ÖAW AUSTRIAN ACADEMY OF SCIENCES

VIENNA INSTITUTE OF DEMOGRAPHY

WORKING PAPERS

01/2018

BOWLING TOGETHER: SCIENTIFIC COLLABORATION NETWORKS OF DEMOGRAPHERS AT EUROPEAN POPULATION CONFERENCES

GUY ABEL, VALERIA BORDONE, RAYA MUTTARAK AND EMILIO ZAGHENI

Vienna Institute of Demography
Austrian Academy of Sciences
Welthandelsplatz 2, Level 2 | 1020 Wien, Österreich
vid@oeaw.ac.at | www.oeaw.ac.at/vid

Vienna Institute of Demography

Abstract

Exploiting a unique database of metadata for papers presented at six European Population Conferences (EPC) for the years 2006-2016, this paper explores: 1) development of research in population studies; 2) trends and patterns of scientific collaboration networks among demographers; and 3) gender differences in demographic research. The data are organised in a panel format whereby each author, institution and country are linked across the six conferences. We find that collaboration among demographers has increased substantially over the past ten years. While there is no gender disparity in the likelihood of co-authoring a paper, men are significantly more likely than women to collaborate with authors from other institutions. Likewise, the fields of research vary considerably by gender where women are particularly overrepresented in the subfield 'fertility and family' whereas men dominate the subfield 'data and methods'. Compared to other subfields, research on 'data and methods' is more likely to involve collaboration across multiple institutions. With respect to collaboration patterns at the institutional level, a chord diagram plot shows that scientific collaborations across institutions are more common between institutions sharing geographical proximity. Finally, using network centrality measures, we identify key demographic research institutes which play a role in driving demographic research in Europe.

Keywords

Centrality measures, demographers, gender, networks, population studies, scientific collaboration.

Authors

Guy Abel, Asian Demographic Research Institute (ADRI), Shanghai University, Shanghai, China and Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW and WU), International Institute for Applied Systems Analysis, Laxenburg, Austria. Email: guy.abel@oeaw.ac.at

Valeria Bordone, Department of Sociology, Ludwig-Maximilians-Universität München, Munich, Germany. Email: valeria.bordone@soziologie.uni-muenchen.de

Raya Muttarak (corresponding author), Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW and WU), International Institute for Applied Systems Analysis, Laxenburg, Austria. Email: muttarak@iiasa.ac.at

Emilio Zagheni, Department of Sociology, University of Washington, Seattle, USA.
Email: emilioz@uw.edu

Acknowledgements

We are grateful to the European Population Association for assistance in obtaining the data used in this study.

Bowling Together: Scientific Collaboration Networks of Demographers at European Population Conferences

Guy Abel, Valeria Bordone, Raya Muttarak, Emilio Zagheni

1. Introduction

Scientific collaboration is a unique setting for studying social networks. Research on scientific collaboration dates back to the 1960s, when sociologists first aimed to provide insight into science as an inherently social and team-based endeavour. The number of co-authors of an academic article, for instance, is an indicator of social capital which can play a crucial role in job mobility and academic success (Bäker 2015). It has been documented that globalization, changing communication patterns and increasing mobility of scientists contributed to a rise in collaborative research, especially international collaboration over the two past decades (Adams 2012; Glänzel and Schubert 2004). Since the 1990s, this upward trend in multi-authorship can be observed in all areas of science from medicine and biosciences to mathematics and social sciences, including law (Adams 2012).

Studies on scientific collaboration networks have gained scholarly interest since the past couple of decades as reflected in the establishment of specific journals such as *Scientometrics* and *Research Policy* which regularly publish articles on the analysis of characteristics of different research fields. However, there are only a few studies that specifically investigate collaboration networks of demographers and the extant ones mainly focus on demographers based in the United States. Merchant (2015) in her PhD thesis provides an extensive historical analysis of the development of demography between 1925 and 2010. Based on the topic modelling method used to identify patterns of word co-occurrence in research articles in the three main English-language demography journals, namely, *Demography*, *Population and Development Review* and *Population Studies*, Merchant (2015) describes how the topics commonly appeared in demographic research changed and evolved over time. For instance, the high frequency of the words 'China', 'Chinese' and 'population' in *Population and Development Review* after 1979 reflects the popularity of research on population in China after the introduction of the 'One-Child' Policy. Merchant (2015) also performs network analysis of connections between journals and authors but the analysis mainly concentrates on authors' connections with the presidents of the Population Association of America (PAA).

Other two studies on the characteristics of demographic research are based exclusively on articles published in the journal *Demography*. The first study analysed published articles from 1964-1991 (Teachman et al. 1993). More than two decades later, the successive study by Krapf et al. (2016) extended the analysis to published articles between 1964 and 2014

with a particular focus on gender differences in authorship by demographic subfields and by order of authors.

As part of the 30th year anniversary of the journal *Demography*, Teachman (1993) performed a content analysis on 1,232 articles published in the journal between 1964 to 1991. It was found that the nature of demographic research had undergone changes since the end of World War II, following many different forces. Development of computer technology and statistical methodology allowed for handling of large databases as well as more complex analysis of demographic data. At the same time, changing women's role which resulted in a greater proportion of women conducting demographic research also shaped the topics being studied. According to Teachman (1993), the surge in published articles focusing on marriage and family was attributable to the rising number of female authors in *Demography*. An increase in co-authored papers was also observed during the study period.

The follow-up study by Krapf et al. (2016) focused specifically on gender disparities in publication patterns in *Demography* over the last 50 years. They found no particular female disadvantages in authorship practices, where women were as equally likely as men to have been either the single author of a paper or the first author in a multi-authored paper. Although it was concluded that *Demography* is a gender-equal journal, Krapf et al. (2016) pointed out that there are substantial gender differences in subfields of publications, with women being more likely to publish in 'family and household' and less so in the 'mortality and health', 'migration' and 'data and methods' categories.

While the results from the two studies give us historical insight into development of research and gendered authorship patterns in demographic research, as noted by Krapf et al. (2016), their analysis of a single journal is not representative of the entire field of demography. Given that *Demography* is the flagship journal of the Population Association of America (PAA), indeed it may not represent patterns found in other geographical contexts. Likewise, focusing on the three key demography journals, the study of Merchant (2015) is also restricted only to published articles in these journals.

Existing studies on scientific collaboration networks commonly use bibliometric methods which involve using bibliometric databases of scientific publications (i.e., journal articles, books and book chapters) to identify authors, their publications, affiliations and co-authors. While such method is useful in analysing collaboration practices such as co-authorship and citation networks, this may be problematic due to sample selection bias. Junior academics and doctoral students typically have a lower number of journal publications as compared to senior academics. Subgroups of demographers therefore can be underrepresented in bibliographic databases. Likewise, compared to natural sciences and engineering, some research subjects in social sciences are more localised with limited target readership (Larivière et al. 2013). Many social science scholars consequently publish more frequently in journals with restricted distribution within a country or region in their own mother tongue or they publish their results in books or book chapters. Since non-English language journals as well as books and book chapters are often not included in a standard bibliographic database, it is thus likely that these scholars are underrepresented.

In addition, published articles by demographers are not strictly limited to journals with solely demographic focus due to interdisciplinary nature of demographic research. This can undermine a comprehensive analysis of collaboration networks as authors may choose to publish in journals from other disciplines such as sociology, economics, geography, political science or statistics, or in top-tier multi-disciplinary journals such as *Science*, *Nature* or *PNAS*.

To the best of our knowledge, the only study using conference participation data in the field of population studies is that of Ediev et al. (2009) based on 15 conferences (1959-2009) of the International Union for Scientific Study of Population (IUSSP). This study provides insight into trends and patterns of collaboration including demographic and geographical differentials. The study however suffers a selection bias since the data on demographic profiles of the authors were taken exclusively from those who were a member of the IUSSP.

Exploiting a unique database of papers presented at the European Population Conferences (EPC), this study reduces the potential sample selection problem of the bibliographic database as well as the membership database. The EPC is the largest demographic conference in Europe, with average attendance of approximately 1,000 participants. This biennial conference organised by the European Association for Population Studies (EAPS) covers a wide range of topics of population studies, including researchers from a variety of disciplines. The database of conference papers presented at the EPC thus provides a unique representation for the study of the networks of demographers. Note however that European based demographers are more likely to be present at the EPC conferences. Therefore, our study does not represent the entire demographers' networks.

To this end, this paper aims to summarise the overall trends in the number of papers, participants, their affiliations and gender by topics in the EPC over the 10 year period for which data are available; investigate the driving factors behind co-authorship on papers presented at the EPC; and explore the network of collaborations between demographic research centres via the co-authorship on papers presented at the EPC.

