# The Contextual Database of the Generations and Gender Programme: Concept, content, and research examples

**Arianna Caporali**[1]

**Sebastian Klüsener**[2]

**Gerda Neyer**[3]

**Sandra Krapf**[4]

**Olga Grigorieva**[2]

**Dora Kostova**[2]

## Abstract

**BACKGROUND**
Differences in demographic behaviours across countries and subnational regions have stimulated interest in studying the relationships between individual characteristics and the contexts in which individuals are embedded. Analytical approaches that include contextual indicators in statistical analyses of demographic behaviour need well-documented comparative data at the national and the subnational regional level. The Contextual Database (CDB) of the Generations and Gender Programme (GGP; http://www.ggp-i.org/data/ggp-contextual-database) supports such analyses by providing comparative data on demographic and socio-economic contexts in up to 60 countries in Europe, North America, Asia, and Oceania.

**OBJECTIVE**
This paper presents conceptual considerations and an overview of the content and the functionality of the CDB. Research examples illustrate how data from this database can increase the analytical potential of demographic analyses.

**CONCLUSIONS**
The CDB is a state-of-the-art research tool that provides well-documented comparative data at the national and the subnational regional level. Although it is conceptually

[1] Corresponding author. Institut national d'études démographiques (INED), Paris, France.
E-Mail: arianna.caporali@ined.fr.
[2] Max Planck Institute for Demographic Research (MPIDR), Rostock, Germany.
[3] Stockholm University Demography Unit (SUDA), Stockholm, Sweden.
[4] University of Cologne, Cologne, Germany.

linked to the Generations and Gender Survey (GGS), it can also be used to analyse data from other surveys, to study macro developments, and for teaching and lecturing. The CDB has a number of valuable features. First, it has a large number of indicators specifically geared towards demographic analyses, which provide extensive temporal and geographic coverage. Second, its dynamic web environment provides a high degree of transparency on data sources, as it offers meta-data for each individual entry. Finally, the CDB supports geocoding schemes that are used by the GGS and other surveys to denote region and country of residence.

## 1. Introduction

In recent decades, the demographic landscape of Europe has become more diverse. Fertility rates have fallen below replacement level in almost all European countries. Yet while some countries have experienced long-term low and lowest-low fertility levels, other countries have managed to stabilise their birth rates, or even to raise them to near-replacement levels. In some countries, the transitions to adulthood, partnership, and parenthood have become more heterogeneous, and family dynamics across the life course – as well as across the genders and the generations – have become increasingly complex. In other countries, family formation patterns have been rather stable, and behavioural changes have been slow and minimal. At the other end of the lifespan, life expectancy has risen, particularly in western Europe; whereas among some groups in some parts of eastern Europe life expectancy has declined. Although longevity has increased in most European countries, there are still huge and sometimes growing disparities in life expectancy across countries and within countries, such as across educational and occupational groups and between women and men (see contributions in Neyer et al. 2013).

These persistent differences in demographic outcomes and demographic behaviour have led both scientists and policy-makers to become increasingly interested in gaining a better understanding of the roles that various economic, social, and political factors play in shaping demographic behaviour and demographic development. The research questions that have been raised include the following: Do the socio-economic, institutional, and cultural contexts in which people live affect their demographic behaviour? Are differences in individual demographic behaviour and demographic patterns across Europe related to differences in contextual conditions? If the context is found to be relevant, which contextual factors matter? To what extent can the various factors that influence demographic behaviour and demographic outcomes be attributed to contextual conditions?

These and similar questions regarding contexts and demographic development are not new: Demographers have long recognised that the spatial and temporal environment in which people live affects their demographic behaviour and shapes demographic outcomes. For example, in his seminal study on European marriage patterns, Hajnal (1965) linked the changes and the differences in marriage patterns across Europe and across time to cultural, social, and economic factors: e.g., to differences in the speed at which the various European economies have developed from the agrarian stage, to the pre-industrial stage, and to the industrial stage; to differences and shifts in the economic prerequisites for household formation and household maintenance; and to changes in gender relationships.

Similarly, some of the classic fertility theories which linked variation in fertility decline and fertility levels across Europe to differences in economic development, modernisation, secularisation, value change, state formation, or gender equity attributed demographic outcomes to contextual features (Notestein 1945; Coale and Watkins 1986; Watkins 1991; McDonald 2000; Lesthaeghe 2010). In most of these cases, contextual factors were used in a narrative manner to describe the specificities of each country and its development, or to characterise the economic, social, cultural, or political circumstances in which people were living at a specific time. Such a narrative description of contextual features can offer valuable indications of the potential impact of macro-level factors on micro-level behaviour. In addition, the use of "thick descriptions" (Geertz 1973)[5] might sometimes be the appropriate way – and in some cases, the only way – to acknowledge the influence of contextual conditions on individual behaviour, and to explain demographic patterns and demographic outcomes (Hoem 2008; Neyer 2013; Klüsener, Neels, and Kreyenfeld 2013). However, "thick" descriptions usually do not provide statistically verified explanations for the effect of contextual conditions on demographic behaviour.

Developments over the past three decades in methodological and statistical techniques, as well as in software, have made it possible for researchers to move from providing a purely descriptive account of contexts to employing statistical approaches that allow them to control for contextual conditions in the analysis of individual-level data. One approach, which has become increasingly applied in demography and other social sciences, is multilevel modelling (Rabe-Hesketh and Skrondal 2008; Snijders and Bosker 2012). These models allow researchers to distinguish between the effects of contexts and the effects of individual characteristics on the behaviour and outcomes of interest. The possibilities of multilevel modelling have led to an increased need for micro- and macro-level data that are linked in a methodologically sound way (Rabe-

[5] Geertz (1973) used this term to point out the need to contextualise individual behaviour in order to understand it.

Hesketh and Skrondal 2008; Snijders and Bosker 2012).[6] In order to improve access to such data, the Generations and Gender Programme (GGP) not only conducts national comparative panel surveys that gather individual-level (micro-level) information on demographic behaviour (the Generations and Gender Surveys, GGS);[7] it also collects macro-level contextual data in the GGP Contextual Database (CDB) (Macura 2002; Festy 2004; Vikat et al. 2007). Beyond providing data that facilitate multilevel modelling, the CDB also represents a useful resource for alternative approaches that aim to account for contextual conditions, or for macro-level investigations: e.g., fixed-effects models (Del Boca 2002), geoadditive and geostatistical approaches (Chaix et al. 2005), fuzzy-set analyses (Murphy 1996; Ragin 2000; Mikkelsen 2015), (agent-based) simulations (González-Bailón and Murphy 2013), and time-series analyses (Engelhardt, Kögel, and Prskawetz 2004).

