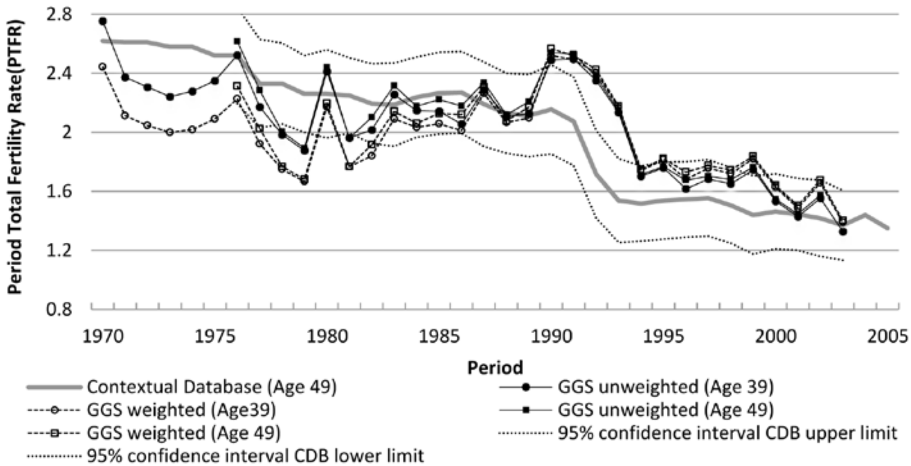
Source: Generations & Gender Survey Bulgaria (Wave 1), Calculations by authors.

Figure 16: Period total fertility rate, France, 1970–2005



Source: Generations & Gender Survey France (Wave 1), Calculations by authors.

Figure 17: Period total fertility rate, Georgia, 1970–2005



Source: Generations & Gender Survey Georgia (Wave 1), Calculations by authors.

Table 11: Period TFR & Period MAC: mean difference, standard deviation of difference and zero-order correlation between GGS estimates and population statistics

Country	Periods	Period TFR			Period MAC		
		Correlation	Mean of differ.	S.D. of differ.	Correlation	Mean of differ.	S.D. of differ.
Australia	1970–2005	0.808**	0.053	0.205	0.668**	0.504	1.293
Austria[1]	1988–2005	0.748**	-0.061	0.076	0.453–	0.469	0.312
Belgium	1978–2005	0.699**	0.048	0.128	0.923**	0.312	0.375
Bulgaria	1977–2001	0.851**	-0.091	0.203	0.451*	0.370	0.459
Estonia	1974–1998	0.940**	-0.198	0.124	0.517**	0.324	0.418
France	1976–2002	0.365–	-0.025	0.134	0.925**	0.036	0.457
Georgia	1976–2003	0.555**	0.051	0.330	0.519**	-0.188	0.524
Germany	1975–2002	0.040–	0.130	0.191	0.841**	0.032	0.512
Hungary	1976–1998	0.897**	-0.007	0.102	0.934**	0.016	0.265
Italy	1989–2000	0.087–	-0.013	0.085	0.847**	0.338	0.360
Netherlands (w0)	1973–2000	0.229–	0.188	0.197	0.923**	0.605	0.439
Netherlands (w1)	1973–2000	0.244–	0.132	0.183	0.937**	0.590	0.419
Norway	1977–2003	0.622**	0.018	0.135	0.951**	0.080	0.309
Romania	1976–2002	0.891**	-0.118	0.271	0.621**	0.125	0.405
Russia	1974–2001	0.946**	0.045	0.130	0.863**	-0.149	0.425

Note: Significance levels: not significant (–), p < .05 (*), p < .01 (**).
[1] for Austria rates between the ages 15 and 25 based on the GGS are compared to rates from vital statistics for the same age range.

Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 12: Period TFR (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year intervals

	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004
Australia	0.0234	0.0446	0.2019	0.0954	0.0621	0.0030
Austria[1]	–	–	–	-0.0065	-0.0634	-0.0063
Belgium	–	-0.0437	0.0211	0.0830	-0.0430	0.1324
Bulgaria	–	-0.2242	-0.1100	0.0630	0.0568	–
Estonia	-0.2756	-0.1240	-0.1942	-0.1671	–	–
France	–	0.0327	-0.0213	0.1212	-0.0425	–
Georgia	–	-0.2095	-0.0328	0.4876	0.2655	–
Germany	-0.1458	0.0443	0.2553	0.1647	0.1833	–
Hungary	–	-0.0311	0.0037	0.0467	–	–
Italy	–	–	–	0.0384	-0.0643	–
Netherlands (w0)	-0.1105	-0.2246	-0.2189	-0.2824	-0.1618	–
Netherlands (w1)	-0.0797	-0.1540	-0.1615	-0.2385	-0.0871	–
Norway	–	-0.0126	0.0879	0.1094	-0.0036	–
Romania	–	-0.3406	-0.1310	0.0293	0.1350	–
Russia	-0.0798	0.1775	0.0542	0.0536	-0.0235	–

Note: [1] for Austria rates between the ages 15 and 24 based on the GGS are compared to rates from vital statistics for the same age range.
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

Table 13: Period MAC (calculated up to age 49): mean difference between GGS estimates and population statistics in subsequent 5-year intervals

	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004
Australia	1.6137	1.5046	0.2554	-0.0266	-0.4580	-0.6986
Austria[1]	–	–	–	0.7176	0.9278	0.6982
Belgium	–	0.4980	0.2353	0.3709	0.3871	0.0003
Bulgaria	–	0.5851	0.5485	0.3736	0.0014	–
Estonia	0.3782	-0.0908	0.3976	0.4916	–	–
France	–	-0.0153	0.1810	-0.0308	-0.2120	–
Georgia	–	0.3315	-0.1792	-0.5461	-0.3904	–
Germany	0.4178	0.1842	-0.1130	-0.0129	-0.0873	–
Hungary	–	-0.2258	0.0893	0.0114	–	–
Italy	–	–	–	-0.2588	-0.1964	–
Netherlands (w0)	-0.5917	-0.6555	-0.4624	-0.8315	-0.3998	–
Netherlands (w1)	-0.5204	-0.5844	-0.4355	-0.8917	-0.4479	–
Norway	–	-0.0412	0.1800	0.3316	-0.0901	–
Romania	–	0.1750	0.2740	0.0949	0.0216	–
Russia	0.1920	-0.0699	-0.2789	-0.2611	-0.3839	–

Note: [1] for Austria rates between the ages 15 and 24 based on the GGS are compared to rates from vital statistics for the same age range.
Source: Generations & Gender Survey (Wave 1), Calculations by authors.

The GGS-based period MAC seems to accurately reflect actual levels and trends registered in vital statistics (deviations <0,50 years) in the case of Belgium, Estonia, France, Germany, Hungary, Italy, Norway, and Romania (Table 13). Period MACs are overestimated in Australia between 1975 and 1985, with the strongest deviations up to 1.5 years occurring in the 1970s. In Austria and the Netherlands the GGS overestimates mean ages up to 1 year in almost all periods considered. Besides overestimations in period MAC, we also find substantive underestimations of the mean age at childbearing. In addition to overestimations in 1970–1985, the Australian GGS shows underestimations between 1995 and 2005, with a deviation ranging up to 2 years in 2005. In Georgia and Russia respectively, underestimations ranging up to 1 year after 1990 and up to 0.5 year after 1995 are found. Nevertheless, the significant positive correlations in Table 11 between GGS-based period TFRs and population statistics indicate that – similar to conclusions drawn for the period MAFFM – the period mean age at childbearing estimated from the GGS closely approximates trends emerging from population statistics.

5. Discussion and conclusion

The aim of this paper is to document the joint effect of various potential sources of survey error on the quality of retrospective data on first marriage and fertility included in the first wave of the GGS. To this end, period and cohort indicators estimated retrospectively from the GGS were compared to population statistics.