The remainder of the paper is organised as follows. The data section describes the data used for the analysis. The subsequent section descriptively presents the development and nature of the EPC over the ten-year period 2006-2016. This is followed by the analysis of the co-authorship patterns. First, we look at which factors best explain the probability of a paper to have multiple authors and second which factors best explain the probability of multiple-authored papers to involve multiple institutions. In the penultimate section, we look at the network structure based on an analysis of papers involving multiple institutions via a series of centrality measures. In the final section we discuss the findings and conclude.

2. Data

With assistance from the EAPS, we obtained the database of metadata for papers presented at the European Population Conferences (EPC) for the years 2006, 2008, 2010, 2012, 2014 and

2016. The data are maintained by Pampa 4.1, hosted by Princeton University, and were supplied to us in the format of an electronic database. For each paper accepted (both for oral and poster presentations), we have information about the authors' names and institutional affiliations, country where the author is based, their email address, co-authors' names and affiliations, title and abstract of the paper, session under which the paper was presented, and the broader theme under which the paper was submitted. The authors' names are anonymised when performing data analysis.

Although the data set came in an electronic format, it required substantial data cleaning. We identified and addressed a number of inconsistencies related to 1) author names across time (e.g., first name and last name were entered in a different order, some authors changed their first or last name, abbreviations rather than full names were entered); and 2) institution names (e.g., institution names were entered in a multitude of different ways, for example using full names, abbreviations, combinations of both or misspelt). Problematic cases were identified: in those situations, author and institution names were harmonised manually. In cases where the authors have more than one affiliation, we selected the first affiliation that the authors indicated in the electronic submission system.

Furthermore, there were also some issues with missing data. Not all authors indicated their affiliation and country where they were based when the paper was submitted to the EPC. The default country in the Pampa abstract submission system was the US (United States). When the information on authors' affiliation was not available, we looked up their affiliations on the Internet based on their names or we derived it from the domains of their email address. A country was then identified based on the location of the institution. In cases where the US was inserted by default and the institution was not located in the US, we also identified a correct corresponding country.

Some information to be used in the analysis such as the city where the institution is located or gender of the author were not directly available in the data. To identify a city, we used the Google Maps API via the *ggmap* package in R (Kahle and Wickham 2013) and geo-coordinates of the institution location. In the cases where the Google Maps API was unable to locate the city of the institution, we manually looked up the location using an Internet search. To identify the gender of the authors, we used the genderize.io API via the *gender* package in R (Mullen et al. 2015). The most probable gender of the authors was identified from the authors' names, based on a dataset of names of user profiles across major social networks. For many non-English names, no probability was available due to a lack of observations in the gernderize.io database. In this case, we assigned gender upon routinely checking with different sources to ensure accuracy.

Upon dealing with missing data and information, a panel dataset was constructed based on author names (n= 3,999). This allowed us to identify the frequency of the author having his/her paper accepted in the EPC. Across the ten-year period of six EPCs, the data covered 2,751 papers presented in the oral sessions, 1,445 posters and 251 sessions in total.

In addition, each paper is classified into different topics in demographic research. Teachman (1993) and Krapf et al. (2016) relied on content analysis and a subjective classification system, respectively, in categorising each paper into different subfields of demography. In our case, the papers submitted to the EPC were already self-assigned to a theme: at the time of submission, the authors have to decide in which theme they would like their paper to be considered out of a list of topics available on the submission form. We thus classify each paper into different topics based on the theme to which the author chose to submit his/her paper. However, the choice of themes has changed over time. In total, there were 31 themes across the six EPCs. Based on similarities of the research area as well as the frequency distribution of the number of papers (See Appendix A), we reclassified the 31 themes of oral presentations into seven topics of research: ageing, health and mortality; data and methods; economic and policy issues; fertility and family; history, development and environment; migration; and life course. Papers presented in a poster session are classified as 'poster sessions'. Although it is possible to use key words from the abstract to classify papers in the poster sessions into one of the seven topics, such method would not account for the different nature of the selection into an oral session versus a poster presentation.

3. Development of the European Population Conference

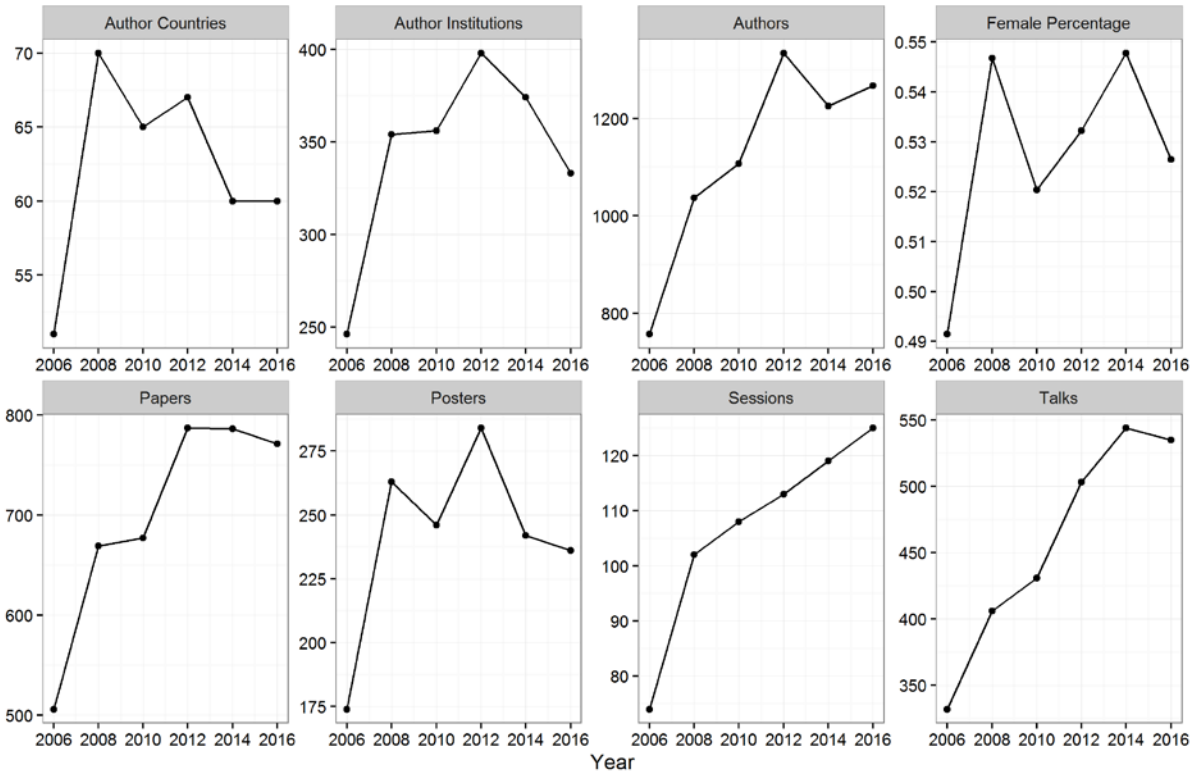
This section describes patterns and trends of the European Population Conference between 2006 and 2016. Figure 1 presents summary statistics of the number of papers presented[1], persons registered as authors, countries of authors' institutions, institutions to which authors have the primary affiliation, posters presented, sessions and oral presentations at each EPC between 2006 and 2016. It should be noted that 'persons' is measured by the number of authors of the papers presented and, as such, it may not reflect the number of people who actually attended the conference.

In the past 10 years, the EPC has become larger as measured by the number of sessions organised, papers presented and authors. The number of authors has increased over the decade considered from below 800 in 2006 to more than 1,200 since 2012. In 2006 around 500 papers were presented in less than 80 sessions. These numbers went up to 800 papers presented in more than 110 sessions in 2012. While the number of papers remained at this level in the later years, the number of sessions has continued to rise to 130 in 2016. This reflects an increase in the diversity of topics studied in demographic research. Indeed, the improvement in (demographic, geographic and socioeconomic) data availability coupled with advancement in statistical techniques and computing power have enabled population researchers to investigate new areas and paradigms such as life course (Ritschard and Oris 2005), multilevel analysis (Bijak et al. 2014), spatial demography (Matthews and Parker

[1] The sum of number of posters and number of talks equals the number of papers.

2013) and climate-induced migration and vulnerability to climate change (Fussell et al. 2014; Héran 2013; Hunter and Menken 2016; Muttarak et al. 2016).

Figure 1: Summary statistics of EPC conferences 2006-2016: number of countries where authors' institutions are, institutions to which authors have the primary affiliation, authors, papers, posters, sessions and oral presentations (i.e., talks); and percentage of female authors over time.



Note: The sum of number of posters and number of talks equals the number of papers.

The number of countries where researchers are affiliated increased sharply from 50 in 2006 to 70 in 2008 and remained between 60 and 70 from 2010 onwards. Authors from 250 institutions had papers accepted for presentation in 2006 and this number has grown to 400 in 6 years then slightly decrease to 330 in 2016.