The aim of this paper is to give an overview of the concept, the content, and the functionality of the CDB, and to provide examples that illustrate how it can be used for research. We start with a presentation of the concepts that guided the construction of the database and the data collection process (Section 2). We then describe the content of the database and its functionalities, as well as the accompanying contextual data collection (Section 3). To illustrate the usefulness of the database, we present two examples of multilevel studies that linked individual data from the GGS with contextual macro data that were in part derived from the CDB (Lappegård, Klüsener, and Vignoli 2014; Neyer, Vignoli, and Lappegård 2011/2016) (Section 4). We conclude with some remarks on the future of the CDB.

## 2. The conceptual framework and the data collection procedure of the GGP Contextual Database

The CDB is intended to serve two main purposes.[8] First, the CDB is designed to provide data that allow researchers to control for contextual factors in investigations of individual-level demographic behaviour. Thus, its data should correspond to the data collected in the GGS. Second, the CDB is designed to allow researchers to test demographic theories related to topics covered by the GGS.[9] Such theories explore, for

[6] For a discussion of the strengths and the limitations of multilevel models, see, e.g., Stegmueller (2013) and Bryan and Jenkins (2016).

[7] For more information on the GGS, see other contributions in this Special Collection, as well as GGP (2016).

[8] For the history of the CDB, see Caporali et al. (2013).

[9] The testing of theories as well as the investigation of the relationships between contexts and individual outcomes touch on issues of causal inference. The CDB could facilitate causal modelling by providing data that allow researchers to better control for contextual factors (for discussions of causal analysis in demography, see, e.g., Ní Bhrolcháin and Dyson 2007; Neyer and Andersson 2008).

example, the links between female economic autonomy, relative economic deprivation, ideational change, wealth flow, gender equity/equality, and generational equity on the one hand, and the transition to adulthood, union/marriage formation, family formation/fertility, union dissolution, and the relationships between the generations on the other. Some of these theories focus on macro-micro linkages; others, like gender equity, address macro-level issues; and still others make use of both. To ensure that it serves such complex and broad purposes, the CDB has to fulfil the following criteria:

1. It should provide data in time series that comply with the GGS's retrospective and prospective structure.
2. To make use of the GGS's panel structure and its main dimensions regarding individuals and families, it should offer data on gender and generational perspectives, as well as on life course perspectives. These gender- and age-specific data must take the time dimension into account.
3. It should provide data for comparative research across countries and within countries, which implies that it should contain national as well as regional data.
4. To ensure that essential dimensions of the contexts in which people live are captured, and to allow researchers to test demographic theories about the relationship between contextual factors, demographic behaviour, and demographic outcomes, it should offer data on economic, social, cultural, and political issues (Neyer 2003).

These aims and criteria go beyond the objectives of most of the other research-oriented databases that existed when the GGP was started in 2000, as most of these databases were constructed to answer specific research questions only. Since the GGP is not limited to a single research topic, but is instead designed for use by the demographic, the sociological, and the economic communities, as well as by other social science research communities, the CDB has exceptionally broad content and data coverage.

In the collection of data, a four-way approach was used as a framework to guide the selection of indicators. First, the content of the GGS questionnaire served as a starting point for determining the relevant contextual domains (Festy 2002). Following a life course perspective, the focus was primarily on the central transitions and careers over the life course, such as the fertility career (becoming a parent, childbearing by parity, step-parenthood), the activity career (comprising education, work, unemployment, retirement), and the partnership career (cohabitation, marriage, dissolution/divorce/widowhood). For each life course career, a corresponding contextual domain for the CDB was identified by looking at which economic, social, cultural, and political

factors influence the respective transition and life course event (Neyer 2003; Spielauer 2004). For instance, the transition to parenthood or to a second child – which is captured in the GGS questionnaire via questions related to the intention to have a(nother) child and to birth events – was contextualised through structural indicators that potentially facilitate or constrain the intention and the decision to have a child. Examples of such macro indicators are the employment/unemployment rate in a region/country (as this influences the mother's income and her ability to continue to work after childbirth), the coverage and the hours per day of childcare services, if possible by age groups (as this indicates the level of public support for childrearing and for reconciling work and care), and family allowances for a first or a second child (since this indicates the level of financial support for families).

The second approach was concerned with demographic theories and hypotheses that can be related to the topics covered in the GGS questionnaire (Neyer 2003; Spielauer 2004, 2007). As was the case above, the theoretical assumptions were linked to contextual indicators which facilitated the testing. For example, the indicators that capture "female economic autonomy" at the macro level include the female labour force participation rate, the share of female part-time workers, the female unemployment rate, and the gender gap in wages.

Since the GGP is designed specifically to investigate gender and inter-generational relations, it was important that both a gender and a generational dimension were maintained in the two approaches mentioned above. While this involved looking for gender- and age-specific indicators over time, it was also clear that collecting gender- and age-specific data alone was not sufficient. The CDB therefore contains indicators which represent qualitative aspects of these relationships. These indicators reflect the main factors that shape gender and generational relationships across the life course: i.e., (1) equality, (2) agency, (3) social rights/social norms, and (4) risks and security (Neyer 2003). For example, levels of equality may be measured based on the income distribution or the representation of different groups of the population in specific areas of public life (e.g., women's labour force participation). Agency may be evaluated based on the levels of social services (e.g., care services) or poverty rates. Social rights/social norms may be captured through entitlement indicators or through provisions (e.g., parental leave). Risks and security may be assessed in terms of the distribution of, for example, health indicators, the unemployment rate, and the level of social expenditures on vulnerable groups (e.g., families).

The third approach was concerned with the methodological issues involved in the data analysis. As was outlined above, in order to enable researchers to conduct comparative studies that can account simultaneously for individual variation in the micro-level data and variation in contextual data, the CDB has to match the retrospective, prospective, and geographical information collected in the panel survey

(Racioppi and Rivellini 2002). It thus has to allow for the linkage over time of individuals and their geographical contexts, and of individuals and their membership groups.[10] Furthermore, the data have to be comparative across countries and other units. In terms of the regional detail, the database has to cover broad subnational regions at the level of aggregation that is most meaningful given the political, social, and/or economic dimensions of a specific country, and that is covered by the GGS in identifying the region of residence of interviewed individuals.[11]

The fourth and final approach was concerned with the practical collection of the data. It was based on a two-step strategy. Since for some countries (e.g., countries in eastern Europe) many of the indicators relevant for the GGP were not available in existing international data collections or had never before been derived, the national GGP partners were asked to collect the data from national sources. This collection process followed a template of collection guidelines for relevant contextual indicators structured around key topics identified via the approaches sketched above (see Spielauer 2004 and Caporali et al. 2013 for a description of the development of the data collection procedure). The work by the national GGP teams produced a rich data collection of many indicators that were not previously available internationally. While the national GGP partners made extensive efforts to collect the appropriate data, the comparability of these data across countries remains somewhat limited because many of these indicators were originally collected by national statistical or governmental offices following national criteria.