5.1 Main findings

The results for cohort and period indicators show considerable variation, both between countries and over subsequent periods/cohorts within countries. The detailed country-specific results included in the appendix may therefore help researchers to decide on the appropriate observation period (calendar time/birth cohorts) to be included in their analysis for each country. By and large, our results suggest that GGS fertility and nuptiality data appear to be fairly accurate from the 1970s onwards – consistent with the design of the GGP – for the period measures in Australia, Austria, Belgium, Estonia, Georgia, Germany, Italy, Norway, and Romania. Other countries (Bulgaria, France, Hungary, the Netherlands, and Russia) show close approximations of period measures beginning from the late 1970s or even the early 1980s. For the latter group it may be advisable to limit the scope of the analysis to years after 1975, or even 1980. Concerning cohort indicators, Australia, Austria, and Estonia show accurate approximations of vital statistics when these are available. For other countries, GGS-based cohort indicators provide an accurate account of demographic trends for cohorts born after 1945 (Belgium and Norway), 1950 (Hungarian fertility), or even 1955–1960

(Germany). For some countries (Bulgarian, French, and Dutch nuptiality, and Georgian, Italian, and Romanian fertility) cohort indicators estimated retrospectively from the GGS show substantial underestimations for all cohorts, suggesting that in these cases caution is required when using the GGS for analyses of cohort first-marriage and fertility patterns.

5.2 Potential sources of survey error in GGS Wave 1

Deviations between GGS estimates and population statistics may result from various sources of survey error. First, selectivity in surveys depends on the representativeness of the sampling frame, the actual sampling procedure, and the amount of unit nonresponse. Given that most countries implemented probability sampling in line with the survey design guidelines, we expect this to be a limited source of survey error in the case of the GGS, except in a limited number of countries where random route procedures were used. However, retrospective collection of partnership and fertility histories relies on survivors in the birth cohorts considered. To the extent that mortality, emigration, and institutionalization have disproportionately affected subgroups with specific nuptiality and fertility patterns, this may also affect GGS estimates. Underestimation of fertility in older cohorts may, for instance, be related to mortality patterns of multiparous women. A more detailed understanding of the determinants of these deviations may be gained from assessing the quality of order-specific fertility data. A comparison between GGS-based order-specific rates and population statistics would make it possible to verify whether specific groups of women are under or overrepresented (e.g., childless women). The GGS sample sizes nevertheless preclude the estimation of stable order-specific time series, and fertility rates for higher-order births in particular are expected to show erratic patterns due to small numbers. Apart from sampling procedures and coverage error, unit nonresponse determines whether sampled individuals are effectively included in the survey. Hence, unit nonresponse – which has been found to vary in terms of various socio-demographic characteristics – may be an important factor driving the found under and overestimations of demographic indicators. Previous research has provided substantial evidence of the relative underrepresentation in surveys of childless women or women who have older children (Kreyenfeld et al. 2010; 2011). Because mothers (and particularly mothers of young children) are more likely to stay at home, they are more easily contacted by interviewers (Groves et al. 2009). In addition, people with young children are frequently more easily motivated to participate in surveys, particularly family surveys. Where available, a detailed analysis of unit nonresponse and final disposition codes in terms of relevant characteristics in the sampling frame may make it possible to improve the quality of GGS estimates through post-stratification in terms of these characteristics. Fokkema and colleagues (see this special volume) perform an evaluation of the representativeness of the GGS data with respect to age, gender, region, household size, marital status, and educational attainment. Apart from respondent selection issues, survey error in partnership and fertility histories may result from item nonresponse, questionnaire design, and reporting error. To the degree that item nonresponse for the questions on partnership and fertility histories is related to nuptiality and childbearing behaviour, this may affect GGS estimates. As a result, analyses and identification of item nonresponse in partnership and fertility histories may provide a feasible strategy to further improve the quality of GGS estimates. Research suggests that the design of the GGS questionnaire was very demanding for both respondents and interviewers due to its complex structure and routings (Beaujouan 2013). Children who have left the parental home are assumed to be underreported as a consequence of the complex nature of surveying non-residential children (Ruckdeschel et al. in this special volume). Similarly, fertility histories are questioned in two blocks, separated by questions on social and educational background, dwelling unit, and the organisation of childcare in the household (Kreyenfeld et al. 2010; Sauer, Ruckdeschel, and Naderi 2012). The impact of questionnaire design is expected to be larger in the case of pencil-and-paper personal interviewing (PAPI), phone interviews, and self-administered paper questionnaires (SAPQ), which have been used in some countries (see Fokkema et al. in this volume). We expect that the complexity of the questionnaire has predominantly affected the quality of partnership and fertility histories in older cohorts (Kreyenfeld et al. 2010; Ní Bhrolcháin, Beaujouan, and Murphy 2011), which is consistent with the underestimation of first marriage intensities and fertility found for older cohorts in this study. Given that reporting histories may be demanding on respondents, some authors have recommended the use of tools such as event-history calendars, which have been found easier for the respondent compared to lists of questions (Belli 1998; Sauer, Ruckdeschel, and Naderi 2012). Finally, retrospective data quality is affected by recall errors concerning dates of births and first marriages (Peters 1988; Belli 1998; Blossfeld and Rohwer 2002; Murphy 2009). Although a marriage or birth entails high emotional involvement, such problems have been detected concerning landmark type of events (Murphy 2009; Kreyenfeld et al. 2010; Ní Bhrolcháin, Beaujouan, and Murphy 2011). Recall problems such as memory loss and telescoping may have contributed to the underestimation of fertility and first-marriage intensities in older cohorts and earlier periods. With respect to partnership histories, current marriages may have been reported more adequately than first marriages. Similarly, children born within a current marriage or partnership may have been reported more accurately than births stemming from former partnerships (Juby and Le Bourdais 1999). In this respect the lack of control questions on current marital status and parity is an important shortcoming of the GGS