Except for 2006, in general the share of female authors has been slightly higher than that of male authors. In fact, female representation in demography is quite prominent. Not only has the number of women attending graduate school and obtaining doctorate in demography been increasing, but the number of publications (at least in the journal *Demography*) involving a female author has also risen (Krapf et al. 2016; Teachman et al. 1993). Compared to other scientific disciplines in both natural and social sciences,

demography occupies the second top rank place with respect to female composition in authorships in the JSTOR corpus between 1990-2011: the proportion of female authors was 46.4% in education, 41.9% in demography, 22.8% in ecology and evolution, 13.7% in economics and 10.6% in mathematics (West et al. 2013).

However, when breaking down by topics of demographic research, gender differentials emerge. Figure 2 shows that the proportion of female authors. The darker the purple colour is in the topic (row) and year (column) considered, the higher the share of male authors. On the contrary, the darker the green colour, the more female authors there are in the topic and year considered. Evidently women are under-represented in 'history, development and environment' and 'data and methods', where the female proportion is below 25% in 2016. Similarly, the study of gender difference by research topics based on papers presented at the IUSSP conferences also shows that the topic related to data and methods has higher share of male first authors (Ediev et al. 2009).

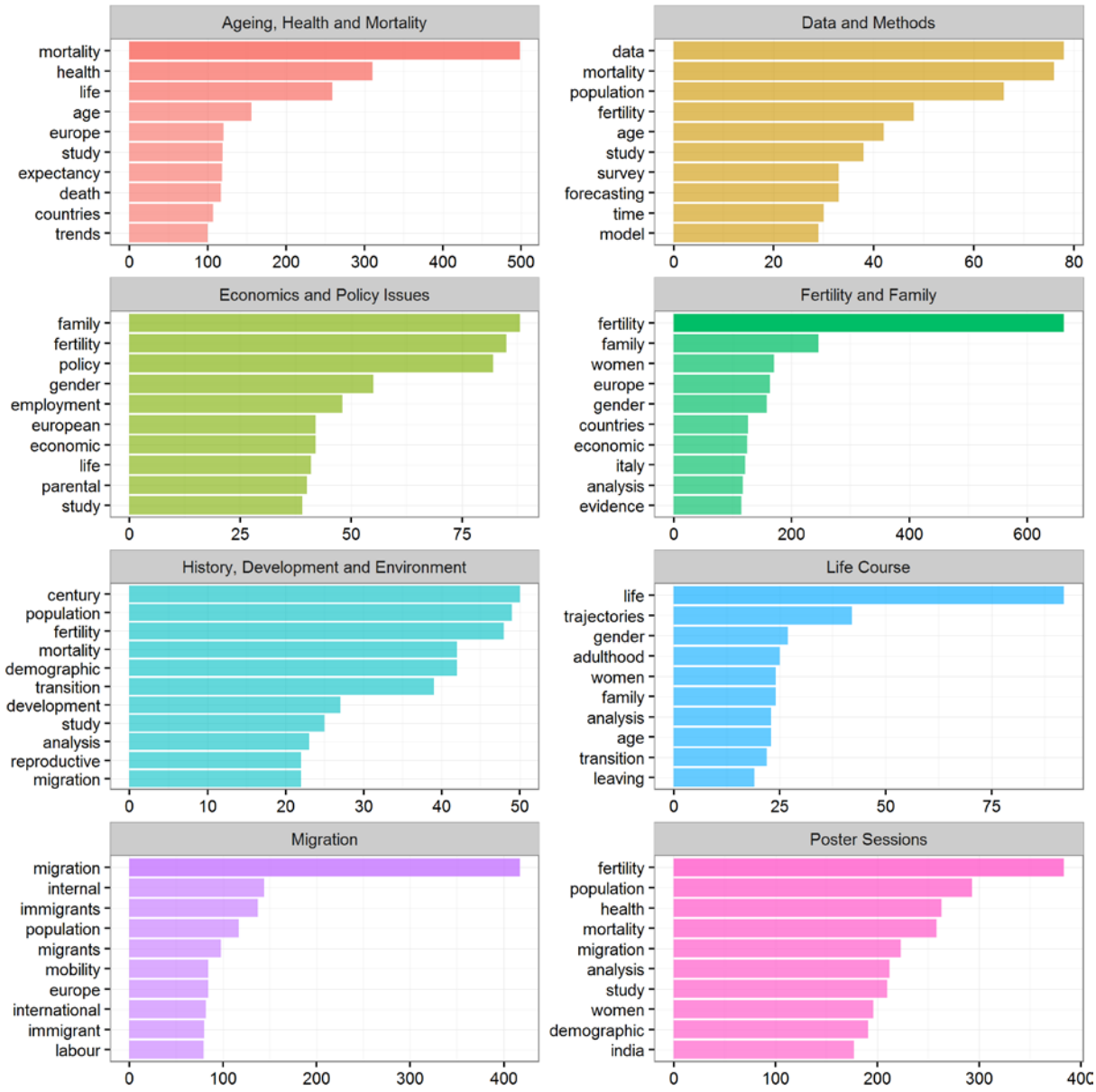
For other topics such as 'ageing, health and mortality', 'migration' and 'economics and policy issues', the proportion of women is slightly higher than that of men (except for the year 2006 for the latter two categories). Krapf et al. (2016), however, found that men are significantly more likely than women to have published papers in the topics 'migration' and 'mortality and health'. The discrepancy in the findings might be due to differences in the classification of papers as well as to the different nature of publications in peer-review journals and conference presentations.

The topic 'fertility and family', especially in 2014 and 2016, shows an exceptionally high proportion of female authors – over 75%. This is in line with Teachman et al. (1993) and Krapf et al. (2016) who reported a similar finding. Although they distinguished between 'fertility' and 'family', both fields display a high prevalence of female authors.

Figure 2: Proportion of female authors by topics

It is also possible to scrutinise in more details the content of the research undertaken in each topic. Figure 3 identifies the ten most commonly used words in the titles for each of the seven topics and in the poster session. Generally, these words reflect the nature of the study which appears to correspond well with the topic where the paper was presented. For instance, 500 papers in the topic 'ageing, health and mortality' contain the word 'mortality' and over 100 papers contain the word 'death' in their abstract. Surprisingly, the word 'ageing' is not among the top ten in this topic while, as expected, the most commonly used word in the theme 'migration' is 'migration' itself.

Figure 3: Top ten most used words in the titles of the paper by topics



Likewise, the top ten commonly used words represent the content of the research focus of each topic. For the topic 'life course', for instance, among the top ten words used in the titles there are 'trajectories', 'adulthood', 'family', and 'leaving', suggesting that these studies mainly focus on transition to adulthood, leaving parental home, family formation, and trajectories into various stages of life. For the topic 'migration', the papers presented include both the words internal and international, though the former is more often mentioned. This hints to a larger amount of studies presented in the subfield of internal

migration. The word 'mobility' is also widely used in the papers presented in this topic, likely as an alternative term to 'migration' (Aybek et al. 2015).

Fertility and/or family are extensively studied across all topics except for 'ageing, health and mortality' and 'migration'. In particular, fertility tops the most commonly used words in the poster sessions accounting for almost 400 papers. Similarly, over 600 papers in the 'fertility and family' topic contain the word 'fertility' in their title. Fertility is also a well-presented research field in the topic 'history, development and environment'. This suggests that fertility remains a prominent research topic in demographic research. Second to fertility, the term 'mortality' commonly appears in the paper titles across all research topics. About 500 papers in the 'ageing, health and mortality' topic and 250 papers in the poster sessions focus on mortality research. The descriptive statistics also suggest that the topic 'ageing, health and mortality' has been, so far, dominated by research on mortality rather than on ageing. Since the number of sessions on intergenerational relationships and care within the family in later life is increasing, future analyses on the EPC data may be able to consider the two subfields ('ageing' and 'health and mortality') separately. Migration is the third most studied topic with over 400 papers in the 'migration' topic and about 200 papers in the poster sessions. With fertility, mortality and migration being fundamental components of population change, no doubt they are core research topics in population studies (Bijak et al. 2014).

Meanwhile, some topics remain understudied in demography. Despite the increase in availability of environmental and spatial data allowing for the integration of environmental issues within population research (Hunter and Menken 2016), it seems that this topic has not yet gained popularity in the EPC. In fact, even within the topic 'history, development and environment, the terms 'environment' or 'climate change' have not appeared as the ten most commonly used words at any EPCs. Also referring to the annual meeting of the Population Association of America in 2015, McDonald (2016) pointed that the papers presented in the sessions containing the word 'environment' were mainly related to localised environmental issues. The study related to global climate change was clearly underrepresented.