To enhance the comparability of indicators across countries/units and over time, the existing international databases – such as the databases available from the European Union (Eurostat), World Bank, UNESCO, OECD, ILO, WHO, UN, or research consortiums (e.g., Human Fertility Database, Human Mortality Database, Comparative Family Policy Database) – were screened (Bisogno 2002; Caporali et al. 2013). Each indicator was checked for cross-country comparability, completeness of the time series, errors, deviations in definitions, notes, and other forms of documentation. This process helped to clarify, for example, the variables and the possible irregularities/breaks in the

[10] Although it was originally planned that the GGS would collect a detailed migration history over each respondent's life course, in the end this was not done in order to avoid overburdening the interviewee. Some information, such as the place of birth, the place of residence at age 15, the place of residence at each wave of the GGS, and the date of immigration into the country (in the case of immigrants) are included in most national GGS questionnaires. Due to the limited information in the respondent's complete migration history, some event-history studies that intend to combine the retrospective individual-level GGS data with regional CDB data may have to rely on the assumption of a "frozen" place of residence of the respondent over his or her life course. This may mean that an "anticipatory" analysis is performed (Hoem and Kreyenfeld 2006a, 2006b). However, since people tend to move less frequently between countries than within countries, the lack of detailed migration histories may be less of an obstacle in country-level studies, provided proper modelling (Stegmueller 2013).

[11] In most cases, at the NUTS 1 level and/or at the NUTS 2 level. For some countries, such as Lithuania, the NUTS 3 level is also covered.

time series, and the completeness of the data sources (see Caporali et al. 2013 for details). To assess these data, the CDB team compared internationally available data with the nationally collected data. Whenever possible, the national and the comparative data were merged to ensure the completeness of the time series and of the subnational regional coverage for all of the countries. The use of this strategy resulted in a large series of comparable indicators across countries, regions, and time. To ensure that the richness of the nationally collected (often non-comparable) data is maintained, and to facilitate the use of those data that are internationally comparative, the collected data are split into two parts: (1) the *Contextual Database* (CDB), which contains the internationally comparable data (see Sections 3.1 and 3.2 below); and (2) the *Contextual Data Collection* (CDC), which contains the (often not comparable) data collected by the national GGP partners (see Section 3.3).

# 3. Content of the Contextual Database and the Contextual Data Collection

## 3.1 Content of the Contextual Database (CDB)

As we have outlined above, the collection procedure for the CDB generated a large series of indicators. For example, the CDB provides rich data at the subnational regional level, long time series for many indicators (as far back as 1970, or even earlier when possible), and extensive coverage of central and eastern European countries. To allow for the testing of theories, data beyond the core European GGS countries were included. Thus, for some indicators, the database covers all of the countries in the UNECE region (Europe, central and western Asia, North America), as well as the GGP countries in Asia and Oceania (Japan and Australia). The majority of the indicators are harmonised at the country level over time, and 12% of them are also harmonised at the regional level.[12]

The CDB contains data on a broad range of demographic and economic indicators, as well as a selection of social and policy indicators. The indicators are organised according to 10 relevant domains: demography, economy and social aspects, labour and employment, unemployment, childcare, education, health, pension, culture, and taxes and benefits. In total, as of June 2016, 74 indicators covering up to 60 countries (Europe, North America, Asia, and Oceania) are available.[13] The CDB includes, for

[12] Regional-level data are provided whenever it was possible to check the reliability and comparability of the data.

[13] An overview of the available indicators for each country is offered in a table downloadable from the web page (http://www.ggp-i.org/data/ggp-contextual-database, last accessed on June 24, 2016) in spread sheet

example, indicators related to fertility (e.g., total fertility rate, mean age at birth, age-specific fertility rates, and completed fertility); indicators on marriage and divorce (e.g., mean age at marriage, cohort ever married, and total divorce rate); life expectancy indicators; indicators related to education (e.g., school entry age and pupil-teacher ratio); various measures of gross domestic product, poverty, and the Gini coefficient; indicators on labour market conditions (e.g., labour force participation rates, average wages, and unemployment rates); and indicators related to pensions (e.g., number of beneficiaries and exit age from the labour market). Some of the indicators are provided by age and by sex. Yet another set of indicators offers information on public expenditures, such as spending on unemployment, childcare, education, health, pensions, family allowances, and social protection.

As we noted above, the data comes from different sources, including international databases of supranational organisations or databases of national institutions. Detailed, in-depth comparisons of these different sources allowed us to create time series that are as complete as possible given the spatial and the temporal availability of the data.[14] This cross-checking of the data guarantees the provision of high quality data from different sources.

The time series are described through detailed meta-information. Each indicator has rich documentation, including a definition of the indicator, a list of all of the national and international sources from which the data have been derived, and general comments about the sources used and the time series provided. In addition, meta-information is available for each data entry. This meta-information includes information on the source, comments about possible breaks in the series due to revisions of data collection methods and/or changes in national and subnational regional boundaries, deviations from the general indicator definition, and/or information on the calculation/estimation procedures used to derive the given number, where applicable. This detailed meta-information provides a high degree of transparency and of assurance of the quality of every single data item.

format ("Overview – Available Indicators per Country"). The table provides a definition for each variable, and informs the user whether the data are available at the national level only or at the subnational regional level as well.

[14] In combining data from different sources, two main strategies were used. First, for indicators for which the GGP national experts had core competences (e.g., for demographic indicators), we preferred national sources provided by these experts. If the time series contained gaps, an effort was made to fill them with data from international sources that were comparable to the data provided by the national experts. The same international sources were used to derive data for missing countries. Second, for indicators that were harmonised across countries by supranational organisations (e.g., macro-economic indicators and labour market variables) these international sources were preferred. To ensure data consistency, an effort was made to avoid using different sources across countries for the same years. Examples showing how the national and international sources were combined are available in Caporali et al. (2013).

### 3.2 Special functions of the CDB

The database environment is set up as a dynamic system, based on a relational database. The web interface offers users a dynamic choice of indicator values across countries, regions, and time (see Figure 1). This means that when selecting an indicator, users have direct access to the data and to all related meta-information. Users can thus make quick and informed choices when extracting data based on their individual needs. Coloured flags signal whether an indicator contains only cross-country comparative data (green flag), or non-comparative data as well (red flag). The latter are, however, the exception in the CDB.