Wave 1 questionnaire. The lack of such controls makes it difficult to detect and remove inconsistent maternal and marital histories from the GGS data. Further control questions (e.g., "Do you really have no children?") have been found to successfully limit inconsistent reporting of histories for GGS data in Austria, France, and the Netherlands (Sauer, Ruckdeschel, and Naderi 2012).

5.3 Challenges for future research

In summary, we conclude that longitudinal micro-data in the GGS provide a useful research infrastructure to gauge the effects of individual and contextual factors on (changing) demographic behaviour, provided that careful consideration is given to the periods or cohorts included in the analysis. Documenting bias draws researchers' attention to period/cohorts and events for which the quality of demographic indicators is poorer, but at the same time it only constitutes a first step in improving the quality of demographic indicators from the GGS. Cases where the GGS overestimates population statistics of first marriage and fertility – predominantly in younger cohorts – are likely to result from topic salience among respondents who have already married or started family formation, while unmarried and/or childless respondents may have been harder to reach or difficult to persuade to participate in a survey such as the GGS. Underestimation of population statistics – the more common problem and frequently found in the older birth cohorts – may result from both selective survival and/or unit response among unmarried and childless women, or from higher item nonresponse for questions concerning partnership histories and fertility. The latter problem in particular may be addressed by detailed analysis of nonresponse patterns, as identification and omission of nonresponse may considerably improve the quality of demographic indicators for older birth cohorts and earlier periods. At the same time, we stress that this paper has only considered first marriage and fertility: it is unclear whether similar conclusions can be drawn for union formation (e.g., unmarried cohabitation) and non-marital fertility, since literature shows additional sources of bias for these processes such as particular forms of social desirability related to the social acceptance and the changing prevalence of cohabitation and non-marital childbirth. This may introduce additional bias by birth cohort for these indicators, and may warrant additional analyses of data quality for indicators concerning these behaviours (Hayford and Morgan 2008).

6. Acknowledgements

This research was supported by grants from the Flemish Research Council (G012011N) and the Research Council of the University of Antwerp (BOF-NOI2010).

References

Abraham, K.G., Maitland, A., and Bianchi, S.M. (2006). Nonresponse in the American time use survey. Who is missing from the data and how much does it matter? *Public Opinion Quarterly* 70(5): 676–703. doi:10.1093/poq/nfl037.

Aleshina, N. and Redmond, G. (2005). How high is infant mortality in central and eastern Europe and the commonwealth of independent states? *Population Studies* 59(1): 39–54. doi:10.1080/0032472052000332692.

American Association for Public Opinion Research (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys [unpublished manuscript]. Ann Arbor, MI: AAPOR.

Auriat, N. (1991). Who forgets? An analysis of memory effects in a retrospective survey on migration history. *European Journal of Population* 7(4): 311–342. doi:10.1007/BF01796872.

Ayhan, H.Ö. and Işiksal, S. (2004). Memory Recall Errors in Retrospective Surveys: A Reverse Record Check Study. *Quality & Quantity* 38(5): 475–493. doi:10.1007/s11135-005-2643-7.