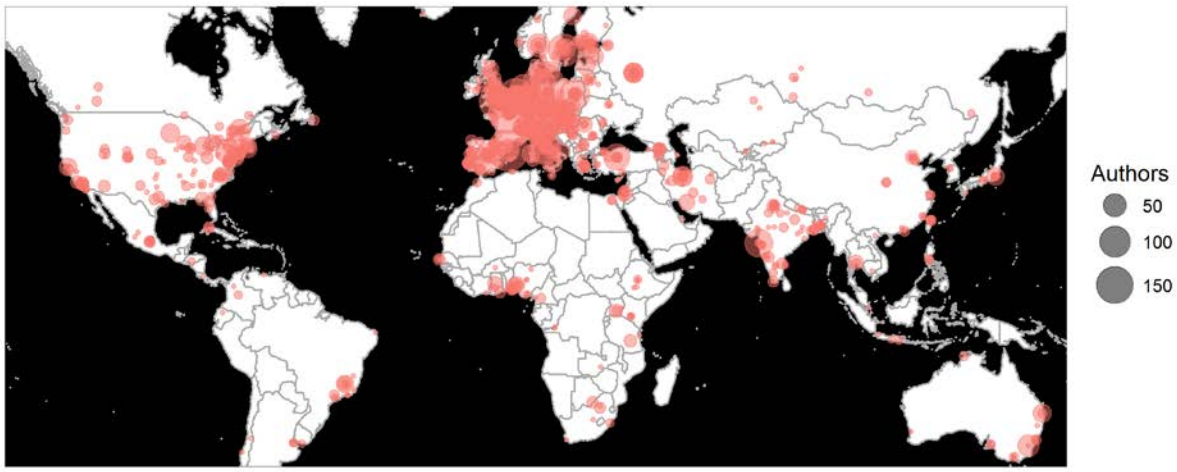
Another aspect that can be inferred from Figure 3 is the geographical coverage of the papers presented at the EPCs. Unsurprisingly, 'Europe' appears amongst the top ten most used words in the titles in most research topics. The word 'countries' appears in the abstract of about 100 papers in 'ageing, health and mortality' suggesting that cross-national or comparative studies are common in this topic. Interestingly, over 100 papers in 'fertility and family' contain the word 'Italy' in their title. This reflects the relevance of Italy in studies on fertility and family due to its low fertility and strong family norms. However, it may also reflect the relatively high representation of demographers from Italian research institutes as compared to other nations (see Figure 5). We cannot account for the country of origin of the researchers, but it is plausible that the number of Italian researchers in demographic institutes around Europe is also quite large. Similarly, almost 200 papers in the poster sessions contain the word 'India'. Indeed, amongst non-European institutions, authors from Indian institutions are highly represented, second only to the United States (see Figure 4).

Note however that papers from the authors based in India were mainly presented in the poster sessions.

Next, we examine geographical locations of the authors presenting papers at the EPCs. Figures 4 and 5 present the geographical distribution of authors by cities where their institutions are located in the world and in Europe, respectively. In particular, the larger the size of the dot on a certain point in the map, the higher the number of researchers having their primary affiliation in the city indicated. Figure 4 shows that the vast majority of the authors at the EPCs are based in European institutions. The non-European countries with the highest number of authors presenting at the EPCs are the United States followed by India, Turkey and Australia. Few authors from Central and South America and Africa are concentrated in Brazil and Mexico for the former and Nigeria, South Africa, Tanzania and Uganda for the latter.

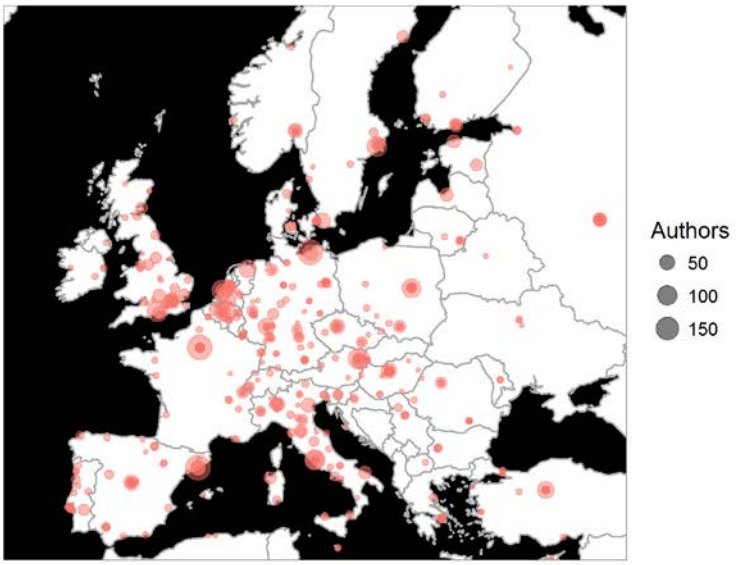
Figure 5 focuses on the distribution of authors in Europe. We note that in Italy and Germany, and to a certain extent in the UK as well as smaller countries like Belgium and the Netherlands, the distribution of authors is widespread across the country. However, there is a clear concentration of authors in large demographic centres such as the Max Planck Institute for Demographic Research (MPIDR) in Rostock and the Centre for Population Change (CPC) in Southampton. For many countries, such as France, Spain, Austria, Hungary and Sweden, we observe that most authors are concentrated in a large city where prominent demographic centres are located. This includes the National Institute for Demographic Studies (INED) in Paris, the Centre for Demographic Studies (CED) in Barcelona, the Vienna Institute of Demography (VID) in Vienna, the Hungarian Demographic Research Institute (HDRI) in Budapest and the Stockholm University Demography Unit (SUDA) in Stockholm.

Figure 4: Locations of authors: global



Note: The size of the circle symbol reflects the number of authors from a particular location.

Figure 5: Locations of authors: Europe



Note: The size of the circle symbol reflects the number of authors from a particular location.

4. Determinants of Collaborative Authorship

In this section, we investigate the drivers of collaboration amongst demographers based on their co-authorship of papers presented over the previous six EPCs. This analysis is carried out in two steps. First, we examine the factors relating to whether a paper involves a single or multiple authors. Second, we explore the driving forces behind multiple authorship that involves co-authors from more than one institution.

4.1 Drivers of Scientific Collaboration

As an initial step to investigate the collaboration of demographers at the EPC, we explore which papers involve collaboration (i.e., papers that have multiple authors) and which papers have a single author. In order to do so, we created a large dataset combining all the Pampa data from all EPCs between 2006 and 2016. Each row in the data was based on an author-paper combination. For instance, papers that involved a single author were assigned only one row. For papers that involved multiple authors, one row for each author on the paper was created. The same author might be present in multiple rows if they are involved in more than one paper. Both the papers presented as a talk and those presented as a poster were considered as a paper in this analysis. The final data frame contained 9,183 author-paper observations (rows). For each observation we created a dummy outcome variable to be used in a logistic regression model, where the author-paper combination was coded as either single- (= 0) or multiple-author (= 1), based on the number of rows in which the same paper was present in the dataset.

Potential explanatory factors were categorised into three areas. The first includes characteristics of the paper: the year of the conference in which the paper was presented at and the topic in which the paper was presented. The second includes characteristics of the author: gender, whether affiliated to a European institution, the number of other papers the author was involved in at the same conference, and the number of colleagues from the same institution who were authors of papers presented at the same conference. The third set of variables also relates to the author, but involves using the harmonized ID [2] of each individual to look at their involvement at the previous EPCs. These include the number of affiliations of the author at the current and past EPCs, the number of past EPCs attended, and a dichotomous variable indicating if the author attended the previous EPC.

From the complete dataset, 60 cases were dropped due to missingness in one or more of the created covariates. In order to select which covariates to include in a logistic model where the outcome variable is the probability of a multiple authorship, we used a Bayesian model averaging approach as implemented in the *BMA* package in R (Raftery et al. 2015). The model averaging approach considers all regression models corresponding to subsets of the covariates described above. As we are unsure of the best model, the model uncertainty inherent in our variable selection problem is dealt with by averaging over the best models in the model class. The best models are based on the evidence in the data in each model, which is used to compute the posterior probability of each covariate being in the model (i.e. the probability that the variable's regression coefficient is non-zero). Unlike classical model selection approaches such as likelihood-ratio tests of nested models, a Bayesian model approach is more adept in handling model uncertainty and providing robust results to alternative model specifications (Montgomery and Nyhan 2010). We assumed equal prior weights for each variable in the model.