If the data deviate from the variable definition for some countries or regions, this is documented in the meta-information. While the database offers the option to immediately download all of the data available for an indicator, users also have the possibility to restrict the output based on certain selection criteria, such as for specific years and geographical units. Depending on the indicators, other selection features may be available (e.g., age and sex). In addition, users can choose the dimensions of the output (e.g., to organise the data columns by regions, by time). Data can be exported in different formats (e.g., CSV, XLS, and XML). All of the available CDB indicators can also be accessed in a single file in SPSS or STATA format.[15]

In Figure 2 we show an output example, with the small pop-up window on the right providing meta-information for a single data entry. Users can access this meta-information either by clicking on the data cell in the output or in the process of defining the dimensions of the output. Here, the user can choose the "Single value column incl. meta-data" output, which displays both the values and the meta-information in a single table.

Furthermore, the CDB allows users to include an identity (ID) column in the output that provides the geocode used in the GGS survey to identify the place of residence of an interviewed person (see Figure 2). With this code, the user should find it easy to match the extracted CDB data to the GGS data. In addition to the GGS codes, other regional coding schemes are also supported, such as NUTS, OECD, and AGS (for German regions only). This allows researchers to match the CDB data to data from other international and/or national surveys (e.g., the European Social Survey, ESS; the Survey on Health, Ageing and Retirement, SHARE).

[15] These files are downloadable from the GGP website at http://www.ggp-i.org/data/download-cdb (last accessed on June 24, 2016).

**Figure 1: Choice of demographic indicators**

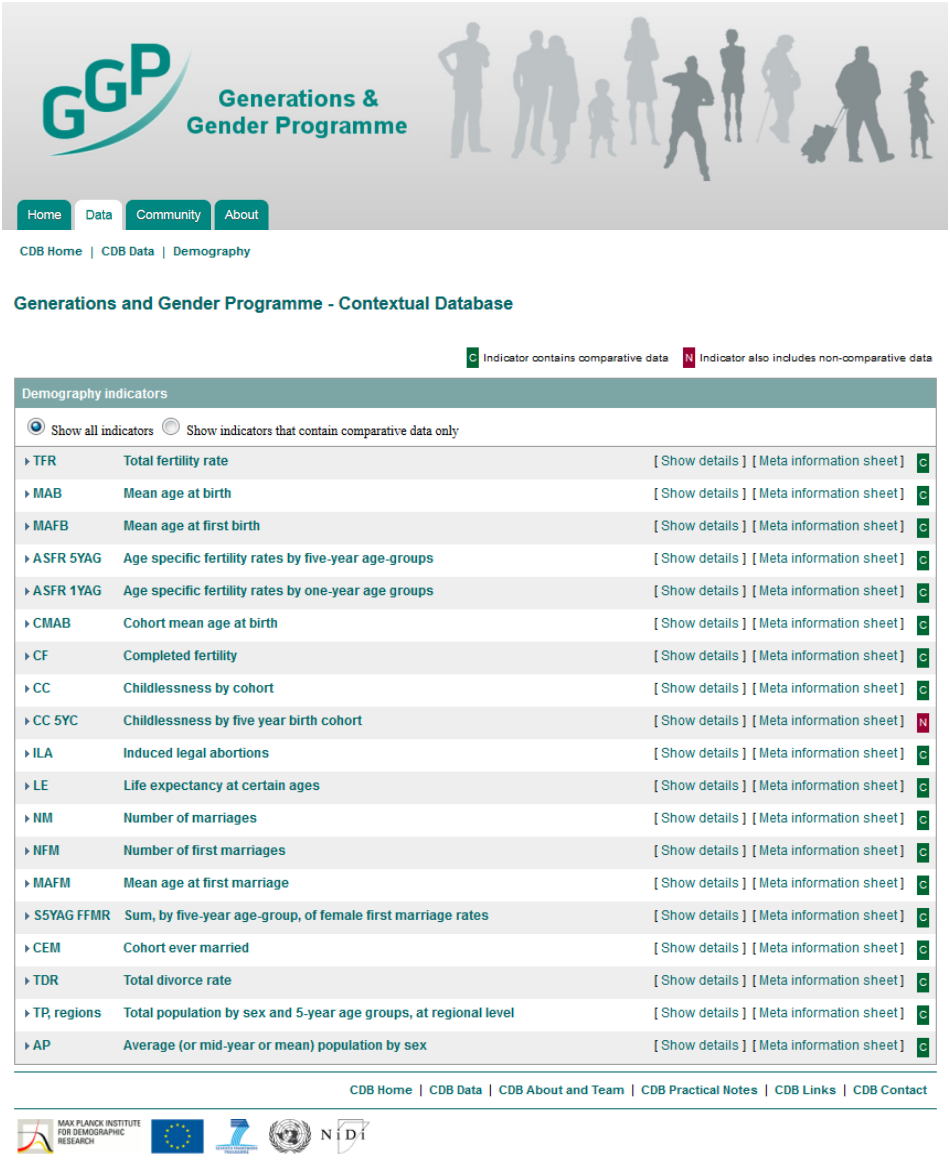
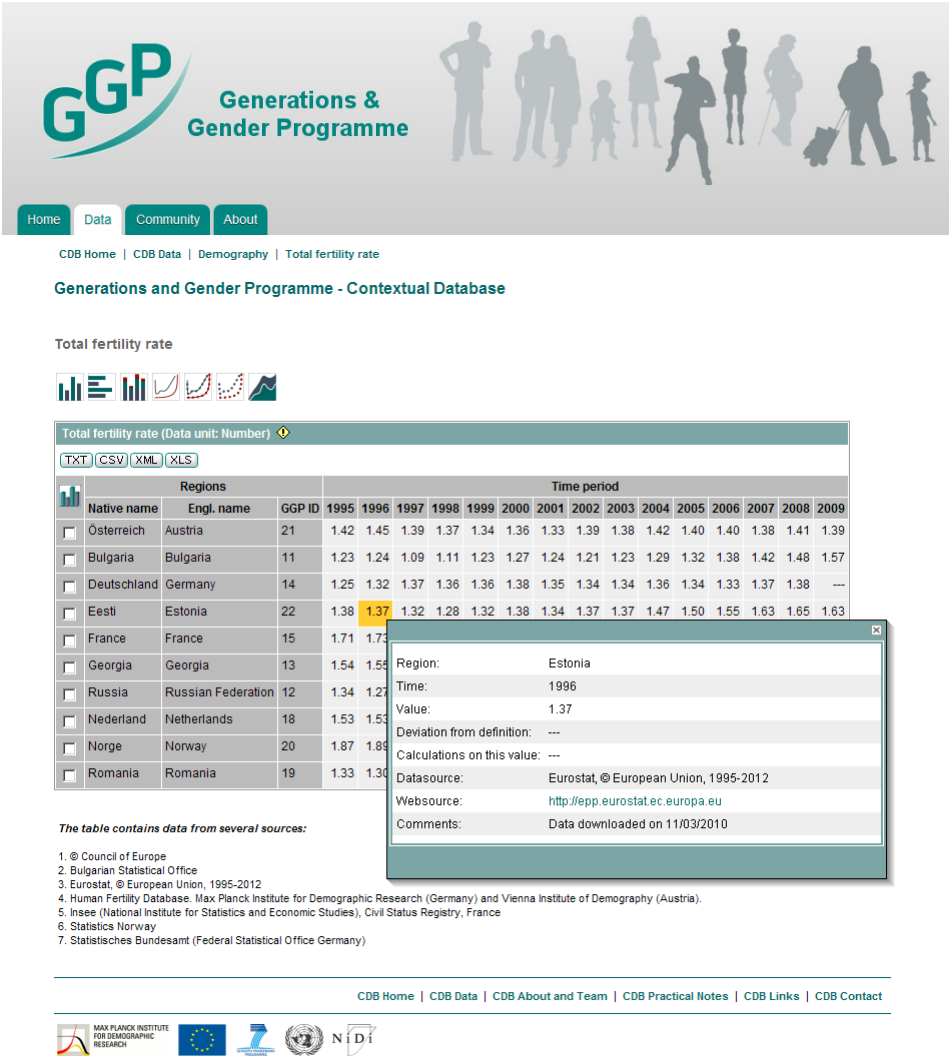
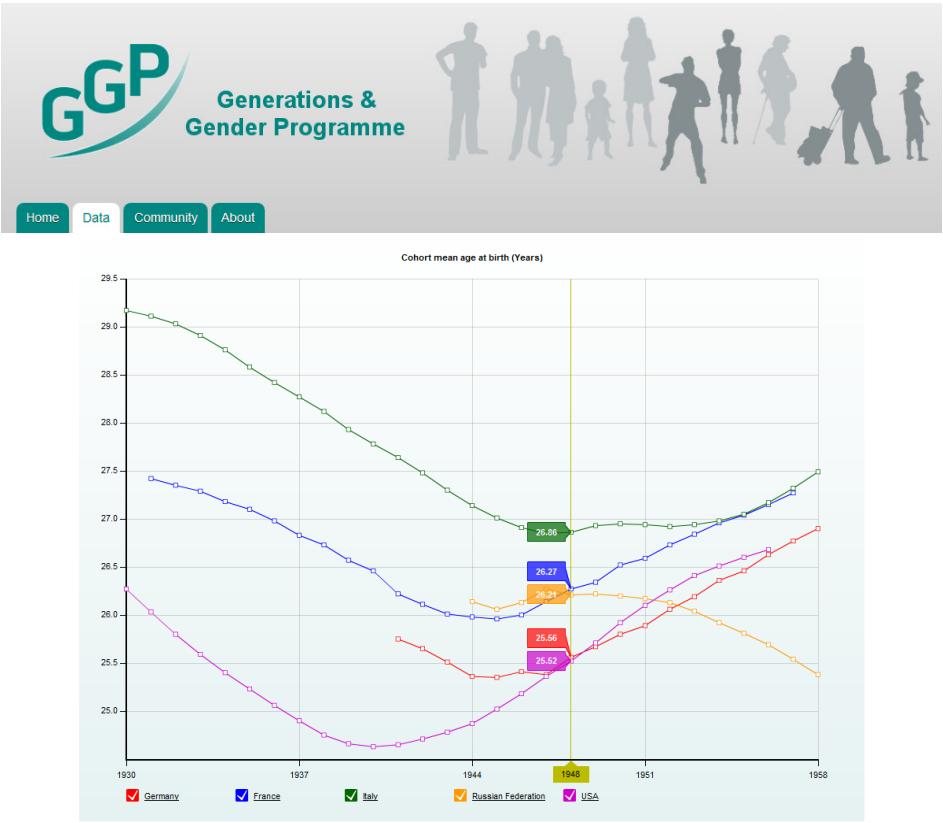