Beaujouan, É. (2013). Counting how many children people want: the influence of question filters and pre-codes. *Demográphia English Edition* 56(3): 35–61.

Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., and Peterson, C. (2001). The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country. *Journal of Human Resources* 36(3): 593–625. doi:10.2307/3069631.

Belli, R.F. (1998). The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory* 6(4): 383–406. doi:10.1080/741942610.

Bergstrand, R., Vedin, A., Wilhelmsson, C., and Wilhelmsen, L. (1983). Bias due to non-participation and heterogenous sub-groups in population surveys. *Journal of Chronical Diseases* 36(10): 725–728. doi:10.1016/0021-9681(83)90166-2.

Blossfeld, H.P. and Rohwer, G. (2002). *Techniques of Event-history Modeling. New Approaches to Causal Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Blossfeld, H.P., Golsch, K., and Rohwer, G. (2007). *Event History Analysis with Stata*. Mahwah, NJ: Lawrence Erlbaum Associates.

Buber, I. (2010). Parity-Specific Weights for the Austrian Generations and Gender Survey. Vienna: Vienna Institute of Demography (VID Working Papers 4/2010).

Buber I. (2013). Attrition in the Austrian generations and gender survey. Vienna: Vienna Institute of Demography (VID Working Papers 10/2013).

Burkimsher, M. (2009). Assessment of fertility indicators derived from GGP samples. Bulgaria, Hungary and Georgia [unpublished manuscript].

Calot, G. (1984). La mesure des taux en démographie: Taux par âge en années révolues et taux par âge atteint dans l'année. *Population (French Edition)* 39(1): 107–146. doi:10.2307/1532114.

Campanelli, P., Sturgis, P., and Purdon, S. (1997). Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates. London: National Centre for Social Research (NCSR paper).

Cherlin, A.J., Griffith, J., and McCarthy, J. (1983). A note on maritally-disrupted men's reports of child support in June 1980 current population survey. *Demography* 20(3): 385–389. doi:10.2307/2061249.

Childs, J.H. and Landreth, A. (2006). Analyzing Interviewer/Respondent Interactions while Using a Mobile Computer-Assisted Personal Interview Device. *Field Methods* 18: 335–351. doi:10.1177/1525822X05286018.

Council of Europe (2005). Recent Demographic Developments in Europe: 2004. Strasbourg: Council of Europe Publishing.

Couper, M.P., Singer, E., and Kulka, R.A. (1997). Participation in the Decennial Census: Politics, Privacy, Pressures. *American Politics Research* 26(1): 59–80. doi:10.1177/1532673X9802600104.

Davis, R.E., Couper, M.P., Janz, N.K., Caldwell, C.H., and Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research* 25(1): 14–26. doi:10.1093/her/cyp046.

De Wulf, K., Van Kenhove, P., and Wijnen, K. (2003). The influence of topic involvement on mail survey response behavior. Gent: Vlerick Leuven Gent Management School (Vlerick Leuven Gent Working Paper Series 2003/16)

Dex, S. and McCulloch, A. (1998). The Reliability of Retrospective Unemployment History Data. *Work, Employment & Society* 12(3): 497–509. doi:10.1177/0950017098123005.

Duan, N., Alegria, M., Canino, G., McGuire, T.G., and Takeuchi, D. (2007). Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats. *Health Services Research* 42(2): 890–907. doi:10.1111/j.1475-6773.2006.00618.x.

Festy, P. and Prioux, F. (2002). *An evaluation of the Fertility and Family Surveys project*. New York: United Nations Publications.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998). The Impact Of Attrition In The Panel Study Of Income Dynamics On Intergenerational Analysis. *Journal of Human Resources* 33(2): 300–344. doi:10.2307/146434.

Gadeyne, S. and Deboosere, P. (2000). Socio-economische factoren in de differentiële sterfte van vrouwen van 40–59 jaar in België. Analyse van de Nationale Databank Mortaliteit. Brussels: Vrije Universiteit Brussel (Working Papers Steunpunt Demografie 2000-8).