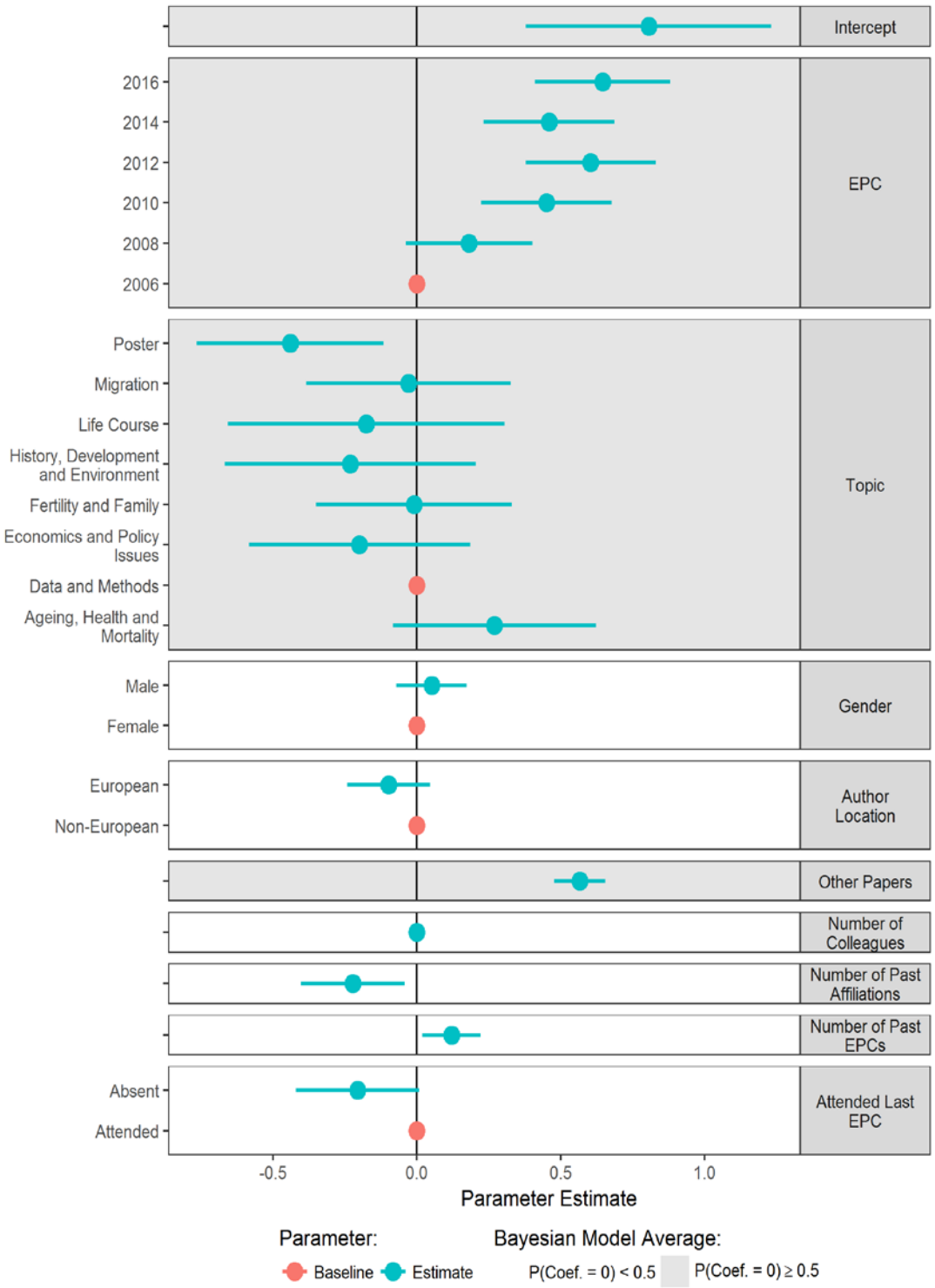
4.2 Probability of Co-Authorship

We begin with the estimation of the probability of multiple authorship. Figure 6 shows the parameter estimates from the full model containing all covariate measures. The uncertainty in the parameter effects is illustrated using the error bars, which represent 1.96 times the standard error. For categorical covariates, there is no uncertainty in the baseline category, where the parameter effect is also plotted and set to zero. No y-axis labels (i.e., labels on the left side of the graph) are provided for continuous covariates, which can be identified from labelled panel on the right side of the graph.

Figure 6 shows a significant effect of the year of EPC on the likelihood of co-authorship. Co-authorship is more likely in more recent EPCs as compared to the 2006 conference. There is no effect of topic, however poster presentations are more likely to be single-authored papers (in particular, this effect is statistically significant as compared to talks in the topic 'data and methods', which is the reference category).

[2] The same author who has attended the EPC more than once is identified through their first and last name and a unique ID is created to link the same author.

Figure 6: Logistic regression estimates of the probability of multiple authorship (coefficients and 95% CI)



Males tend to be more likely than females to co-author papers presented at the EPC, however the parameter in the model is not statistically significant (at $p<0.05$), as shown by the CI including zero in Figure 6.

Authors from European institutions are less likely than those from non-European affiliations to present co-authored papers, although the coefficient is not statistically significant at $p<0.05$. If authors have more papers at the same EPC, it is more likely that they present co-authored papers, while the number of colleagues from the same institution presenting at EPC does not affect their likelihood to present co-authored papers.

The higher the number of past affiliations, the less likely it is to be involved in a co-authored paper as compared to presenting a single-author paper. The number of past EPCs attended and whether the author has attended the last EPC seem to positively affect the likelihood of presenting co-authored papers as compared to single-author papers.

The posterior inclusion probabilities for the covariates are indicated on Figure 6 via the background shading of the panel. The covariates with posterior probabilities greater than 0.5 include: the EPC year, the topic of the paper and the number of other papers presented at the same EPC by the author. This suggests that there is strong evidence in the data for each of these covariates as predictors of multiple authorship. The 0.5 inclusion probability is chosen as Barbieri and Berger (2004) showed that the single regression model with the predictors whose posterior inclusion probabilities are above 50% is predictively optimal.

4.3 Probability of Co-Authorship Across Institutions

In a secondary step, to investigate the collaboration network of demographers at the EPCs, we filtered the set from the previous section to consider only the papers that involved multiple authors. This reduced the dataset to 7,817 author-paper combinations. A new dichotomous outcome variable, to be used in a logistic regression model, was created to indicate if the multiple-authored paper included authors from the same institution (= 0) or from multiple institutions (= 1). The covariates used in the previous section were expanded to include additional measures on the gender composition of the authors in the paper (i.e., mixed, all males or all females) and the number of co-authors each author has in each paper. Figure 7 shows the parameter estimates from the full model containing all covariate measures.

The year at which the co-authored paper was presented does not affect the likelihood of the authors being from multiple institutions, with the exception of 2016, when it was less likely to present papers with multiple authors from multiple institutions as compared to 2006. However, the topic shows a statistically significant effect. In particular, papers with multiple authors presented in 'ageing, health and mortality' are more likely to have co-authors from multiple institutions in comparison to the reference category, which are papers presented as posters. On the contrary, papers in the topic 'migration' are less likely to be co-authored across multiple institutions than papers presented in almost all the other

subfields (with the exception of 'history, development and environment' and 'data and methods' for which the CIs partly overlap).

With one person increases in the number of co-authors, the log odds of co-authoring with authors from other institutions increases by more than 50%.

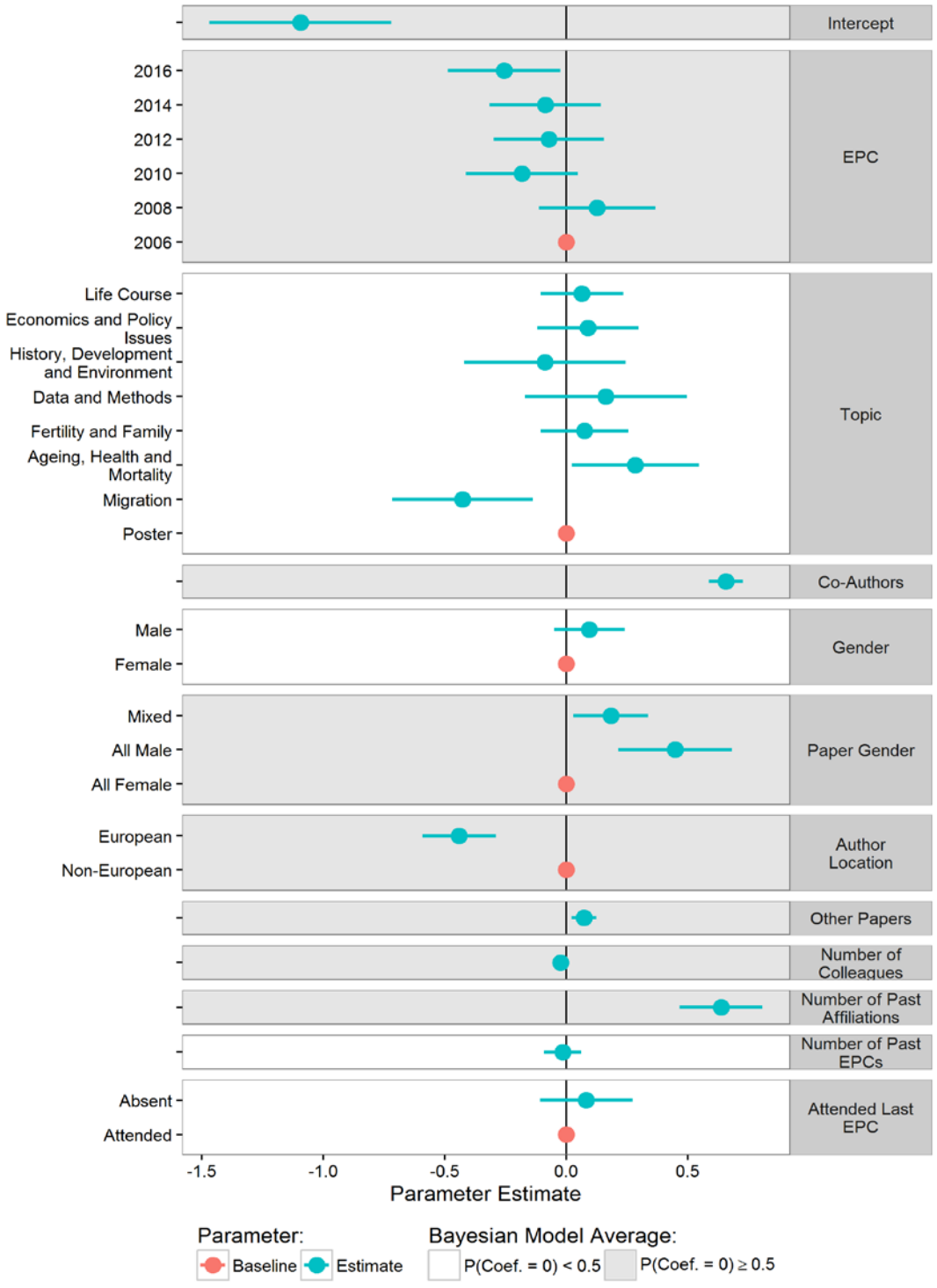
There is no gender difference in the likelihood of multiple institutions collaborating on co-authored papers. However, if the co-authored paper is mixed in terms of gender and even more so if the co-authors are all males, the likelihood of co-authoring across institutions is higher than in the case of all female co-authors.