**Figure 2: Data output (with meta-data for a single data entry and GGP geocodes)**



The plot function gives the user some initial insight into the data (see Figure 3). Several dynamic options are available, including bar, line, and pie plots. These plots are interactive, allowing the user to zoom in on specific time periods, or to include or exclude countries and/or regions.

**Figure 3: Plot example**



The combination of all these functions and its content make the CDB a unique contextual research tool that goes beyond and/or complements other contextual databases of international research infrastructures.[16]

### 3.3 Content of the Contextual Data Collection (CDC)

The CDC provides detailed data for GGP member countries. As of June 2016, it covers 12 countries (Austria, Belgium, Bulgaria, Canada, France, Georgia, Germany, Hungary, Lithuania, Norway, Romania, Russia). Most of the data were collected by teams of national statistical offices, research institutes, or research departments within statistical offices that were involved in the GGP. Data for approximately 253 indicators are available. Among these indicators are around 127 national-level time series, 67 subnational regional variables, and 59 policy histories that contain standardised descriptions of policy reforms.[17] Whenever possible, the data go back to 1970. As it might be particularly difficult to obtain long time series for subnational regional indicators, the focus of the data collection activities for these indicators is on the period after 2000. In contrast to the CDB, in which the data are largely harmonised, the CDC contains a large number of indicators that are not always comparable across countries. Additionally, not all indicators are available for all countries. However, the data in the CDC are very rich in terms of the national sources used, and comparable across the regions within each of the countries.[18]

[16] For example, we compared the CDB with the Multilevel Data of the 6th round of the European Social Survey (ESS MD) available at: http://www.europeansocialsurvey.org/data/multilevel/ (last accessed on July 10, 2015; ESS ERIC 2016; Rydland, Arnesen, and Østensen 2007). This database has goals similar to those of the CDB: namely, to provide contextual data for the combined analysis of individual-level and contextual-level data by using the ESS. The ESS MD offers indicators on demography, economy, education, composite measures, political institutions, health, and crime. Some of the indicators in the ESS MD and in the CDB overlap, but most of the indicators are either only in the ESS MD or only in the CDB. The ESS MD contains a somewhat larger number of indicators than the CDB and regional information down to NUTS 3 (if available), but the time frame generally covers the 1990s and/or the 2000s only. The CDB offers longer time series, regional information down to the NUTS 2 level (for some countries, down to the NUTS 3 level), and gender- and age-specific data. The structures of the ESS MD and the CDB are similar. The CDB additionally provides detailed meta-information for each data entry and a dynamic web environment, while the ESS offers the additional option to calculate correlations and to run regressions. Both databases allow to display data in figures. With respect to research utilisation, our in-depth comparison showed that the two databases complement each other.