Generations and Gender Programme (2014). Data Documentation [electronic source]. GGP. [http://www.ggp-i.org/data/data-documentation.html]

Gibson, J. and Kim, B. (2010). Non-Classical Measurement Error in Long-Term Retrospective Recall Surveys*. *Oxford Bulletin of Economics and Statistics* 72(5): 687–695. doi:10.1111/j.1468-0084.2010.00599.x.

Goldberg, M., Chastang, J.F., Leclerc, A., Zins, M., Bonenfant, S., Bugel, I., Kaniewski, N., Schmaus, A., Niedhammer, I., Piciotti, M., Chevalier, A., Godard, C., and Imbernon, E. (2001). Socioeconomic, demographic, Occupational, and Health Factors Associated with Participation in a Long-term Epidemiologic Survey: A Prospective Study of the French GAZEL Cohort and Its Target Population. *American Journal of Epidemiology* 154(4): 373–384. doi:10.1093/aje/154.4.373.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley-Interscience. doi:10.1002/9781118490082.

Groves, R.M., Fowler, F.J., Couper, J.M., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2009). *Survey Methodology*. Hoboken: Wiley.

Groves, R.M. and Lyber, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* 74(5): 849–879. doi:10.1093/poq/nfq065.

Hayford, S.R. and Morgan, S.P. (2008). The Quality Of Retrospective Data On Cohabitation. *Demography* 45(1): 129–141. doi:10.1353/dem.2008.0005.

Hertrich, V. (1998). Are men's and women's answers to be equally trusted? A dual collection of birth and marriage histories in a population in Mali. *Population: An English Selection* 10(2): 303–318.

Hill, P.B. (2005). Methodenprobleme in der Ehe- und Familiensoziologie. In: Busch, F.W. and Nave-Herz, R. (eds.). *Familie und Gesellschaft. Beiträge zur Familienforschung*. Oldenburg: BIS-Verlag: 165–186

Jacobs, S.C. (2002). Reliability and recall of unemployment events using retrospective data. *Work, Employment & Society* 16(3): 537–548. doi:10.1177/095001702762217489.

Joyner, K., Peters, H.E., Hynes, K., Sikora, A., Taber, J.R., and Rendall, M.S. (2012). The Quality of Male Fertility Data in Major U.S. Surveys. *Demography* 49(1): 101–124. doi:10.1007/s13524-011-0073-9.

Juby, H. and Le Bourdais, C. (1999). Where have All the Children Gone? Comparing Mothers' and Fathers' Declarations in Retrospective Surveys. *Canadian Studies in Population* 26(1): 1–20.

Kessler, R.C., Little, R.J., and Groves, R.M. (1995). Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiologic Reviews* 17(1): 192–204.

Klein, T. and Fischer-Kerli, D. (2000). Die Zuverlässigkeit retrospektiv erhobener Lebensverlaufsdaten: Analysen zur Partnerschaftsbiografie des Familiensurvey. *Zeitschrift für Soziologie* 29(4): 294–312.

Klijzing, E. and Cairns, H. (2000). On the Quality of FFS Event History Data: Of Critical Importance In Life Course Research. Bielefeld: University of Bielefeld. (Globalife Working Paper No. 13).

Kreyenfeld, M., Hornung, A., Kubisch, K., and Jaschinski, I. (2010). Fertility and Union Histories from German GGS Data: Some Critical Reflections. Rostock: Max Planck Institute for Demographic Research (MPIDR Working Paper 2010-023).

Kreyenfeld, M., Zeman, K., Burkimsher, M., and Jaschinski, I. (2011). Fertility data for German-speaking countries What is the potential? Where are the pitfalls?. Rostock: Max Planck Institute for Demographic Research (MPIDR Working Paper 2011-003).

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5(3): 213–236. doi:10.1002/acp.2350050305.

Kvåle, G., Heuch, I., and Nilssen, S. (1994). Parity in Relation to Mortality and Cancer Incidence: A Prospective Study of Norwegian Women. *International Journal of Epidemiology* 23(4): 691–699. doi:10.1093/ije/23.4.691.

Kveder, A. (2005). A Note on the Definitions of Final Disposition Codes. In: United Nations (eds.). *Generations and Gender Programme: Survey Instruments*. Geneva: United Nations: 117–118.