A European researcher is less likely than non-European ones to be involved in papers with multiple authors from multiple institutions. This may partially be the result of a selection process whereby non-European co-authors are more likely to submit to the EPC if they have European co-authors.

Authors of more than one paper presented at the same EPC as well as those who have had several affiliations in the previous years have a higher likelihood to collaborate across institutions. However, the CIs including zero show no statistically significant effect for the number of colleagues from the same institution presenting at the same EPC, number of the past EPCs attended and whether the researcher has attended the last EPC on his/her likelihood to have a co-authored paper with colleagues from other institutions.

The posterior inclusion probabilities for the covariates are indicated via the background shading of the panel. Seven covariates all had covariates with posterior probabilities greater than 0.5: the EPC year, the topic of the paper and the number of co-authors on the paper, the gender makeup of the authorship team, if the author's institution is based in Europe or not, the number of other papers of the author, the number of colleagues from the same institution also attending the EPC and the number of past affiliations. The high model averaged probabilities for each of these covariates suggest a strong posterior support for their inclusion in a predictive model for multiple institution papers.

Figure 7: Logistic regression estimates of the probability of co-authorship with authors from other institutions (coefficients and 95% CI)



5. Collaboration Network Structure

5.1 Collaboration Networks Visualised

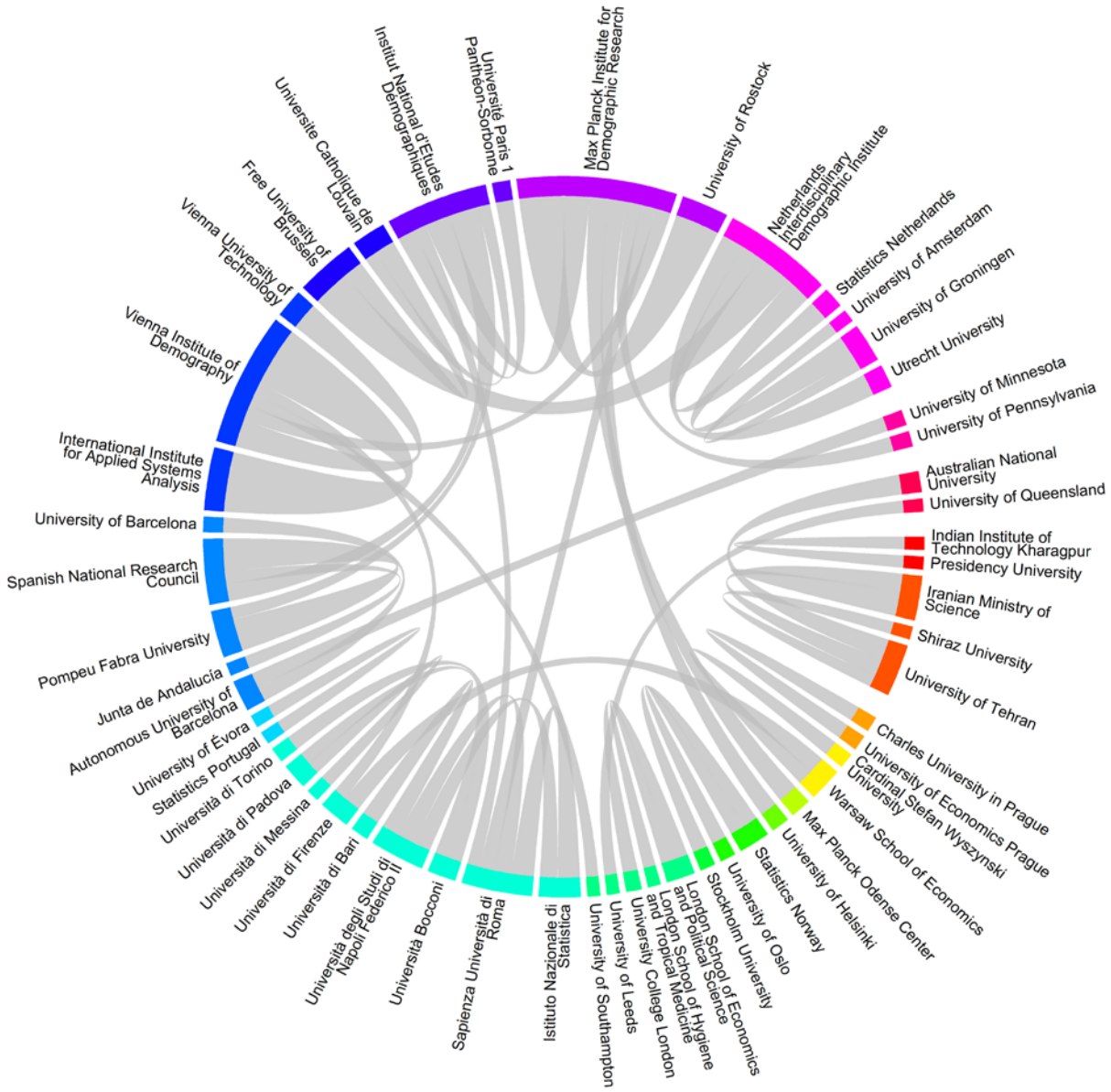
In this section we investigate the networks of collaborations across institutions from papers that involved multiple authors. We first visualise the collaboration networks across institutions to investigate where the strongest linkages are and subsequently employ centrality measures of network analysis to identify important institutes within the collaboration network structure.

In order to obtain a better understanding of the inter-relations between some of the major institutions represented at the EPCs, we used the *circlize* package in R to produce a chord diagram plot shown in Figure 8 (Gu 2016). The plot contains institutions sharing authorship of five or more papers over the ten year period considered. They are ordered alphabetically within United Nations regions. Institutions in the same country share the same colour of their outer sector, although with different shades. The size of the chord at its base represents the number of papers connecting the two institutions. The size of the chord elsewhere is scaled so as to minimize visual clutter.

The largest co-authorship links between any two institutions connect the VID and the International Institute for Applied Systems Analysis (IIASA) (24 papers), the MPIDR and the University of Rostock (18), and the Free University of Brussels and the Netherlands Interdisciplinary Demographic Institute (NIDI) (15). The largest collaboration within the same country is between the VID and IIASA. This strong collaboration between VID and IIASA reflects a particularly close link between the two institutes, especially after the founding of the Wittgenstein Centre for Demography and Global Human Capital (WIC) in 2010 where the VID and IIASA are two of the three core pillars of WIC (Goujon et al. 2015).

The majority of the chords in the plot are connecting institutions in the same country and hence do not pass through the centre of the circle. However, international collaborations such as between the INED and the University Pompeu Fabra, the MPIDR and the University of Helsinki or the Max Planck Odense Center, and between the Warsaw School of Economics and the University of Florence are not uncommon. Some of the institutions outside Europe have strong collaboration links with those in Europe such as the University of Minnesota with the Autonomous University of Barcelona, and the University of Queensland with the University of Leeds. Institutions in Iran and India tend to have strong collaborations within their country but use the EPC as a venue to present their research. Collaboration across institutions located in different countries is the largest between the Free University of Brussels and NIDI. Meanwhile, MPIDR and INED have cross-institute collaborations with many different institutes.