[17] A description of most of the indicators included in the CDC is provided in Caporali et al. (2013), Figure 1, pp. 7–9. An overview of all the indicators in the CDC is provided in the guidelines for national data collectors downloadable from the website (http://www.ggp-i.org/data/ggp-contextual-database, last accessed on June 24, 2016) in spread sheet format ("CDB_Templates_v1.10").

[18] The data are available and downloadable in spread sheet format from the GGP-website: http://www.ggp-i.org/data/ggp-contextual-database (last accessed on June 24, 2016).

## 4. The Contextual Database in practice

In order to demonstrate the potential of contextual data from the CDB for demographic research, we present two studies by Lappegård, Klüsener, and Vignoli (2014) and Neyer, Vignoli, and Lappegård (2011/2016).[19] Lappegård, Klüsener, and Vignoli (2014) examined the inconsistencies in the existing explanations of the recent rise in childbearing in cohabitation. While some authors have argued that this trend is largely attributable to a pattern of progress primarily driven by the increasing economic autonomy of women (e.g., Lesthaeghe 2010), others have attributed it to a pattern of disadvantage driven by economic uncertainties (Perelli-Harris et al. 2010). In their paper, Lappegård, Klüsener, and Vignoli (2014) argued that these inconsistencies may have arisen because the existing studies either focused on the differences between countries, between the subnational regions, or between individuals. Processes such as women's increasing economic autonomy are often linked to welfare state developments (Sainsbury 1999; Esping-Andersen 2009). Thus, these processes might be particularly relevant for explaining between-country variation, whereas processes of economic uncertainty might be more relevant for explaining variation at the regional level, as a "general milieu of social disorganization" (Billy and Moore 1992: 982) might emerge in regions with high structural unemployment. The CDB is very helpful in testing such propositions, as it provides both national-level and regional-level data.

For their study, the authors used individual-level data from the Harmonized Histories project (Perelli-Harris, Kreyenfeld, and Kubisch 2010) to investigate whether first births among partners who were sharing the same household occurred within cohabitation or marriage. The Harmonized Histories data set comprises individual-level data from the GGS as well as from national surveys in countries not covered by the GGP. Lappegård, Klüsener, and Vignoli (2014) linked their individual-level data with contextual data from the CDB and other sources. In total, 16 European countries were covered. These countries were then subdivided into 116 regions. The contextual measures included social disapproval of cohabitation, the importance of religious norms, social norms related to the economic autonomy of women, and economic conditions. The first three measures were constructed from aggregated survey responses of the GGS, the ESS, and the European Value Survey (EVS). The data on economic conditions (measured as the adult unemployment rate) were obtained from the CDB. The data were analysed in a multilevel logistic regression model with a random intercept. The first level consisted of the surveyed individuals, who were nested in their region of residence and in their country of residence. This set-up allowed the authors to

[19] Other studies that used data from the CDB are, e.g., Arpino, Esping-Andersen, and Pessin, (2015); Lyons-Amos (2015); Wood, Neels, and Vergauwen (2016).

make full use of the analytical potential offered by the CDB data, by simultaneously controlling for variation in the contextual indicators at the regional and country levels, and also accounting for individual characteristics.

The results of the analysis provide support for the assertions made by Lappegård, Klüsener, and Vignoli (2014), i.e., that the relevance of the explanations varied by geographic scale. Considerations related to the increasing autonomy of women were shown to be particularly relevant for explaining variation between countries. Arguments related to a pattern of disadvantage were found to be more useful for explaining variation within countries, as regions with higher unemployment had significantly higher shares of births in cohabitation. Overall, the study provides us with an interesting example of how the linking of survey data with contextual information at the regional and the national levels can contribute to our understanding of recent changes in family formation behaviour in Europe.

The second study investigated the impact of economic opportunities and of economic uncertainties on childbearing intentions from a gender perspective (Neyer, Vignoli, and Lappegård 2011/2016). The authors were interested in the question of whether women's and men's childbearing intentions were affected differently by these economic contextual conditions. A large number of studies have shown that economic uncertainty, measured as individual unemployment, is associated with reduced childbearing intentions and childbearing (see: Kreyenfeld, Andersson, and Pailhé 2012). By contrast, economic opportunities, measured as being in employment, have been found to have varying effects on women's fertility intentions and childbearing (Matysiak and Vignoli 2008), but an elevating effect on men's fertility intentions (Neyer, Lappegård, and Vignoli 2013). Neyer, Vignoli, and Lappegård (2011/2016) explored the question of whether these patterns still hold if the labour market structures – that is, the economic opportunities and the economic constraints – in the region and in the country in which a person lives are taken into account. The authors captured economic opportunities via female and male labour force participation rates at the regional and the national levels, as these rates can be regarded as indicators of the degree of economic security felt by women and men, and of the options for finding work in a given region or country. Economic constraints were measured by the regional and the national unemployment rates, as these indicators reflect whether employment is scarce in a given region or country.

Neyer, Vignoli, and Lappegård (2011/2016) employed a two-level logistic regression model with random effects at the regional level, controlling for fixed country effects. They found that including regional and national indicators of employment opportunities and employment uncertainty explained a substantial portion of the unexplained variance in a model that included only individual-level information. These indicators applied to men more than they did to women, and to the childless more than

they did to parents. In general, the economic performance of the country was found to have a stronger effect than the economic performance of the region. However, regional labour market conditions mattered for women's childbearing intentions. This may be an indication that women are more constrained by local labour market conditions because of work-family tensions. The study nicely showed that economic opportunities and economic constraints in the region or the country affect women's and men's childbearing intentions differently. It is possible to conclude from the study that in order to understand the relationship between economic factors and childbearing, it is necessary to take a gender perspective and to consider the structure of both economic opportunities and economic uncertainties, and at both the regional and the country levels.

## 5. Conclusion

This paper provided an overview of the conceptual considerations, content, and functions of the GGP CDB. Although the main purpose of the database is to increase the analytical potential of individual-level survey data of the GGS by providing contextual data, the database may also be useful for researchers who analyse individual-level data from other surveys, or who are interested in studying macro-level trends. The features of the CDB make it a unique support tool for researchers interested in micro-macro linkages, as well as for researchers concerned with socio-economic structures and macro-level processes. It can also be employed in teaching and lecturing, and in preparing (general public) presentations. We demonstrated the potential of the CDB using two research examples. These examples showed how an analysis that combines individual-level survey data and regional and national contextual-level data can make important contributions to our understanding of demographic behaviour in highly developed countries.