Lauwereys, G., Neels, K., and De Winter, T. (2011). GGS Wave 1 Belgium: Final Disposition Codes & Standardised Response Rates. GGP Belgium Paper Series (GGP Belgium Paper Series No. 3).

Lievesley, D. (1988). Unit Non-Response in Interview Surveys. London: Social and Community Planning Research (Working Paper).

Lucas, C.P., Fisher, P., Piacentini J., Zhang H., Jensen P.S., Shaffer D., Dulcan M., Schwab-Stone M., Regier D., and Canino G. (1999). Features of Interview Questions Associated with Attenuation of Symptom Reports. *Journal of Abnormal Child Psychology* 27(6): 429–437. doi:10.1023/A:1021975824957.

Lynn, P., Beerten, R., Laiho, J., and Martin, J. (2001). Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys. Essex: University of Essex, Institute for Social and Economic Research (ISER Working Papers 2001-23).

Mitchell, C. (2010). Are divorce studies trustworthy? The effects of survey nonresponse and response errors. *Journal of Marriage and Family* 72(4): 893–905. doi:10.1111/j.1741-3737.2010.00737.x.

Morgan, S.P., Botev, N., Chen, R., and Huang, J. (1999). White and non-white trends in first birth timing: Comparisons using vital registration and current population surveys. *Population Research and Policy Review* 18(4): 339–356. doi:10.1023/A:1006245612218.

Murphy, M. (2000). The Evolution of Cohabitation in Britain, 1960–1995. *Population Studies* 54(1): 43–56.

Murphy, M. (2009). Where have all the children gone? Women's reports of more childlessness at older ages than when they were younger in a large-scale continuous household survey in Britain. *Population Studies* 63(2): 115–133. doi:10.1080/00324720902917238.

Neels, K. (2006). *Reproductive strategies in Belgian fertility, 1960–1990*. Brussels/The Hague: NIDI-CBGS.

Neels, K. and Gadeyne, S. (2010). Het effect van opleidingsniveau op tempo en kwantum van rangspecifieke vruchtbaarheid in België: 1960–2000. In: Matthijs, K., Putte, B., Van de Kok, J., and Bras, H. (eds.). *Leven in de lage landen: historisch-demografisch onderzoek in Vlaanderen en Nederland*. Leuven/Den Haag: Acco: 211–238.

Ní Bhrolcháin, M., Beaujouan, É., and Murphy, M. (2011). Sources of error reported childlessness in a continuous British household survey. *Population Studies* 65(3): 305–318. doi:10.1080/00324728.2011.607901.

Nicoletti, C. and Peracchi, F. (2005). Survey response and survey characteristics: micro-level evidence from the European Community Household Panel. *Journal of the Royal Statistical Society: Series A* 168(4): 763–781. doi:10.1111/j.1467-985X.2005.00369.x.

O'Connell, M. (2007). The visible hand: Editing marital-history data from Census Bureau surveys. In: Hofferth, S.L. and Casper, L.M. (ed.). *Handbook of measurement issues in family research*. Mahwah, NJ: Lawrence Erlbaum Associates: 145–158.

Peters, H.E. (1988). Retrospective Versus Panel Data in Analyzing Lifecycle Events. *The Journal of Human Resources* 23(4): 488–513. doi:10.2307/145810

Piacentini, J., Jensen, P., and Schaffer, D. (1991). *Attenuation of child p.sychiatric diagnostic information: a detailed analysis*. Paper presented at the Scientific Proceedings of the Annual Meeting of the AACAP, VII. San Francisco, CA.

Pickery, J., Loosveldt, G., and Carton, A. (2001). The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel surveys: A Multilevel Approach. *Sociological Methods & Research* 29(4): 509–523. doi:10.1177/0049124101029004004.

Poulain, M., Riandey, B., and Firdion, J.M. (1992). Data from a life history survey and from Belgian Population Register: A comparison. *Population: An English Selection* 4: 77–96.

Régnier-Loilier, A. (2007). Conditions de passation et biais occasionnés par la présence d'un tiers sur les résponses obtenues à l'enquête Erfi. *Economie et statistique* 407(1): 27–49. doi:10.3406/estat.2007.7071.