Figure 8: Co-authorship across institutes (only institutes with at least five papers together are presented)



Note: Colour on the outer sector indicates a country where the institution is located.

5.2 The Network Matrix

In order to gain a further insight into the collaboration network of demographers at the EPC, we constructed a (symmetric) square matrix of the counts of papers involving collaboration between any two institution pairs. Our count was based on the number of distinct

combinations between any two institutions on each multiple author paper over all EPCs. For example, consider a paper that involved three authors. If all authors were from different institutions, three distinct institution pairs would be created. If two authors were from the same institution, then only two institution pairs would be created (the pair between the two distinct institutions and one pair that would account for the collaboration within the same institution). If the paper involved three authors from same institution, then only one institution pair would be created (the pair between any of the two authors from the same institution).

The network matrix was used to calculate a number of network centrality measures and determine the most important institutions based on different centrality measures involved in the EPC. The results of the rankings of the institutions by the five different centrality measures (degree, closeness, betweeness, Eigen and subgraph) are shown in Table 1. The ordering of the rows is determined by the degree centrality measure. Only institutions that are ranked in the top ten of at least one of the centrality measures are shown. Computations were performed using the *igraph* package in R (Csardi 2015). Below, we explain the five centrally measures used to create the institutions' rankings.

Table 1: Centrality measure of collaborative network structure amongst institutes

Institute	Centrality Measure				
	degree	closeness	betweeness	Eigen	subgraph
Institut National d'Etudes Démographiques (INED)	1	4	2	1	1
Max Planck Institute for Demographic Research	2	8	3	2	2
Vienna Institute of Demography (VID)	3	1	1	4	3
Stockholm University	4	2	4	10	4
Sapienza Università di Roma	5	5	7	8	6
Autonomous University of Barcelona	6	11	8	3	16
University of Southampton	7	12	10	5	5
Université Catholique de Louvain	8	3	6	9	12
Università Bocconi	9	33	47	14	7
Netherlands Interdisciplinary Demographic Institute (NIDI)	10	28	19	7	9
Pompeu Fabra University	10	39	21	13	10
International Institute for Applied Systems Analysis (IIASA)	11	9	5	11	19
Statistics Norway	15	10	20	17	11
University of Rostock	16	6	11	6	14
Hungarian Demographic Research Institute (HDRI)	20	7	13	38	23
University of Tehran	20	168	9	181	206

Degree centrality ranks the institutions by their number of multiple author papers. Those with the highest scores are considered the most central to the network. It is calculated as the sum of the combined row and column totals of our collaboration matrix. The top ranking institute according to degree centrality is the INED, with 95 multiple author papers over the last 10 years' EPCs. Second in the ranking is the MPIDR (with 73 multiple author papers) and the VID scores as third (72).

Closeness centrality quantifies how central or peripheral an institution is. While degree centrality simply captures the number of collaborations, closeness centrality identifies a central institution that is close, on average, to other institutions. An important institution is typically 'close' to and has multiple collaborative papers with the other institutions in the network. Based on the ranking of the closeness measures, VID is the top ranking institute, followed by Stockholm University and the Catholic University of Louvain. Other institutions that were outside the top ten rankings based on the degree centrality measure become more central under the closeness measure, including IIASA, University of Rostock, HDRI and Statistics Norway. The higher rankings of these institutions once closeness centrality is used are due to their high number of collaborations with other central institutions.

Betweeness centrality quantifies the number of times an institution acts as a bridge along the shortest path between two other institutions, where there is no direct co-authorship link. For instance, Institute B lies between Institute A and Institute C and there are no alternate paths for Institute A and Institute C to be connected except through Institute B. Based on the rank of betweeness centrality, the VID is again the highest ranking institute, followed by the INED and the MPIDR. The University of Tehran has a particularly high betweeness ranking (ninth) compared to other centrality measures. Their researchers typically work with both researchers from other institutes in Iran and outside Iran on the same paper.

Eigen centrality measures the influence of an institution in the collaboration network. An institution will have a higher score if it is connected to other important neighbours. The INED has the highest Eigen centrality measure, followed by the MPIDR and the Autonomous University of Barcelona. Both NIDI and the University of Rostock do particularly better under the Eigen centrality measure than in some of the other measures.

Subgraph centrality is a measure based on the participation of each institution in all subgraphs with in a network, where a subgraph is a smaller subset based on a collection of institutions with strong links between them. Smaller subgraphs are given more weight than larger ones in the subgraph centrality measure, and hence those institutions with a wide range of collaborations are ranked higher. The ranking of the top four follows that of the degree centrality. The Autonomous University of Barcelona and the Catholic University of Louvain have lower rankings under the subgraph measure than in other measures.

6. Discussion and Conclusion

In this study, we aimed at improving our understanding of the discipline of demography by using a novel data set that provides a unique window into the network of collaborations among demographers. We evaluated ties among institutions, at six European Population Conferences over the course of a decade. We believe that a scientific reflection on the discipline is not only an academic exercise. It is informative about current trends and about the likely future of demography and of the institutions that foster population research. Our analysis also provides a mirror of societal concerns that translate into basic and applied population research. The results presented thus are relevant not only for demographers and the scientific community that focuses on the 'science of science', but also for policy makers and the institutions that connect policy analysis with fundamental research.

We observed that the level of engagement at European Population Conferences has been increasing steadily between 2006 and 2016. More papers are presented, more authors contributed to the papers presented at the conferences and a more diverse set of topics are represented. This is a sign that a robust community is growing and that the European Association for Population Studies in gaining prominence on a global stage.

Compared to other scientific disciplines, women are indeed well represented in demography (Merchant 2015; West et al. 2013). For the large majority of sub-fields within demography, women are either equally represented as men or over-represented. However, women are under-represented in areas like 'data and methods'. This is a fact worth considering because most of the top posts in demographic institutions are often occupied by formal or statistical demographers, who are often men. Although overall trends in women representation in demography is a welcome sign of gender equality, the composition by area of research indicates that there is still a disadvantage for women in the core areas of formal methods. Importantly, there has not been a trend towards increased female participation in the areas of 'data and methods' over the course of the past decade.

Furthermore, multiple authorship has increased in the past decade but not necessarily collaborations across institutions. Although there is no gender difference in the likelihood of having multiple authors in a paper, men are more likely than women to collaborate with researchers from other institutions. This may reflect the fact that male researchers are more mobile than female researchers since family obligation typically restrict duration of international collaboration and geographic extension for females (Abramo et al. 2013). Indeed, it was found that the presence of children and a working husband reduce international mobility of female academics (Shauman and Xie 1996). Correspondingly, the authors with higher number of past affiliations – a proxy for mobility – also have higher probability of collaborating across institutions.

Key institutions play a central role in driving collaboration in demographic research in Europe. Some notable examples include INED, NIDI, MPIDR and VID. This is consistent with anecdotal evidence. Interestingly, there are a number of less obvious institutions that rank well in certain specific measures of our network analysis. The University of Tehran is

an example of an institution that may be well positioned to create bridges between researchers, although it is not a central actor in the European landscape of demographic research. We believe that one of the merits of our study is that it provides quantitative evidence that goes beyond anecdotes and reveals some unexpected patterns.

This article not only offers insights about demographic research in Europe. It is also an example itself of an emerging approach to demographic research and beyond. For this study, we used the so-called 're-purposed' data. The Pampa database was not created for research purposes. It was produced to meet the need of organising and managing information for large events like population conferences. We 're-purposed' the data to address the research questions of interest. This is a process similar to what many researchers across disciplines do with data like social media, mobile phone records, Web searches, blog posts, e-mails, etc. This developing area of research benefits from new data sources, that were unimaginable just a decade ago. But the data also come with limitations: typically, these data sets are messy, include inconsistencies and require pre-processing and data cleaning before they can be used for research. For this article we followed a modern workflow that reflects the 'data science revolution': the workflow includes data cleaning, advanced statistical approaches, tools borrowed from a number of disciplines (e.g., social network analysis), and modern visualisations. One of the interesting features of this data is that it is not a sample. It is a complete record of presentations and collaborations at the European Population Conferences over the course of a decade.

New, often re-purposed data, offer unprecedented opportunities for research, but there is more than that. The same technology that allows conference organisers to manage the flow of information also enables researchers to quickly get a sense of what other scholars are working on and potentially contact them. The data in this sense, are not neutral. In the field of social media data analysis, researchers use digital crumbs as data as well as they evaluate the implications of use of social media for social processes. In the context of our data, we believe that there are important opportunities to both use the data to understand the community of demographers, as well as to study the impact of the European Population Conferences on the discipline and beyond.