The future activities of the CDB team are to focus on increasing the number of harmonised policy indicators to further enhance the analytical potential of the database, and to intensify their collaboration with other database projects, including projects such as Anne Gauthier's Comparative Family Policy Database (Gauthier 2016), the Multilinks Database on Intergenerational Policy Indicators (Multilinks Project and Wissenschaftszentrum Berlin für Sozialforschung 2016), the InGRID (Inclusive Growth Research Infrastructure Diffusion 2016) database activities, and the Population Europe Resource Finder and Archive (PERFAR; Max Planck Society for the Advancement of Science on behalf of the collaborative network Population Europe 2016). The possibility of deriving aggregate national- and regional-level data on values from individual-level survey data of the GGS, the ESS, and the EVS is also explored.

We conclude with an assessment of the conditions that should be met by databases of macro-level indicators to fulfil future social science needs. With the development of the internet, a vast number of databases of contextual indicators have become available. These databases have often been collected and compiled by national or international administrative offices and organisations for purposes other than research. While these databases are excellent resources for researchers, they may not meet the needs of all research projects. Research-oriented databases must fulfil specific criteria: i.e., the databases should be theory-driven, and their content should conform to basic research principles; and the data should be verifiable, reliable, and replicable. Meeting these criteria involves a thorough documentation of the collection process and the data sources, a definition of the indicators, and a harmonisation of procedures. Another issue that arises in this context is the almost insurmountable tension between the need to create a theory-driven, empirical research-oriented database on the one hand; and the need to serve the multiple theoretical and methodological interests of the broader research community on the other. We believe that in the long run, this tension can be resolved only by the establishment of research-oriented databases through joint European research efforts, like the GGP.

## 6. Acknowledgements

The work on the Contextual Database has been made possible by a design study grant from the European Union as part of the 7th Framework Research Programme (grant number 212749) and by financial support from the Max Planck Institute for Demographic Research (MPIDR, Rostock, Germany). Arianna Caporali acknowledges support from the Institut national d'études démographiques (INED, Paris, France). Gerda Neyer acknowledges financial support from the Swedish Research Council (Vetenskapsrådet) via the Linnaeus Center for Social Policy and Family Dynamics in Europe (SPaDE, Stockholm), grant registration number 349-2007-8701, and from the Riksbankens Jubileumsfond, grant IN15-0511:1. We would like to thank the previous CDB working groups, the national GGP teams who collected data in their countries, and Anne Gauthier (Netherlands Interdisciplinary Demographic Institute (NIDI, The Hague) for allowing us to reproduce the data in her Comparative Family Policy Database through the CDB. We also wish to thank Fred Heiden and the IT staff of the MPIDR for programming the web environment, and the student assistants at the MPIDR for their support with data collection and data entry. We are grateful to Jan M. Hoem for providing comments to previous versions of the paper and to Miriam Hils for language editing.

## References

Arpino, B., Esping-Andersen, G., and Pessin, L. (2015). How do changes in gender role attitudes towards female employment influence fertility? A macro-level analysis. *European Sociological Review* 31(3): 370–382 (plus Supplementary Material). doi:10.1093/esr/jcv002.

Billy, J.O.G. and Moore, D.E. (1992). A multilevel analysis of marital and nonmarital fertility in the U.S. *Social Forces* 70(4): 977–1011. doi:10.2307/2580198.

Bisogno, E. (2002). UNECE data of possible interest to GGP Contextual Database. In: *GGP Contextual database group: Activities and conclusions:* 22–26. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed on April 15, 2016).

Bryan, M.L. and Jenkins, S.P. (2016). Multilevel modelling of country effects: A cautionary tale. *European Sociological Review* 32(1): 3–22. doi:10.1093/esr/jcv059.

Caporali, A., Klüsener, S., Neyer, G., Krapf, S., and Grigorieva, O. (2013). Providing easy access to cross-country comparative contextual data for demographic research: Concept and recent advances of the Generations & Gender Programme Contextual Database. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2013-001).

Chaix, B., Merlo, J., Subramanian, S.V., Lynch, J., and Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology* 162(2): 171–182. doi:10.1093/aje/kwi175.

Coale, A.J. and Watkins, S.C. (eds.) (1986). *The decline of fertility in Europe*. Princeton: Princeton University Press.

Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in Italy. *Journal of Population Economics* 15(3): 549–573. doi: 10.1007/s001480100089.

Engelhardt, H., Kögel, T., and Prskawetz, A. (2004). Fertility and women's employment reconsidered: A macro-level time-series analysis for developed countries, 1960–2000. *Population Studies* 58(1): 109–120. doi:10.1080/0032472032000167715.

Esping-Andersen, G. (2009). *The incomplete revolution: Adapting to women's new roles*. Cambridge: Polity Press.

ESS ERIC (2016). European Social Survey Multilevel Database [electronic resource]. Bergen: Norwegian Social Science Data Services. http://www.europeansocialsurvey.org/data/multilevel/.

Festy, P. (2002). From GGS to GGP; from micro- to macro- via meso- levels: Variables in the survey [GGS] and in the contextual database [CDB]. In: *GGP Contextual database group: Activities and conclusions*: 18–21. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed on April 15, 2016).

Festy, P. (2004). GGP Contextual database group: First discussions, first conclusions. Available from the author: festy@ined.fr.

Gauthier, A.H. (2016). Comparative Family Policy Database [electronic resource]. Rostock: Max Planck Institute for Demographic Research. http://www.demogr.mpg.de/cgi-bin/databases/FamPolDB/index.plx.

Geertz, C. (1973). *The interpretation of culture: Selected Essays*. New York: Basic Books.

GGP (2016). Generations and Gender Programme [electronic resource]. The Hague: Netherlands Interdisciplinary Demographic Institute. http://www.ggp-i.org/.

González-Bailón, S. and Murphy, T.E. (2013). The effects of social interactions on fertility decline in nineteenth-century France: An agent-based simulation experiment. *Population Studies* 67(2): 135–155. doi:10.1080/00324728.2013.774435.

Hajnal, J. (1965). European marriage pattern in perspective. In: Glass, D.V., and Eversley, D.E.C. (eds.). *Population in History*. Chicago: Aldine Publishing Company: 101–143.

Hoem, J.M. (2008). Overview chapter 8: The impact of public policies on European fertility. *Demographic Research* 19(10): 249–260. doi:10.4054/DemRes.2008.19.10.

Hoem, J.M. and Kreyenfeld, M. (2006a). Anticipatory analysis and its alternatives in life-course research. Part 1: The role of education in the study of first childbearing. *Demographic Research* 15(16): 461–484. doi:10.4054/DemRes.2006.15.16.

Hoem, J.M. and Kreyenfeld, M. (2006b). Anticipatory analysis and its alternatives in life-course research. Part 2: Two interacting processes. *Demographic Research* 15(17): 485–498. doi:10.4054/DemRes.2006.15.17.

Inclusive Growth Research Infrastructure Diffusion (InGRID) (2016) [electronic resource]. Leuven: Research Institute for Work and Society. http://inclusive growth.be/.