Régnier-Loilier, A. (2011). Présentation, questionnaire et documentation de la troisième vague de l'Étude des relations familiales et intergénérationnelles (Erfi-GGS 2011). Paris: INED (Documents de travail 187).

Rendall, M.S., Clarke, L., Peters, H.E., Ranjit, N., and Verropoulou, G. (1999). Incomplete reporting of men's fertility in the United States and Britain: A research note. *Demography* 36(1): 135–144. doi:10.2307/2648139.

Sauer, L., Ruckdeschel, K., and Naderi, R. (2012). Reliability of retrospective event histories within the German Generations and Gender Survey. Bundesinstitut für Bevölkerungsforschung (BiB Working Paper 1/2012).

Shachar, R. and Eckstein, Z. (2007). Correcting for bias in retrospective data. *Journal of Applied Econometrics* 22(3): 657–675. doi:10.1002/jae.939.

Simard, M. and Franklin, S. (2005). Sample Design Guidelines. In: United Nations (eds.). *Generations and Gender Programme. Survey Instruments*. Geneva: United Nations: 5–14.

Smith, J.P. and Thomas, D. (2003). Remembrances of things past: test-retest reliability of retrospective migration histories. *Journal of the Royal Statistical Society: Series A* 166(1): 23–49. doi:10.1111/1467-985X.00257.

Sorensen, E. (1997). A national profile of nonresident fathers and their ability to pay child support. *Journal of Marriage and Family* 59(4): 785–797. doi:10.2307/353782.

Spielauer, M. (2007). The Contextual Database of The Generations and Gender Programme. In: United Nations (eds.). *Generations and Gender Programme: Concepts and Guidelines*. Geneva: United Nations: 35–44.

Sudman, S. and Bradburn, N.M. (1973). Effects of time and memory factors on response in surveys. *Journal of American Statistical Society* 68(344): 805–815. doi:10.1080/01621459.1973.10481428.

Sudman, S. and Bradburn, N.M. (1974). *Response Effects in Surveys: A review and synthesis*. Chicago, IL: Aldine Publishing Company.

Swicegood, C.G., Morgan, S.P., and Rindfuss, R.R. (1984). Measurement and Replication: Evaluating the Consistency of Eight U.S. Fertility Surveys. *Demography* 21(1): 19–33. doi:10.2307/2061024.

Van Loon, A.J., Tijhuis, M., Susan, H., Picavet, J., Surtees, P.G., and Ormel, J. (2003). Survey Non-response in the Netherlands: Effects on Prevalence Estimates and Associations. *Annals of Epidemiology* 13(2): 105–110. doi:10.1016/S1047-2797(02)00257-0.

Velkoff, V.A. and Miller, J.E. (1995) Trends and Differentials in Infant Mortality in the Soviet Union, 1970–90: How Much Is Due to Misreporting? *Population Studies* 49(2): 241–258. doi:10.1080/0032472031000148496.

Vikat, A., Spéder, Z., Beets, G., Billari, F. C., Bühler, C., Désesquelles, A., Fokkema, T., Hoem, J.M., MacDonald, A., Neyer, G., Pailhé, A., Pinnelli, A., and Solaz, A. (2007). Generations and Gender Survey: Concept and Design. In: United Nations (eds.). *Generations and Gender Programme: Concepts and Guidelines*. Geneva: United Nations: 3–32

Wagenaar, W.A. (1986). My Memory: A Study of Autobiographical Memory over Six Years. *Cognitive Psychology* 18(2): 225–252. doi:10.1016/0010-0285(86)90013-7.

Weisberg, H.F. (2009). *The Total Survey Error Approach: A Guide To The New Science Of Survey Research*. Chicago, IL: The University of Chicago Press.

Wu, L.L., Martin, S.P., and Long, D.A. (2001). Comparing Data Quality of Fertility and First Sexual Intercourse Histories. *Journal of Human Resources* 36(3): 520–555. doi:10.2307/3069629.

Wunsch, G.J. and Termote, M.G. (1978). *Introduction to demographic analysis. Principles and Methods*. New York: Plenum Press. doi:10.1007/978-1-4684-2373-0.

Zabel, C. (2009). Do imputed education histories provide satisfactory results in fertility analysis in the western German context? *Demographic Research* 21(6): 135–176. doi:10.4054/DemRes.2009.21.6.