This paper is written against a backdrop of transformative changes in the way people collaborate and communicate in their countries as well as across borders. The extent to which collaborations develop across countries can be considered an indicator of European integration. We observed that, for the most part, co-authorship happens across institutions located in the same country. There are notable exceptions, especially in the context of the leading institutions. We did not observe structural patterns of isolation in the collaboration network, but we observed that we are still quite far from a model of integrated European collaboration. We think that it is relevant to monitor trends over time. Our data set could also be leveraged to study patterns of geographic mobility of researchers across institutions and countries. That information is particularly relevant in the context of understanding European integration.

In this article, we reflected about demography, demographers and internal dynamics of the discipline. In the future, we are interested in expanding our work in various directions, including the geographic and academic scope. For example, a comparative analysis with meta data from the Population Association of America Annual Meetings would provide insights into differences and similarities between the American and the European contexts. Similarly, by comparing our results with those obtained for conferences in other disciplines could offer new perspectives about similarities and differences of demography, with respect to other fields. For instance, evaluating the composition of demographers in terms of areas (e.g., statistics, sociology, geography, etc.) and collaborations, and how those metrics compare with the ones for other disciplines, could provide insights into similarities of paths across fields and inform our judgement about the likely future of demography as a discipline.

References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics*, *7*(4), 811–822. doi:10.1016/j.joi.2013.07.002

Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, *490*(7420), 335–336. doi:10.1038/490335a

Aybek, C. M., Huinink, J., & Muttarak, R. (2015). Migration, spatial mobility, and living arrangements: An introduction. In C. M. Aybek, J. Huinink, & R. Muttarak (Eds.), *Spatial Mobility, Migration, and Living Arrangements* (2015 edition., pp. 1–19). Cham: Springer. http://link.springer.com/book/10.1007%2F978-3-319-10021-0

Bäker, A. (2015). Non-tenured post-doctoral researchers' job mobility and research output: An analysis of the role of research discipline, department size, and coauthors. *Research Policy*, *44*(3), 634–650. doi:10.1016/j.respol.2014.12.012

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870–897. doi:10.1214/009053604000000238

Bijak, J., Courgeau, D., Silverman, E., & Franck, R. (2014). Quantifying paradigm change in demography. *Demographic Research*, *30*, 911–924. doi:10.4054/DemRes.2014.30.32

Csardi, G. (2015). *Network analysis and visualization*. https://cran.r-project.org/web/packages/igraph/igraph.pdf

Ediev, D. M., Goujon, A. V., Lutz, W., & Speringer, M. (2009). *From Vienna to Marrakech 50 Years of Independent IUSSP Conferences 1959-2009*. Vienna: Vienna Institute of Demography. https://www.oeaw.ac.at/fileadmin/subsites/Institute/VID/download/iusspbrochure.pdf. Accessed 1 December 2018

Fussell, E., Hunter, L. M., & Gray, C. L. (2014). Measuring the environmental dimensions of human migration: The demographer's toolkit. *Global Environmental Change: Human and Policy Dimensions*, *28*, 182–191. doi:10.1016/j.gloenvcha.2014.07.001

Glänzel, W., & Schubert, A. (2004). Analysing Scientific Networks Through Co-Authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 257–276). Springer Netherlands. doi:10.1007/1-4020-2755-9_12

Goujon, A., Fürnkranz-Prskawetz, A., & Eder, J. (2015). *40 years of the Vienna Institute of Demography 1975-2015. From an Austrian to a European to a Global Player*. Vienna: Vienna Institute of Demography. http://www.oeaw.ac.at/fileadmin/subsites/Institute/VID/PDF/Publications/diverse_Publications/VID_40years_Web_Final.pdf

Gu, Z. (2016). *Circular visualization*. https://github.com/jokergoo/circlize

Héran, F. (2013). Demography and its vocabulary over the centuries: a digital exploration. *Population and Societies*, *505*.

Hunter, L. M., & Menken, J. (2016). Will climate change shift demography's "normal science"? *Vienna Yearbook of Population Research*, *2015*(13), 23–28. doi:10.1553/populationyearbook2015s23

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161.

Krapf, S., Kreyenfeld, M., & Wolf, K. (2016). Gendered Authorship and Demographic Research: An Analysis of 50 Years of Demography. *Demography*. doi:10.1007/s13524-016-0482-x

Larivière, V., Gingras, Y., & Archambault, É. (2013). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, *68*(3), 519–533. doi:10.1007/s11192-006-0127-8

Matthews, S., & Parker, D. M. (2013). Progress in Spatial Demography. *Demographic Research*, *S13*(10), 271–312. doi:10.4054/DemRes.2013.28.10

McDonald, P. (2016). Engagement of demographers in environmental issues from a historical perspective. *Vienna Yearbook of Population Research 2015*, *13*, 15–17.

Merchant, E. K. (2015). *Prediction and Control — Global Population, Population Science, and Population Politics in the Twentieth Century*. University of Michigan, Ann Arbor, MI. Retrieved from http://www.emilyklancher.com/digdemog/tmod/topicmod.html

Montgomery, J. M., & Nyhan, B. (2010). Bayesian Model Averaging: Theoretical Developments and Practical Applications. *Political Analysis*, *18*(2), 245–270. doi:10.1093/pan/mpq001

Mullen, L., Blevins, C., & Schmidt, B. (2015). *Predict Gender from Names Using Historical Data*. https://github.com/ropensci/gender

Muttarak, R., Lutz, W., & Jiang, L. (2016). What can demographers contribute to the study of vulnerability? *Vienna Yearbook of Population Research 2015*, *13*, 1–13. doi:10.1553/populationyearbook2015s1

Raftery, A. E., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2015). *BMA: Bayesian Model Averaging*. ftp://cran.r-project.org/pub/R/web/packages/BMA/BMA.pdf

Ritschard, G., & Oris, M. (2005). Life Course Data In Demography And Social Sciences: Statistical And Data-Mining Approaches. *Advances in Life Course Research*, *10*, 283–314. doi:10.1016/S1040-2608(05)10011-2

Shauman, K. A., & Xie, Y. (1996). Geographic mobility of scientists: sex differences and family constraints. *Demography*, *33*(4), 455–468.

Teachman, J. D., Paasch, K., & Carver, K. P. (1993). Thirty Years of Demography. *Demography*, *30*(4), 523–532. doi:10.2307/2061804

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The Role of Gender in Scholarly Authorship. *PLOS ONE*, *8*(7), e66212. doi:10.1371/journal.pone.0066212

VIENNA INSTITUTE OF DEMOGRAPHY

Working Papers

Nitsche, Natalie, *Partners' Educational Pairings, Work Divisions, and Fertility: Evidence from Germany*, VID Working Paper 19/2017.

Spahl, Wanda, Sabine Weiss, Judith Kohlenberger and Isabella Buber-Ennser, *Immigration and the Social Welfare State in Austria, Germany, and Switzerland: A Comparative Meta-Study*, VID Working Paper 18/2017.

Hoffmann, Roman, *Following the Peers: The Role of Social Networks for Health Care Utilization in the Philippines*, VID Working Paper 17/2017.

Brzozowska, Zuzanna and Monika Mynarska, *Fertility Intentions and Their Realisation: Insights from the Polish Generations and Gender Survey*, VID Working Paper 16/2017.

Yildiz, Dilek, Peter G.M. van der Heijden and Peter W.F. Smith, *Estimating Population Counts with Capture-Recapture Models in the Context of Erroneous Records in Linked Administrative Data* VID Working Paper 15/2017.

Brzozowska, Zuzanna, Éva Beaujouan and Kryštof Zeman, *Why Has the Share of Two-Child Families Stopped Growing? Trends in Education-Specific Parity Distribution in Low-Fertility Countries*, VID Working Paper 14/2017.

Rengs, Bernhard, Isabella Buber-Ennser, Judith Kohlenberger, Roman Hoffmann, Michael Soder, Marlies Gatterbauer, Kai Themel and Johannes Kopf, *Labour Market Profile, Previous Employment and Economic Integration of Refugees: An Austrian Case Study*, VID Working Paper 13/2017.

Beaujouan, Eva and Caroline Berghammer, *The Gap between Lifetime Fertility Intentions and Completed Fertility in Europe and the United States: A Cohort Approach*, VID Working Paper 12/2017.

Philipov, Dimiter, *Rising Dispersion in Age at First Birth in Europe: Is it related to Fertility Postponement?* VID Working Paper 11/2017 and Human Fertility Database Research Report 2017-005.

Lima, Everton E. C., Kryštof Zeman, Mathias Nathan, Ruben Castro and Tomáš Sobotka, *Twin Peaks: The Emergence of Bimodal Fertility Profiles in Latin America*, VID Working Paper 10/2017 and Human Fertility Database Research Report 2017-004.

Goujon, Anne, Sandra Jurasszovich and Michaela Potancoková, *Religious Denominations in Vienna & Austria: Baseline Study for 2016 - Scenarios until 2046*, VID Working Paper 9/2017.