Klüsener, S., Neels, K., and Kreyenfeld, M. (2013). Family policies and the Western European fertility divide: Insights from a natural experiment in Belgium. *Population and Development Review* 39(4): 587–610. doi:10.1111/j.1728-4457.2013.00629.x.

Kreyenfeld, M., Andersson, G., and Pailhé, A. (2012). Economic uncertainty and family dynamics in Europe: Introduction. *Demographic Research* 27(28): 835–852. doi:10.4054/DemRes.2012.27.28.

Lappegård, T., Klüsener, S., and Vignoli, D. (2014). Social norms, economic conditions and spatial variation of childbearing within cohabitation across Europe. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2014-002).

Lesthaeghe, R. (2010). The unfolding story of the Second Demographic Transition. *Population and Development Review* 36(2): 211–251. doi:10.1111/j.1728-4457.2010.00328.x.

Lyons-Amos, M. (2015). Multilevel models for cross-national comparisons: The association between individual and national-level demographic characteristics in fertility and partnerships. ESRC Centre for Population Change (CPC Working paper 71).

Macura, M. (2002). Executive Summary. The Generations and Gender Programme: A study of the dynamics of families and family relationships. New York/Geneva: United Nations. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocProgrExecSumm.pdf, last accessed on April 15, 2016).

Matysiak, A. and Vignoli, D. (2008). Fertility and women's employment: A meta-analysis. *European Journal of Population* 24(4): 363–384. doi:10.1007/s10680-007-9146-2.

Max Planck Society for the Advancement of Science on behalf of the collaborative network Population Europe (2016). Population Europe Resource Finder and Archive (PERFAR) [electronic resource]. Rostock: Max Planck Institute for Demographic Research. www.perfar.eu.

McDonald, P. (2000). Gender equity in theories of fertility transition. *Population and Development Review* 26(3): 427–439. doi:10.1111/j.1728-4457.2000.00427.x.

Mikkelsen, K.S. (2015). Fuzzy-set case studies. *Sociological Methods & Research* [Published online before print] 1–34. doi:10.1177/0049124115578032.

Multilinks Project and Wissenschaftszentrum Berlin für Sozialforschung (WZB) (2016). Multilinks Database on Intergenerational Policy Indicators [electronic resource]. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). http://multilinks-database.wzb.eu/.

Murphy, M. (1996). The dynamic household as a logical concept and its use in demography. *European Journal of Population* 12(4): 363–381. doi:10.1007/BF01796913.

Neyer, G. (2003). Gender and generations dimensions in welfare-state policies. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP 2003-022).

Neyer, G. (2013). Welfare states, family policies, and fertility in Europe. In Neyer, G., Andersson, G., Kulu, H., Bernardi, L., and Bühler, C. (eds.) (2013). *The Demography of Europe*. Dordrecht: Springer: 29–53. doi:10.1007/978-90-481-8978-6_3.

Neyer, G. and Andersson, G. (2008). Consequences of family policies on childbearing behavior: Effects or artifacts? *Population and Development Review* 34(4): 699–724. doi:10.1111/j.1728-4457.2008.00246.x.

Neyer, G., Andersson, G., Kulu, H., Bernardi, L., and Bühler, C. (eds.) (2013). *The Demography of Europe.* Dordrecht: Springer. doi:10.1007/978-90-481-8978-6.

Neyer, G., Lappegård, T., and Vignoli, D. (2013). Gender equality and fertility: Which equality matters? *European Journal of Population* 29(3): 245–272. doi:10.1007/s10680-013-9292-7.

Neyer, G., Vignoli, D., and Lappegård, T. (2011/2016). *Gender, economic conditions and fertility. Does context matter for fertility intentions?* Paper presented at the Population Association of America Annual Meeting, Washington, DC, March 31–April 2, 2011 (forthcoming as SRRD: Stockholm University Research Report in Demography 2016).

Ní Bhrolcháin, M. and Dyson, T. (2007). On causation in demography: Issues and illustrations. *Population and Development Review* 33(1): 1–36. doi:10.1111/j.1728-4457.2007.00157.x.

Notestein, F.W. (1945). Population – The long view. In: Schultz, T.W. (ed.). *Food for the World*. Chicago: Chicago University Press: 36–57.

Perelli-Harris, B., Kreyenfeld, M., and Kubisch, K. (2010). Harmonized histories: Manual for the preparation of comparative fertility and union histories. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2010-011).

Perelli-Harris, B., Sigle-Rushton, W., Kreyenfeld, M., Lappegård, T., Keizer, R., and Berghammer, C. (2010). The educational gradient of childbearing within cohabitation in Europe. *Population and Development Review* 36(4): 775–801. doi:10.1111/j.1728-4457.2010.00357.x.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modelling using Stata*. 2nd ed. College Station: Stata Press.

Racioppi, F. and Rivellini, G. (2002). The contextual dimension in GGP: Some methodological issues about data collection and sampling procedures. In: *GGP Contextual database group: Activities and conclusions:* 11–17. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed on April 15, 2016).

Ragin, C.C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.

Rydland, L.T., Arnesen, S., and Østensen, Å.G. (2007). Contextual data for the European Social Survey. An overview and assessment of extant sources. Bergen: Norwegian Social Science Data Services. (http://www.europeansocialsurvey.org/docs/methodology/ESS_context_sources.pdf, last accessed on April 15, 2016).

Sainsbury, D. (ed.) (1999). *Gender and welfare state regimes*. Oxford: Oxford University Press. doi:10.1093/0198294166.001.0001.

Snijders, T.A.B. and Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. 2nd ed. London: Sage.

Spielauer, M. (2004). The Contextual Database of the Generations and Gender Program: Overview, conceptual framework and the link to the Generations and Gender Survey. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP 2004-014).

Spielauer, M. (2007). Part two: The Contextual Database of the Generations and Gender Programme. In: *Generations and Gender Programme: Concepts and Guidelines*. New York/Geneva: United Nations: 33–44.

Stegmueller, D. (2013). How many countries for multilevel modelling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57(3): 748–761. doi:10.1111/ajps.12001.

Vikat, A., Spéder, Z., Beets, G., Billari, F., 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 (GGS): Towards a better understanding of relationships and processes in the life course. *Demographic Research* 17(14): 389–440. doi:10.4054/DemRes.2007.17.14.

Watkins, S.C. (1991). *From provinces into nations: Demographic integration in western Europe*, *1870-1960*. Princeton: Princeton University Press.

Wood, J., Neels, K., and Vergauwen, J. (2016). Economic and institutional context and second births in seven European countries. *Population Research and Policy Review* 35(3): 305–325. doi:10.1007/s11113-016-9389-x.